THE UNIQUE PHYLOGENETIC DISTRIBUTION OF VAULT PARTICLES REVEALS ITS FUNCTIONAL ROLES

ASFA ALLI SHAIK

(B.TECH, ANNA UNIVERSITY)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

NATIONAL UNIVERSITY OF SINGAPORE

2013

DECLARATION

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Asfa Alli Shaik

28 March 2013

Dedicated to

My Mom & Dad

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my family for their constant encouragement through all the years of my PhD journey. This thesis would not have been possible but for their continual motivation and understanding. I would like to thank my dad for his immense support through all my decisions and for all his practical advices. I thank my mom for nurturing me all these years and for all her love. I have always admired my mom's confidence in me and it surely did motivate me through every path of my life. I thank my little sister for maintaining a lively environment at home with all her laughter and gags. I would like to express my heartfelt gratitude to my Grandparents for always being there for me. My grandfather has always remained a man of values and principles and he has been a great inspiration for me. I always look up to him for his blessings which I utmost cherish. I spent a lot of my childhood days with my grandmother and she has been a great company. At this moment I would also like to extend my thankfulness to all my family members for their immense love and care.

I thank my supervisor Chris for giving me an opportunity to work on an exciting project. Through all these years, he has been constantly there to provide me with all the guidance and support that motivated me to work towards my goal. His ambitious projects made me look at science from a different perspective and not merely adhere to the canonical ways. I would also like to thank my collaborator Dr.Cynthia He for accommodating me in her lab and for all her scientific advices.

I would like to thank my labmate Soumya for her patient guidance and help in sailing me through the wet-lab experiments. I thank Yin-ru for all the helpful inputs and project related discussions. I also thank Sowmya K P, our lab executive for her support. I would like to extend my thanks to all my lab members for keeping the environment lively and fun-filled. I surely had fun during all our get-togethers.

A special thanks to Ladan, from Dr. Cynthia's lab, for helping me kick-start the *Trypanosoma* project. She has been a wonderful mentor and friend. I learnt a great deal working with her. I also would like to thank Omar for all the useful discussions. I thank all the other members in the lab for their friendly support.

Through the years, I have made some amazing friends. I am deeply indebted to each one of them in some way or the other. Divya, Vasanth and Lokesh, my best friends from undergraduate days, have always been there through thick and thin. I always cherish our wonderful times together. I am truly grateful to Suhas, also my labmate, for his immense support during the last stages of my PhD. He has been an awesome friend. I would also like to thank Madhuvika, my friend and roommate since the first day I came to NUS for making my days wonderful. I would also like to extend my thankfulness to Srirama, for all his encouragement; Parakalan, for all the fun and entertainment; Karthik, for being a good company; Arun, also my labmate, for being a good friend; Srinath for all the interesting discussions and Shaveta for being an amazing roommate and company.

I would like to express my sincere gratitude to my department, DBS for their funding support through my doctoral research. I would like to thank Priscilla and Reena for their immense help in various admin-related matters.

Last but not least, I would like to extend my heartfelt gratitude to all my teachers who have nurtured me through all these years.

TABLE OF CONTENTS

SUMMARY

Since its discovery in 1986, the function of the 13 million Dalton vault complex has remained perplexing. Approximately three times larger than the ribosome, it is the largest ribonucleoprotein complex within the eukaryotic cell. With an enigmatic barrel-shaped structure and hollow interiors, vaults exhibit an intriguing 39 fold dihedral symmetry. About 70% of the complex is composed of the Major Vault Protein (*MVP*) which forms the exterior shell structure, while some vaults also harbor minor vault proteins, vault poly-ADP ribose polymerase (*VPARP*) and telomerase-associated protein 1 (*TEP1*), along with untranslated vault RNAs. This massive complex is conserved in a wide range of eukaryotes, but absent from model organisms like yeast, *Arabidopsis, C.elegans* and *Drosophila* raising speculation about its origin, evolution and function. Various hypotheses - that vaults could be a part of the nucelopore complex, could act as a molecular cargo-carrier, could effectively shuttle out drugs in multidrug resistance cancer cells or could play an important role in signaling cascades and immunity - have been postulated. But the lack of a distinct phenotype in knockout models has hindered researchers from defining a precise function for vaults.

The current work focuses on unraveling the unique evolutionary history of vaults to track its origin, and also identify any subtle unifying traits in organisms with vault genes. A broad phylogenetic analysis of the vault particle genes including *MVP, VPARP* and *TEP1* reveals a complex evolutionary pattern and clues to the ancestral roles of vault. It is evident that all eukaryotes and bacteria with vault genes have had an ancestral loss of

essential amino acid biosynthesis or nitrogen fixation, and hence are heterotrophic and largely autophagic, including one of the bacterial species. The *MVP* gene appears to have arisen in an ancient cyanobacterium that lost genes pertaining to nitrogen fixation and acquisition of vaults into distinct clades of heterotrophic eukaryotes and bacterial species appears to be the result of multiple events of horizontal transfer. The observations are consistent with an ancestral function for vaults as a stable amino acid storage complex complimenting loss of nitrogen fixation and amino acid biosynthesis.

The efficacy of vault as an amino acid storage polymer was tested using compositional and theoretical biochemical analysis. There is a high ATP and carbohydrate energy equivalent of vault after amino acid recycling. A proposed synthesis-turnover based nutrient absorption function fits well with a number of reported vault expression and turnover patterns published in literature and other high throughput expression data. Key experimental observations associated with vault, including neuronal transport, accumulation in oocytes, and intrinsic immunity, can now be explained by this new proposed functional role of Vault.

Though homologs of *MVP* are known to be present in single-celled organisms including *Paramecium* and Kinetoplastids, a majority of the vault characterization studies have been undertaken in multicellular eukaryotes. The only single-celled eukaryote in which vaults have been studied is the slime mold *Dicytostelium discoideum,*that displays growth defects on vault gene disruption under conditions of nutrient stress. To evaluate if the nutrient-related roles of vaults extend to even those organisms that may have acquired

vault gene later in evolution through independent horizontal gene transfer events, the current study focused on characterizing vaults in ancient single-celled eukaryote Kinetoplastid. It is shown that vault genes have undergone paralogous expansion in Kinetoplastids into three differentially diverging sequences (*TbMVP1-3).Trypanosoma brucei* can form vault-like structures and *TbMVP1* sequence carries the inherent information that favors vault assembly. RNAi knockouts experiments reveal that*TbMVP1*is not required for *T. brucei* under normal nutrient conditions, but affects the density of cell growth only under nutrient limiting conditions, similar to that displayed by *Dictyostelium* on vault gene knockout. The nutrient-phenotype that becomes obvious on vault gene knockdown in an evolutionarily distant single-celled eukaryote reiterates that the proposed roles for vaults in nutrient accumulation and storage may be an evolutionarily conserved feature.

LIST OF PUBLICATIONS

1. Asfa Alli Shaik and Christopher W. V. Hogue. Unraveling the Evolutionary History of the Vault Complex
   (To be resubmitted)
2. Asfa Alli Shaik and Christopher W. V. Hogue. The medium is the message: A Consistent Theory for the Function of Vault Complex
   (To be submitted)
3. Asfa Alli Shaik, Ladan Gheiratmand,Cynthia Y. He and Christopher W. V. Hogue. A Role for Vaults in Nutrient Accumulation in *Trypanosoma brucei*
   (Manuscript in preparation)
4. Asfa Alli Shaik and Christopher W. V. Hogue. A new functional understanding ofexperimental observations of Vaults.
   (Manuscript in preparation)

Conference Presentations

1. The Society for Molecular Biology and Evolution, 2012 Conference, Dublin, Ireland (June 2012)
2. The 4th Asian Young Researchers Conference on Computational and Omics Biology, Singapore (Dec 2010)
3. The Structural Biology and Functional Genomics Conference, Singapore (Dec 2010)
4. The14th Biological Sciences Graduate Congress, Bangkok (Dec 2009).
5. The Singapore Symposium on Computational Biology (SYMBIO), Singapore (Sept 2009)

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ADP	Adenosine Diphosphate
ActA	Actin assembly-inducing protein
AIC	Akaike information criterion
Ala	Alanine
AMP	Adenosine Monophosphate
Arg	Arginine
Asp	Aspartate
ATP	Adenosine Triphosphate
BCAA	Branched Chain Amino Acids
BCKDH	Branched Chain Ketoacid Dehydrogenase
BCKDK	Branched Chain Ketoacid Dehydrogenase Kinase
BIC	Bayesian information criterion
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CAI	Codon Adaptation Index
CDS	Coding Sequences
CFTR	Cystic Fibrosis transmembrane conductance regulator
COP1	Constitutively Photomorphogenic 1
cpMVP	Cysteine rich peptide-MVP
DC	Dendritic cells
DNA	Deoxyribonucleicacid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraaceticacid
EGF	Epidermal growth factor
EGFR	Epidermal growth factor receptor
EM	Electron Microscopy
ER	Endoplasmic reticulum
Erk	Extracellular regulated kinases
FADH2	Flavin Adenine Dinucleotide
FAZ	Flagellar Attachment Zone
FRAP	Fluorescence recovery after photobleaching
FRET	Fluorescence resonance energy transfer
GAS	Gamma Activated Sites
GEO	Gene Expression Omnibus
GFP	Green Fluorescent Protein
Gln	Glutamine
Gly	Glycine
GTP	Guanosine Triphosphate
HBSS	Hank's Balanced Salt Solution
HCV	Hepatitis C Virus
HDAC	Histone Deacetylase
HDF	Human Diploid Fibroblast
His	Histidine
HRP	Horseradish Peroxidase

IF	Immunofluorescence
ICAM	Intercellular Adhesion Molecule
Ile	Isoleucine
IRF	Interferon Regulatory Factor
JAK	Janus Kinase
JNK	c-Jun-N-terminal kinases
Leu	Leucine
LPS	Lipopolysaccharide
LRP	Lung Resistance Protein
Lys	Lysine
MAPK	Mitogen Activated Protein Kinase
MDR1	Multidrug Resistance Protein 1
MEF	Mouse Embryonic Fibroblasts
Met	Methionine
mINT	MVP Interaction Domain
ML	Maximum Likelihood
mRNA	messenger RNA
MRP	Multidrug Resistance-associated Protein
MTOC	Microtubule Organizing Center
MtQ	Microtubule Quartet
MVP	Major Vault Protein
NAD	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
NCS	Non Crystallographic Symmetry
NJ	Neighbor Joining
NMR	Nuclear Magnetic Resonance
NPC	Nuclear Pore Complex
PARP	Poly (ADP-ribose) Polymerase
PBS	Phosphate Buffer Saline
PFR	ParaFlagellar Rod
Phe	Phenylalanine
Pro	Proline
PRPP	5-phospho-α-D-ribosyl 1-pyrophosphate
PTEN	Phosphatase and Tensin Homolog
PVDF	Polyvinylidene difluoride
RNA	RiboNucleicAcid
RNAi	RNA interference
SDS	Sodium Dodecyl Sulphate
Ser	Serine
SH2	Src Homology 2
siRNA	Small Interfering RNA
SMAD5	Mothers against decapentaplegic homolog 5
STAT1	Signal Transducer ad Activator of Transcription 1
TbMVP	Trypanosama brucei MVP
TBST	Tris Buffered Saline (Tween20)
TEP1	Telomerase Associated Protein 1

TGFB1	Transforming Growth Factor Beta 1
Thr	Threonine
TNF	Tumor Necrosis Factor
tRNA	transport RNA
TROVE	Telomerase Ro and Vault Module
Trp	Tryptophan
tTERT	Telomerase Reverse Transcriptase
UTR	Untranslated Region
UV	Ultra Violet
Val	Valine
VIT	Vault protein Inter-Alpha-Trypsin domain
VPARP	Vault PARP
vRNA	Vault RNA
VSG	Variable Surface Gycoproteins
vWA	van Willebrand type A domain
YFP	Yellow Fluorescent Protein
YPEL4	Yippee like 4 protein

Chapter 1

Introduction

While fractionating small subcellular structures including coated vesicles, ribosomes, smooth vesicles and ferritin from liver microsomal extracts using preparative agarose gel electrophoresis, Kedersha and Rome in 1986 identified a discrete fraction that appeared as uniform structures with ovoid morphology under negative staining and transmission electron microscopy (negative EM) (Kedersha and Rome 1986). Under negative EM, these particles formed highly regular unique barrel-shaped structures measuring 65-70 nm and 35-40 nm in dimensions. These unusual subcellular structures were originally observed as contaminants in clathrin-coated vesicle preparations from rat liver tissues (Kedersha et al. 1986). With more than 70% of the structure dominated by a single polypeptide of mass 104 kDa that co-migrates with a 100 kDa coated vesicle-associated polypeptide, it was speculated that these structures could be functionally related to clathrin-coated

vesicles. Though their fenestrated morphology bore some resemblance to those of clathrin-coats, it was established that these purified structures contained no clathrin and indeed were novel structures whose functions were unknown (Kedersha and Rome 1986). With a symmetric morphology reminiscent of the multiple arches of gothic cathedrals, the particles were named ***'Vaults'*** (Kedersha and Rome 1986).

Vaults, predominantly a protein complex with no detectable membranes, also harbor a single species of 5.6S RNA along with distinct vault associated proteins, making the complex the largest subcellular ribonucleoprotein complex. Its composition and structure mimics an RNA virus. It is approximately three times the size of a ribosome, yet remained unnoticed until 1986. Even if vaults were found in cell preparations, they could have been mistaken for coated vesicles due to their ovoid morphology (Kedersha and Rome 1986). Conventional stains for EM are highly attracted to charged components of membranes and nucleic acids but show low affinity to protein material, thus particles like vaults are almost invisible using a positive stain. Vaults became apparent only after enrichment and negative staining as shown in Figure 1.1 (Rome et al. 1991). Vault structures of similar size and morphology have been isolated and observed from several eukaryotic species including the invertebrate sea urchins and the evolutionarily distant amoebozoan slime mold *Dictyostelium*(Hamill and Suprenant 1997; Kedersha et al. 1990). While their wide distribution and conserved morphology is suggestive of an important function for vaults in eukaryotic cell, vaults are found to be conspicuously missing in certain other model eukaryotes including

plants, insects (*Drosophila melanogaster*) and nematodes (*Caenorhabditis elegans*).

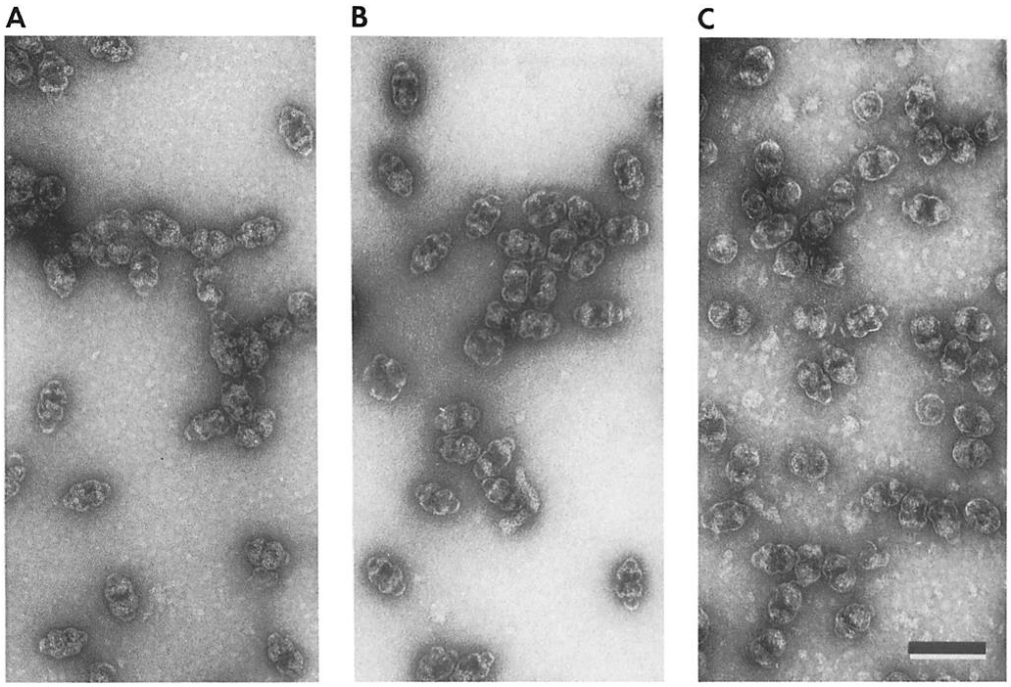


Figure 1.1 Negative EM of highly structured vaults purified from various eukaryotic species
(A) Vaults from rat liver; (B) Vaults from bullfrog liver; (C) Vaults from rabbit liver. Scale bar 100 nm. Figure reprinted from permission from (Kedersha et al. 1990)© Kedersha et. al and published by Rockefeller University Press.

1.1 THE DYNAMIC VAULT SHELL

With a mass of about 13 million Daltons, vaults represent the largest ribonucleoprotein complex found in eukaryotic cells. Interest in characterizing its components and unveiling its function has been steady since the complex came to limelight. Mammalian vaults are composed of three proteins and also contain several copies of small-untranslated RNA termed the vault RNA (*vRNA*). The predominant 104 kDa polypeptide, named the Major Vault Protein (*MVP*), was found to be the same protein as the previously identified Lung Resistance Protein (LRP) that is overexpressed in multidrug resistant cancer models. (Tanaka et al. 2009). Naturally occurring vaults also enclose minor vault proteins, 193 kDa vault poly-ADP ribose polymerase (*VPARP*) and 240 kDa telomerase-associated protein 1 (*TEP1*), along with a few copies of untranslated *vRNA* (vault RNA) that accounts for less than 5% of the entire structure(Kickhoefer et al. 1999; Kickhoefer et al. 1999; Kickhoefer et al. 1993; van Zon et al. 2001).

1.1.1 Structural Journey – Unveiling the Vault Cage

Vault has a very unique structure, unlike any previously known macromolecule and efforts on understanding its architecture, structural complexity and assembly have been carried out using various techniques including TEM, cryo-EM, nuclear magnetic resonance (NMR) and X-ray crystallography. Electron micrographs revealed that vault particles display a barrel-shaped structure with an invaginating thin-walled 'waist' region and protruding 'cap' regions at either end. Studies based on quantitative scanning transmission EM suggested that each vault particle was symmetric and composed of two shell-like complexes joined together at the middle. Half-

vaults were known to appear alongside intact vaults in preparations from *Dictyostelium* (Kedersha et al. 1990). Half-vaults can open into flower-like structures with each flower consisting of eight rectangular 'petals' based on observations using freeze-etch microscopy (Kedersha et al. 1991).

Since, *MVP* accounts for more than 70% of the protein mass it was speculated that it constitutes the exterior shell of the particle (Rome et al. 1991). Cryo-EM studies showed that the complex was indeed hollow with density around the central-barrel shaped cavity (Kong et al. 1999). The study also proposed an eightfold-symmetry for the vault complex based on three-dimensional image reconstructions, however at low resolution of 31 Å. Based on stoichiometric analysis, it was suggested that 96 molecules of *MVP* assemble to form the exterior shell of an intact vault complex.

MVP monomers expressed in insect cells that do not have the *MVP* or related minor genes can assemble spontaneously into intact vault complexes, demonstrating that *MVP* is the primary molecule responsible for the structure of vaults (Kong et al. 1999; Stephen et al. 2001). These *MVP*-only structures were found to be hollow and displayed properties similar to that of native vaults. Baculovirus expression in insect cells has since been routinely used in obtaining recombinant vaults in quantity. Thus the *MVP* protein sequence contains all the inherent structural information that governs the multimerization and assembly of the entire vault structure.

Deducing the structure of recombinant vaults with *MVP* N-terminal peptide tags using cryo-EM reconstruction techniques suggested that the internal density within the vault cage varied according to the length and size of the peptide tag. Reconstituting the structure using vaults with N-terminal

cysteine rich tags (cp*MVP* vaults) at 16-Å resolution also revealed a 48–fold rotational symmetry. Accordingly, it was suggested that 48 copies of *MVP* constitute each half vault and that two halves interact via non-covalent interactions between the N-terminal regions of *MVP* monomers at the vault waist region (Mikyas et al. 2004). These cryo-EM reconstructions also identified that the N-terminal region localize within the vault particle with their ends pointing towards the interior of the particle along the midsection (Mikyas et al. 2004). Thus, fusion tags engineered to the N-terminal region of *MVP* are generally packed within the interior space of the vault particle.

MVP is characterized by short sequence pseudo-repeats of about 55 amino acids occurring at the N-terminal half and also includes a long coiled-coil domain towards the C-terminal end. The coiled-coil domain has been found to be important in the interaction of individual *MVP* monomers with each other, and hence considered essential for *MVP* multimerization and vault assembly (van Zon et al. 2002). NMR studies suggested that these sequence repeats constitute structural units that adopt β-sheet-rich-folds. Based on this analysis, it was suggested that the barrel structure was built from at least six repeating structural domains. NMR analysis on a two-domain *MVP* fragment revealed a three-stranded antiparallel β-sheet with β2-β1-β3 architecture for each domain, with a flexible inter-domain linker region between β1 and β2 and a more structured loop region between β2 and β3 (Kozlov et al. 2006).

The first atomic model for vaults was proposed based on X-ray crystal structure data and computational model-building at 9-Å (Anderson et al. 2007). Like the cryo-EM observations, a cage-like structure with invaginated waist and two protruding caps were proposed for recombinant cp*MVP* vaults.

Six distinct structural features namely ‘Waist’, ‘Barrel’, ‘Shoulder’, ‘Cap Helices’, ‘Crossover’ and ‘Double-layer cap disk’, were defined. A 48-fold dihedral symmetry was proposed with each *MVP* monomer folding into 14 domains. However, at the ‘Crossover’ region, which forms the interface between the cap helix and the C-terminal cap disk region, it was suggested that the symmetry gets halved and this 24-fold symmetry is maintained in the C-terminal cap disk region as well. Each half vault, representing an asymmetric unit of the crystal, was proposed to be built from 24 identical pairs of *MVP* chains A and B, the two chains being different conformations of the *MVP* protein. The two identical chains were suggested to assemble differently at the cap region, with the type A chains pointing outwards to form the cap region and the type B chains folding inwards, hence shaping to 24-fold symmetry at the cap region. A whole-vault model was assembled as a 41.7 nm x 41.7 nm x 67.5 nm macromolecule.

Several attempts to analyze the crystal structure of vaults followed. In one study, analysis of two dimensional crystals of naturally occurring vaults isolated from murine cells revealed 6-fold dihedral symmetry with vault particles arranged hexagonally and interacting with each other along the central barrel region (Querol-Audí et al. 2005). However, a 3-fold dihedral symmetry was revealed by three dimensional crystals of vaults from monkey cells with flat triangular morphology that diffracted to about 10-Å (Querol-Audí et al. 2005). Concurrent with the reported 8-fold symmetry previously reported, it was suggested that the symmetry of vaults occurs in multiples of 24-fold rotational symmetry.

While most of the studies pointed to a 48-fold dihedral symmetry for the vault particles, a high resolution structure demonstrated that rat liver vaults have an odd 39-fold dihedral symmetry based on crystal analysis at 10-Å resolution (Kato et al. 2008). Based on a tightly packed crystalline lattice with no overlap between individual particles, a dimension of about 40 nm x 40 nm x 70 nm was suggested for each vault complex. On phase refinements by non-crystallographic symmetry (NCS) averaging assuming 2-fold to 48-fold rotational symmetry, it was found that structures exhibiting 3-fold, 13-fold and 39-fold symmetries revealed significantly lower reliability factors (R factors) and higher correlation coefficients. The deduced 39-fold dihedral symmetry, a unique observation for a macromolecular complex, was consistent with previously reported 3-fold symmetry observed in vaults isolated from monkey cells but challenged many earlier studies suggesting an eightfold or 48-fold dihedral symmetry. The detailed contours of the massive complex came to light after the X-ray crystal structure of rat liver vaults was resolved at 3.5-Å (Tanaka et al. 2009). The high resolution structure confirmed the 39-fold dihedral symmetry observed previously and also revealed that vaults assemble from 78 copies of *MVP* monomers, with 39 *MVP* chains in each half-vault and not from 96 copies as was believed earlier (Kato et al. 2008). This high resolution structure shows a graceful twist and packing of the coiled-coiled domain rather than the strange zig-zag structure in the previous low resolution model. The particle measures about 67 nm from top to the bottom with an internal cavity measuring about 62 nm x 35 nm (Figure 1.2). The cap region measures about 20 nm in diameter.

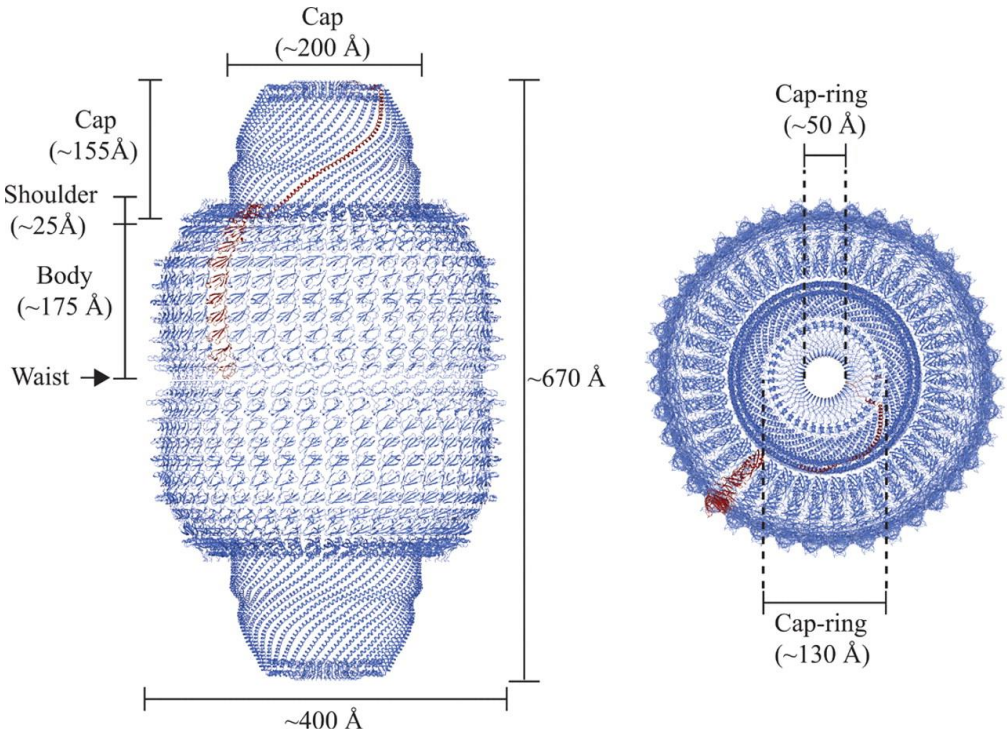


Figure 1.2 Structure of the Vault Complex
(Left) The particle exhibits a 39-fold dihedral symmetry with 78 *MVP* chains assembling to form an entire complex. The particle displays a barrel-shaped structure with two protruding caps, two shoulders and an invaginated waist region. The two halves of the vault particle interact at the waist mediated by N-terminal domains of *MVP*. A single *MVP* chain is represented in red. (Right) Top view of the complex showing the cap ring region. From (Tanaka et al. 2009). Reprinted with permission from The American Association for the Advancement of Science.

In contrast to the proposed 14 structural domains, the 3.5- Å structure shows that each *MVP* chain is comprised of nine structural repeat domains, a shoulder domain, a cap-helix domain and a cap-ring domain (Figure 1.3). End-to-end association of structural repeat domain 1 was found to form the waist region. Except for domains 8 and 9 which assemble from five antiparallel β-strands, the other repeat domains have two antiparallel β-strands, consistent with the NMR sub-structure (Kozlov et al. 2006). The structural domains end at the shoulder domain and assemble as four-stranded α helices on one side and four-stranded antiparallel β sheet on the other side, thus folding into a

single α/ β globular domain. This is followed by the cap-helix domain forming a spiraling 42-turn-long α helix that terminates as a U-shaped structure at the top of the cap forming the cap-ring domain. The self-assembly of MVP is attributed to 41 interactions that occurred specifically at the cap-helix region between two *MVP* subunits. The bulk of interaction strength arises from hydrophobic residues that appear at the interface between two helices contributing to specific hydrophobic interactions between two *MVP* chains. Although the N-terminal associations at the waist regions between two half-vaults were found to be predominantly hydrophilic, the interactions were particularly weaker than the hydrophobic interactions stabilizing the cap structure, thus explaining the appearance of half-vault structures that appear as flower-like structures (Kedersha et al. 1991).

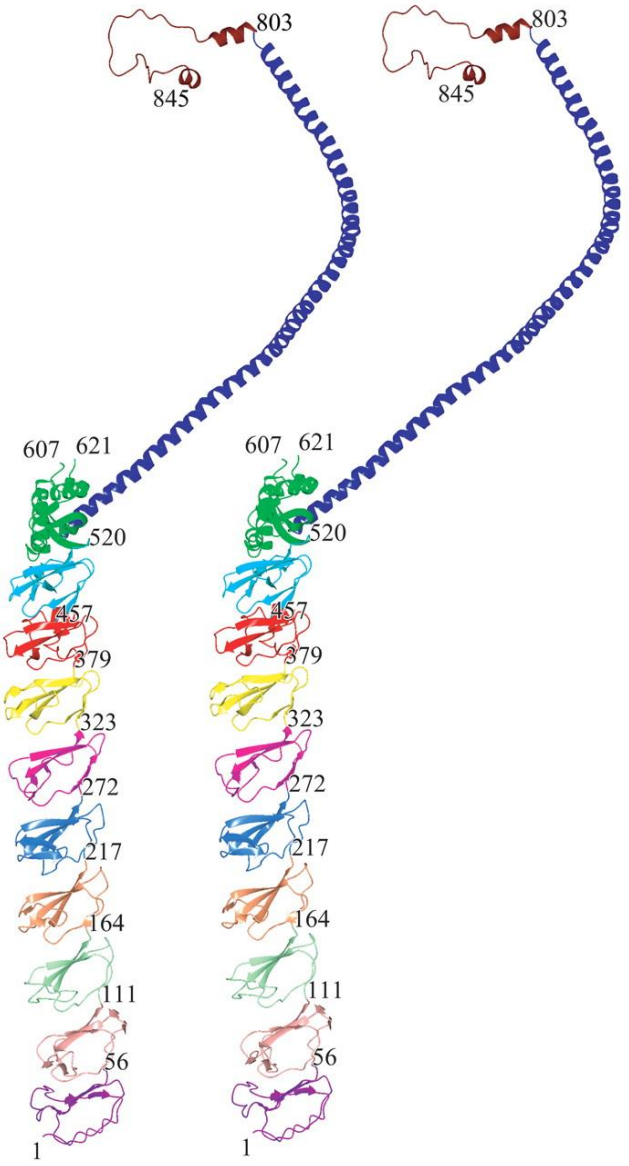


Figure 1.3 Overall fold of an *MVP* chain
The nine structural folded repeat domains in a single *MVP* chain from rat are depicted as follows: domain 1 (Met1-Pro55), purple; domain 2(Arg56-Thr110), pink; domain 3 (Pro111-Ile163), light green; domain 4 (Gln164-Val216), coral; domain5 (Asp217-Val271), light blue; domain 6 (Pro272-Asp322), magenta; domain 7 (Val323-Gln378), yellow; domain 8 (Ala379-Arg456), red; domain 9 (Val457-Gly519), cyan. The shoulder domain (Pro520-Val646),green; cap-helix domain (Asp647-Leu802), purple; and cap-ring domain (Gly803- Ala845), dark red. From (Tanaka et al. 2009). Reprinted with permission from The American Association for the Advancement of Science.

Though the vault complex exhibits structural stability over a wide range of pH, temperature and cellular conditions, studies have established that the exterior shell is dynamic in nature (Esfandiary et al. 2008; Kedersha et al.

1991; Yang et al. 2010). At low pH conditions (pH 3.4) the complex loosens its associations along the midsection and dissociates into half-vaults (Figure 1.4). In one of these studies, FRET revealed that individual *MVP* monomers can be exchanged between vault particles. The complex, as a whole, can spontaneously disassemble at the waist region and reassemble back into intact vault particle without compromising its structure, suggestive of a half-vault exchange. The ability of the complex to dissociate into half-vaults may be an important mechanism through which vaults interact with their cellular environment. However, the functional significance of this dynamic behavior is unexplained.

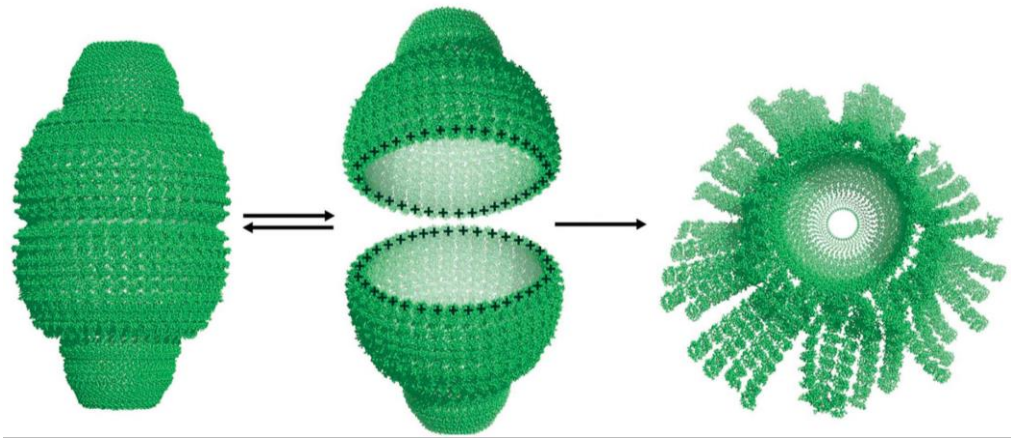


Figure 1.4 Vault dissociation into two halves
At low pH conditions, as in lysosomes, charge repulsion induces the disassembly of the particle along the waist region to form two half-vaults. The half-vault moiety can open up into a flower like structure as shown in right side of the figure. Figure reprinted with permission from (Querol-Audí et al. 2009).

1.1.2 Looking inside – Exposing Locations of Minor Vault Constituents

While initially it was known that *vRNA* does not contribute to the structural integrity of the vault complex, the exact location and roles of the minor vault proteins *VPARP* and *TEP1*remainedunclear (Kedersha and Rome 1986; Rome et al. 1991). Though the exact stoichiometry of the minor vault proteins associating remains uncertain, it has been estimated that about 4-16 *VPARP* and 2-4 *TEP1* molecules are found in intact vaults (Berger et al. 2008). Expression of recombinant vaults in insect cells shows that *VPARP* and *TEP1*, when co-expressed separately or in combination with *MVP*, are incorporated into intact vault particles, and exhibit independent interaction with *MVP*. One yeast two-hybrid assay did not reveal any interaction between *TEP1* and *MVP* or *VPARP* (van Zon et al. 2002). This hinted that *TEP1*only binds to intact vaults, instead of individual *MVP* chains. The ratio of incorporation of *VPARP*to*TEP1*in recombinant vaults is higher than that in endogenous vaults (Mikyas et al. 2004). Co-expression of *MVP* along with minor vault proteins resulted in formation of recombinant vaults that were more structurally stable and regular.

In spite of an internal cavity large enough to enclose other macromolecules, the vault shell does not accumulate other proteins within its hollow interior excepting the minor vault proteins. The minor vault proteins can accumulate after the assembly of MVP is complete, suggesting a dynamic opening of vault exterior shell (Poderycki et al. 2006). Despite this observation, there has been no evidence of non-specific proteins being enclosed within the vault interior.

Cryo-EM images or TEM of negatively stained vault preparations consistently reveal regions of interior density, believed to arise from minor vault proteins. The production of recombinant vaults in large quantities proved useful in determining the location of *VPARP*, using various single-particle reconstruction techniques. Using differential mapping of cryo-EM reconstructions of various recombinant and tissue-derived vaults, *VPARP* molecules appear tucked within the inner waist surface of vault particles.

A C-terminal sequence fragment of *VPARP* from 1562-1724 aa, termed the *MVP* interaction domain (mINT), was established as its assembly domain with *MVP* as established by a yeast-hybrid screen (Kickhoefer et al. 1999). When various non-vault associated proteins including luciferase, GFP or mCherry were fused with mINT domain, the expressed recombinant vaults successfully enclosed the fusion protein within its hollow cavity(Kickhoefer 2005; Kar et al. 2011). Cryo-EM and single-particle image reconstruction revealed that the recombinant vaults containing mINT fusion proteins exhibit properties similar to those of endogenous vaults, however, display additional density along the central barrel region (Kickhoefer 2005). Interaction studies using various *MVP* and *VPARP* truncated proteins established that N-terminal part of *MVP* could interact with the C-terminal region of *VPARP* containing the mINT domain (van Zon et al. 2002). Thus, the increased density found along the vault barrel region, particularly immediately above and below the waist regions, has been attributed to the interaction between mINT region of *VPARP* and N-termini structured region of *MVP* monomers. that the mINT region interacts with purified recombinant vaults and gains access to its interior surface even in the absence of cell extracts (Poderycki et al. 2006).

NMR experiments further suggested that *VPARP* mINT specifically binds to the exposed folded domains spanning 113-221 aa in *MVP* and that binding of *VPARP* does not impose major conformation changes in *MVP* molecule (Kozlov et al. 2006). Based on electrospray mobility analysis of vault complexes, it was calculated that about 9.5 *VPARP* molecules could be incorporated into preformed vaults similar to the calculated 8.3 enclosed copies of *VPARP* arising from co-expression in insect cells.

The second minor vault protein, *TEP1*, with a notable RNA binding domain, is a constituent of another ribonucleoprotein, the telomerase complex. Since*TEP1*is known to interact with telomerase RNA, it was speculated that the vault bound *TEP1*was responsible for binding *vRNA*s. This was confirmed by a yeast three-hybrid assay (Kickhoefer et al. 1999). It was found that *TEP1*residues 270-871, which comprise the p80 homology region, bind to both telomerase RNA and *vRNA*s (Poderycki et al. 2005).

Comparisons of intact vault reconstructions with RNase-treated vaults suggested that *vRNA*s are contained within the cap region on both ends. Cryo-EM reconstructions of RNase-treated vaults identified a dense region within the vault cap with a proposed 16-fold symmetry (Kong et al. 2000). *TEP1*contains 16 WD40 repeats at its C-terminus. The WD40 repeats are known to fold together as organized beta-propeller structures. This was thought to correlate to the 16-fold WD40 repeat of *TEP1*and was modeled accordingly along the cap region in close associations with *vRNA*, confirming the proposed 8-fold symmetry (Kong et al. 2000). However, another study reported 5 additional WD40 repeat regions and proposed that *TEP1* may structurally fold as three connected seven-bladed propellers, as the seven-

bladed β-propeller structure was a known feature of many WD-repeat containing proteins (Mikyas et al. 2004). Though cryo-EM reconstructions of vaults isolated from *TEP1*$^{-/-}$ mice display overall structural features similar to those from wild type, the cap region density attributed to *vRNA* was missing. On comparing reconstructions of RNase-treated vaults with *TEP1*$^{-/-}$ vaults, it was apparent that while the 16-fold dense region around the outermost edge was preserved in both reconstructions, regions of less density were identified within an intermediate ring in vaults isolated from *TEP1*$^{-/-}$ mice. This led to the mapping of *TEP1* to the protruding cap region, as was previously predicted based on structural modeling (Kickhoefer et al. 2001). Differential density mapping of various recombinant and tissue derived vaults, however, revealed very weak differential density within the cap cavity region and thus it was suggested that only a small region of *TEP1* could possibly be involved in interaction with the vault interior (Mikyas et al. 2004).

Recombinant vaults produced by co-expression of various *TEP1*truncations, revealed that only those proteins that retained the p80 homology domain were successfully encaged within vault complexes (Poderycki et al. 2005). Though *vRNA* is found in close associations with *TEP1*, it does not mediate association of *TEP1* with vaults. Akin to the mINT domain for *VPARP*, the p80 homology domain in *TEP1* facilitates its association with both intact vault complexes and bound *vRNA*. It is also known that *TEP1* is responsible for stable association of *vRNA* with vault particles as vaults purified from *TEP1*$^{-/-}$ mice showed only small traces of *vRNA*(Kickhoefer et al. 2001). Approximately, three- to fivefold-reduced levels of *vRNA* have been reported in *TEP1*$^{-/-}$ mice, suggestive that stability

and accumulation of *vRNA* within the vault is dependent on *TEP1*. A direct correlation between *vRNA* stability and *TEP1*was shown by comparing the half-lives of *vRNA* from wild type and *TEP1*$^{-/-}$ mice. The *vRNA* half-life was reduced to about 0.5-1 hour in *TEP1*$^{-/-}$ mice as against 4-6 hour in wild type mice. Thus*TEP1*acts as a *vRNA* stabilizing molecule, facilitating its localization within the internal cavity of the vault caps. However the role of this accumulated and stable *vRNA* within the vault complex remains unclear.

1.1.1 Vault Components – A Perspective

1.1.1.1 *MVP*

The protein responsible for the cage-like structure of vault, *MVP*, is a unique and highly conserved protein through evolution. *MVP* can spontaneously assemble to form intact vault particles without minor vault constituents, implying that *MVP* primarily exhibits its cellular function as structured vault particles. With equilibrium favoring vault assembly against free *MVP* monomers, experiments show it takes about 4 hours for newly synthesized *MVP* chains to assemble into an intact vault particle (Zheng et al. 2005). Pulse-chase experiments have established that *MVP* assembled as vault particles represent a highly stable macromolecule structures with an apparent half-life of about 3 days (Zheng et al. 2005). The structural similarity of vaults isolated from a wide range of organisms including single-celled slime molds and multicellular metazoans, can be attributed to the conserved sequence of *MVP*.

The conserved coiled-coil domain at the C-terminal region of *MVP* makes up the stable cap-helix region in vault particles and mediates interactions between individual *MVP* monomers, thus enabling vault

assembly. Abolishing the coiled-coil domain, either entirely or partly, affects interactions between *MVP* monomers and disrupts assembly (van Zon et al. 2002). Motif analysis shows at least two putative EF-hands at the N-terminal region of *MVP* hinting at possible calcium or magnesium binding sites. EF-hands are marked by a loop structure separated by two alpha helices, with residues in the loop involved in calcium binding. Calcium-binding assays have confirmed that the N-terminal region of *MVP* effectively binds calcium through these EF-hand motifs (Yu et al. 2002). *MVP* also interacts with PTEN, a tumor suppressor gene involved in the regulation of cell-cycle, in a calcium-dependent manner and also mediates its nuclear translocation (Yu et al. 2002; Minaguchi et al. 2006).

MVP is a widely expressed protein in many different cell types. The high expression of *MVP* in certain metabolically active cells provides vaults in high copy numbers. Mammalian cells have up to 10^4 vaults and sea urchin oocytes harboring around 10^7 vaults (Kickhoefer et al. 1998; Hamill and Suprenant 1997). Both the mRNA and protein levels of *MVP* have been found to be elevated in a certain cellular conditions including cancer related multidrug resistance, cell aging, rapidly dividing tumors, cell regeneration, oocyte/embryo development and intracellular infections. Mammalian *MVP* expression is up-regulated by the cytokine interferon-gamma, displaying increased levels of transcription and translation (Steiner et al. 2006). The proinflammatory cytokine primarily secreted by macrophages, tumor necrosis factor-alpha, negatively regulates expression of *MVP* at both mRNA and protein levels (Stein et al. 1997). A relatively small increase in transcription level magnifies the protein levels of *MVP* by many folds, suggestive of high

translation efficiency. This phenomenon has been observed in interferon-gamma induced cells and also in actively regenerating cells (Steiner et al. 2006). This suggests that the expression of *MVP* may be regulated at the post-transcriptional level and the *MVP* mRNA is stabilized to allow enhanced translation resulting in elevated protein levels. Paradoxically, the high level of expressed cellular *MVP* is often accompanied by an increased *MVP* turnover rate, resulting in rapid degradation of the produced protein. Thus, protein stability seems to be compromised when *MVP* accumulates in the cell in response to various cellular conditions (Li et al. 1999; Steiner et al. 2006; Sutovsky et al. 2005).

The human *MVP* gene maps to chromosome position 16p11.2 and the promoter is TATA-less, akin to the murine counterpart. The regulatory human promoter region is marked by several transcription factor binding sites including an inverted CCAAT box, a GATA box, an E box and a GC box element, and is also defined by specific activating or inhibitory regions (Lange et al. 2000; Steiner et al. 2004). The expression of *MVP* at the transcriptional level involves binding of specific transcription factors during different cellular conditions. Deletion analysis confirmed the GC-box element as a necessary region for basal *MVP* promoter activation. It has also been reported that the upstream gene region is positively stimulated by binding of several Sp-family transcription factors (Steiner et al. 2004). Additional consensus binding for STAT1 (GAS element), activated upon interferon-gamma induction, has also been found on the core activating sequences in the promoter region. The inhibition of histone deacetylase (HDAC) leads to both transcriptional and translational upregulation of *MVP*.

It has been established that p53 binds to a response element within the Y-box region in the human *MVP* promoter region and negatively regulates vault expression. The binding of the $HDAC_2$-p53 transcriptional repressor complex to the Y-box region has been found to repress the expression of *MVP* via interaction with YB-1 (Tian et al. 2011). The expression of *MVP* is also believed to be controlled by a mechanism of alternative splicing. An alternative 3'-splice site in intron 1 results in a longer splice variant within the 5'- untranslated region of *MVP* mRNA (Lange et al. 2000). The longer variant appears to contain a small upstream open reading frame that represses expression of *MVP*, both *in vitro* and *in vivo*, and hence suggested a role in regulating *MVP* expression, particularly during malignant transformations (Holzmann et al. 2001). In murine promoters, essential regulatory elements have been mapped to regions of the first exon spanning all the way until the 5'-end of first intron (Mossink et al. 2002).

1.1.1.2 *VPARP*

Characterization of the minor vault proteins that constitute naturally occurring vaults led to the identification of a 193 kDa protein that interacts with *MVP* in a yeast two-hybrid screen and also shares 28% identity with the catalytic domain of poly (ADP-ribose) polymerase (Kickhoefer et al. 1999). The novel protein termed the Vault Poly (ADP-ribose) Polymerase was found to catalyze the formation of poly (ADP-ribose) polymers, hence has been regarded as a novel PARP family protein. The PARP family of proteins represents a group of protein-modifying and nucleotide-polymerizing enzymes that play key roles in DNA repair, genomic stability, cell death, transcriptional

control and epigenetic regulation. Predominantly nuclear enzymes, they catalyze the attachment and elongation of poly-ADP-ribose units to glutamic and aspartic residues to target proteins and onto themselves using NAD+ as a substrate (Citarelli et al. 2010).

At least 17 different members of this family, each encoded by different genes yet with overlapping functions, have been documented thus far. Of the distinctly identified members, including PARP-1, PARP-2, PARP-3, PARP-4 (*VPARP*), tankyrase 1 and tankyrase 2, the nuclear PARP-1 involved in genomic stability remains the best-characterized protein. Poly (ADP-) ribosylation has been implicated in a wide range of cellular conditions including apoptosis, cancer, inflammation, neurodegeneration and brain damage. Evolutionary analysis of the PARP family of proteins suggested that an ancient eukaryotic ancestor harbored at least two PARP genes, one among them being highly similar to PARP-1 functioning in DNA repair (Citarelli et al. 2010). The expansion of PARP encoding genes into a broad family of PARP proteins over evolutionary time underscores their functional significance in diverse cellular processes. Because of its involvement in DNA repair and the observed tumor-suppressor effects of PARP inhibitors, the PARP proteins, particularly PARP-1, have been regarded as a potential target for cancer therapy (Wang et al. 2012; Basu et al. 2012; Lavarone et al. 2013).

The *VPARP* gene mapping to chromosome 13q11 encodes a multi-domain protein characterized by a PARP domain, BRCT domain, a putative van Willebrand type A domain (vWA) and a vault inter-alpha-trypsin domain (VIT) apart from the mINT domain that mediates interaction with *MVP*. A subset of total cellular *VPARP* is vault-associated and co-localizes with *MVP*,

observed as punctate patterns, in the cytoplasm of many cell types. Owing to its dynamic exterior shell, it has been found that vaults incorporate *VPARP* within their interior surface in about 1.5 hours (Zheng et al. 2005; Poderycki et al. 2006). This is unlike many PARP proteins which display a predominant nuclear localization. In addition to ribosylating themselves, it has been found that vault-associated *VPARP* are also capable of ribosylating *MVP*, albeit with lower ribosylation efficiency compared to other PARP proteins. Whether poly (ADP-) ribosylation affects *MVP* or the conformation of intact vault complexes is yet to be explored.

A fraction of *VPARP* is also localized within the nucleus but shows no co-localization with *MVP*, suggestive of a non-vault associated fraction. A portion of the nuclear *VPARP* also aligns along the mitotic spindle. In spite of belonging to the PARP family of proteins, which respond to DNA damage, no change in distribution of *VPARP* was detected following UV-treatment of cells (Kickhoefer et al. 1999). Contrary to the yeast two-hybrid assay, which failed to detect any interactions between *VPARP* and *TEP1*, co-immunoprecipitation of *TEP1* along with non-vault associated *VPARP* was observed in cells transfected with both genes. However, this interaction is dubious as no such interactions could be observed between the endogenous minor vault proteins. Overexpressed *VPARP*, lacking both vWA and mINT domain, is also reported to display telomerase activity. It should be remembered that PARP proteins are responsible for chromosomal stability and also directly control telomere length (d'Adda di Fagagna et al. 1999). In spite of the reported telomerase activity, *VPARP*$^{-/-}$ mice displayed no significant telomere abnormalities. The length and structure of telomere, telomerase activity and stability was

comparable to those of wild-type mice and remained unchanged in knockout mice, thus questioning the reported telomerase activity in *VPARP* transfected cells. It has also been established that *VPARP* plays no role in maintaining chromosomal stability or processes related to DNA damage repair. This suggests that vault-associated *VPARP* has distinct roles to play and may not be directly involved in genome stabilization or repair like other conventional PARP proteins. The only reported phenotype for *VPARP* knockout mice is its increased susceptibility to carcinogen induced colon and lung tumorigenesis (Raval-Fernandes et al. 2005).

Wrapping within the interiors of vault significantly increases the stability of *VPARP*, giving it an apparent half-life of at least 40 hours. While evolutionary analyses have revealed existence of *VPARP* based on sequence homology in single-celled *Dictyostelium*, no evidence on its expression or function has been described in any lower or single-celled eukaryotes (Citarelli et al. 2010). For instance, the single-celled protist *Trypanosoma* encodes proteins that display significant homology to *MVP* from higher eukaryotes; however no protein homologous to *VPARP* is encoded in its genome. It should be noted that *VPARP* is not essential for vault assembly and does not influence the incorporation of other minor vault constituents. This hints that incorporation of *VPARP* into vaults is an event that occurred late in evolution. However, the exact functions of *VPARP* being associated with vaults or even otherwise are unknown as yet. Thus, *VPARP* joins the likes of its host complex, the intriguing vault particle, and remains elusive thus far.

1.1.1.3 *TEP1*

The 240 kDa minor vault protein that appeared along with *VPARP* in vaults isolated from rat liver was found to be identical to the previously described mammalian telomerase-associated protein, *TEP1*(Harrington et al. 1997; Kickhoefer et al. 1999). The genes towards the ends of a chromosome are inherently maintained intact by addition of new telomeres to the existing ends of chromosome catalyzed by the telomerase ribonucleoprotein. The mammalian homolog to the Tetrahymena p80 telomerase protein, *TEP1* potently interacts with mammalian telomerase RNA and hence is alleged to be a part of the telomerase complex (Harrington et al. 1997). Based on co-purification of Tetrahymena p80 protein with telomerase activity on immunoprecipitation with anti-serum, it was believed that p80 also exhibits significant interaction with the catalytic protein component, the telomerase reverse transcriptase (tTERT). However, interaction studies in Tetrahymena ruled out any association between p80 protein and tTERT, raising speculations on the role of p80 in telomerase activity (Mason et al. 2001). Hence, it was suggested that p80 protein was not a core telomerase specific component and displayed affinity to other RNA as well (Mason et al. 2001).

*TEP1*interact as well as stabilize the association of *vRNA* within the interior cap region of the vault ribonucleoprotein (Poderycki et al. 2005; Kickhoefer et al. 2001). In spite of the association of *TEP1*, purified vaults neither display telomerase activity nor associate with telomeres (Kickhoefer et al. 1999). *TEP1*$^{-/-}$ embryonic stem cells (ES cells) or mice do not differ significantly in their telomerase activity and distribution or mean length of telomeres from their wild type controls and remain fertile with no

developmental defects (Liu et al. 2000). A double-knockout *VPARP*$^{-/-}$*TEP1*$^{-/-}$ mouse strain displayed no chromosomal abnormalities and appeared normal with unaltered telomerase activity, excluding any functions related to telomerase catalysis for the minor vault proteins. Consequently, a more structural role of *TEP1* in the assembly of the complexes was proposed based on its association with the two unrelated ribonucleoprotein complexes, vaults and telomerase.

Whether *TEP1* influenced the assembly of the ribonucleoproteins was studied in a series of experiments using a *TEP1*$^{-/-}$ knockout model (Kickhoefer et al. 2001; Liu et al. 2004). *TEP1*plays no role in telomerase RNA processing or stability and does not influence its association with telomerase complex. This pointed to additional telomerase RNA binding proteins that may aid assembly of telomerase complex, making *TEP1*a functionally redundant protein. In contrast, *TEP1* in vaults stabilizes *vRNA*, as evident from reduced half-life of *vRNA* from *TEP1*$^{-/-}$ mice (Kickhoefer et al. 2001).

Transcribed from 14q11.2 genomic loci, the *TEP1* protein is characterized by several conserved domains including the p80 homology domain that mediates interaction with telomerase RNA and *vRNA*. Interestingly, the p80 homology domain is also sufficient to allow its interaction with intact vault complexes. Based on domain analysis, an evolutionarily conserved binding domain potentially involved in RNA-binding activity termed the TROVE module was identified in several RNA-interacting proteins including Tetrahymena p80, *TEP1* and Ro RNP protein component, Ro 60, that interacts with the Y RNA (Bateman and Kickhoefer 2003). A vWA domain is also commonly present in all these RNA binding proteins. It

has been found that removal of the vWA domain interferes with the ability of *TEP1* to bind *vRNA* and it is believed that vWA domain could probably maintain the conformation of the RNA-binding domain in the TROVE module (Poderycki et al. 2005)Notably, a vWA domain is also present in*VPARP*.

In spite of the fact that *TEP1* can interact with both telomerase and vault RNA, purified vault preparations only include *vRNA* and no traces of telomerase RNA. These details point to distinct mechanisms for*TEP1*in different complexes, binding to either *vRNA* or telomerase RNA and allowing for specific targeting towards vaults or telomerase complex, respectively. While *TEP1* is stably enclosed within the interior surface of vault cap regions, no specific interactions have been established between *MVP* monomers and *TEP1*(Mikyas et al. 2004; Poderycki et al. 2005; van Zon et al. 2002). This emphasizes that the interaction between *TEP1* and *MVP* occurs by virtue of the assembled vault structure cap region, within which *TEP1* binds. Finally, WD40 repeat regions are found towards the C-terminus of *TEP1* and they have been modeled within the cryo-EM density of the vault cap regions (Kong et al. 2000). Whether vault-associated *TEP1* has additional roles to play, apart from stabilizing *vRNA* within vault interiors, remains ambiguous.

1.1.1.4 *vRNA*

In addition to the minor vault proteins, vaults isolated from a number of multicellular eukaryotes also featured untranslated vault RNAs that localize within the cap region and constitute about 5% of the total mass of a vault particle (Kedersha and Rome 1986). While in rats and mice only a single *vRNA*, 141 bases long, has been described, human vaults have been found to accommodate three *vRNA* sequences that vary from 86- 89 bases (Kickhoefer

et al. 1993; Kickhoefer et al. 1999). The *vRNA* genes, named hvg1-3, are arranged as a triplet-repeat structure on chromosome 5q33.1 (van Zon et al. 2001). Based on homology search, a fourth *vRNA* gene, *hvg4* has also been described and it is mapped on to Xp11.2 genomic loci. The *vRNA* gene is composed of a novel RNA pol III promoter containing two different classes of promoter elements, namely the external 5'flanking type-3 and internal type-2 promoter elements (Vilalta et al. 1994). The two promoter elements function synergistically and mediate transcription by RNA polymerase III (Kickhoefer et al. 1993; Vilalta et al. 1994). The conserved internal promoter elements are marked by one A box and two B boxes (B1 and B2) (Vilalta et al. 1994; Kickhoefer et al. 2003). The region upstream of the transcription start site in all vertebrate *vRNA* genes includes a TATA box as well as a conserved proximal and distal sequence element defining the external promoter region (Kickhoefer et al. 2003).

The hvg1*vRNA* transcript associates with the vault complex (Kickhoefer et al. 2001). A yeast three-hybrid screen demonstrated that *vRNA*s, *hvg1, hvg2* and *hvg4* could interact with *TEP1*, hinting at the association of the varied *vRNA* within intact vault complexes (Kickhoefer et al. 1999). Expression analysis of various human cell lines demonstrated that hvg1 was much more effectively expressed while hvg2 and hvg3 show low expression levels (Walker et al. 2004). Kickhoefer et al. demonstrated that hvg1 and hvg4 consistently co-purify with vaults as opposed to*hvg2* and *hvg3* (Kickhoefer et al. 1999). Contrary to this, another study established that *hvg4* gene is not expressed in many analyzed cell lines (van Zon et al. 2001). It is interesting to note that the promoter regions of both *hvg4* and a pseudo gene

described in mouse, *mvg2*, do not share significant regions of identity with the upstream regions composed of type-3 TATA and other conserved regions defining the external promoter elements of *hvg1-3* (Kickhoefer et al. 2003). Given the importance of upstream promoter elements in transcribing *vRNA*, the expression and association of *hvg4* with vaults is highly speculated and *hvg4* is widely regarded as a pseudo gene(van Zon et al. 2001; Stadler et al. 2009).

It should be noted that a majority of *vRNA* is cytoplasmic in nature and only a small fraction of the expressed *vRNA* associates with vaults (van Zon et al. 2001). In spite of the fact that multiple human *vRNA*s *hvg1-3* can associate with vault complexes, it has been experimentally determined 80% of the *vRNA* in vault complexes is constituted by *hvg1*, making it the predominantly associated species (van Zon et al. 2001). Evolutionary analysis suggested that *hvg2* and *hvg3* are genes that arose out of a recent duplication event (Stadler et al. 2009). In multidrug resistant cell lines in which vaults are more often overexpressed, the association of *hvg3* with vaults increases (van Zon et al. 2001). This highlighted the varied nature of the multiple *vRNA*s and suggested that they exhibit different levels of affinities to vault particles depending on cellular conditions.

The known *vRNA* genes in all gnathostomes map to a conserved region linked to the protocadherin-*α* cluster encoding a family of synaptic adhesion molecules (Stadler et al. 2009). A novel non-coding RNA termed *CBL-3* displayed elevated expression levels in addition to canonical *vRNA*s (*hvg1-3*) in human lymphocytes exposed to Epstein-Barr virus infection (Nandy et al. 2009). Interestingly, the *CBL-3* transcript was found to co-purify with vaults

and hence was proposed as a novel vault complex-associating RNA. Based on homology search it was later identified as a new *vRNA* gene that syntenically maps to the TGFB1-SMAD5 locus in all eutherian mammals as against the canonical locus for most of the described *vRNA* genes (Stadler et al. 2009).

The *vRNA*s do not mediate structure assembly of the vault particles as their digestion with ribonucleases still maintains the unique morphology of the vault particle, albeit with less density within the cap regions (Kong et al. 2000; Kickhoefer et al. 2001). Apart from its interaction with *TEP1*, *vRNA* has also been found to interact with *La* RNA-binding protein *in-vivo* and *in-vitro*(Kickhoefer et al. 2002). The *La* phosphoprotein is known to associate with small-untranslated RNAs prior to their incorporation into larger ribonucleoproteins via the polyuridine-rich sequences found at the 3'end of RNAs transcribed by RNA polymerase III (Maraia and Bayfield 2006; Ford et al. 2001). In contrast to its usual transient associations with RNAs, *La* binds more stably with *vRNA* as the latter remains unprocessed and retains the polyuridylate tail at the 3'end intact. Since the *La* RNA-binding protein loosely purifies with vaults, it is believed that it could promote the *vRNA-TEP1* interaction within vault complexes (Poderycki et al. 2005). The functional significance of the multiple copies of *vRNA*s in many eukaryotic species and why they are targeted to the interior of the vault complex remains mysterious thus far.

1.2 CELLULAR FUNCTIONS ASCRIBED TO VAULTS

Vaults were initially believed to diffuse freely through the cytoplasm (van Zon et al. 2006). However, on observations using enhanced-video

microscopy and FRAP analysis, vaults were found to move with a velocity of about 10 μm/s, suggestive of fast active transport (Slesina et al. 2006). This phenomenon showed high coherence with previous reports on fast anterograde and retrograde axonal transport of vaults, along the axons of electric ray, akin to the transport of synaptic vesicles and mitochondria along microtubules (Li et al. 1999). The movement of vaults within the cell is attributed to their association with cytoskeletal elements, particularly microtubules, and the involvement of molecular motors (Kedersha and Rome 1990; van Zon et al. 2006). The co-localization and co-purification of microtubules with *MVP* have been observed in varied cell types and even after microtubule disassembly, specific interactions between vaults and tubulin dimers/oligomers remain intact (Herrmann et al. 1999; van Zon et al. 2006; Hamill and Suprenant 1997; Slesina et al. 2006; Eichenmüller et al. 2003). It has been observed that about 5-6 vaults can associate with 1 μm-long microtubule, and the binding interaction has been found to occur along the vault cap regions (Eichenmüller et al. 2003). It has been found that microtubule depolymerization, in addition to resulting in slower vault movements, also causes vaults to aggregate into cylindrical dynamic structures termed 'vault-tubes'(van Zon et al. 2006; van Zon et al. 2003). Though a number of studies have found vaults to co-localize with actin, a direct interaction between the two molecules has not been described (Kedersha and Rome 1990; Herrmann et al. 1999; Slesina et al. 2006).

In spite of their ubiquitous cytoplasmic distribution, vaults have been found to particularly accumulate along filamentous actin-rich lamellipodia of spreading fibroblasts, stress fibers and cell adhesion sites, neuritic tips,

presynaptic compartments, lipid rafts and along growth cones, in response to various cellular signals by virtue of their cytoplasmic transport mechanisms (Kedersha and Rome 1990; Herrmann et al. 1999; Herrmann et al. 1996; Kowalski et al. 2007).

The observation of vault's hollow structure, transported along the microtubule over long distances within the cells, provided an intriguing possibility for function. In a "form follows function" approach, it was hypothesized that vaults acted as molecular cargo-carriers. Vaults began to be regarded as molecular cargo-carriers that shuttle proteins or other molecules within cells. Some assumptions about vault functions are indeed based on the cargo-carrier hypothesis. Are vaults really involved in transporting cargo? Do they have other significant cellular roles to play? Are they really important to cell viability? In the next section, hypotheses about vault functions and their experimental support are discussed.

1.2.1 Do Vaults Mediate Drug Efflux and Multidrug Resistance?

Apart from the multidrug resistance protein (*MRP*) and P-glycoprotein (*MDR1*), chemotherapy resistant tumor cells were also found to frequently overexpress a lung resistance-related protein *LRP*, the expression of which was believed to serve as a marker to predict response in acute myeloid leukemia and ovarian cancer. *LRP*is in fact *MVP*(Scheffer et al. 1995). The observation that *MVP* was overexpressed in a number of cancer models, particularly those displaying multidrug resistance, and that it mapped proximal to *MRP* positioned at chromosome 16p13.1, raised questions concerning the role of vault expression in drug resistance (Slovak et al. 1995; Kickhoefer et

al. 1998; Steiner et al. 2006; Rimsza et al. 1999; Sasaki et al. 2002; Izquierdo et al. 1998; den Boer et al. 1998).

Vaults have been known to be highly expressed in metabolically active cells including tumor cells, epithelial cells of lungs and digestive tracts, regenerating cells, macrophages and dendritic cells and also developing embryo (Hart et al. 1997; Berger et al. 2001; Izquierdo et al. 1996; Pan et al. 2013; Schroeijers et al. 2002; Stewart et al. 2005; Sutovsky et al. 2005). The high expression of *MVP* in the above-mentioned tissues, which are frequently prone to attack by xenobiotics, suggested that vaults may be involved in protecting the cells from toxic compounds. Elevated levels of *MVP* persist in tumor cells resistant to a wide spectrum of chemotherapeutic drugs (Berger et al. 2008). HDAC inhibitor driven overexpression of *MVP* in colon carcinoma cells was reported to correlate with less sensitivity of cells to doxorubicin, etoposide, vincristine and paclitaxel, thus associating *MVP* with drug sequestration and efflux (Kitazono et al. 1999). It was pointed out that doxorubicin, which is known to accumulate in nucleus, was rapidly effluxed in cells overexpressing *MVP* and that inhibition of *MVP* using ribozymes altered drug efflux.

Along with *MVP*, overexpression of *vRNA*, particularly *hvg1*, in glioblastoma and leukemia derived multidrug resistant cell lines and increased incidences of carcinogen induced tumors in *VPARP* deficient mice seemed to convince researchers all the more about mediation of drug resistance by vaults (Gopinath et al. 2005; Raval-Fernandes et al. 2005). Increase in levels of vaults and vault-associated proteins were constantly being associated with non-P-glycoprotein mediated multidrug resistance. The theory of ‘vault-

mediated drug efflux' became quite popular and prompted several clinical studies to correlate *MVP* expression to observed drug resistance, heralding vaults as a promising prognostic marker for chemotherapy outcome.

While it seemed convincing that vaults are involved in multidrug resistance and function as drug sequesters, Mossink *et al.*showed that vaults are not involved in mediating resistance to cytostatic agents (Mossink et al. 2002). They found that the embryonic stem cells and bone marrow cells derived from *MVP*-knockout mice, which are viable, healthy and display no obvious abnormalities, did not reveal any significant increase in sensitivity to various cytostatic drugs they were exposed to, compared to wild-type control cells. The actions of other multidrug resistance mediators, *MDR1* and *MRP*, were also found to be unaffected on *MVP* depletion, ruling out notions that they function as redundant counterparts. While *MVP* knockdown by siRNA in human bladder cancer cells was shown to result in increased doxorubicin sensitivity and nuclear accumulation, and hence increased cytotoxicity, no difference was found between *MVP*-knockout mice (*MVP*-/-) and wild type with regard to anthracyline doxorubicin response (Herlevsen et al. 2007; Mossink et al. 2002). These contradictory results cannot be attributed merely to the difference in the *MVP* gene silencing methodologies as lung carcinoma cells subjected to siRNA mediated *MVP* knockdown neither altered intracellular localization of doxorubicin nor showed increase in drug sensitivity or drug export efficacy (Huffman and Corey 2005).

While knockdown/knockout experiments provided controversial results, several other experiments focused on studying whether *MVP* could confer improved resistance to drug-sensitive cells. Drug sensitive SW1573

cells overexpressing *MVP* (fourfold increase in *MVP* levels) display similar daunorubicin efflux kinetics as SW1573 cells not overexpressing *MVP* (van Zon et al. 2004). This suggested that increased levels of *MVP* did not confer any resistance to the drug sensitive cell line. Also, the elevated levels of *MVP* did not result in vault accumulation near daunorubicin-filled vesicles at nuclear membranes challenging views that vaults were engaged as drug shuttles in pumping out drugs from nucleus. Drug sensitive HeLa cells, induced for *MVP* overexpression, also did not show any improvements in resistance to the drug and displayed similar drug efflux rates with respect to doxorubicin differently (Huffman and Corey 2005). Even reintroduction of *MVP-GFP* fusion protein into Mouse Embryonic Fibroblasts (MEF) derived from $MVP^{-/-}$ mice (MEFs from $MVP^{-/-}$ mice showed no changes in drug efflux rates compared to MEFs from $MVP^{+/+}$ mice) did not confer any additional resistance to daunorubicin compared to $MVP^{-/-}$ MEFs, strongly suggesting that vaults do not mediate efflux or sequestration of drugs (van Zon et al. 2004).

With contradictory data regarding vault function in mediating drug resistance, there may be a simple guilt-by-association assumption underlying this assignment of function. Metabolically active cells, more often in a state of stress or undergoing rapid proliferation, have been known to overexpress vaults. Thus, it should be remembered that tumor cells in a broad sense, not only multidrug resistant cells, exhibit elevated *MVP* levels. Unlike the functions of other known multidrug resistance mediators, no molecular mechanisms explaining how vaults may mediate drug efflux or resistance have survived experimental scrutiny. There has been no study yet that

unequivocally confirms drug binding to or being encaged within vault complexes. Negative results in this regard seem to have falsified the vault drug efflux hypothesis. The high expression of vaults may merely reflect a subsequent consequence of multidrug resistance but may not be the actual causative factor.

1.2.2 Does Vault Shuttle Cargo In and Out of Nucleus?

A subset of vaults, about 5% or less, has been reported to associate with the nucleus. In sea urchins, vaults were found to be concentrated around the nucelolus and around the nuclear envelope regions (Hamill and Suprenant 1997). Localization of vaults along the nuclear envelope was also observed in numerous tumor cells and also in cortical neurons (Kickhoefer et al. 1998; Slesina et al. 2005; Paspalas et al. 2009). The nuclear envelope is marked by macromolecular pores formed by the nuclear pore complex (NPC) that responds to changes in cisternal calcium levels. Since, the size and symmetry of vaults were comparable to what might actually constitute the central plug of the NPC, it was proposed that vaults may represent a part of the central mass (Chugani et al. 1993). Except for a change in FRET signal on varying cisternal calcium, no direct evidence supports the notion of vaults constituting the central plug (Dickenson et al. 2007). Instead, vaults suggested to play a role in the biogenesis of NPCs, thus explaining their association with nuclear envelope (Vollmar et al. 2009).

Vaults have been found to associate with cytoskeletal elements; in particular, they exhibit interactions with microtubules along their caps (Kedersha and Rome 1990; Eichenmüller et al. 2003). They have also been

found to be enriched within cholinergic nerve terminals and exhibit anterograde and retrograde transport within axons of the electromotor neurons (Herrmann et al. 1996; Li et al. 1999). The claims of a molecular cargo-carrier like function (Szaflarski et al. 2011) combined with their juxtaposition to the exterior of the nucleus, led to the hypothesis that vaults may participate in nucleocytoplasmic transport and mediate shuttling of proteins and mRNAs between the nucleus and cytoplasm (Hamill and Suprenant 1997). Yet apart from the distinct vault-associated molecules *MVP*, *VPARP*, *TEP1* and *vRNA*, a wider range of cargo has eluded researchers thus far. Co-purification of *MVP* along with ribosomes in sea urchin embryos was suggestive of vaults functioning as carriers for ribosomes (Hamill and Suprenant 1997). However, the interaction was weak and did not remain intact after vault purification, and vaults do not enclose ribosomes. Despite vaults being implicated in the biogenesis and assembly of NPCs, mouse embryonic fibroblasts lacking *MVP* genes displayed no significant changes in nuclear import/export kinetics compared to control cells expressing *MVP*, suggesting no roles for vaults in nuclear trafficking (van Zon et al. 2006). The structure and mechanism of the nuclear pore is now well understood and requires no vaults to function, thus this function appears to have been also falsified.

1.2.3 Are Vaults Important for Immune Responses?

Dendritic cells (DCs) are an integral part of the immune system and function at the interface of innate and adaptive immunity. Functioning as antigen-presenting cells, they are ubiquitously found in tissues that are constantly exposed to external environment and play key roles in initiating

primary immune responses. Immature DCs process antigens and on maturation stimulate antigen specific T cells, eliciting an adaptive immune response. A role for vaults in dendritic cell survival was postulated based on clear upregulation of vaults during maturation of dendritic cells derived from varied sources including blood monocytes, CD34+ mononuclear cells or chronic myeloid leukemia (Schroeijers et al. 2002). Based on overexpression of *MVP* in reactive monocytosis and chronic myelomonocytic leukemia, *MVP* as a putative marker of monocytic lineage was also proposed (Sunnaram et al. 2003). Addition of antibodies against vaults in DC cultures severely challenged cell survival and resulted in reduced viability of LPS- or TNF-alpha–matured DCs, underscoring the importance of vaults in immune response. Blockade of *MVP* also seemed to take a toll on expression of other critical differentiation and maturation markers including CD1a and CD83 and was reported to affect induction of T cell proliferation and subsequent interferon-gamma release (activators of macrophages) from T cells.

Doubts about this role of vault in dendritic cell maturation arose when *MVP* knockout mice failed to reveal any DC-related immune impairment. The *MVP* deficient DCs showed no signs of altered surface marker expressions and displayed normal development (Mossink et al. 2003). The antigen uptake, processing and maturation of DCs from *MVP* deficient mice were comparable to those of control mice. Knockout mice are capable of inducing T cell proliferation and also efficiently elicit T cell mediated immune responses or T cell dependent humoral response, when challenged with varied T cell antigens, suggesting that DC maturation or their migration in vivo does not depend on *MVP* or vaults (Mossink et al. 2003). With no role for vaults in DC

development, the underlying reason for their high expression in these cells remains unanswered.

MVP is however clearly an interferon-gamma inducible gene that involves JAK/STAT pathway based interaction of STAT1 within GAS elements in proximal regions of *MVP* promoter. Interestingly, interferon-gamma induction not only led to efficient transcription and translation of *MVP* but also decreased the stability of vaults (Steiner et al. 2006). Of note, overexpression of *MVP* profoundly down regulates expression of interferon-gamma induced expression of ICAM-1, the receptor on respiratory epithelial cells responsible for entry of infectious human rhinovirus, pressing on a role for *MVP* in mediating infection resistance.

Another line of evidence comes from *MVP* expression suppressing hepatitis C virus (HCV) replication and protein synthesis by inducing type-I interferon expression via translocation of interferon regulatory factor IRF7 and NF-κB into the nucleus. It has been found that *MVP* expression is also effectively driven by other viruses, including the Influenza A virus and vesicular stomatitis virus. Apart from *MVP*, intense upregulation of *vRNA*s up to a 1000 fold on infections by Epstein-Barr virus or Kaposi's sarcoma virus in human lymphocytes stressed that vaults as an entire structure effectively mediate antiviral response. Thus, vaults are regarded as a virus induced host factor that somehow mediates antiviral responses.

A profound role for *MVP* in reducing bacterial-burden and epithelial cell-mediated resistance to infection by *Pseudomonas aeruginosa* arises from studies conducted on *MVP*-deficient mice. *MVP* knockout mice appeared to display a threefold increase in mortality attributed to reduced bacterial

internalization and increased bacterial accumulation in lung. With the aid of cystic fibrosis transmembrane conductance regulator (CFTR), *MVP* mediates pathogen clearance by being rapidly recruited to lipid rafts.

The evolutionary recruitment of vault binding proteins on the surface of the invasive bacteria *Listeria monocytogenes* leads to the accumulation of intact vaults on its surface, which helps these bacteria to evade autophagic recognition. The surface protein internalin InlK binds to vaults and mediates an evasion process of host defenses. This happens without the involvement of ActA (Actin assembly-inducing protein), the primary surface protein that hijacks actin assembly, propelling the bacteria through the mammalian host.

From these lines of experimental evidence, an involvement of vaults in mediating infection resistance is apparent, yet mechanisms remain ambiguous.

1.2.4 Are Vaults Major Players in Signaling Cascades?

The increased expression of *MVP* in various tumor lines and its unrestrained movement within the cytoplasm implied that *MVP* could potentially be involved in cellular signaling related to growth and proliferation. Accordingly, it was reported that *MVP* could possibly regulate the epidermal growth factor (EGFR)-induced MAPK pathway. It has been found that *MVP* could be tyrosine phosphorylated and hence, establish interaction with many Src homology2 (SH2) domain-containing tyrosine phosphates including SHP-2 and Src kinase (Kolli et al. 2004; Kim et al. 2006). It is interesting to note that SHP-2, a signal enhancer of EGF, plays crucial roles in embryonic development and directly mediates neural stem/progenitor cell differentiation and proliferation (Qu et al. 1999; Ke et al. 2007). In response to EGF, tyrosyl-phosphorylated *MVP* also associates with

extracellular-regulated kinases (Erks) and the SHP-2 to form a protein complex, suggestive of a cellular scaffold-like function for *MVP* in Erk pathway. However, on EGF stimulation, *MVP*-deficient MEFs displayed no significant change in overall activation of Erks except for a delayed Erk activation response. (Kolli et al. 2004)The transactivation of Elk1, another molecule positively regulated by SHP-2 on EGF stimulation, was found to be reduced by 50% in *MVP*-deficient MEFs compared to those of wild-type MEFs. Based on these observations, *MVP* was believed to modulate or fine-tune the Erk pathway at the level of Ras or even downstream, rather than being a specific activator and/or inhibitor.

MVP has also been found to interact with Src in an EGF-mediated manner (Kim et al. 2006). Src is an important signaling molecule and proto-oncogene involved in malignant transformation, controlling various aspects of cell development, migration, synaptic transmission and plasticity, immune response and cell adhesion (Parsons and Parsons 2004). It has been found that Src activity and *MVP* phosphorylation are both important in mediating this interaction. Yet contrary to the observed activation of MAPK pathway by *MVP* interacting with SHP-2 and Erk, *MVP* overexpression was found to reduce EGF-dependent Erk activation in Src overexpressing cells. Purified *MVP* was also reported to quench tyrosine kinase activity of Src in vitro. The impeding effect of *MVP* on MAPK pathway has also been emphasized by another study that suggests that *MVP* interacts with and inhibits YPEL4's ability to activate Elk-1 (Liang et al. 2010). Of note, YPEL4 a nuclear protein that localizes to the centrosome and nucleolus, is associated with functions related to cell division events (Hosono et al. 2004).

The tumor suppressor gene, p53, has been known to negatively regulate the expression of *MVP*, both at the mRNA and protein level (Marroni et al. 2003). Interestingly, another study suggested that *MVP* is up regulated in p53 overexpressing young human diploid fibroblasts (HDFs), contrary to the belief that *MVP* overexpression arose from defects in p53-related suppression of *MVP*(An et al. 2009). Senescent HDFs and aged organs also display increased levels of *MVP* and *MVP* has been believed to mediate resistance to apoptosis by regulating expression of anti-apoptotic Bcl-2 via the JNK pathway (Ryu et al. 2008; Ryu and Park 2009). A role for *MVP* in modulating anti-stress response has also been underscored by suppression of c-Jun mediated AP-1 transcriptional activity, possibly via the interaction of *MVP* with Constitutively Photomorphogenic 1 (COP1) (Yi et al. 2005). Interestingly, COP1, an E3 ubiquitin ligase, is also an important negative regulator of p53 (Dornan et al. 2004). *MVP*-deficient cells displayed elevated levels of c-Jun and AP-1 transcription activity, highlighting the importance of *MVP* in cell proliferation, differentiation and apoptosis (Yi et al. 2005). However, it should be noted that while COP1 is an essential gene for plant and regulates photomorphogenic development, there has been no report of vaults in plants.

It is obvious that *MVP* is being implicated in a wide variety of signal transduction roles by virtue of its association with several key players in the signaling cascades. Interestingly, most of the reported interactions support functions relating to cell proliferation, survival and tumor progression, including association of vaults with PTEN (Yu et al. 2002). An important role for *MVP* in signal transduction pertaining to immune responses, particularly in

modulating interferon-gamma mediated JAK/STAT pathway, was suggested when *MVP*-induced attenuation of STAT1 phosphorylation and subsequent STAT1 translocation to the nucleus was observed (Steiner et al. 2006). The ubiquitous expression of *MVP* and the unrestrained movement of vaults through the cytoplasm gave credence to the notion that vaults could effectively function in cellular signaling. Though, the interactions of *MVP* with various key players are being suggested, it should be remembered that purified vault preparations do not reflect the presence of any of these additional interacting players. The observed *MVP* interactions with various signaling protein/complexes, in most cases, have been identified primarily via yeast two-hybrid assay or immunoprecipitation experiments and hence, may involve association of individual *MVP* monomers and not the complex as a whole. Even if intact vault complexes could possibly mediate such associations, the transient interactions, most likely, should occur at the periphery and not within the complex. While there may be a role for vault in signaling, it may be late evolved functionality arising from non-specific phosphorylation and not the primary function of the entire complex.

1.3 'PRECISE ROLES' – DOES VAULT HAVE ANY?

Ever since the vault complex was discovered in 1986, various functional roles have been attributed to the vault complex. But when challenged to describing a clear-cut functional and mechanistic role, the answer remains elusive. Some functions ascribed to vaults are based on appearance and guilt-by-association, and have not been substantiated by experimental evidence, and are often confounded by contradictory experimental evidences. Most of the observations made *in vitro* could not be

recapitulated *in vivo*. *MVP*-knockout mice appear completely normal with no defects in development, suggesting it is dispensable.

While the barrel-shaped structure with hollow interiors emphasize a molecular cargo-carrier-like function, vaults purified from a wide variety of eukaryotes reveal no additional cargo other than the conventional vault-associated proteins. If cells depended on vaults for shuttling important cargo like ribosomes or transcription factors through various regions within the cells, particularly into the nucleus, the *MVP*-knockout mice would have displayed obvious phenotypes pertaining to failures in transcription or translation machinery.

The clearest phenotypes observed, are those pertaining to cell survival during nutritional stress and mediation of antiviral or antibacterial responses during infection resistance. Disruption of vaults in *Dictyostelium* displayed no defects under normal nutrient conditions, but showed reduced survival rates on nutritional stress (Vasu and Rome 1995). Similar observations have been made on MEFs lacking vaults and subjected to growth factor deprivation (Kolli et al. 2004). It should be noted that all cells that display elevated expressions of *MVP* both at the transcript and protein levels are those that are metabolically active and undergo rapid proliferation, including transformed cells, neuronal cells, regenerating cells, developing oocytes or embryos, stressing on the possible roles for vaults pertaining to cellular metabolism and survival (Li et al. 1999; Hamill and Suprenant 1997; Sutovsky et al. 2005; Yoshinari et al. 2009; Izquierdo et al. 1996; Berger et al. 2001; Rao et al. 2009).

1.4 OBJECTIVES AND SCOPE OF THIS WORK

While vaults are dispensable for normal development in eukaryotes, a vast amount of cellular resources are invested in translating and constructing a molecular complex of this size. If vaults served no purpose in eukaryotic cellular physiology, they should have been lost through years of evolution. The fact that they have been conserved and are present in a wide range of eukaryotes underscores the importance of vaults as cellular assets that engage in functional roles in a not so evident manner.

The main objective of this thesis is to find the evolutionarily relevant, ancient and perhaps original cellular functional role for the vault complex. With the exception of slime-molds, it should be noted that a majority of vault characterization studies focus on multicellular eukaryotes. Although homologs of *MVP* have been identified in other single-celled eukaryotes including *Paramecium* and Kinetoplastids, there has been no study on elucidation of vaults in such organisms. The aims of this thesis are described as follows.

i. Chart the evolutionary history of the vault complex to understand its origin and identify common traits unifying all organisms harboring vault genes.
ii. Use phylogenetic information and compositional data of vault sequences to arrive at a putative novel cellular function for vaults and re-analyze the experimental literature in light of the proposed function.
iii. Characterize vaults in the evolutionarily distant single-celled eukaryote *Trypanosoma brucei* to gain insights onto vault origin in Kinetoplastids and shed light on its possible ancestral functions.

Chapter 2

Unraveling the Evolutionary History of the Vault Complex

2.1 INTRODUCTION

Vaults have been isolated from numerous eukaryotic species including mammals, bullfrog, chicken and the slime mold *Dictyostelium* (Kedersha et al. 1990).With advances in genomic and proteomic technologies, *MVP* has been described in many invertebrates including tubeworms and molluscs (Sanchez et al. 2007; Dondero et al. 2006). The absence of vaults in other prominent experimental organisms including *Caenorhabditis elegans, Drosophila melanogaster, Saccharomyces cerevisiae* and *Arabidopsis thaliana* has confounded researchers in understanding its functional evolution and its selective exclusion from certain lineages. Understanding the evolutionary origin of vault and its components may provide insights into their puzzling phylogenetic distribution and also help identify common traits between all organisms harboring vault genes, thus hinting on its primary function. In this

section, comparative genomics and phylogenetics-based approach have been used to understand the distribution of vault genes across all phyletic groups and also to reconstruct the evolutionary history of the vault complex.

2.2 MATERIALS AND METHODS

2.2.1 Sequence retrieval

The human *MVP* sequence (Accession No. NP_059447.2) of length 893 amino acids, was used as a query in a protein BLAST search to identify homologs in the GenBank non-redundant protein database. The BLAST hits were manually assessed and sequences smaller than 600 amino acids were not included for phylogenetic analysis. In addition, pre-computed sequence alignment list for the query sequence generated by an all-against-all BLAST search (BLINK) was also inspected to screen for additional sequences across the taxonomy. Position-specific-iterated BLAST (PSI-BLAST) was performed to improve sensitivity and capture potential orthologs from distant evolutionary groups. *MVP* homolog of the cyanobacteria *Moorea producens 3L* (Accession ZP_08432100.1) annotated as 'membrane protein, colicin uptake' showed high levels of similarity with *MVP* sequences from molds. The gene mapped to the *Lyngbya majuscula 3L* genomic scaffold scf52120 (*Moorea producens* was misclassified as *Lygnbya majuscula* previously). The sequence was retained for further analysis since the presence of the expressed protein was confirmed by MudPIT analysis. The protein annotated as *MVP* from *Hordeum vulgare* (barley) was dismissed from the dataset since the predicted protein, with accession BAK00750, was derived from the complete CDS (coding sequences) clone and the gene could not be located on the physical map of the barley genome assembly. The protein was excluded from

the analysis suspecting possible contamination from mold in the same environment.

In the compiled list of homologous proteins, sequences displaying duplicated copies within the same species were encountered. In such cases, the genome locus information was extracted for each of the proteins to confirm the presence of multiple genes transcribing similar proteins. If two proteins (isoforms or redundant proteins) have the same chromosomal information (or scaffold information), only one was retained for subsequent analysis. This resulted in a refined dataset of 109 full-length proteins.

A similar search was also conducted for human *TEP1* (Accession No. NP_009041) and human *VPARP* (NP_006428), comprising the minor vault proteins, using the human sequences as query. Redundant sequences and proteins that map to same genomic loci were discarded.

2.2.2 Sequence alignment and phylogenetic analysis

The sequences were aligned using the MUSCLE multiple sequence alignment algorithm (Edgar 2004). An initial alignment was constructed to remove truncated or poorly aligned proteins. For further refinement and subsequent phylogenetic analysis, the alignment was manually edited using MEGA5 (Tamura et al. 2011). Poorly conserved regions, which could obscure phylogenetic information, were deleted. The process of editing and realigning was repeated several times until large gaps were removed. Many regions in sequences of basal deuterostomes and kinetoplastids were not information rich and were based on predictions from genomic data. These possibly included non-coding regions or misidentified exon boundaries. Such blocks of

ambiguous regions were removed and not used for subsequent phylogenetic analysis.

The phylogenetic analysis was performed using MEGA5 package. The refined alignment was used to identify the amino acid substitution model that fits the data best. It is known that the Bayesian information criterion (BIC) has higher accuracy and precision compared to the Akaike information criterion (AIC) for appropriate model selection (Luo et al. 2010). Hence, the model with the lowest BIC score was used for phylogenetic reconstruction. The evolutionary history was inferred using the Maximum Likelihood method (ML). To account for variable substitution rate among sites, the gamma distribution was defined using 6 substitution rate categories and the tree was built using the best amino acid substitution model. Positions containing gaps and missing data were ignored for tree construction. The reliability of the tree was tested using 100 bootstrap replicates. The ML tree was established using a heuristic search based on nearest neighbor interchange. The Neighbor Joining (NJ) analysis based on equal input distances with 1000 bootstrap replicates was also performed in indicated datasets in addition to the ML inferences. In most cases both ML and NJ analyses gave similar topologies and hence, in the interest of clarity, only the topology obtained using ML will be presented.

2.2.3 Essential amino acid analysis

The pathways corresponding to synthesis and metabolism of common essential amino acids – Leu, Val, Ile, His, Trp, Met, Lys, Thr, Phe and conditionally essential amino acids Pro and Arg in the Kyoto Encyclopedia of Genes and Genomes (KEGG) were used for the analysis. The biosynthetic pathways for the above-mentioned amino acids were analyzed in the

organisms included for phylogenetic analysis for their completeness. This was done by manually confirming the presence of all the enzymes required for its synthesis. For other organisms of interest including *P. pacifica, M. marina* and *M. producens 3L* for which data was not available on KEGG, the sequences of all enzymes involved in the amino acid biosynthesis pathways of closely related organismswere retrieved. The sequences of all enzymes obtained from *Sorangium cellulosum* were used to seed for enzyme orthologs in *P. pacifica* and the sequences obtained from *Trichodesmium erythraeum* were used to identify enzyme orthologs from *M. marina* and *M. producens 3L.* The BLAST hits thus obtained were manually screened for protein coverage, identity and e-value to check for the presence of true enzyme orthologs. Amino acids, for which the organism retained enzyme orthologs required for synthesis, were considered non-essential. If the identified orthologs were found insufficient to make an intact synthetic pathway, the amino acid was considered essential in that organism. In cases where there are multiple or alternate pathways to synthesize an amino acid, all the subsidiary pathways were also considered. As a control, analysis was made on the eukaryote, *Saccharomyces cerevisiae,* which can synthesize all of the above-mentioned amino acids.For the purpose of analysis, protists including *Entamoeba histolytica, Toxoplasma gondii*, *Tetrahymenathermophil* and *Trichomonas vaginalis* that do not harbor vault genes were also included.

2.3 RESULTS

2.3.1 Unique Phylogenetic Distribution of *MVP*

To reconstruct the evolutionary history of the vault complex, proteins homologous to the human *MVP* were extracted from all available taxonomic

groups and manually screened to include highly similar and full-length proteins. The initial dataset of 132 sequences comprised of more than one sequence from certain organisms. Some of these were redundant sequences, while others were paralogs that arose out of a gene duplication event. To identify the paralogs from the redundant sequences, a multiple sequence alignment was built from the initial dataset followed by constructing a neighbor joining tree, hence pruning the dataset to 109 sequences. The paralogs were also confirmed by mapping onto genomic loci whenever data was available.

A Maximum Likelihood (ML) tree was generated for *MVP* proteins across all the lineages (Figure 2.1). The unrooted phylogenetic tree thus obtained suggested that there is but one autotroph with vault gene, a filamentous cyanobacterium *Moorea producens 3L*, and all sequenced plant, algae, fungi and archaea lack vault genes. Xenologs of *MVP* were also found in certain species of heterotrophic bacteria with gliding motility. The appearance of vault like sequences in discrete bacterial species is suggestive of independent horizontal gene transfer events. Single copies of the *MVP* gene were observed in all chordates including the invertebrate tunicates (*Oikopluera dioica* and *Ciona intestinalis*) and cephalochordates represented by lancelet *Branchiostoma floridae*. The heterotrophic single celled eukaryotes, choanoflagellates, considered to be the closest relatives of animals and represented by *Monosiga brevicollis* and *Salpingoeca sp.* also were found to have only one *MVP* homolog(King et al. 2008). *Capsaspora owczarzaki,* an amoeboid symbiont and a putative sister-organism to metazoans and choanoflagellates, of class Filasterea has two homologs of*MVP*(Ruiz-Trillo et

al. 2006).The Cnidarians represented by *Nematostella vectensis* and *Hydra magnipapillata* and the ambulacrarians comprising the hemichordate *Saccoglossus kowalevskii* and the echinoderm *Strongylocentrotus purpuratus* have at least two copies of the *MVP* gene. The parazoan sponge, *Amphimedon queenslandica*, has multiple paralogs while the simplest multicellular metazoan and the only representative of the phylum placozoa *Trichoplax adherans* has one copy of the vault gene.

Among the protists, the ciliate represented by *Oxytricha trifallax* and the amoebozoan slime molds *Dictyostelium discoideium* and *Polysphondylium pallidum* have two *MVP* paralogs that clearly separate out in the unrooted tree. Another ciliate, *Paramecium tetraurelia* seems to have undergone two rounds of duplication, which is evident from the branching within each paralog node. The paralogous proteins arise from different genomic scaffolds and hence are considered products of gene duplication events. As mentioned earlier, in the kinetoplastids represented by genus *Trypanosoma* and *Leishmania*, two of the three homologs (*MVP2* and *MVP3*) have longer branches (rapidly evolving) and in the unrooted tree they cluster along with the ciliates, while *MVP1* clearly separates out.

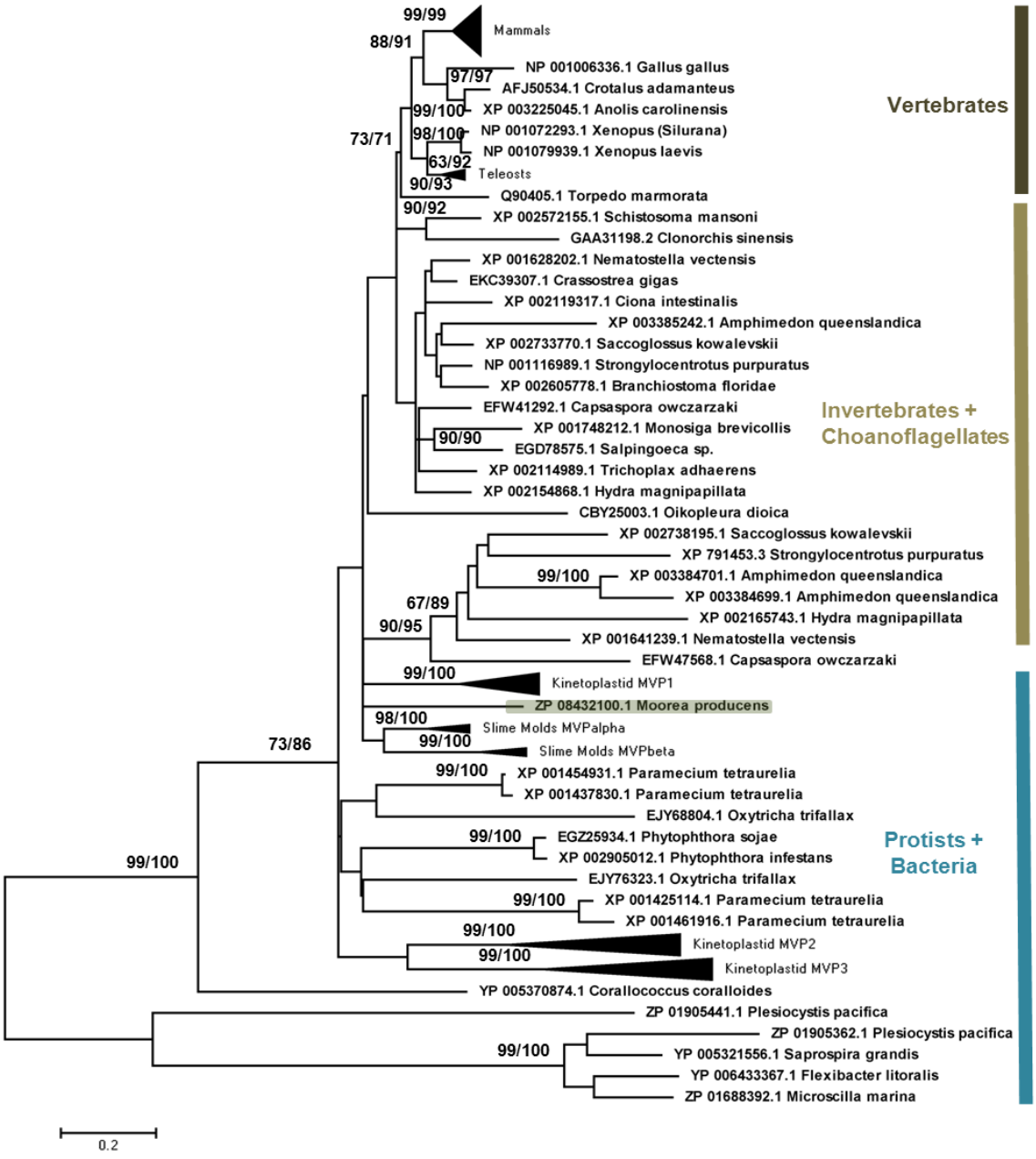


Figure 2.1 Phylogeny of all retrieved *MVP* homologs across the taxa
The shown topology is a bootstrap consensus tree obtained by maximum likelihood method based on the WAG model. Collapsed branches represent those that were produced in less than 50% bootstrap replicates. In this and subsequent phylogeny figures, the node support values are given in the order of ML and NJ bootstrap percentages unless otherwise stated. Values are given for all nodes supported by >70% bootstrap support in either of the methods. The branch length is directly proportional to the number of substitutions per site. Sequences from 23 organisms are clustered into the "Mammals" clade. The "Kinetoplastid" clade represent sequences from both *Trypanosoma* and *Leishmania* and the "Slime Molds" represent sequences from both *Dictyostelium* and *Polysphondylium*. The cyanobacterial *MVP* xenolog from *Moorea producens* is highlighted in green. The tree is modeled with sequences from diverse evolutionary groups evolving at different rates and hence, suffers from long-branch attraction.

2.3.1.1 *MVP* in a Non-Nitrogen-Fixing Cyanobacterium

A well-conserved *MVP* gene is found in the very recent genome sequence of the marine filamentous cyanobacterium *Moorea producens 3L* (previously misidentified as *Lyngbya majuscula 3L* due to morphological similarities) (Engene et al. 2012). The filamentous cyanobacterium forms sheaths composed of exopolysaccharide and harbors heterotrophic bacteria and other microbes on its exteriors. The *MVP* gene was not only annotated at the genome level, but its protein product was confirmed to be abundant by mass spectrometry (Jones et al. 2011). It is an ancient vault gene similar to the gene in molds and ciliates, and close to where one might anticipate the root of the *MVP* gene tree to be placed. It is worth noting that *Moorea producens 3L* has lost genes involved in nitrogen fixation and was shown, using heavy nitrogen labeling, to compensate through intracellular protein recycling. Interestingly, there are no *MVP* genes in the nitrogen-fixing filamentous cyanobacterial genomes of *Nostoc punctiforme* and *Trichodesmium erythraeum* or even in the closely clustering genera *Coleofasciculus* and *Symploca* (Jones et al. 2011; Engene et al. 2012).In the following phylogenetic reconstructions, the cyanobacterial sequence is used to root the trees.

2.3.1.2 *MVP* Xenologs in Certain Gliding Heterotrophic Bacteria

The search for *MVP* homologs also retrieved a small group of heterotrophic bacteria that encode full-length proteins and align along vault proteins from eukaryotes. The bacterial species are obligate heterotrophs with gliding motility. Two of the species *Plesiocystis pacifica* and *Corallococcus coralloides* belong to the Myxococcales, a family of social eubacterial predators that transition from single-celled to multicellular fruiting bodies

upon starvation, in a mechanism analogous to *Dictyostelium discoideium* (Iizuka et al. 2003; Huntley et al. 2012). *P. pacifica* has two *MVP* paralogs while *C. coralloides* has only one homolog. Vault coding genes, however, are absent from the complete genomes of other myxobacteria including *Myxococcus xanthus* or *Sorangium cellulosum* (Schneiker et al. 2007).*Microscilla marina* and *Flexibacter litoralis*, which only has one copy of the gene, belong to Cytophagales that are marine unicellular gliders. The *Saprospira grandis*, a multicellular filamentous Sphingobacteriale has one vault homolog. It is interesting to note that this organism preys on other bacteria and protists since amino acid biosynthetic pathways for the essential amino acids are incomplete (Saw et al. 2012).

The topology of the ML tree constructed with only bacterial proteins clearly delineates the *C. coralloides* and *M. producens 3L* sequences from the rest of the sequences (Figure 2.2). While the rest of the proteins cluster into one clade, the paralog from *P. pacifica* clearly branches out. This may be an organism specific duplication event as no paralogs could be detected in other bacterial species.

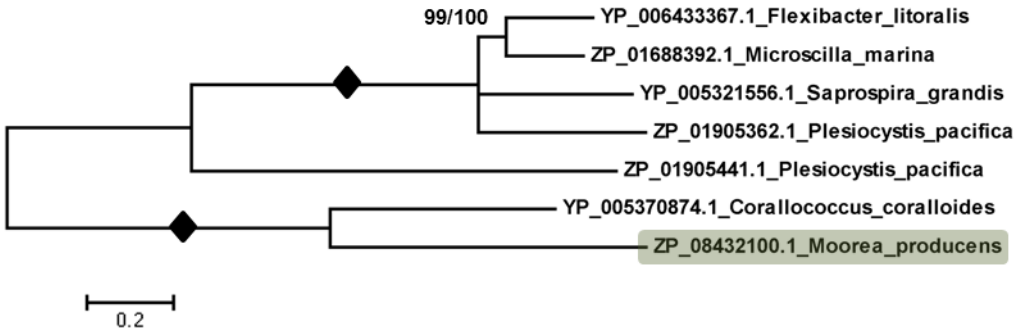


Figure 2.2 Phylogenetic relationships among all the bacterial *MVP* xenologs
Shown is the unrooted ML bootstrap consensus topology obtained based on the WAG+G+F model of evolution. *C. coralloides* is evolutionarily closer to the cyanobacterial ortholog (indicated in green) than to the rest of the bacterial sequences that seemed to have undergone substitutions over the period of evolution. The long branches and lack of large numbers of other related bacterial species with *MVP* genes suggests these xenologs correspond to independent horizontal transfer events, rather than common descent. Black diamond indicates 100% bootstrap support for the indicated node by both ML and NJ methods.

2.3.2 Evolutionary Origin of *MVP* and Independent Horizontal Gene Transfer Event into Eukaryotes

While the initial phylogenetic tree gives an overview on the distribution of vaults across the taxonomy, it does not clearly define the evolutionary origins of vaults. The evolutionary relationships among the unicellular protists are, not surprisingly, ambiguous due to crowding from paralogs that have particularly long branches. Intriguingly, looking at the span of *MVP* genes across the phyletic groups, it is clear that while *MVP* is predominantly represented in the Unikont clade that comprise the amoebozoans and the opisthokonts (metazoans and choanozoa), there seems to be a limited representation of organisms with *MVP* from the superphylum Excavata (kinetoplastids) and from the chromalveolate clade (*Phytophthora*) that also includes the alveolata (ciliates) (Figure 2.3 Top). A comprehensive search for the all homologs of *MVP* in the Excavates using BLAST or PSI-BLAST retrieved no other sequences except for the Kinetoplastids homologs

from phylum Euglenozoa and the closely related *Naegleria gruberi* homolog from phylum Percolozoa. Closely related groups including the Diplomonads or Jakobids belonging to the Excavates did not have any *MVP* homologs. Even organisms that are known to branch closely with the Kinetoplastids within the Euglenozoa phylum, *Diplonemapapillatum* or *Euglenagracilis*, did not have any putative *MVP* homologs (Simpson et al. 2006). Though a common ancestor of kinetoplastids had an *MVP* gene, it seems unlikely that it emerged in common ancestor of Euglenazoa.

The situation is quite similar with respect to occurrence of *MVP* in *Paramecium* or *Phytophthora.* Except for the ciliates *Paramecium* and *Oxytricha trifallax,* other close-branching alveolates like *Toxoplasma gondii, Tetrahymena thermophile* or *Plasmodium* do not encode any vault genes. Even within the heterokonts, apart from the homologs of *MVP* in genus *Phytophthora,* there seems to be no other organism carrying genes similar to *MVP*.

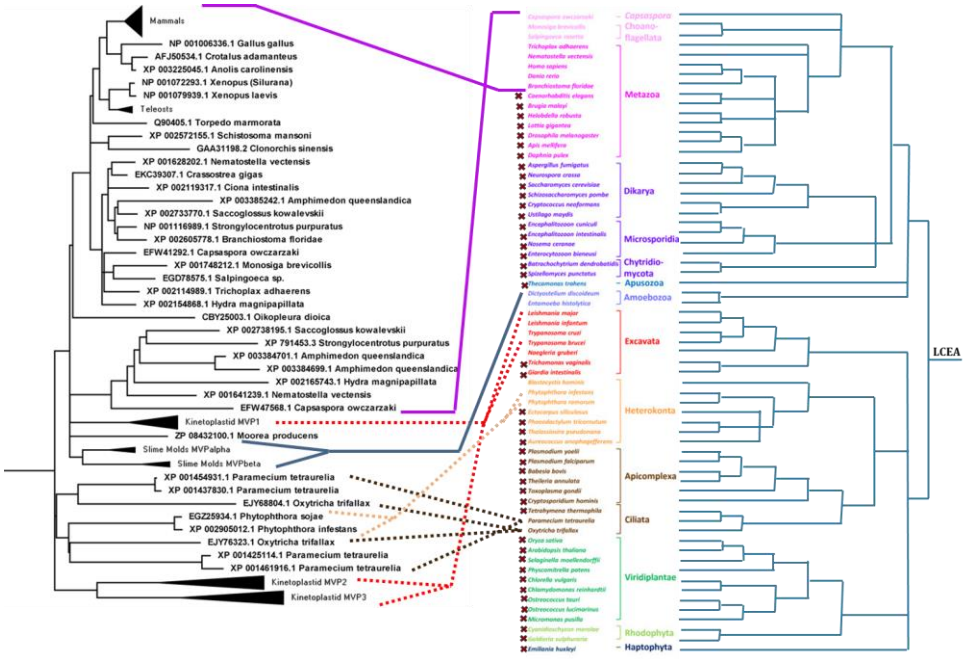
If *MVP* indeed arose in a single-celled eukaryote ancestral to both unikonts and bikonts, has the gene been specifically lost multiple times across many phyla over evolution so as to be underrepresented in only a sparse set of organisms in bikonts?

Given the large amount of sequence data now available, as EST datasets and complete genomes, a scenario of repeated gene loss seems unlikely. Instead, a parsimonious interpretation would suggest multiple horizontal gene transfer events into the bikonts from an ancestral unikont. This may be compared to similar gene transfer events that led to the occurrence of vault genes in certain specific groups of bacteria. Taking into account the

current distribution of *MVP* in sequenced genomes, the existence of vault genes in specific organisms of excavates, ciliates or heterokonts could be parsimoniously explained by only five independent ancestral horizontal gene transfer events - two into excavates and ciliates each and one into *Phytophthora* (Figure 2.3 bottom). Apart from these horizontal transfer events, secondary loss events at the level of fungi, Icthyosporea (a sister clade to metazoans, choanoflagellates and *Capsaspora*) and ecdysozoan protostomes would best describe the most parsimonious evolutionary scenario that led to the origin of vault and its distribution among eukaryotes. Thus it appears that multiple horizontal transfer and loss best explains the anomalous phylogenetic distribution of *MVP*.

The ancestral emergence of *MVP* genes in the unikont clade is shown in Figure 2.4 (top). The ancestral unikont could have acquired the gene by horizontal transfer from a cyanobacterium (*Moorea producens 3L)* that relied on intracellular amino acid storage capability to account for the loss of nitrogen fixation genes. Hence, the tree has been rooted with this sequence. The clear delineation of amoebozoans from the rest of the opisthokonts can be observed. This tree is in acceptance with well-established organismal relationships except for the placement of *Capsaspora* within the invertebrate clade. The invertebrates seem to have undergone lineage specific duplications and form long branching paralogs. The long branches that skew the other branches were removed in (Figure 2.4 bottom)

A



B

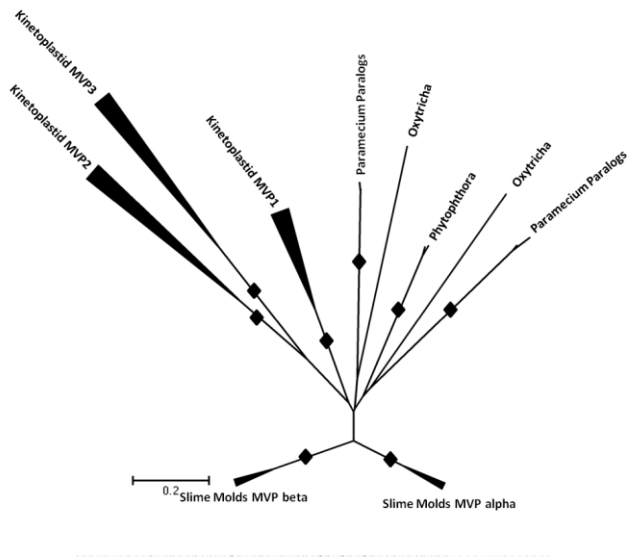


Figure 2.3 Limited representation of *MVP* genes in Protists

(A) Comparision of distribution of phylogenetic distribution of *MVP* genes as against a Reference phylogenetic tree (adapted from Eme et al. 2011) rooted between Unikonta and Bikonta showing the relationships between eleven major eukaryotic lineages for which genome sequencing are complete or nearly complete. Organisms with *MVP* genes in both trees are connected by coloured lines. Dotted lines represent organisms that possibly acquired *MVP* by horizontal gene transfer events. Red crosses in Reference Tree points to organisms that positively lack *MVP* genes. (B)Shown is the bootstrap consensus tree obtained by ML analysis based on rtREV+G+F model of evolution. The independent lineage-specific duplications that occurred in the protists are strongly supported. The slime molds that belong to unikonts display tightly supported branching. The bikont protists are marked by exceptionally long branching sequences suggesting considerable evolutionary changes. Black diamond indicates 100% bootstrap support for the indicated node by both ML and NJ methods.

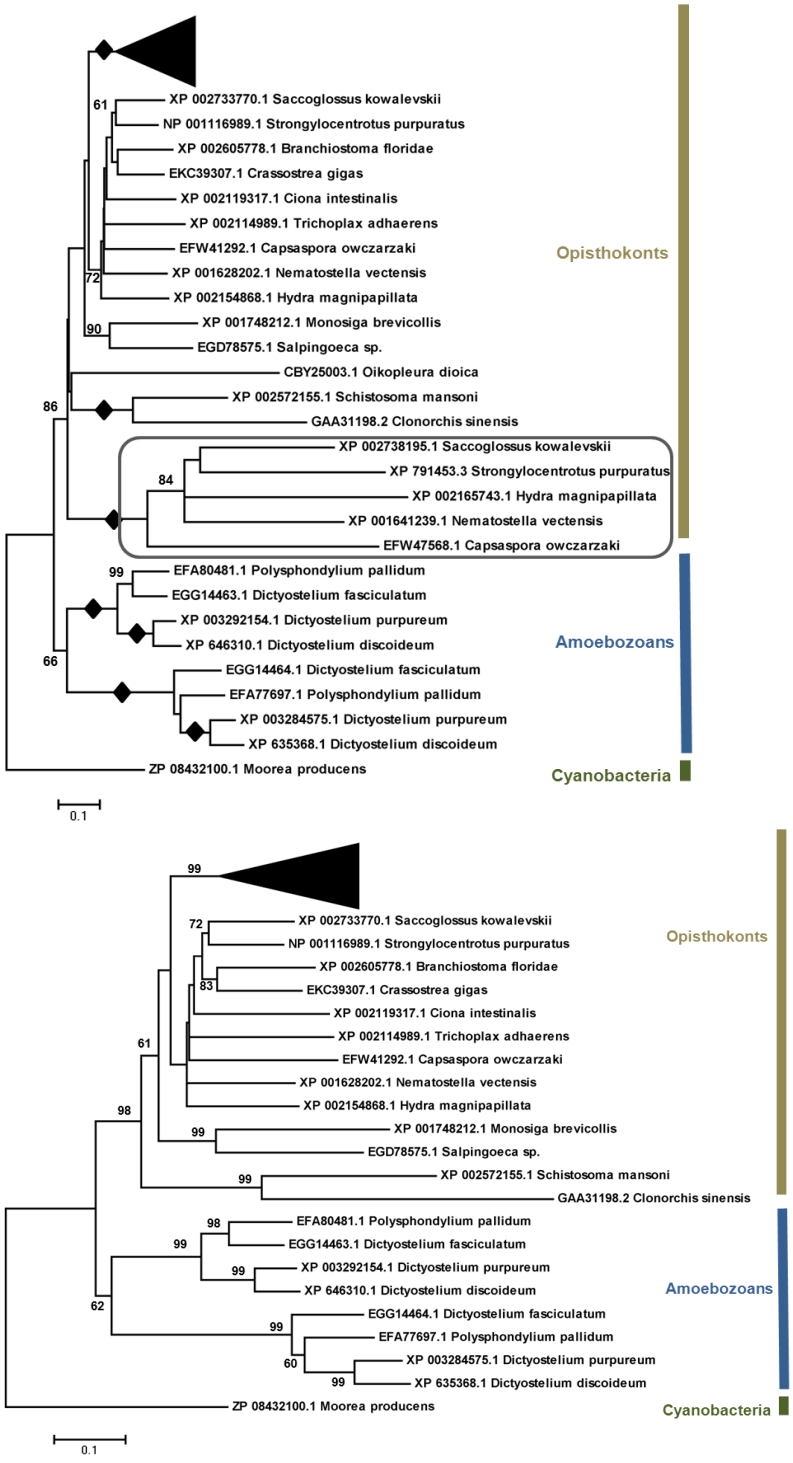


Figure 2.4 Evolutionary origin of *MVP*

The tree that likely represents the origin of *MVP* genes in eukaryotes. The bacterial xenologs and long branching protists sequence have been removed. The shown topology is a bootstrap consensus tree obtained by ML method based on the rtREV +G (top) and JTT+G (bottom) model of evolution rooted with cyanobacterial sequence. The long branching paralogs (gray box in top) have been removed from the topology in bottom. Values are given for all nodes supported by >60% bootstrap support. The branch length is directly proportional to the number of substitutions per site. The black triangle represents all the vertebrate sequences that are clustered into a single clade. Black diamond indicates 100% bootstrap support for the indicated.

Although horizontal gene transfer is a rare event in multicellular opisthokonta and Plantea, it is well documented in Amoebozoa, Excavata and Chromalveolata (Nosenko and Bhattacharya 2007; Henze et al. 2001). Sparse phylogenetic distributions of specific gene families have been previously reported in protist genomes (Andersson et al. 2006). Horizontal gene transfer of certain enzymes involved in glycolytic pathway to protists (Excavates) from an ancestral cyanobacterium has also been observed (Henze et al. 2001). Lateral gene transfer is a recurrent attribute of protists that particularly follow a phagotrophic lifestyle (Andersson 2005). The acquired genes more often may offer undue advantages to these eukaryotes to explore new environments. Thus it is speculated that these bikont protists also acquired *MVP* gene in a similar scenario.

Accordingly, individual trees for depicting gene transfer into Kinetoplastids, *Phytophthora* and ciliates were constructed and the trees were rooted with the cyanobacterial sequence. The choanoflagellate sequences were also included for analysis as a representative from opisthokonts. In all the trees shown in Figure 2.5, it is evident that the choanoflagellates are closer to the cyanobacterial sequence than to any other protists, suggesting that the long branching and rapidly evolving protist sequences were a result of early horizontal gene transfer events. The ancestral kinetoplastid had one copy of the vault gene, even before the divergence of *Trypanosoma* and *Leishmania* likely occurred. It is known that gene duplicates in *Trypanosoma* arose out of tandem duplication events (Jackson 2007). In *Trypanosoma* chromosomes 4 and 8 are revealed to be partial duplicons that arose out of a large-scale

duplication event (Jackson 2007). The occurrence of multiple copies of *MVP* gene could be attributed such large scale duplication events.

Of the ciliates *Paramecium* has four *MVP* paralogs, likely arising from whole genome duplication events, since at least three successive whole-genome duplication events have been attributed to the appearance of most of the genes in *Paramecium* (Aury et al. 2006). The paralogs of *MVP* that arose out of the second round of duplication have not undergone many substitutions through evolution given the small branch lengths corresponding to the bifurcating branches. The duplicated paralogs map to different loci on the genome and hence, are not the same genes. However, another ciliate, *Oxytricha trifallax*, has only two paralogs. In *Oxytricha trifallax,* a chimeric chromosome that arose out of duplication events in a process akin to exon shuffling has been described (Zhou et al. 2011). Since, there are no other ciliate representatives it is difficult to assess if the paralogous expansion occurred in the common ancestor of these two organisms or if the duplicates arose independently in these two organisms.

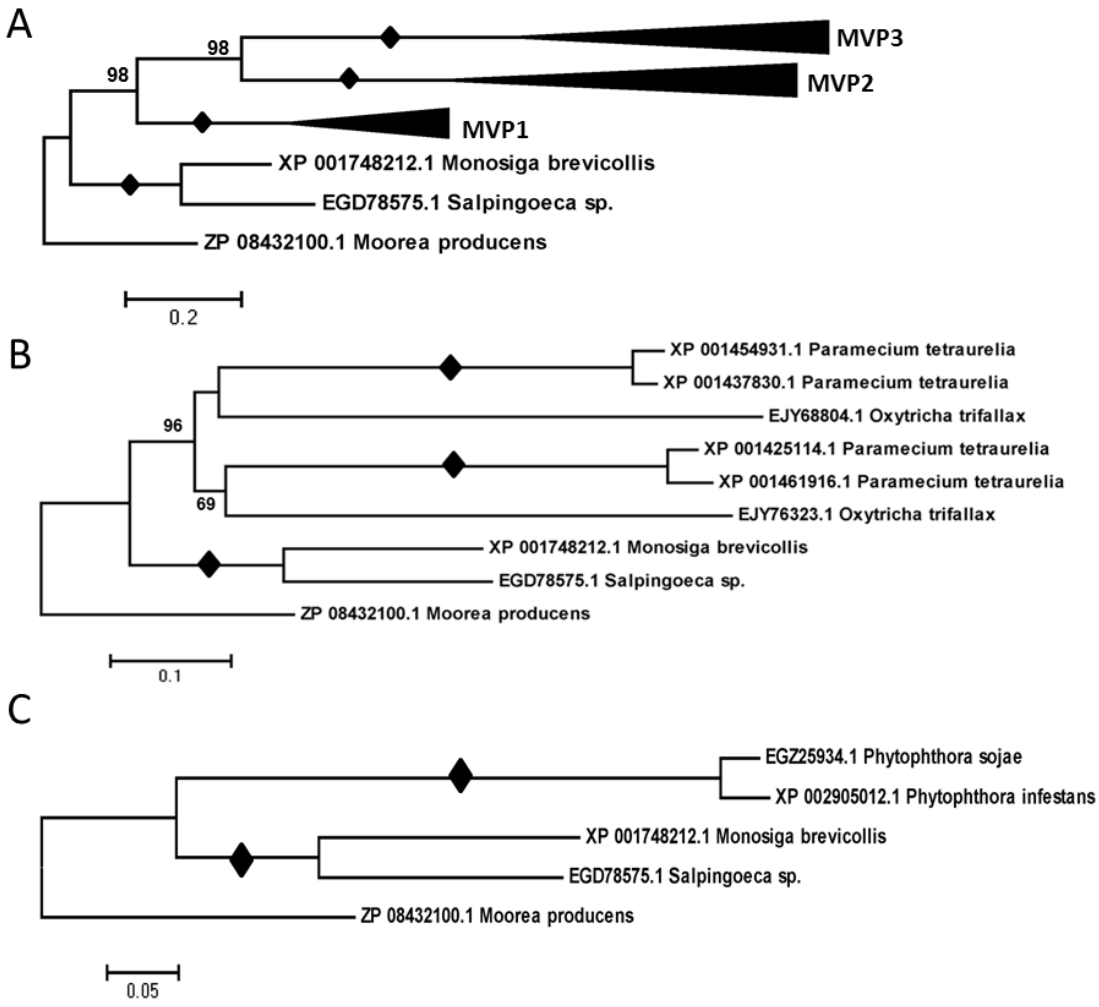


Figure 2.5 Horizontal gene transfer events into eukaryotic protists belonging to Bikonta
Shown are the topologies obtained using the ML method for (A) Kineotoplastids and NJ method for (B) *Paramecium* (C) *Phytopthora.* The ML tree was built using the WAG+G model of evolution. All the trees have been rooted with the cyanobacterial sequence and the choanoflagellates are included for the purpose of analysis. The tight branching of the cyanobacteria with the choanoflagellates away from the long branching bikont protist sequence in all topologies is suggestive of a horizontal gene transfer event into these ancient protists.

Given the observation of vault gene and protein in a cyanobacterium, it could be argued that a common ancestor of both the cyanobacteria and the gliding heterotrophic bacteria had vaults. However, to claim such a common decent, massive numbers of loss events of the *MVP* gene in a large number of bacterial species must be accounted for. Hence, few independent horizontal

transfer events, possibly from ancestral unicellular protists which may have been in the same habitat, are more parsimonious. Also, unlike the cyanobacterium, the *MVP* genes in the heterotrophic bacterial species have undergone significant evolutionary changes adapting to G+C in both nucleic acid and protein sequence indicating the direction of horizontal transfer from eukaryote to bacteria. It is yet unknown if these gliding heterotrophic bacterial *MVP* genes produce protein or form functional vaults.

2.3.3 Divergence of *MVP* in Opisthokonts

From the previously discussed phylogenies, it is clear that *MVP* either originated or emerged by horizontal gene transfer into the unikonts clade. Apart from the amoebozoans, the unikonts comprise the choanozoa, fungi and metazoans. This section charts the spread of *MVP* in the opisthokont clade with emphasis on the metazoans.

2.3.3.1 Evolution of *MVP* in Deuterostomes

41 protein sequences representing deuterostomes from chordates, echinoderms and hemichordate were used to reconstruct a comprehensive ML phylogeny of the vault proteins. The JTT was selected as the best amino acid substitution model and gamma distribution parameter was estimated to be 1.043 corresponding to moderate variation. The analysis also included the small branching Capsaspora homolog along with choanoflagellates as an out-group (Figure 2.6). The deuterostome topology reveals a clear separation of the vertebrate and the invertebrate sequences with well-supported clades. As mentioned earlier, the chordates have only one *MVP* homolog. It is seen that the *MVP* homolog of the cephalochordates and tunicates (invertebrate chordates) clusters along with the ambulacrarians with good branch support in

the consensus tree, forming a sister clade to the vertebrates. The tunicate *Oikopleura dioica*, however, does not cluster along with the invertebrates, instead branches out separately as a long-branch attraction artefact. Though it looks like the sequence has undergone significant evolutionary changes, there is a possibility that non-coding or intron regions have been included, since the sequence is predicted from whole genome shotgun assembly. The paralogs of the ambulacrarians, sister group of chordates and represented by sea urchins and acorn worms, were distinctly grouped out to form a separate clade with long branches with 100% support. Analysis of the dataset suggests that these paralogs are likely the result of independent lineage-specific duplications that could have occurred in the invertebrates after divergence of vertebrates. It is difficult to distinguish whether the additional copies have altered or paralogous functions. Removing the paralogs and the long branching *Oikopleura dioica* sequence did not alter the topology of the deuterostomia tree. In both the reconstructed topologies, the *Capsaspora* sequence was found to cluster within the invertebrate clade instead of branching out as expected.

The choanoflagellate has only one *MVP* homolog. Hence, it is prudent to consider that the ancestral deuterostome had only one *MVP* homolog and the appearance of paralogs occurred after the divergence of the chordates from ambulacrarians. The depicted phylogeny corresponds well to major events in vertebrate evolution including the teleost-tetrapod split that occurred around 450 Mya and the origin of birds from a common ancestor of reptiles (Ravi and Venkatesh 2008). From the presented phylogeny it is evident that vault proteins are found in three of the deuterostome phyla except for the phylum Xenacoelomorpha, previously associated with Platyhelminthes, that includes

the Xenoturbellids and acoemomorph worms (Telford 2008). Searching the NCBI trace archive did not reveal any homologs in this newly classified phylum and hence, it is concluded that vault genes are lost in this phylum.

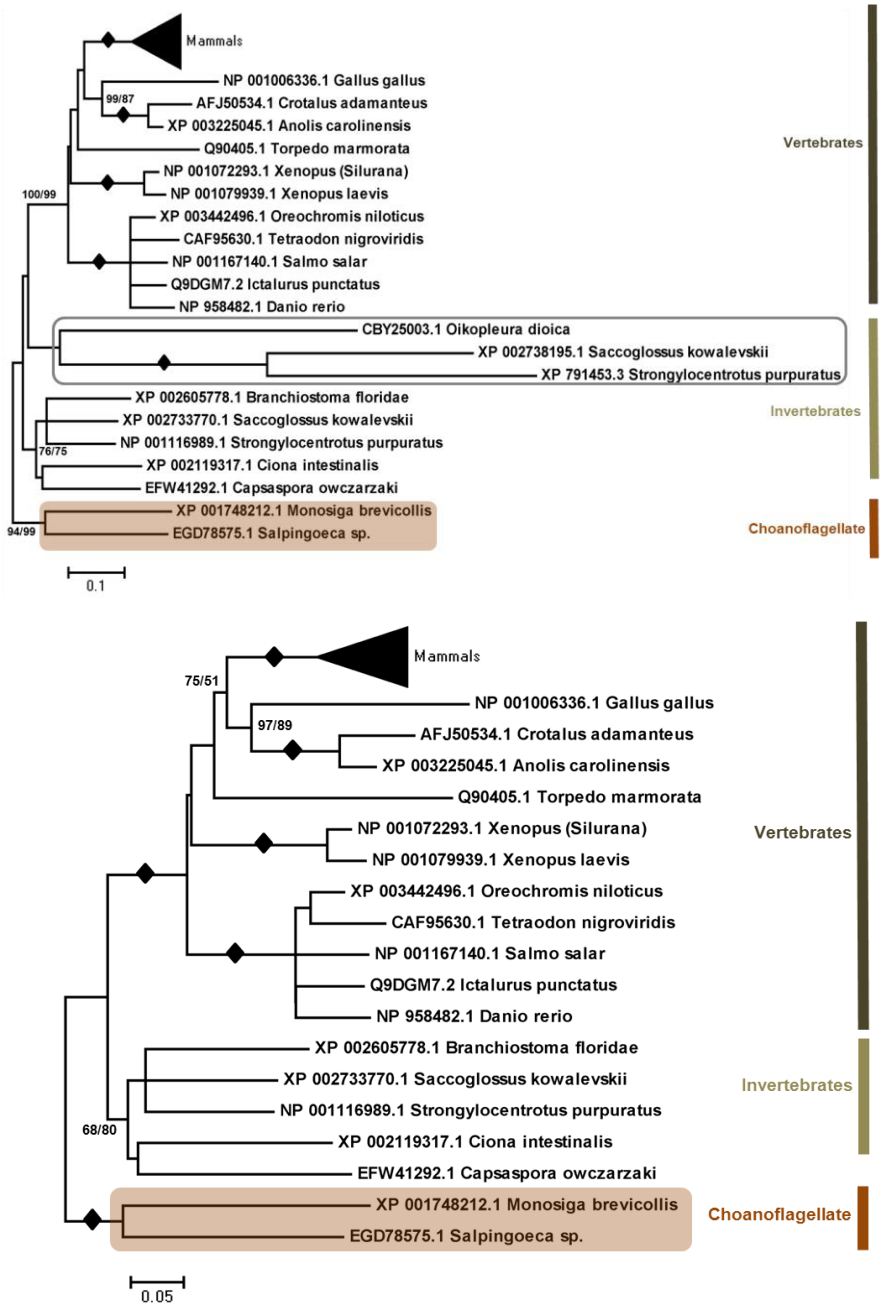


Figure 2.6 *MVP* evolution in deuterostomes
Maximum Likelihood (ML) tree describing evolutionary relationships among *MVP* genes in deuterostomes rooted with choanoflagellates with long branches included (top) and removed (bottom). The shown topology is a bootstrap consensus tree based on the JTT+G model of evolution. Black diamond indicates 100% bootstrap support for the indicated node by both ML and NJ methods. Values are given for all other nodes supported by >70% bootstrap support in either of the methods. The branch length is directly proportional to the number of substitutions per site. Sequences from 23 organisms are clustered into the "Mammals" clade.

2.3.3.2 *MVP* in Non-Deuterostome Opisthokonts

Homologs from cnidarians, protostomes, choanoflagellates, parazoan organisms (including *T. adherans* and *A. queenslandica*), and sequences from molds were used to make an initial ML tree with the cyanobacterial species out-group (Figure 2.7). The tree showed a clear delineation of the non-deuterostome opisthokonta from the molds. Paralogs from cnidarians, amoeboid symbiont *C. owczarzaki* and porifera *A. queenslandica* cluster together with long branches. The *A. queenslandica* sequence also had a long branch and hence was removed from subsequent analysis. Distinct paralogs were removed and the tree was reconstructed with only one species in each organism. Cnidarians, placozoan *T. adherans*, oyster *C. gigas* and *C. owczarzaki* formed a clade sister to the choanoflagellates *Salpingoea sp*. and *Monosiga brevicollis*. Sequence from molluscs clustered along with cnidarians and placozoa. The platyhelminthes (*Clonorchis sinensis* and *Schistosoma mansoni*) formed a paraphyletic group with strong support. The position of *C.owczarzaki* relative to metazoan and choanoflagellate has always remained unclear and is found to vary depending on the genes being analysed (Ruiz-Trillo et al. 2008; Ruiz-Trillo et al. 2004). *MVP* orthologs and paralogs of the invertebrate deuterostomes including the ambulacrarians, lancelets and tunicates cluster near the orthologs and paralogs of non-deuterostome opisthokonta. There is a delineation in the evolution of vaults from the unicellular choanoflagellates to invertebrates and protostomes and finally to vertebrates. Protostomes, like the chordates, have only one *MVP* homolog and the fact that they cluster near choanoflagellates, invertebrate deuterostomes

and cnidarians with strong support underscores the lineage specific duplications that occurred after the protostome-deuterostome split.

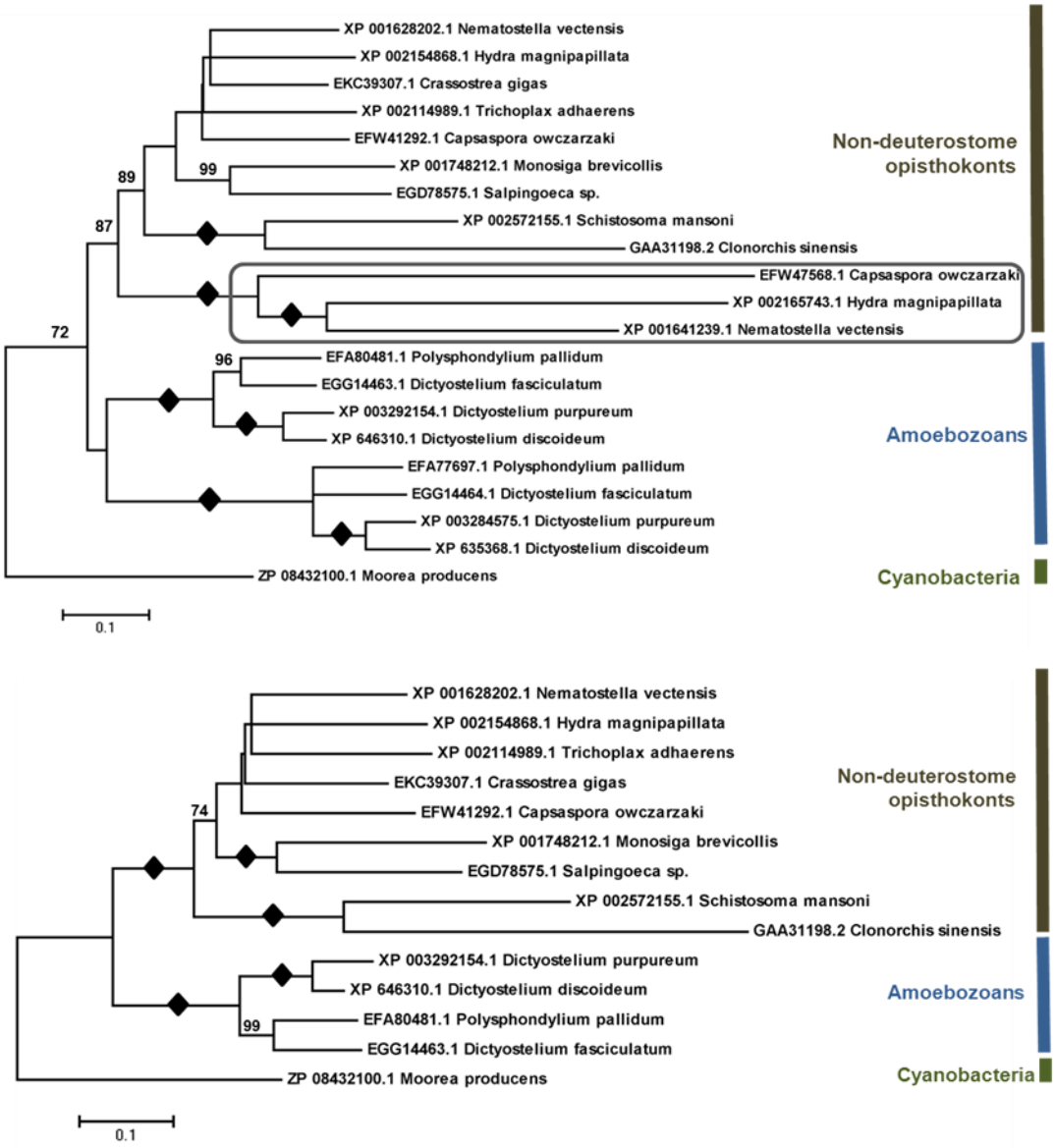


Figure 2.7 Evolutionary relationships between non-deuterostome opisthokonts
ML tree showing evolutionary relationship among *MVP* genes in non-deuterostome opisthokonts and amoebozoans with long branches included (top) and removed (bottom),rooted with the cyanobacterial *MVP* sequence. Shown is a bootstrap consensus topology obtained based on the rtREV+G model of evolution. Black diamond indicates 100% bootstrap support for the indicated node by both ML and NJ methods. Values are given for all other nodes supported by >70% bootstrap support in either of the methods. The branch length is directly proportional to the number of substitutions per site.

2.3.3.3 Only Lophotrochozoan Protostomes have *MVP*

From the depicted phylogenetic tree it is evident that a few protostomia, including the parasitic flukes and oyster that harbor vault genes (Figure 2.7). *MVP* transcripts have also been identified experimentally in annelid *Riftia pachyptila* (Sanchez et al. 2007). While lophotrochozoan protostomes represented by flatworms, annelids and molluscs have vault genes, many complete genomes from insects and nematodes representing ecdysozoan protostomes lack vault genes (Stephen et al. 2001; Berger et al. 2008). The phylogenetic pattern of *MVP* gene distribution suggests that it appeared as a full-length gene in a single celled ancestor of eukaryotic animals and molds but it was subsequently lost after the metazoan radiation from the common ancestor of fungi and ecdysozoa explaining why vault complexes are missing from model organisms *S. cerevisiae*, *Drosophila* and *Caenorhabditis elegans*.

2.3.4 Co-evolution of *VPARP* and *TEP1* with *MVP*

The minor vault proteins, *VPARP* and *TEP1*, associate with naturally occurring vaults and co-purify along with the *MVP* gene product. In this section the evolution of these minor vault proteins are also analysed to trace their evolutionary timeline.

A previous phylogenetic analysis aimed at understanding the evolution of poly-ADP-ribose family of proteins, classified organisms into two clades based on *VPARP* sequences (Citarelli et al. 2010). The analysis was extended by including more sequences and explaining the phylogeny in the context of origin of vaults. *VPARP* has been known to interact near the N-termini of *MVP* with its C-terminal binding domain (1562-1724 aa) (Kickhoefer et al.

1999). This minimum interaction domain (mINT) binds to *MVP* on the inner vault surface and when this region was excluded, it was shown that *VPARP* remains in the soluble fraction and cannot be packaged into intact vault particles (Kickhoefer 2005; Goldsmith et al. 2009). Unlike *MVP*, homologs of *VPARP* had a limited phylogenetic distribution. Except for slime molds and *Capsaspora owczarzaki*, homologs of *VPARP* were not found in any of the unicellular organisms including choanoflagellates, ciliates, kinetoplastids or bacteria. This suggests that the *VPARP* co-evolved along with *MVP* from a common ancestor of unikonta comprising the opisthokonta and amoebozoans, but was lost in clades representing fungi and nucleariids. It was also noticed that while the vertebrates had good conservation through the length of the sequence, other organisms showed very poor similarities beyond position 1200 corresponding to human *VPARP* protein sequence. This trend was observed throughout the invertebrate and unicellular sequences. Interestingly, multiple homologs of *VPARP* of different lengths in the invertebrates including sea urchin, sea anemone and oyster, the placozoan *Trichoplax* and the opisthokont *Capsospora*were observed. The homologs from each of the above-mentioned organisms could not be successfully mapped onto the genome, since information on complete genome assembly is not available. One of the homologs from sea anemone and *Capsapsora* lacked the PARP-like domain and hence was dismissed from analysis, along with the *Trichoplax* homolog that was smaller and hence likely does not represent a full-length protein. Of the two full-length homologs identified in sea urchin and oyster, the true *VPARP* homolog could not be clearly distinguished and hence both were retained for analysis. The invertebrate sequences, in many cases, were

predicted by conceptual translation from genomic scaffolds. Whether the multiple homologs truly represent alternative splice forms of the same protein, or if they actually belong to different genomic loci and code for different proteins or if they are an artefact due to inclusion of intronic or other non-exonic regions in some of the annotated proteins is unclear.

From the unrooted ML tree a clear delineation of the vertebrates and invertebratesis evident (Figure 2.8). *C. owczarzaki* distinctly separates out from the metazoans, while the molds cluster together to form a separate clade. While the two sea urchin sequences included in the analysis clustered into a single clade, the position of the oyster sequence is ambiguous, as one of them clusters with other invertebrate sequences and the other branches out separately. Although, *C. owczarzaki* and choanoflagellates, both are considered to be the closest unicellular relatives to multicellular metazoans, there are no homologs of *VPARP* in choanoflagellates. This suggests that *VPARP* originated in a common ancestor prior to the divergence of *C. owczarzaki* from choanoflagellates and metazoans, but may have been lost in the choanoflagellate lineage.

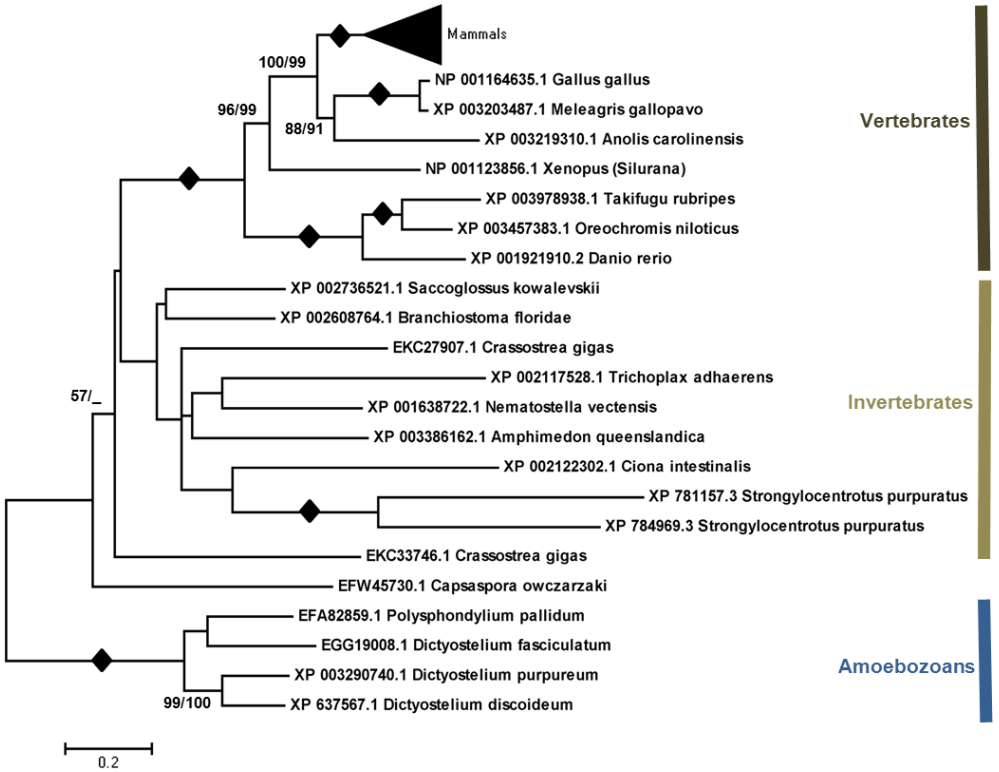


Figure 2.8 Evolutionary origin of *VPARP*

ML tree describing the evolutionary origin of vault poly-ADP ribose polymerase (*VPARP*) across all the taxa. The bootstrap consensus topology was obtained based on the JTT+G+I model of evolution. The clade "Mammals" is representative of 24 organisms. The '_' implies that the depicted branching was not reconstructed by the NJ method. The *Capsaspora* clearly clusters out from the common ancestor of metazoans. The placement of the amoebozoans as out-group to the clade Filozoa that includes *Capsaspora*, choanoflagellates and all metazoans is strongly supported. The branching within the invertebrates remains poorly resolved. The placement of one of the *C. gigas* homolog is likely a long-branch attraction artifact. Black diamond indicates 100% bootstrap support for the indicated node by both ML and NJ methods. Values are given for all other nodes supported by >70% bootstrap support in either of the methods. The branch length is directly proportional to the number of substitutions per site.

The phylogenetic distribution of *TEP1* using the human protein sequence as a query revealed that the span of organisms was almost similar to *VPARP* except there are no orthologs of *TEP1* in birds and in the hemichordate *Saccoglossus* (Figure 2.9). Homologs of *TEP1* in both *C. owczarzaki* and choanoflagellate were found, suggesting a similar origin to that of *VPARP*. While the *C. owczarzaki* aligns through the length of the sequence, the choanoflagellate *M. brevicollis* was found to align only in parts.

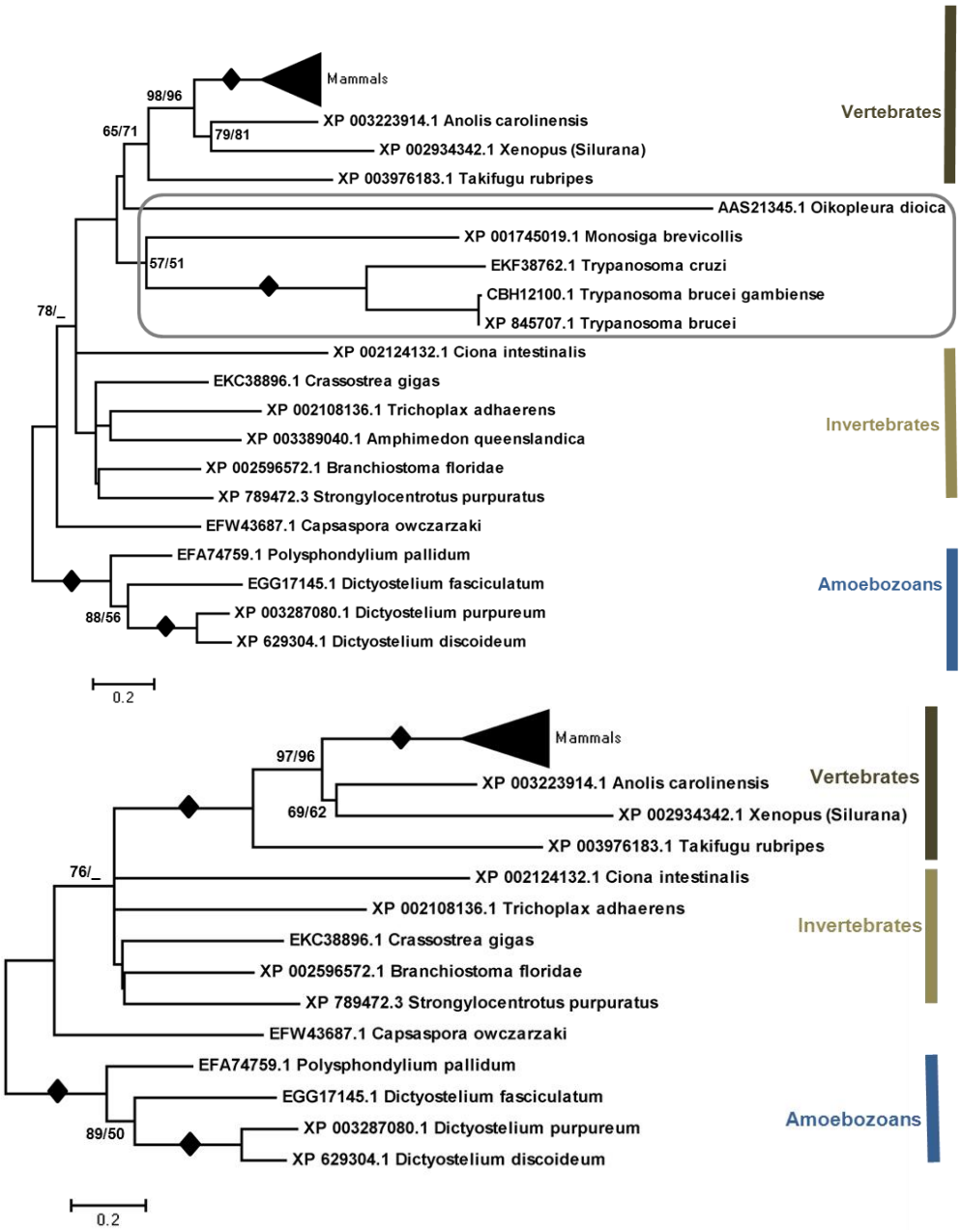


Figure 2.9 Evolutionary origin of *TEP1*

(Top) ML tree describing the evolutionary origin of *TEP1* across all the taxa retrieved. The bootstrap consensus topology was obtained based on the WAG+G model of evolution. The grey box points to the long branching sequences that are not likely to represent true *TEP1* homologs due to their incorrect placement on the evolutionary tree. (Bottom) The bootstrap consensus topology was obtained based on the JTT+G+I model of evolution. The long branching *O. dioica*, *Trypanosoma* and *M. brevicollis* sequences have been removed. The clade "Mammals" represents sequences from 22 organisms. The '_' implies that the depicted branching was not reconstructed by the NJ method. The placement of *Capsapsora* and amoebozoans with respect to metazoans is similar to that of *VPARP*. Black diamond indicates 100% bootstrap support for the indicated node by both ML and NJ methods. Values are given for all other nodes supported by >70% bootstrap support in either of the methods. The branch length is directly proportional to the number of substitutions per site.

TEP1 homologs were found in *Trypanosoma*, but not in *Leishmania*, albeit with poor amino acid conservation and alignment. To identify if the choanoflagellate and *Trypanosoma* sequences are true homologs, an initial ML tree was constructed. The unrooted tree identified that these sequences along with the sequence from *O. dioica* form exceptionally long branches. Hence, these sequences were removed and the ML tree was reconstructed. The topology was similar to that of the *VPARP* unrooted tree with *C. owczarzaki* grouping out of the metazoans and molds forming a distinct sister clade.

Vault isolates from sea urchin have identified heavy molecular weight bands around 200 kDa that might correspond to *TEP1* and *VPARP*, and associated *vRNA* (Stewart et al. 2005). However, vaults isolated from *Dictyostelium* show no evidence of vault-associated proteins (Kedersha et al. 1990). From this observation it is likely that the association of minor vault proteins within vault complexes could have evolved with the origin of multicellular eukaryotes.

2.3.5 Organisms with *MVP* do have intact biosynthetic pathways for essential amino acids

The phylogenetic distribution of vault genes forms a unique pattern, wherein some distinct eukaryotic branches corresponding to plants and fungi are missing. From a broad glimpse at the species distribution of vaults, it is seems that vault genes are present only in heterotrophs and absent from autotrophs including plants and species of fungi (non-auxotrophic organisms) that retain their capability to synthesize all amino acids. The only outlier in the list of organisms is the cyanobacterium, *Moorea producens3L*. Hence, the amino acid biosynthetic pathway of each species included in the phylogenetic analysis was analyzed, focusing on the metabolism of the nine essential amino

acids. The pathway data for each organism was accessed in the KEGG Pathway database. Interestingly, almost all of the eukaryotes had incomplete pathways for the synthesis of the nine essential amino acids – Leu, Val, Ile, Lys, Met, His, Thr, Phe, Trp. All the eukaryotic organisms had important enzymes missing for the synthesis of branched chain amino acids, namely, Leu, Val and Ile. In addition, the synthetic pathways of the aromatic amino acids, Phe and Trp, were also incomplete in all the eukaryotes analyzed (Figure 2.10). The unicellular choanoflagellate has retained the enzymes involved in the synthesis of Met and His, but has lost the ability to synthesize other essential amino acids. Other single celled protists including the *Dictyostelium* and the organisms that may have horizontally acquired *MVP* including kinetoplastids, *Phytophthora* and organisms within the ciliate clade also display similar nutritional demands. While the kinetoplastid *Trypanosoma* has lost its ability to synthesize Met, another kinetoplastid *Leishmania*has Met biosynthetic enzymes intact. It is interesting to note that *Phytophthora* that are closer to plants also lack enzymes for synthesis of certain essential amino acids.

The pathways corresponding to amino acids regarded as conditionally essential, including Pro and Arg, were also analyzed for the enlisted organisms. All organisms could synthesize Pro except for the ciliate *Paramecium*. It was observed that in all deuterostomes, cnidarians and the placozoan *T. adherans* the enzymes for the biosynthesis of Arg were intact; however in the unicellular protists, including the choanoflagellate, Arg had to be supplemented in diet.

The heterotrophic gliding bacterial species, which may have acquired vault genes from eukaryotes, also have amino acid nutritional requirements. Out of the nine essential amino acids analyzed, the *C. coralloides* can only synthesize Met, Thr and Lys. The multicellular filamentous *S. grandis* has an intact biosynthetic pathway for only Thr and it needs to prey on other protists to meet its nutritional needs. The *F. litoralis*, however, can synthesize two branched chain amino acids, Val and Ile, in addition to His, Thr and Met. The amino acid synthetic pathway information corresponding to *M. producens 3L*, *P. pacifica* and *M. marina* were not available on KEGG and hence, ortholog enzymes that render the amino acid biosynthetic pathway complete was identified using protein BLAST search. The enzymes from myxobacterium *Sorangium cellulosum* was used to seed for *P. pacifica* orthologs. The filamentous cyanobacterium *Trichodesmium erythraeum*, which is closely associated with *M. marina*, was used to identify orthologous enzymes from both *M. producens 3L* and *M. marina* (Hopkinson et al. 2008; Jones et al. 2011). It was identified that Trp, Phe and Lys are essential in both the marine bacterium and *P. pacifica.* In addition, *P. pacifica* also lacked certain key enzymes involved in the synthetic pathways of Met and His. The two organisms analyzed indirectly may have incomplete biosynthetic pathways for other amino acids as well, since the identified enzyme orthologs more often displayed poor conservation through the length of the protein.

The autotroph *M. producens 3L* on the other hand, has orthologs for all the key enzymes involved in synthesis of essential amino acids except for Met. This cyanobacterium is unable to fix nitrogen and relies on recycling its internal sources of carbon and nitrogen, particularly, cyanophycin, a storage

polymer, made by non-ribosomal peptide synthesis of Arg and Asp, for its energy needs (Jones et al. 2011).

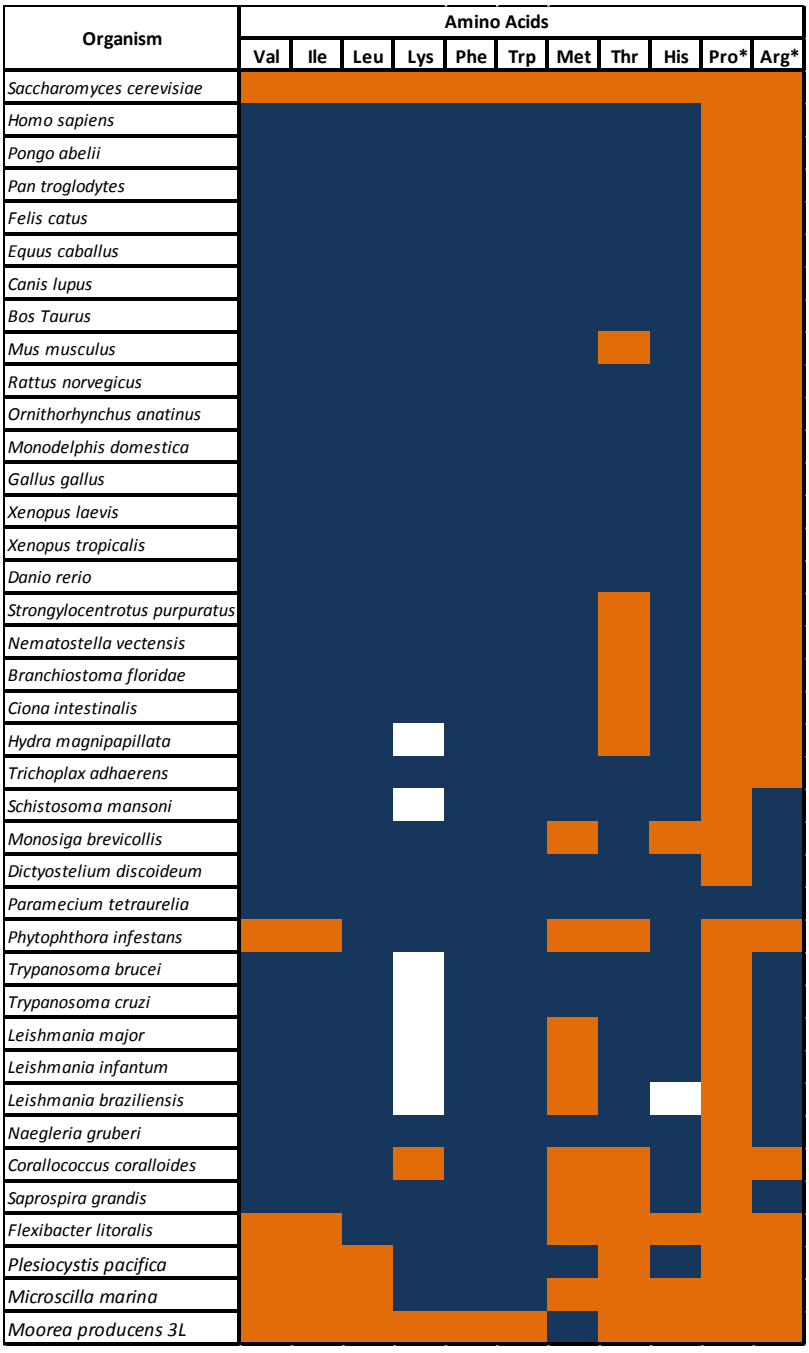


Figure 2.10 Analysis of essential amino acid biosynthetic pathways across all organisms that harbor vault homologs
Amino acids marked with '*' point to conditionally essential amino acids. Dark Blue coloring indicates that the enzymes leading to synthesis of a particular amino acid are missing and that the amino acid is therefore essential. Orange coloring indicates amino acids that can be synthesized by the organism. White areas denote that the data was unavailable in KEGG for the particular organism. For *P. pacifica*, *M. marina*, *M. producens 3L*, the presence of enzyme orthologs for the synthesis of each amino acid was accounted for based on a protein homology search (for further details, see main text). *Saccharomyces* is included as a control organism that can synthesize all amino acids

In addition, other protists including ciliate *Tetrahymena thermophila*, apicomplexan *Toxoplama gondii* (both belong to alveolates), excavate *Trichomonas vaginalis* and also *Entamoeba histolytica* (amoebozoan) that do not harbor vault genes were also analyzed (Figure 2.11). The analysis revealed that these protists also display defects in essential amino acid biosynthesis and hence are heterotrophic protists. This leads to the conclusion that while all organisms carrying vault genes display loss of amino acid biosynthetic capability (except for cyanobacterium), not all heterotrophic protists have vault genes. The selective emergence of vaults in many of these heterotrophic protists including *Paramecium* and Kinetoplastids reiterates the notion that these protists acquired vaults by horizontal gene transfer from their environment.

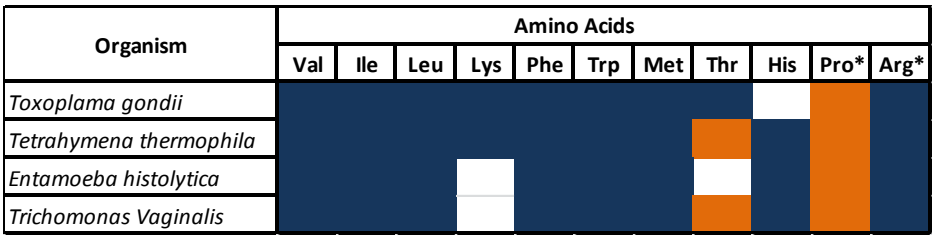


Figure 2.11 Amino acid synthesis analysis on other eukaryotic protists
The closely branching organisms that do not have vaults but branch closely to those eukaryotic protists that harbor vault genes were subjected to similar pathway analysis. Amino acids marked with '*' point to conditionally essential amino acids. Dark Blue coloring indicates that the enzymes leading to synthesis of a particular amino acid are missing and that the amino acid is therefore essential. Orange coloring indicates amino acids that can be synthesized by the organism. White areas denote that the data was unavailable in KEGG for the particular organism.

2.4 DISCUSSION

Vaults are known to be conserved in a wide range of eukaryotes; however, their absence in distinct eukaryotic branches including plants, fungi, nematodes, and insects has always been a puzzling aspect in understanding the evolution of vaults. The analysis on the evolution of vaults and its components sheds light on its unique phylogenetic distribution. Based on the conclusions made an evolutionary model that depicts the possible evolutionary timeline for the origin of vaults and their subsequent phylogenetic distribution has been proposed (Figure 2.12)

From the phylogenetic distribution it is clear that vaults are present only in organisms that lost certain key enzymes pertaining to synthesis of essential amino acids. Loss of amino acid biosynthetic capability occurred independently across different lineages over the course of evolution. A Great Genomic Deletion model has been implicated in eukaryotic clades that lack essential amino acid biosynthetic enzymes (Guedes et al. 2011). Interestingly, fungi that are closely related to metazoans have no vault genes; yet, retain all enzymes essential for amino acid biosynthesis. A more reasonable interpretation, based on evolutionary reconstructions, would be that *MVP*, *VPARP* and *TEP1*, all co-evolved together in an ancestral unikont that evolved independently into the amoebozoans and opisthokonta clade and was subsequently lost in the fungi, nucleariida and also in class Icthyosporea. This common ancestor may have had enzymes for synthesis of amino acids intact. However, during course of evolution, as it diverged into organisms representing metazoans or choanozoans, the presence of vault genes may have compensated in some way for the subsequent loss of amino acid biosynthetic

capability. However, the fungal clade with no selective pressure to retain vault genes underwent a secondary loss. It is also interesting to note that except for metazoans, other eukaryotes that lack essential amino acid biosynthetic pathways (including those that have no *MVP* genes), all retain enzymes related to nitrogen assimilation (Guedes et al. 2011). Thus it seems like vault genes originated primarily in unikonts to compensate for the loss of both essential amino acid synthetic and nitrogen assimilative enzymes.

It is interesting to note that even organisms that acquired vault genes by lateral gene transfer events, including certain heterotrophic bacteria with gliding motility, display defects in synthesis of essential amino acids. It is known that adaptation to specific environments is a major trigger for horizontal gene transfer events. Different eukaryotic lineages can independently acquire the same beneficial gene under certain circumstances. The horizontal gene transfers that contribute significantly to protist genomes are attributed to the phagotrophic or parasitic lifestyle of the organisms involved (Nosenko and Bhattacharya 2007). Not just the protists, even the bacterial xenologs that acquired vault gene have a lifestyle akin to those of slime molds and occupy similar niches. They also form multicellular aggregates under starvation. This reiterates that these organisms may have succumbed to some selective pressure that led to the acquirement of genes from organisms that ancestrally harbored vault gene.

While vaults are observed in only organisms that lost synthesis of essential amino acids capability over evolution, there is one autotrophic cyanobacterium that harbors a clear conserved *MVP* homolog. An early eukaryotic *MVP* may have been acquired by Moorea producens 3L.

Alternatively, *MVP* originated in older filamentous cyanobacteria and was acquired by an ancestral single celled eukaryotic unikont, which also lost its core amino acid biosynthesis pathways. The key question is - In which direction did this obvious horizontal transfer occur?

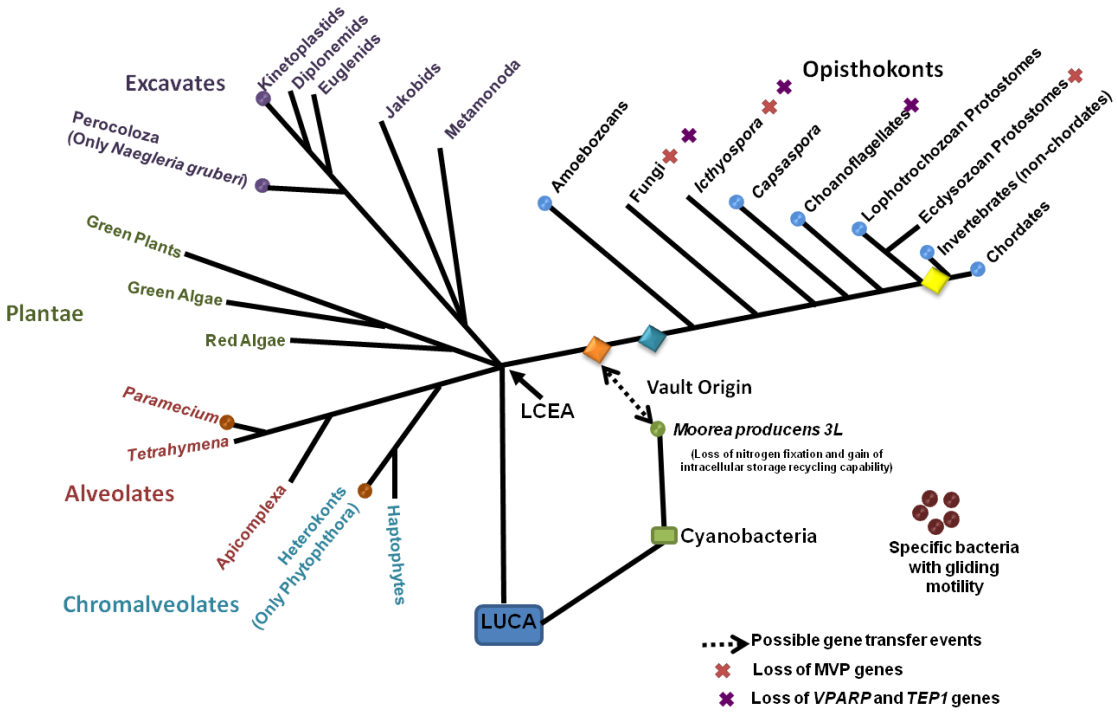


Figure 2.12 Proposed evolutionary model for the origin of Vault Complex
The *MVP* gene harboring organisms are indicated with colored circles. The organisms are grouped according to well-established evolutionary relationships. The emergence of *MVP* in the Unikont clade comprising the Opisthokonts and Amoebozoa is depicted with the orange diamond.*MVP* could have originated in a non-nitrogen fixing cyanobacterium that recycles amino acids and been transferred to an ancient unikont ancestor. The minor vault proteins, *VPARP* and *TEP1*also emerged in the Unikont clade and co-evolved with *MVP*as depicted with blue diamond.*vRNA* evolved later in deuterostomes as shown as yellow diamond. The clades Excavates, Plantae, Chromalveolates together comprise the Bikonts. The specific group of bacteria with gliding motility, Kinetoplastids, *Paramecium*, *Naegleria* and *Phytophthora*, all have acquired *MVP* gene through horizontal gene transfer events.

Bacterial species seem less capable of evolving complex multi-domain proteins like *MVP*, however filamentous cyanobacteria like *Moorea producens 3L* are an exception, with domain-shuffled polyketide synthetases/non-

ribosomal peptide synthetases, as well as abundant recombination and DNA repair enzymes (Jones et al. 2011). So *MVP* could have very well originated in such an organism, as mechanisms for domain duplication and shuffling are well supported. Vaults likely originated with a simple *MVP* gene in such an ancestral cyanobacterial species through the selective pressure of loss of nitrogen fixation process; a pressure which has most certainly influenced other genes and gave rise to complex strategies in *M.producens 3L* for amino acid storage. Therefore it should not be surprising that a long multidomain repeat protein like *MVP* could have emerged in such a cyanobacterial lineage, perhaps simply as a conventionally transcribed amino acid storage molecule.*MVP* may have been later acquired in a fully formed state by a cyanobacterium-feeding single-celled eukaryote through horizontal transfer from an ancestral cyanobacterium that lost nitrogen fixation. The acquisition of a cyanobacterial *MVP* may have helped complement the subsequent loss of amino acid biosynthesis genes in the ancient eukaryote.

MVP knockout studies performed in multicellular eukaryotes have revealed no obvious phenotypic changes except under nutritional stress conditions (Vasu and Rome 1995; Mossink et al. 2003; Sutovsky et al. 2005; Kolli et al. 2004). The correspondence of vaults with both lost essential amino acid biosynthetic capability and lost nitrogen fixation capacity, and the exclusion of vaults in cases of amino-acid synthesizing intracellular endosymbiosis, suggests that the function of vault may be most simply related to metabolism of amino acids. Evolution of nutrient-absorption systems to compensate for the loss of essential amino acids has already been hypothesized (Guedes et al. 2011). This reiterates that vaults may have an

important but overlooked role to play in cell survival under stress and starvation conditions.

Within the multicellular metazoans, it is observed that vault genes have been specifically lost in the ecdysozoan protostomes. Although insects and nematodes lack amino acid synthesis genes, they often have widespread and evolutionarily ancient relationships with obligate intracellular endosymbiotic bacteria like *Buchnera* and *Wolbachia*, which are an internal source of essential amino acid synthesis. Primary obligate endosymbionts have been associated with their insect hosts for millions of years and co-speciated with their insect hosts (Douglas 1998; Wernegreen 2002). The intracellular endosymbionts may compensate for the nutritional requirements in ecdysozoan protostomes. In the Lophotrochozoan mouth- and gut-less annelid tubeworm *Riftia pachyptila*, it was found that *MVP* mRNA is conspicuously and completely down-regulated in the specialized tissue trophosome that harbors nutrient providing intracellular endosymbionts (chemolithoautotrophic *γ- Proteobacterium*) but remains enriched in the branchial plume that is involved in exchange of metabolites with environment (Sanchez et al. 2007). It is interesting that though vaults and intracellular endosymbiosis are found in the same organism they remain segregated in different tissues and appear to be mutually exclusive.

The notion of horizontal gene transfer events into specific organisms in bikonts is also supported by missing sequences corresponding to *VPARP* and *TEP1* in these organisms as opposed to those of the opisthokonts and amoebozoans. It is also worth mentioning that the *VPARP* homologs from unicellular organisms (along with those from invertebrates) are very different

from their vertebrate counterparts and suffers from poor conservation through the length of the alignment, particularly along the mINT domain that is essential for interaction with *MVP* monomers. Therefore, it is not very clear if they are capable of being enclosed within the vault complex. Interestingly, both the minor vault genes are present in *Capsaspora* that diverged prior to that of choanoflagellates and metazoans, but have been specifically lost in choanoflagellates indicative of a lineage specific gene loss

The results from the phylogenetic reconstructions shed light on the evolutionary history of the vault complex, alongside the minor vault components. Appearance of vault genes in independent lineages of organisms with lost amino acid biosynthetic capability suggests a possible and convincing correlation between loss of amino acid synthesis and vaults. The absence of vaults in fungi or ecdysozoan protostomes is likely compensated by amino acid availability either by intact biosynthetic pathways or through beneficial intracellular endosymbiosis. The puzzling phylogenetic distribution and selective occurrence in a few species of unicellular protists is parsimoniously explained as a horizontal gene transfer event that may have occurred due to selective pressure in adapting to specific nutrient niches where the gene was available. Based on the provided evidence it is reasonable to speculate that function of vault may be related to basal heterotrophic nutrient requirements.

Chapter 3

The Medium is the Message – Vaults as Nutrient Sequesters

3.1 INTRODUCTION

The detailed phylogenetic reconstruction tracing the evolutionary history of the vault complex reveals a connection between loss of essential amino acid biosynthetic capability and vaults. While the phylogenetic analysis distinctly points to the emergence of vault in a single-celled eukaryotic ancestor of unikonts, the specific loss of vault genes in fungi that has intact pathways for amino acid synthesis is interesting. Even organisms which possibly acquired vault later by horizontal gene transfer events, including ciliates, *Paramecium* and also specific species of gliding bacteria, are compromised in their ability to synthesize amino acids, especially branched chain amino acids Ile, Val, and Leu. Thus it seems possible that the function of vaults may have compensated for the loss of amino acid synthesis in

eukaryotes and those organisms that can synthesize their amino acids including plants, algae and fungi did not display any selective pressure to retain vault genes and hence, lost it through evolution.

Proteins, apart from playing their specific cellular functions, also provide valuable amino acids once they are recycled. The cellular system represents dynamic machinery as opposed to a static system, where new organelles and proteins are constantly being synthesized while the old ones are being recycled from time to time. Autophagy, in particular macroautophagy, is an evolutionarily conserved self-digestive and recycling process in all eukaryotic cells, whereby the cells degrade older proteins and organelles to promote cell survival (Klionsky and Emr 2000). The process is more pronounced during limiting nutrient conditions, when the bulk of old proteins are recycled and the resulting amino acids are released back into the cytosol. Autophagy is known to be up regulated following nitrogen starvation in many multicellular eukaryotes from yeast through mammals (Mizushima 2007). Autophagy also plays a major role in starvation- induced development of the social amoeba *Dictyostelium* (Otto et al. 2003). Defects in autophagy have been implicated with many physiological conditions including cancer, neurodegeneration and infections (Mizushima et al. 2008; Amano et al. 2006). While the evolutionarily evolved autophagic process promotes cell survival in most cases, autophagy induced cell-death has also been reported.

While most mammalian tissues reach autophagic maxima within 24 hours, brain tissue does not exhibit autophagy even after 46 hours of nutrient starvation. Autophagy is also pronounced in embryos where degradation of maternally derived proteins provides energy and nutrients for development

(Tsukamoto et al. 2008; Tsukamoto et al. 2008). Though autophagy-deficient mice appear normal during birth, they are marked by dramatically reduced amino acid profiles in plasma and other tissues as early as 10 hours within birth and display higher mortality rates (Kuma et al. 2004). This emphasizes that amino acids degraded from the bulk of tissue proteins serve as prime energy sources during early phases of development (Onodera and Ohsumi 2005).

It is known that vaults are highly expressed in developing embryos and also during other physiological conditions involving cancer, infection or neurodegeneration. As mentioned above, all of these conditions involve autophagy and mobilization of amino acids derived from degraded proteins (Mizushima and Klionsky 2007). Are the elevated levels of vaults in highly autophagic tissues merely coincidental? Or is there a relationship between the large numbers of polymerized amino acids in vaults and autophagic turnover?

In higher eukaryotes, the liver serves as the prime organ of protein metabolism including assembly of new proteins and dismantling of the old ones to generate useful amino acids. Intriguingly, the liver can sense amino acid concentrations and efficiently trigger necessary catabolic or anabolic processes. Accordingly, a dramatic decrease in protein degradation rate from 4.5%/hr to 1.5%/hr has been reported in response to an increase in amino acid concentrations of up to tenfold (Schworer et al. 1981). Apart from protein catabolism, the liver also houses a bulk of glycogen, oligomerized glucose molecules, that represents secondary long-term energy storage. Vaults have been primarily isolated from and are also known to be enriched in livers. Given their massive size and protein content, could vaults have evolved as

nutrient (amino acid) stores and represent a form of secondary storage for polymerized amino acids?

In this chapter, the possibility of vaults functioning as a nutrient store has been analyzed by using a combination of compositional and theoretical metabolic analysis. Also, experimental evidence pertaining to vault expression and expression profiles of vaults in several microarray experiments are re-analyzed in light of a proposed nutrient storage function.

3.2 RESULTS

3.2.1 Conserved compositional bias of *MVP* and *vRNA*

Regions that define folded protein domains or non-structural linkers in a structural protein are marked by distinct bias in their amino acid compositions. This has been established previously by Dumontier et al. based on analysis of a large number of crystal structures (Dumontier et al. 2005). A protein domain is most likely expected to fit into the amino acid compositions defined for folded and linker region depending on its structure. Interestingly, comparison of this expected 'folded' protein composition to *MVP*, revealed an excess of Glu, Val and Gln, and elevated amounts of Leu, Pro and Arg (Figure 3.1A). The compositional bias appears to be a consequence of the unique vault structure. The cap region is Glu rich, and the waist is Glu + Val rich. Val residues play an important structural role in the stable interface between the cap helices. The long cap helix has no helix-breaks or turns, which may explain its simpler composition as against a folded domain. Consistent with this view, the long cap helix has a narrow amino acid composition with high numbers of Ala, Ser, Glu, and Gln residues.

From the phylogenetic analyses discussed in the previous chapter, it is evident that vault genes are conserved across a wide range of eukaryotes. Some species have acquired full length vault genes by horizontal gene transfer events and have undergone lineage-specific paralogous expansion. To determine if the observed compositional bias is conserved across all phyletic groups harboring vault genes, the sequence composition of vaults was analyzed across all the taxa. With over seventy-five amino acid sequences for *MVP* analyzed, it is shown that this amino acid bias is conserved across the phylogeny, with minor variations expected due to environmental conditions and G+C content.

The *vRNA* represents a very simple structure with a loop region and a stem region. Though multiple copies *vRNAs* are reported in a wide variety of organisms, it has been difficult to curate the entire set of *vRNA* due to its poor sequence conservation and high variability in sequence length. Apart from the hairpin region, the loop region is highly variable among the studied sequences. It is known that the expression of *vRNA* is elevated during Epstein-Barr virus infections by almost 1000 fold (Nandy et al. 2009). A *vRNA* bias towards U, G and C was originally reported by Rome ((Kedersha and Rome 1986; Kickhoefer et al. 1996). To evaluate if the nucleotide bias itself could be conserved feature across the *vRNA* sequences, a comprehensive collection of computationally predicted *vRNA* sequence dataset was used (Stadler et al. 2009). The nucleotide bias of *vRNA* across this dataset was analyzed and compared against basal genome compositions in organisms predicted to harbor *vRNA* genes (Table 3.1). Basal genome compositions were computed on downloaded complete genome assembly data for each organism from the

UCSC Genome Bioinformatics database. Based on the analysis, it was found that the U+G+C bias of *vRNA*s in mammals scores about 85.2% to that of basal eukaryotic genome composition of around 66.6%. The difference between the *vRNA* and basal eukaryote genome compositional bias was also pronounced in other groups – tetrapods, teleosts and basal deuterostomes. Akin to the paralogous expansion of *MVP* genes, the *vRNA* genes also seem to have duplicated into pseudo genes. Interestingly, the increased bias in nucleotide composition seems to be compromised in the pseudo genes (Table 3.2). Out of the 7 predicted *vRNA* pseudo genes, U+G+C percentage is lower compared to that of functional *vRNA*s in 5 genes, suggestive of a possible evolutionary pressure that maintains the *vRNA* sequence bias. Given that the sequence conservation is poor and highly variable, the conserved nucleotide bias is interesting. It also suggests that there has been an active evolutionarily pressure that maintains the U+G+C (but not adenine) bias in *vRNA*.

Remarkably, a compositional enrichment of amino acids and nucleotides is evident for both *MVP* and *vRNA*, as evolution proceeded from sea to land (Figure 3.2). This indicates that vault contains two evolutionarily conserved and distinctly compositionally biased biopolymers: a protein *MVP* and a nucleic acid *vRNA*.

Table 3.1 Comparison of U+G+C bias between *vRNA* and basal genome composition

	Organism	***vRNA* (%)**	**Genome (%)**
Mammals	*Homo sapiens*	85.51	65.57
	Pan troglodytes	86.57	60.84
	Pongo pygmaceus	85.66	63.15
	Macaca mulata	85.79	65.05
	Mus musculus	79.58	68.254
	Rattus norvegicus	84.62	64.26

	Cavia porcellus	89.3	68.43
	Oryctolagus cuniculus	83.76	68.4
	Equus caballus	83.41	70.33
	Canis familiaris	86.73	66.127
	Bos taurus	85.42	66.37
	Loxodonta africana	85.61	68.65
	Dasypus novemicinctus	85.44	70.01
	Monodelphis domestica	83.93	66.93
	Ornithorhynchus anatinus	87.25	67.12
Other Tetrapods	*Xenopus tropicalis*	81.08	62.88
	Gallus gallus	86.57	67.06
Teleost	*Danio rerio*	79.3	68.06
	Oryzias latipes	82.04	56.6
	Gasterosteus aculeatus	87.33	69.7
	Takifugu rubipres	82.65	63.78
	Tetraodon nigroviridis	88.37	61.7
Basal Deuterostomes	*Petromyzon marinus*	82.96	59.28
	Ciona intestinalis	80.83	63.32
	Branchiostoma floridea	81.51	64.99
	Strongylocentrotus purpuratus	78.897	61.2

Table 3.2 Comparison of U+G+C bias between *vRNA* (actual) and *vRNA* pseudo genes

Organism	Common Name	Pseudogenes (%)	*vRNA* (%)
Spermophilus tridecimlineatus	Squirrels	70.09	83.91
Procavia capensis	Rock Hyrax	80.39	86.73
Bos Taurus	Cow	78.49	85.42
Canis familiaris	Dog	76.19	86.73
Tursiops truncates	Dolphin	76.00	83.16
Oryctolagus cuniculus	Rabbit	85.57	82.41
Latimeria menadoensis	Coelacanth	81.44	74.75

3.2.2 *MVP* is a Unique Protein with High CAI

Translational efficiency, an important measure of protein expression within cells, depends on efficient use of codons. A significant correlation between mRNA levels and codon bias has been established in literature (Tuller et al. 2010). Codon bias corresponds to the differential usage of synonymous codons based on levels of corresponding tRNAs. Proteins that are

efficiently translated and expressed at elevated levels have the optimal codons that are recognized by an abundance of corresponding tRNA and are marked by high Codon Adaptation Index (CAI). The human *MVP* that is expressed in many cell types has a high CAI of 0.78 correlating with high copy numbers of vaults observed in purifications (Sharp and Li 1987; Kedersha and Rome 1986).

Apart from its abundant expression, vaults have also been reported to undergo turnover or degradation during various cellular conditions including interferon-gamma induction, embryonic development or following an axonal accumulation. The vault particle can be broken down into its constituent amino acids via ubiquitin-mediated proteasomal proteolysis or lysosome-mediated autophagy which consequently releases its ribonucleotides (*vRNA*) (Kedersha et al. 1990; Mortimore and Pösö 1987; Sutovsky et al. 2005). Given that naturally occurring vaults are polymerized structures constructed from 78 copies of *MVP* and also harbor few copies of minor vault constituents, an intact vault complex is a huge store of entrapped amino acids. With nearly 100,000 amino acids in a single vault complex and high reported copy numbers, the metabolic fate of degraded vault is significant because concomitant increases in *MVP* transcription, translation and turnover have also been observed (Steiner et al. 2006; Li et al. 1999). Taken in combination, these simultaneous increases in transcription, translation and turnover would represent a futile metabolic cycle of amino acid/ribonucleotide polymerization and release. In the absence of a clear distinct cellular function for vaults, the futile cycle itself could prove useful in controlling the concentrations of the free amino acid pool within the cell.

Any protein could essentially be turned over to derive useful amino acids to combat cellular stress. But what makes Vaults special? The most straightforward reason is its amino acid content and high copy numbers within cells. The vault composition bias is towards amino acids that are most readily utilized as nutrients in metabolism, with close to 50% contributed by Glu, Gln, Arg, Pro, Leu and Val. These amino acids could be routed through various metabolic pathways for cellular energy or may serve as precursors for synthesis of other useful molecules. The biased amino acids in vaults, Arg, Pro and Gln, can be efficiently converted into Glu and enter the tricarboxylic acid (TCA) cycle (Owen et al. 2002). Amino acids including Glu, Asp and Arg also serve to be important as neurotransmitters and are found in excess at axon terminals.

The compositional bias and CAI as a function of protein length was plotted for all annotated human protein coding sequences from the NCBI Consensus CDS protein data set (Figure 3.1B). Remarkably, *MVP* is one of the two proteins (the other being CARD10, a caspase recruitment domain family member 10 protein with a known function) that holds a nutritional amino acid bias in combination with a high CAI, and at the same time forms a large, stable and folded protein with no large regions of low complexity sequence and no intrinsically disordered regions. Naturally all proteins can be recycled but the observed bias in amino acid composition complemented with a high CAI appears to be a rare combination and makes *MVP* unique as a molecule. Also, the fact that it polymerizes into a stable and neutral macromolecular structure makes it an ideal storehouse of nutrient amino acids.

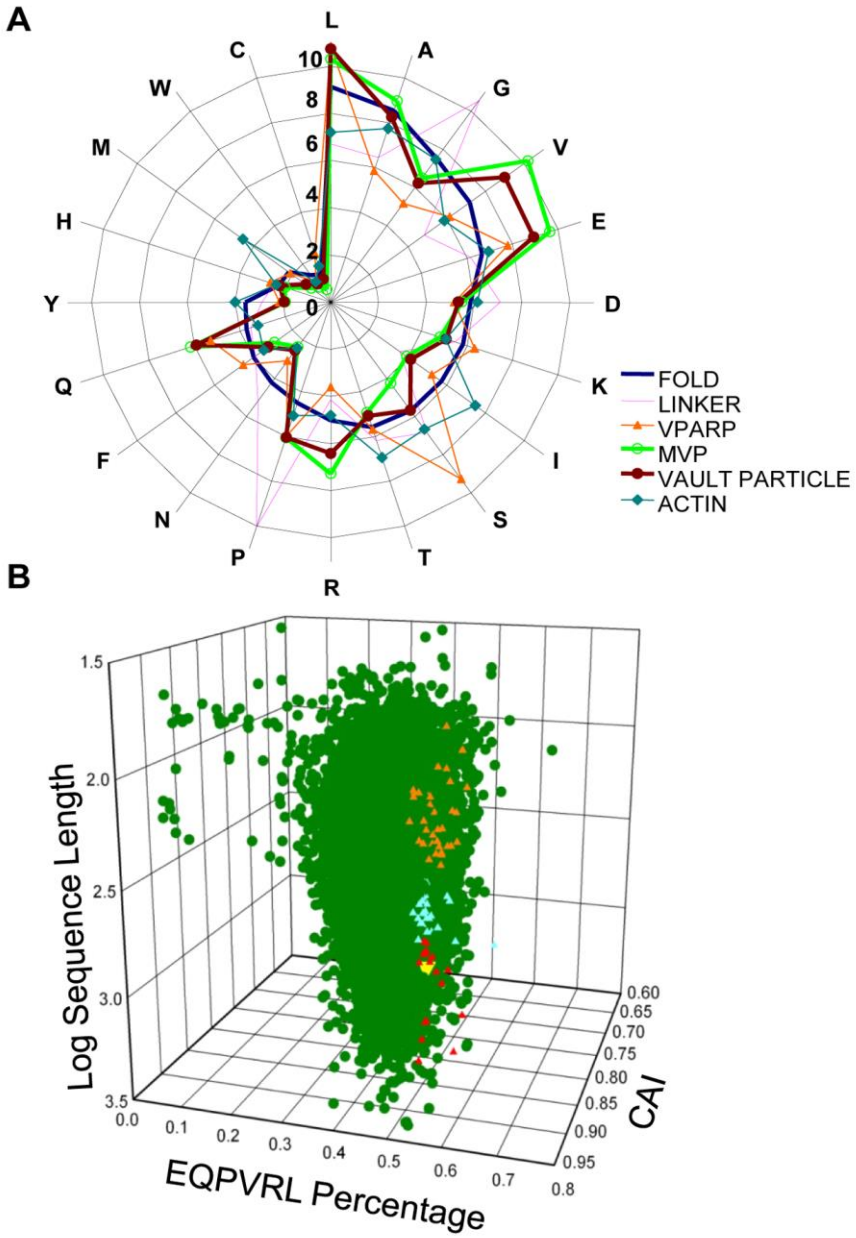


Figure 3.1Conserved compositional bias of *MVP* protein
(A) Composition of vault and vault genes compared to the average composition for folded proteins (FOLD) and unstructured linker regions (LINKER)(Dumontier et al. 2005). Amino acids are shown in clockwise order sorted from highest to lowest composition from the FOLD set, which appears as a smooth blue spiral starting at the top and going clockwise towards the centre. The human vault particle composition is estimated by including 78 copies of *MVP*, 12 copies of *VPARP* and 3 copies of *TEP1*.*MVP* exhibits higher than expected Leu, Val, Glu, Gln, Arg and Pro compared to FOLD, LINKER and actin (control). (B)Composition bias (EQPVRL Percentage) compared against Codon Adaptation Index (CAI) and length for all human protein coding regions obtained from the CCDS database. *MVP* is in yellow. Red represent sequences with length >600 amino acids, equal or higher CAI and compositional bias than *MVP* with significant low complexity regions (LCR). Blue and orange represent sequences that are similar in CAI and composition, but smaller than *MVP* (300-600 amino acids and less than 300 amino acids, respectively). The remaining proteins are represented by green dots.

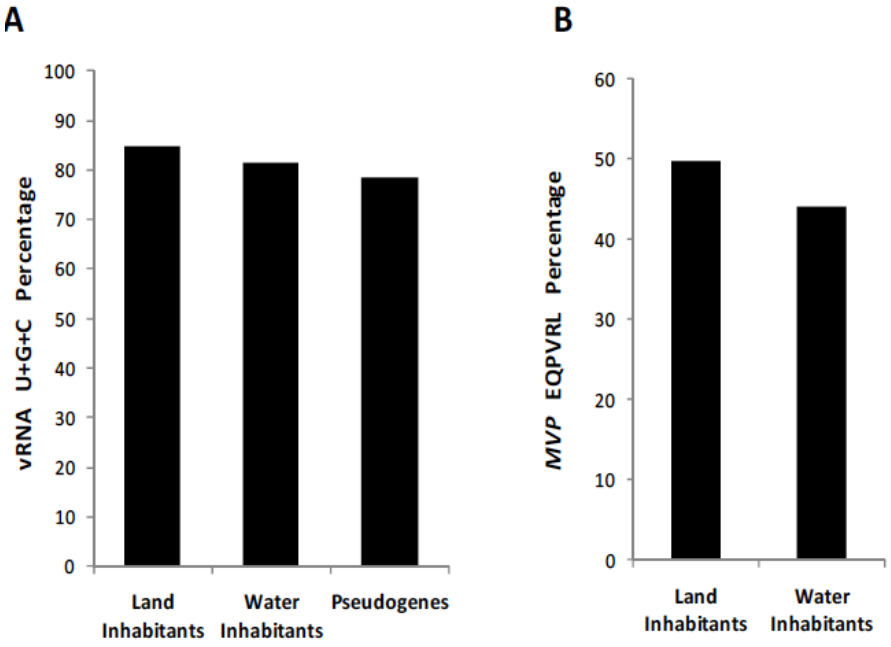


Figure 3.2 Compositional enrichment of vaults through evolution
Compositional enrichment of (A) U+G+C content in *vRNA* and (B) amino acids Glu, Gln, Pro, Val, Leu and Arg in *MVP* as organisms evolved from sea to land. Only deuterostome sequences were considered for this analysis. Also, (A) shows compositional enrichment of true *vRNA* compared with *vRNA* pseudogenes. For (B) only full length *MVP* sequences were used.

3.2.3 Recycling Vaults – A Reserve of Useful Precursors

Maintenance of cellular energy is an important determinant that controls cell growth and proliferation. However, during cellular stress, particularly starvation response, balancing cellular energy to promote cell survival becomes a challenging task. The cell tackles starvation primarily by burning its glucose and fat reserves. The glycogen reserves in muscles and liver cells come handy during this process. But during prolonged starvation, the cell resorts to breaking down proteins in muscles and other tissues to derive useful cellular energy, primarily by autophagy. The process is initiated when a portion of the cytoplasm is sequestered into a double-membrane structure called the autophagosome and fuses with a lytic compartment that hosts materials to be degraded. It has been known that mature ribosomes also

undergo rapid degradation upon nutrient starvation in yeast by a process termed ribophagy, a type of selective autophagy (Kraft et al. 2008). Other than autophagy, the ubiquitin-proteasome system can also rapidly degrade proteins when fast adaptation is needed. The turnover of vaults by such a mechanism has already been proposed in developing embryo (Sutovsky et al. 2005).

Vaults, being present in high copy numbers, particularly in metabolically active cells, could potentially serve as an energy reserve in response to various cellular stimuli including stress and nutrient starvation. Since the *MVP*-knockout models display no apparent phenotype and vaults are considered not essential for normal physiological functions, bulk degradation of vaults may not affect normal cellular functioning but could serve to provide a lot of amino acids within the cells. It has been well-established that amino acids can serve as a direct energy source for the cell and can also drive gluconeogenesis. It has also been established that vaults dissociate at low pH suggesting that the acidic nature of the lysosomes could trigger the process (Goldsmith et al. 2007). Hence, it could be speculated that vaults are engulfed into such lytic compartments and subjected to degradation.

To mimic such a process and account for the usability of vaults as a direct energy source or as precursors for important metabolic reactions, theoretical metabolic degradation of vaults and their constituent amino acids was performed. All the amino acids were subjected to canonical degradation pathways and the end products were routed through various metabolic pathways to determine the theoretical energy equivalent of one vault particle. This theoretical study assumed complete degradation of all amino acids constituting the vault complex. The total amino acid count for one vault

complex takes into consideration 78 copies of *MVP*, 12 copies of *VPARP* and 3 copies of *TEP1* (Table 3.3). The exact copy numbers of *VPARP* and *TEP1* within intact complexes remain unknown and hence, the indicated values are estimates (Anderson et al. 2007). The details of the theoretical metabolic degradation and routing of vault amino acids through the various catabolic or anabolic pathways are detailed in Appendices.

Table 3.3 Amino Acid compositions of *MVP* chain and structured Vault Complex

Amino Acid	1 *MVP*	78 *MVP* Chains	Composition of *MVP*	1 Vault Complex	Composition of Vaults
A	80	6240	8.96	8106	8.25
C	5	390	0.56	1002	1.02
D	49	3822	5.49	5241	5.34
E	**86**	**6708**	**9.63**	**8760**	**8.92**
F	26	2028	2.91	3189	3.25
G	58	4524	6.49	6117	6.23
H	18	1404	2.02	2208	2.25
I	35	2730	3.92	4053	4.13
K	43	3354	4.81	4992	5.08
L	**92**	**7176**	**10.30**	**10542**	**10.73**
M	9	702	1.01	1254	1.28
N	21	1638	2.35	2487	2.53
P	**54**	**4212**	**6.05**	**5928**	**6.04**
Q	**55**	**4290**	**6.16**	**5808**	**5.91**
R	**65**	**5070**	**7.28**	**6315**	**6.43**
S	38	2964	4.26	5571	5.67
T	44	3432	4.93	4983	5.07
V	**91**	**7098**	**10.19**	**8841**	**9.00**
W	7	546	0.78	918	0.93
Y	17	1326	1.90	1908	1.94
Total	893	69654	100	98223	100

3.2.3.1 Vault Amino Acids as Substrates for Gluconeogenesis

While glycogen serves as the primary reserve for glucose, nutrient deprivation leads to depletion of glycogen sources within one day. Hence, pyruvate or other tricarboxylic acid cycle intermediates in the liver are driven into gluconeogenesis to maintain glucose homeostasis. A bulk of the vault amino acids, except for leucine and lysine, on complete degradation either form pyruvate or other important intermediates that could be efficiently routed via the tricarboxylic acid cycle (Appendices, Table A.1). Pyruvate or oxaloacetate can be shuttled into the gluconeogenesis pathway for production of glucose. Taking into account total energy required for degradation of each amino acid in vault and total energy required for routing the intermediates through the gluconeogenesis pathway, in terms of ATP and other energy carriers including NADH, NADPH and $FADH_2$, it is found that vaults can economically prime the synthesis of almost 44248 molecules of glucose with net ATP to spare (Figure 3.3). Molecular biology grade glycogen purified from oysters (Fermentas Inc., Canada) has a maximal molecular weight of about 8x106, comprising about 50,000 glucose molecules. Thus, these two particles are nearly equivalent in carbohydrate energetic value.

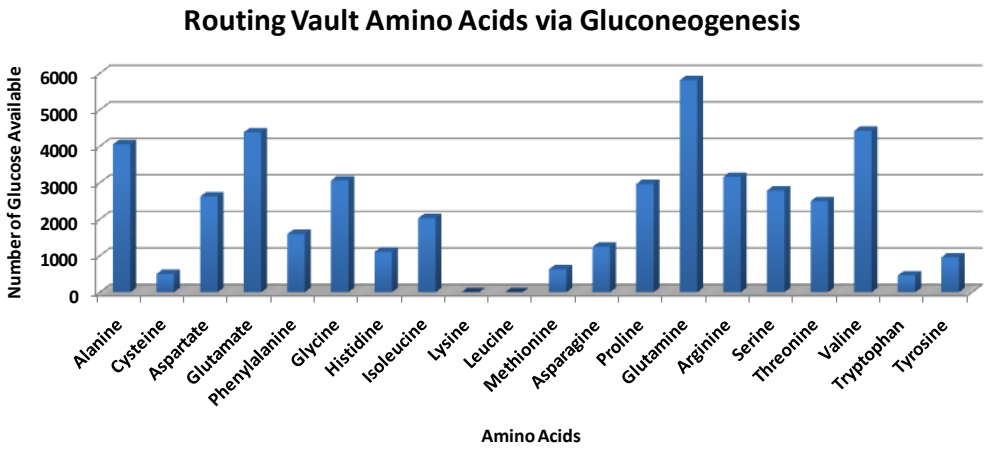


Figure 3.3 Glucose molecules recycled from one vault
Total amino acids from one vault particle are theoretically degraded into TCA cycle intermediates and routed into gluconeogenesis pathway for synthesis of glucose. Total energy involved in the form of ATP and other energy equivalents are taken into consideration. Refer Appendix for details on calculation.

3.2.3.2 ATP Equivalents Regenerated from a Degraded Vault Complex

Extending the above analysis and completely routing all the amino acid degradation products through pathways of tricarboxylic acid cycle and oxidative phosphorylation via electron transport chain, results in the formation of more than a million molecules of useful chemical energy in the form of ATP (Figure 3.4 and Appendices, Table A.2).

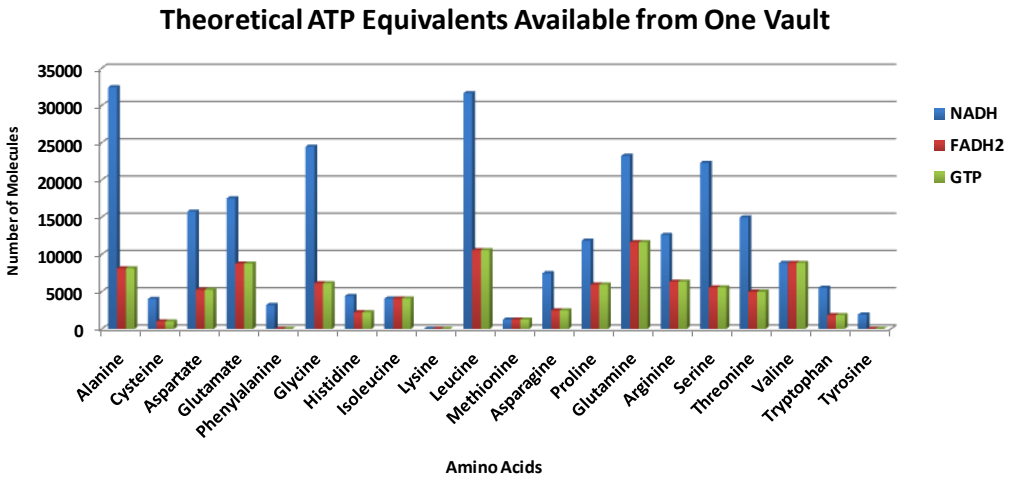


Figure 3.4 ATP Equivalents available from one vault
Total amino acids in one vault particle are subjected to degradation and intermediates routed to TCA cycle and oxidative phosphorylation. Each NADH molecule is equivalent to 3 ATP and each $FADH_2$ molecule is equivalent to 2 ATP molecules. Refer Appendix for details on calculation.

3.2.3.3 Vaults as precursors for *de novo* Nucleotide Biosynthesis

De novo synthesis of nucleotides necessitates the involvement of Asp, Glu and Gly in several of the key reactions. A bulk of the vault amino acids including Gln, Glu, Pro and Arg could be converted into either Asp or Glu and hence, be driven into the synthesis pathway. Cys and Ser could also be routed into the pathway once they are converted into Gly (Appendices, Table A.3 (Part 1). The *de novo* synthesis of both purine and pyrimidine involves the utilization of an activated sugar intermediate termed the 5-phospho-α-D-ribosyl 1-pyrophosphate (PRPP). While a subset of amino acids is directly involved in the de novo synthesis, the other amino acids can be efficiently routed for the synthesis of PRPP (Appendices, Table A.3 (Part 2). Such an efficient usage of amino acids in one vault particle could drive the synthesis of almost 9200 purine or 12500 pyrimidine skeletons theoretically (Appendices, Table A.3 (Part 3). Hypothetically speaking, the resources from about 5.6 x 10^5 vaults are sufficient to build an entire haploid human genome containing 3

billion base pairs. This appears to be also suited to scale, given the copy number of vaults reaching as high as 10^7 in embryos.

3.2.3.4 Assembling New Proteins from One Vault Particle

Protein recycling may become particularly important when nutrients are sparse and basal proteins promoting cell survival have to be constantly synthesized (Mizushima and Klionsky 2007). It has been established that the median length of eukaryotic proteins is about 361 amino acids (Brocchieri and Karlin 2005). Assuming assembly of average sized proteins with folded domain compositions, it is found that amino acids in vaults are capable of forming as many as 146 proteins and also provide for all the ATP required for protein translation machinery, including initiation, elongation, translocation and translation. (Appendices, Table A.4)

It is evident based on the current study that vaults display prime attributes to serve as a cellular energy reserve and could be regarded as a dependable source for energy, amino acid and also for release of various other metabolic precursors during periods of cellular stress.

3.2.4 Syntenic Conservation of *MVP* with BCKDK

Comparison of entire genomes in charting the evolutionary history of eukaryotes, particularly mammals, has led to the identification of a limited number of syntenic segments. Syntenic segments, termed the 'conserved linkage group', are marked by regions of highly conserved gene orders in the chromosome (Kemkemer et al. 2009). These segments signify a selection pressure that has been active in maintaining the gene order among the different organisms over the years of evolution. The *MVP* gene maps to the 16p11.2

genomic locus. To screen if *MVP* is also subjected to such an evolutionary pressure and maintains a specific gene order with its neighbors, the chromosomal regions in as many as 16 genomes were compared. The genomic regions neighboring to *MVP* were scanned using the Ensembl Genome Browser. On analysis, it was found that the particular locus in which *MVP* is present is syntenically conserved across all mammalian genomes including platypus. However, the conserved gene order, centered on *MVP*, is lost in the chicken genome. Synteny analysis revealed that at least 83 genes from mouse and 93 genes from chimpanzee, shared a conserved specific gene order upstream or downstream of *MVP*. Syntenic conservation of *MVP* gene locus between human, mouse and rat is shown in Figure 3.5.

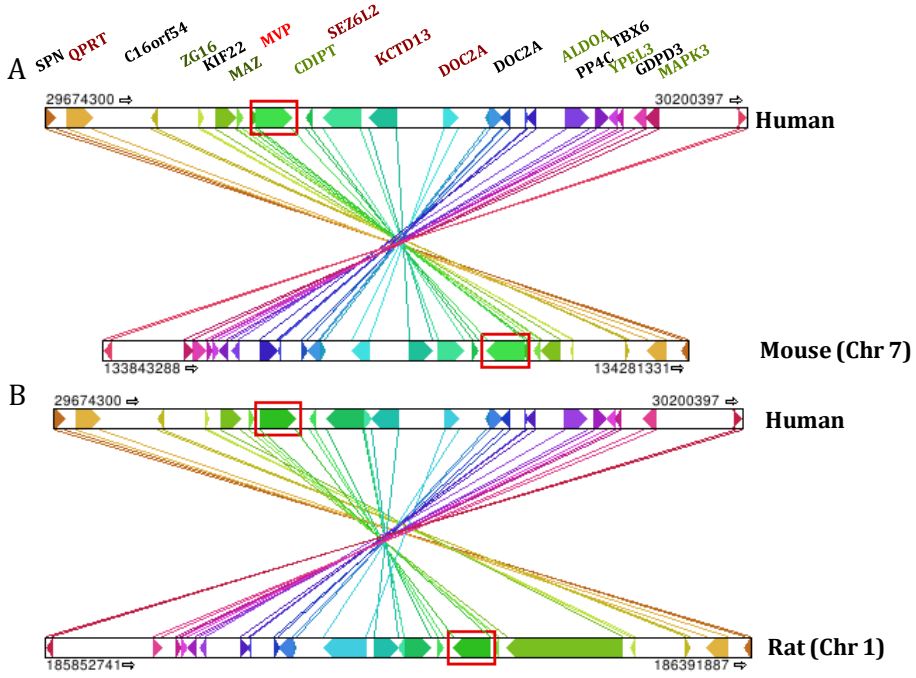


Figure 3.5 Synteny analysis of *MVP* gene locus
The conserved gene order of *MVP* and its neighbors along the chromosome is shown between (A) human and mouse and (B) human and rat. The *MVP* gene is shown in red box. The colored arrows and lines connecting them represent genes that are syntenically conserved across the locus. Important genes that function in metabolism are highlighted in green. Genes that play a role in neural development or neurodegeneration are highlighted in red.

Interestingly, *MVP* is also syntenically conserved with gene *BCKDK* (Branched Chain Ketoacid Dehydrogenase Kinase) within this region. Branched Chain Ketoacid Dehydrogenase (*BCKDH*) complex is an important regulator of the branched chain amino acid (BCAA) catabolism. BCAA are indispensable amino acids and account for almost 40% of the total body weight (Shimomura et al. 2004). *BCKDK* binding to the *BCKDH* complex reduces breakdown of BCAA by inactivating the complex, thus increasing BCAA pool within the cell. The proposed nutrient synthesis-turnover function for vaults agrees well with the observed gene association of *MVP* with *BCKDK*. The activity of the *BCKDH* complex is known to be up regulated during periods of nutritional stress to supply cellular energy. Conversely, amino acid deprivation enhances translation of *BCKDK* to increase the BCAA pool (DOERING and DANNER 2000). Concerted gene regulation, in the form of a negative feedback, or functional complementation of *MVP* and *BCKDK* may potentially determine the amino acid concentrations, particularly BCAA, within cells.

3.2.5 The overlooked vault function – Clues from expression patterns

Over the years, vaults have been implicated in various cellular functions. However, describing one precise function that can clearly explain all the observed expression patterns has remained complicated. The existing data, when subjected to meaningful analysis, reveals several clues that seem to correlate with the proposed function of vaults as a nutrient absorption particle that could efficiently function in a synthesis-turnover based manner in response to various cellular stimuli and remain dormant or neutral under

normal conditions. In this section, the expression pattern of vault in various tissues or cellular conditions has been re-analyzed in light of the proposed novel function. Vault expression data hidden in various microarray experiments has also been re-visited.

3.2.5.1 High Expression of Vaults in Nutrient Absorbing/Storage Tissues

Vault protein expression patterns in human tissues have been characterized previously using monoclonal antibodies(Izquierdo et al. 1996; Sugawara et al. 1997). The *MVP* protein expression data was normalized and plotted along with observed *MVP* mRNA expression profiles in tissues of mouse, rat and human, retrieved individually from NCBI GEO (Gene Expression Omnibus) database (Su et al. 2002; Walker et al. 2004; Ge et al. 2005). On analyzing the pattern of expression, it was found that intestines consistently exhibited highest vault mRNA or protein expression across all the four independent datasets (Figure 3.6). Another microarray survey comparing expressions of mouse intestinal mesenchymal and epithelial cells reveals that *MVP* mRNA expression is pronounced in intestinal epithelial cells as shown in Figure 3.9 (Li et al. 2007). Intestinal epithelial cells serve as prime sites for nutrient absorption and assimilation and the fact that *MVP* expression is highest in the intestine and specifically elevated in the epithelial cells accentuates the role for vaults as important players in nutrient absorption. Transcriptional profile data of Zebrafish intestine reveals that expression of *MVP* and *VPARP* is enriched in the anterior and middle intestine but fall rapidly towards the posterior end (Wang et al. 2010). The expression patterns of the vault genes also seems to follow trends displayed by other small molecule digestion and

uptake genes including *apoa1*, *vil1l*, *fabp2* among others (Figure 3.7). Thus, the expression evidence data indicates that vault particles are expressed to a greater extent in the vertebrate digestive tract, particularly in the absorptive tissues of the intestine and also in the liver, where they may be involved in secondary storage.

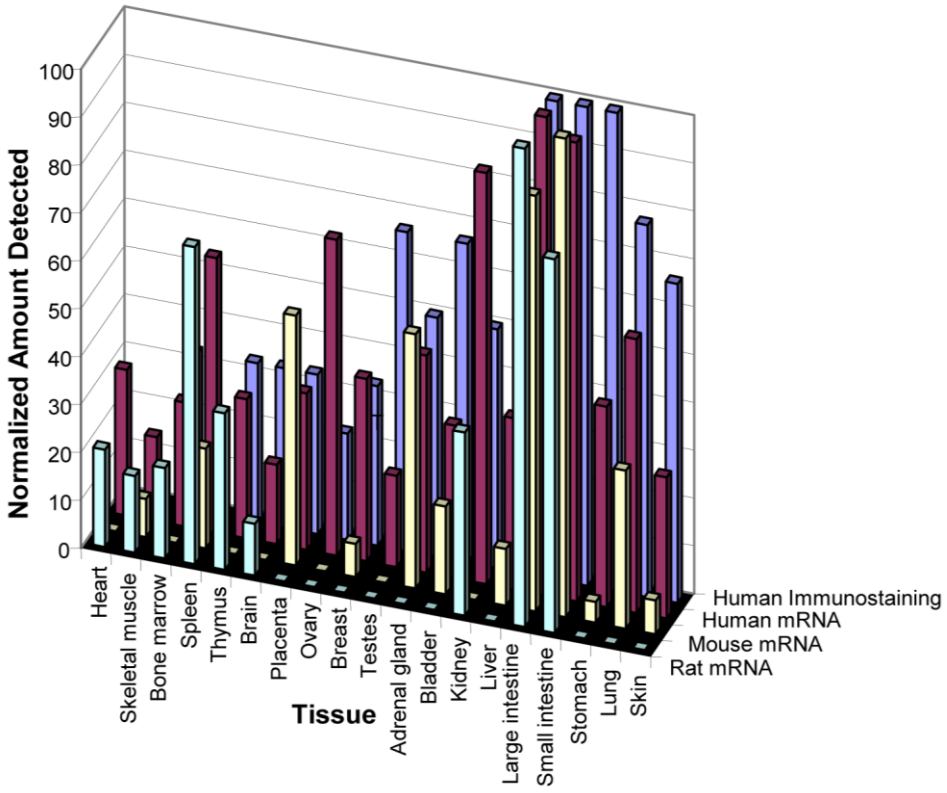


Figure 3.6 Comparison of tissue expression of vault
Normalized graph of tissue specific expression levels of vault particles obtained from human tissue immunostaining (Sugawara et al. 1997) and mRNA expression of the *MVP* gene from three separate microarray surveys (Su et al. 2002; Walker et al. 2004; Ge et al. 2005). Unweighted averages of cell and tissue types belonging to a particular tissue are reported in cases where multiple samples or cell types were measured in the original study. Zero values reflect tissue results not available from the original dataset.

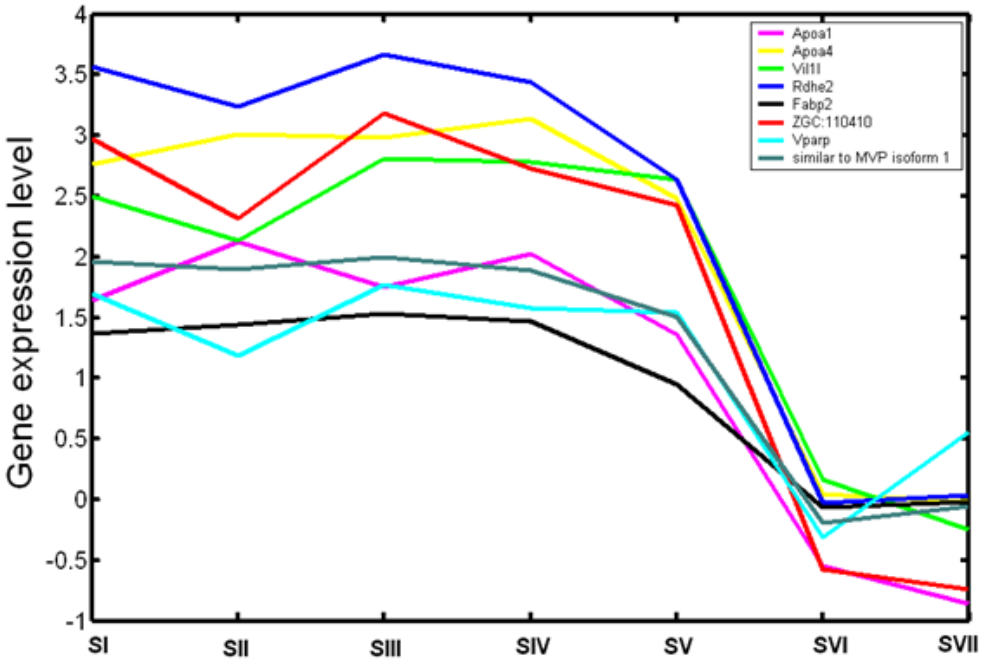


Figure 3.7 Expression of *MVP* and *VPARP* in Zebrafish intestine
Expression profiles *of MVP, VPARP* and other intestinal genes along the anterior-posterior intestine based on microarray results(Wang et al. 2010).Magenta: *Apoa1* - Apolipoprotein A-1; Yellow: *Apoa4* - Apolipoprotein A-4; Green: *Vil1l* - Villin 1 Like protein; Blue: *Rdhe2* - Short chain dehydrogenase/reductase family 16C, member 5; Black: *Fabp2* - fatty acid binding protein 2, intestinal; Red: *ZGC:110410* - similar to glutamate receptor, ionotropic, N-methyl D-aspartate-associated protein 1 (GRINA, glutamate binding); Light Blue: *VPARP* – Vault Poly ADP Riobose Polymerase; Grey: *MVP* - similar to *MVP* isoform 1.

3.2.5.2 Starvation and Vaults – Clear Patterns from *Dictyostelium*

Early *MVP* gene knockout experiments in *Dictyostelium discoideum* revealed a clear nutrient defect phenotype (Vasu and Rome 1995). *Dictyostelium* undergoes starvation-induced differentiation, beginning with single-cells in the vegetative state which then progress through many stages to form a multicellular fruiting body. Macroautophagy is an important determinant controlling this multicellular development of *Dictyostelium.* Autophagy mutants displayed reduced bulk protein degradation during starvation-induced development (Otto et al. 2003). Vault protein expression has been examined during *Dictyostelium* differentiation, and was found to decline from 0 to 34 hours after differentiation, with the latter time point

corresponding to the terminal fruiting stage (Kedersha et al. 1990). The protein expression pattern was plotted alongside mRNA expression trend of *MVP* observed by transcriptional profiling (Figure 3.8). The two *MVP* genes, *MVPA* and *MVPB* genes are expressed in the vegetative state and subsequently down regulated at the onset of starvation-induced differentiation (Iranfar et al. 2003). Thus, the data from both the experiments suggest that *Dictyostelium* vault particles become more concentrated during vegetative growth and that starvation halts vault expression. Since *Dictyostelium* depends primarily on protein catabolism to support cell survival during starvation-induced development phase, it is reasonable to speculate that vaults may contribute to a bulk of the protein-turnover process considering the amount of useful energy that could be obtained by vault degradation.

In *Dictyostelium*, it is very well established that degraded amino acids function as substrates for the TCA cycle and provide energy as cells aggregate and form fruiting bodies (Shiraishi and Savageau 1993) Since differentiation in *Dictyostelium* results in the catabolism of about 50% of the cellular protein, the simplest explanation for the previously reported nutritional stress phenotype in *Dictyostelium MVPA-*/*MVPB-* knockouts is that the highly abundant vault particles function as feedstock proteins and offer valuable amino acids that could fuel differentiation and development.

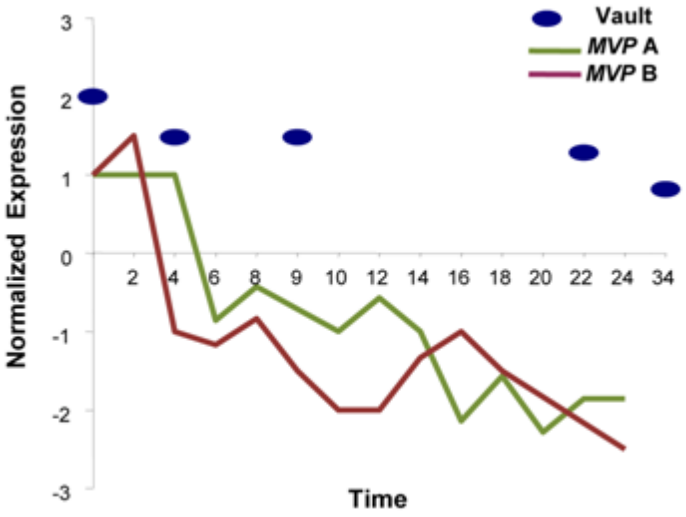


Figure 3.8 Dynamics of *DictyosteliumMVP* expression at the transcript and protein levels**
MVP A/*MVP* B mRNA expression (Iranfar et al. 2003) and vault particle expression approximated by pixel density counting from the western blot illustrated in Ref. (Kedersha et al. 1990) with the Unscan-IT-gel software (Silk Scientific Inc. Utah, USA). Time is counted after induction of differentiation at T=0 h. Vault particle western blot spot intensity is arbitrarily normalized to the value 2 at T=0 h.

A nutrient particle theory for vault function would imply that a vault-deficient *Dictyostelium* model would display defects in sequestering sufficient amino acids during vegetative phase and that it would have an adverse effect on cell survival during starvation-induced development stage. Consistent with this assertion, the*MVP A*$^{-}$/*MVP B*$^{-}$ knockout model of *Dictyostelium* displayed reduced growth rate and reached final cell densities of one-half to one-third to those of wild-type cells (Vasu and Rome 1995). Interestingly, this defect in cell survival or the observed nutrient phenotype became apparent only when the cells were subjected to limited nutrient medium (starvation) as the *MVP*-knockout cells appeared to grow normally in nutrient rich medium. Similar nutrient phenotypes have also been noted in *MVP*$^{-/-}$MEFs that display increased cell death on serum deprivation but show no obvious defects

otherwise. (Kolli et al. 2004). While other more complex vault functions could explain the limited growth during nutrient deprivation, the current theory still proves consistent with an impaired nutrient uptake in *MVP*-deficient models. These studies also heighten the argument that phenotypes of vault knockout models become evident only during conditions of cell stress or starvation. This provides a reasonable explanation as to why most of the studies employing *MVP*-knockout models failed to see a clear phenotype.

3.2.5.3 Explore the Unexplored – Hidden Clues from Microarray Profiles

The gene expression profiles of *MVP* from various transcriptional profiling experiments were manually mined from the Gene Expression Omnibus (GEO) repository and analyzed in light of the proposed nutrient sequestration function of vaults (Figure 3.9 and 3.10). On analysis, it was found that *MVP* consistently displayed a higher percentile of expression when the study involved effects of nutrition or development. Consistent with the proposed function, transcriptome profiles of several studies pointed to low levels of *MVP* expression in muscles and liver on starvation. The differential expression profile of *MVP* in various tissues during a fasting response over various time points is illustrated below in Figure 3.9 (Hakvoort et al. 2011). A distinct upregulation of *MVP* in intestines even during starvation response is remarkable (Figure 3.9). The down regulation of *MVP* in liver and muscles is suggestive of protein degrading as a starvation response to mobilize amino acids. In addition, transcript level changes of *MVP* in response to several other experimental or physiological conditions have also been analyzed. The fold

expression of *MVP* was determined by comparing the extracted gene expression measurements for the samples involved. A few of the significant or interesting patterns obtained are plotted in Figure 3.10.

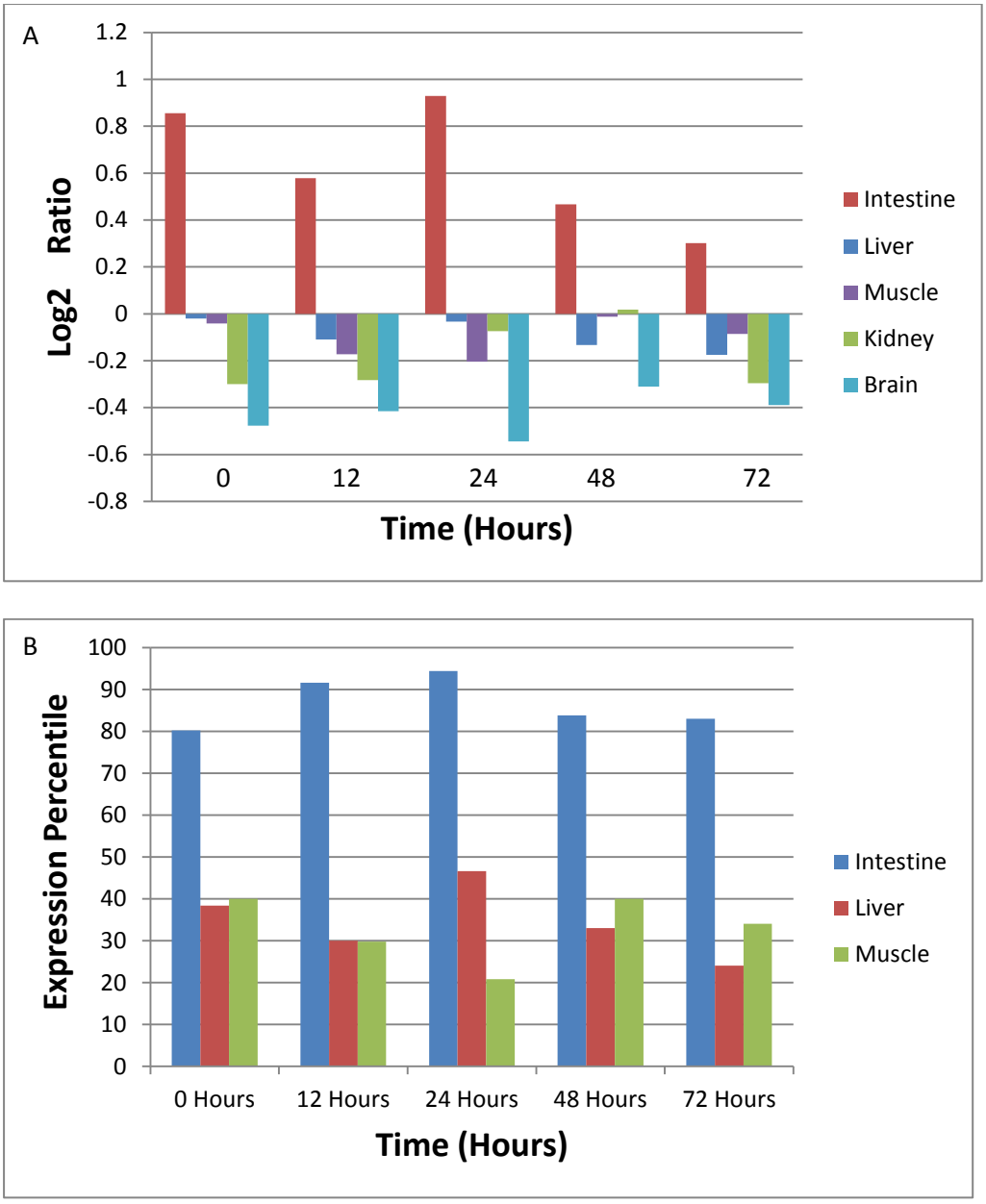


Figure 3.9 Changes in transcript profile of *MVP* expression across various tissues in response to fasting

(A) Differential expression of *MVP* during the indicated time points in response to fasting is indicated as log2 ratio. A log2 ratio of zero indicates no change in expression while a log2 ratio of 1 represents a 2-fold change with respect to control. (B) The rank order of expression measurements across the indicated time points for the various tissues during fasting. The percentile measurement is indicative of the expression of the gene with respect to other genes. A higher percentile ranks the gene high in terms of expression. Data obtained from (Hakvoort et al. 2011).

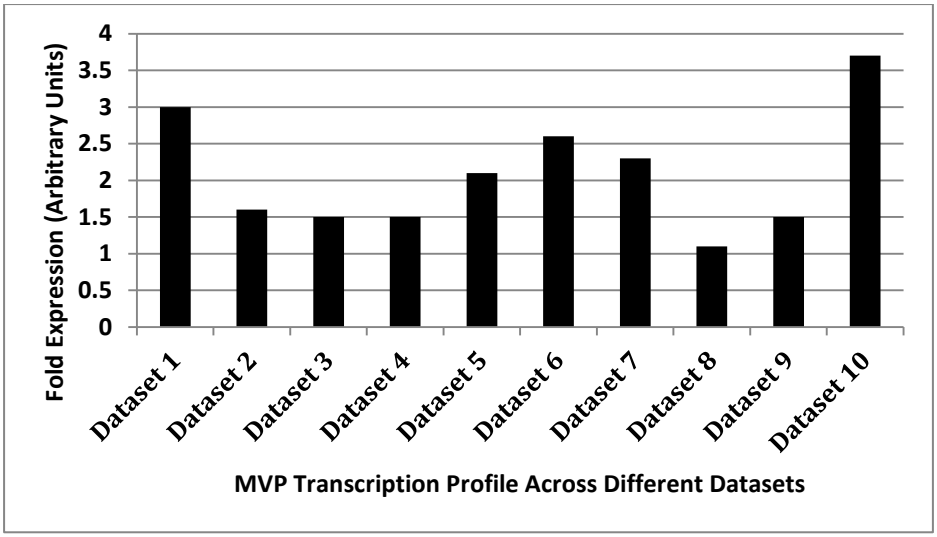


Figure 3.10 Expression of *MVP* across various transcriptomic profiles in the GEO database of NCBI

The experimental details of the various datasets are given below in legend. The fold expression was determined by comparing the averaged expression measurements between two samples. 'A/B' for various experimental conditions listed in the legend (below) refers to average signal count of sample A divided by average signal count of sample B.

Legend for Figure 3.10

GEO Profile	Experimental/Physiological Condition	Reference
Dataset 1	Mouse maternal mRNA utilization Embryo / Oocyte 1-cell	(Potireddy et al. 2006)
Dataset 2	B-lymphoblastoid cell line autophagy after 24 hours Starvation-induced autophagy / Control	(Dengjel et al. 2005)
Dataset 3	Hepatitis C virus (HCV) core protein effect on hepatocytes: Infection / Control	(Nguyen et al. 2006)
Dataset 4	Selective degradation of transcripts for oocyte maturation: Immature germinal vesicle stage / metaphase II stage (mature)	(Su et al. 2007)
Dataset 5	Human dermal enodothelial cell response to herpes virus: Control / Kaposi sarcoma-associated herpes virus (KSHV)	(Hong et al. 2004)
Dataset 6	Effect of ketogenic diet on liver: Control / Ketogenic Diet	(Kennedy et al. 2007)
Dataset 7	Skeletal myoblast differentiation to myotubes Myotube / Myoblast	(Chen et al. 2006)
Dataset 8	Effect of dietary supplement on liver: Liver / Control	(Kiela et al. 2005)
Dataset 9	Myotube starvation model of atrophy: Control / Starved	(Stevenson et al. 2005)
Dataset 10	Dietary effects on hepatic and hippocampal gene expression: Liver / Hippocampus	(Berger et al. 2002)

In addition, a large number of published vault experiments were re-examined to determine whether any observations would falsify the proposed function as a nutrient amino acid sequestering and storage particle that undergoes regulated synthesis-turnover cycles. High expression of vaults in rapidly proliferating and metabolically active tumor or regenerating cells, unusually high concentrations of vault particles in sea urchin oocytes and in mammalian oocytes and embryos, and their observed depletion throughout embryogenesis are observations that are readily compatible with the proposed nutrient absorption function for vaults(Sugawara et al. 1997; Rao et al. 2009; Yoshinari et al. 2009; Sutovsky et al. 2005; Stewart et al. 2005).

3.3 DISCUSSION

As originally noted by Rome in 1990, vault appears in all highly autophagic mammalian tissues (Kedersha et al. 1990). Autophagic recycling of amino acids is a well-established essential cellular function required for nutrient processing, and for surviving a number of conditions including starvation, muscle wasting, pathogenic infection, inflammatory bowel disease, neurodegeneration and cancer (Mizushima et al. 2008). Normal rat liver cells show 20-40% increase in protein content after a daily meal, which thereafter is released through lysosome-based autophagy back into the blood in a diurnal cycle (Mortimore and Pösö 1987). Protists including *Dictyostelium* rely on phagocytosis and autophagy to digest and recycle essential amino acids, and perhaps not coincidentally, the only filamentous cyanobacterium with *MVP* also relies on amino acid recycling to compensate for a genomic deficiency in nitrogen fixation (Otto et al. 2003; Jones et al. 2011). Based on vault expression patterns it could be concluded that vaults are preferentially distributed in tissues and organisms with a requirement for high protein turnover to support amino acid recycling. Hence, it appears likely that vault particles probably compensate for certain deficiencies in amino acid synthesis during periods of starvation, and probably this was the original selective pressure for the evolutionary origin of the vault particle.

In many studies where vaults were reported to be highly expressed at the transcript or protein level, an increase in turnover rate of vaults was also observed. This is a paradox corresponding to high Transcription, Translation and Turnover. Vaults are known to be relatively stable structures with an apparent half-life of about 3 days (Zheng et al. 2005). However, in response to

specific stimuli including interferon-gamma induction or accumulation following axonal crush, vault particles have been reported to display reduced stability (Steiner et al. 2006; Li et al. 1999). While, interferon-gamma induced the expression of vault by threefold at the transcript and by six- to eleven-fold at the protein level, pulse-chase experiments following interferon-gamma treatment revealed that the labeled methionine remained higher at 12^{th} hour than at the 24^{th} hour, pointing to a reduced stability of vault particles and hence, increased turnover (Steiner et al. 2006).

In the case of axons a bulk of vaults were reported to be accumulated at the nerve terminals along with synaptic vesicles by anterograde transport. Based on the signal obtained during retrograde transport it was speculated that more than half of them were degraded at the nerve terminals as only a few returned towards the cell body (Li et al. 1999). Intriguingly, vaults displayed an accelerated accumulation at the nerve terminal while the accumulation of synaptic vesicles, the store house of neurotransmitters decelerated. An increased turnover following accumulation accentuates that vaults are bound to release their amino acids at the site of degradation. Given that one vault complex is a massive store of amino acids, the turnover of vaults at specific sites may prove beneficial to the cell. As mentioned earlier, vaults are enriched in amino acids Glu (8760/vault), Gln (5808/vault) and Asp (5241/vault) that serve as prime neurotransmitters and the degradation of vault is bound to be accompanied by a concomitant increase in useful amino neurotransmitter or neurotransmitter precursors. These neurotransmitters can later be loaded into the existing synaptic vesicle via neurotransmitter transporters. Of note, in the

PC12 cell line, *MVP* and secretory organelles like synaptic vesicles demonstrate co-localization at the developing neurites (Herrmann et al. 1999).

Recent transcriptome profiling identified *MVP* to be up regulated in both larvae and in the adult during Zebrafish fin regeneration (Yoshinari et al. 2009). Regeneration is a complex process that necessitates various coordinated cell events including wound closure by epithelial cells and formation of wound epidermis followed by formation of blastema, a mass of proliferating cells that makes up for the lost or damaged parts. (Yoshinari and Kawakami 2011). While Zebrafish finfold regeneration reported 2.7 fold elevated levels of *MVP* in wound epidermis and blastema, a proteomic profiling data of adult urodele limb regeneration reported reduced protein levels immediately following amputation. Interestingly, the protein levels came back to normal once the regeneration process was completed. (Yoshinari et al. 2009; Rao et al. 2009). The two independent datasets are suggestive of an increased transcription and increased turnover of vaults, consistent with the paradox mentioned previously. Current literature data point to an indispensable role for vaults in regeneration as *MVP*-knockdown resulted in compromised locomotor functions and reduced growth of axons following spinal cord injury in Zebrafish(Pan et al. 2013). Given the indispensable role of *MVP* during regeneration, the proposed role of vault as a nutrient particle may be prudently implicated in such a process.

The accumulation and turnover function may also prove advantageous during embryogenesis as degradation of maternally derived proteins has been known to supply nutrients to the developing embryo. The fact that *MVP* is accumulated in poor quality oocytes and embryos and also that such an

accumulation could be triggered in the presence of proteasomal inhibition in developing zygotes heightens the current argument that vaults could be effectively turned over to provide for valuable amino acids.

MVP is strategically located at a chromosomal region, the deletion or duplication of which has been implicated in various neurodegenerative disorders including Schizophrenia and Autism and also energy imbalance (Guha et al. 2013; Zufferey et al. 2012). Upregulation of *MVP* has also been reported in other neurological conditions like frontal lobe epilepsy (Liu et al. 2011). About 1% of patients with ASD have displayed microdeletions of about 27 genes in this region and hence, these deletions are regarded as moderate risk factors for ASD. Recently a genomic study has narrowed on a critical deletion region of 5 genes that contributes to autism disorders (Crepel et al. 2011). Intriguingly, *MVP* is one among the 5 identified genes along with *CDIPT1*, *SEZ6L2*, *ASPHD1* and *KCTD13*that are syntenically linked. The SEZ6l2 (seizure related 6 homolog-like 2), like the *MVP* is also overexpressed in lung cancers and is regarded as a prognostic marker.

Defects in functioning of the syntenically linked gene mentioned earlier, *BCKDK*, is also found to be the underlying cause for a rare form of autism (Novarino et al. 2012). The symptoms of this particular form of autism, arising from significant low levels of plasma and brain BCAA, could be completely abolished within one week of nutrient supplementation with BCAA enriched diet. Since amino acids serve as important precursors for neurotransmitters and given the accumulation of *MVP* in axons and nerve terminals, it is reasonable to suggest that vaults are important for brain function, with a more pronounced effect during specific cellular conditions

including stress or starvation. The evolutionarily conserved gene order at this particular chromosomal location and its implications in various neurodegenerative disorders and energy imbalance possibly hints at a controlled gene regulation mechanism that could be active during early phases of development.

Thus, from the theoretical analysis of the vault particle in terms of its energy equivalents and from revisiting the huge amount of evidence reported in literature, it seems very likely that vaults could function efficiently in sequestering and recycling amino acids under appropriate cellular conditions.

Chapter 4

Proposed Roles of Vault through Evolution

Describing one cellular function for vault that is consistent with all experimental evidence established thus far has proven to be complicated. The proposed synthesis-turnover based nutrient absorption function for vaults seems to fit well with their observed high expression patterns in various metabolically active cells and also provides credible reasoning as to why vaults display selective accumulation and high turnover rates in response to specific stimuli.

In this section, a cellular model has been described that puts the various roles of vaults in light of the proposed novel function (Figure 4.1). Experimental evidence on the functions of vault relating to autophagy, protein turnover, essential amino acid transport, interferon-gamma induced immune response and signal transduction in axons have been described through an evolutionary perspective to cumulatively sketch out the functions of vaults.

4.1 STARVE THE INVADER – SAVE THE CELL

The earliest bacterial and archaeal ancestors of eukaryotes that harbored the ancestral proteasome possessed core metabolic pathways. Later, eukaryotes evolved to include features like the diversification of membrane trafficking and lysosome-based autophagy, reorganization of protein synthesis, marked by complexation of aminoacyl-tRNA synthetases and enlarging of the ribosome. Early large eukaryotic cells may have been prone to intracellular invasion from viruses and bacteria. In some cases these invaders may have become endosymbionts, as in the case of the chloroplast. In other cases, the development of an innate immune system to guard against these invaders could have provided selective advantages. A simple mechanism whereby the host removes free nutrient amino acids from the cytoplasm by increasing its protein synthesis could serve as an effective deterrent against invading bacteria. Nutrient starvation as an effective mechanism against infections has already been described for tryptophan (Kane et al. 1999; Leonhardt et al. 2007). The polymerization of amino acids into a particle like vaults could have effectively served to limit the concentration of amino acids in large cells and at the same time store the precious amino acids in a non-functional but structurally stable protein. Thus vaults could have evolved as effective host mechanism against intracellular invaders. The later evolutionary inclusion of *vRNA* could have provided a similar mechanism to limit the concentration of free ribonucleotides, with the exception of A, for ATP dependent functions. This could be effective in starving RNA viruses of free ribonucleotides.

4.2 ESTABLISH AMINO ACID GRADIENTS WITHIN CELLS

The loss of essential amino acid biosynthesis pathways for Ile, Val, Leu, Phe, Trp, Met, Thr, His appears to have occurred prior to the eukaryotic appearance of vault. However, it is not possible to make a phylogenetic

analysis of lost genes for a direct comparison. Most of the essential amino acids including the important BCAA set (Leu, Val and Ile) transit through exchangers into the cell. Therefore, cellular intake of essential amino acids follows a gradient established by external and internal amino acid concentrations (Hyde et al. 2003; Mortimore and Pösö 1987; Bröer 2008). Maintenance of intracellular amino acid concentrations higher than that of extracellular environment depends on active membrane transport mechanisms. *Primarily relying on amino acid exchangers, how does the cell effectively maintain a gradient for a constant supply of free essential amino acids?* Amino acid exchangers such as System L function as a 1:1 amino acid exchanger and couple the uptake of essential amino acids with efflux of other neutral amino acids like glutamine. Because efflux of a neutral amino acid is necessary for an uptake, these exchangers rely on activity of secondary active transporters to maintain a continuous supply of neutral amino acids in the cell. Most essential amino acids are indispensable for the normal functioning of the cell. This stresses on the need for a judicious mechanism to retain them within cell membranes for use when they are insufficiently supplemented through food intake. Intracellular protein synthesis can alter the extracellular to intracellular amino acid gradient and hence it might be reasonable to speculate that some specific proteins may have emerged to take on a passive role as storage proteins. By evolving to achieve high rates of synthesis and precisely controlled mechanisms for turnover, such proteins could play a more active role in amino sequestration and release during various cellular conditions. It is possible that vaults may have originated to serve this function in a unicellular ancestor, and later evolved as efficient nutrient supplements for zygotes beginning with the first multicellular animals.

4.3 MEDIATE IMMUNE COMBAT

In metazoan tissue, regulation of vault genes and the establishment of interferon-gamma induced expression may have offered finer regulation of vault expression as required during normal development or while under attack from bacterial or viral invaders. A cruder multi-copy strategy for *MVP* may have served this purpose in the non-bilaterian deuterostomes. The later establishment of *vRNA* and *TEP1* may have been responsible for additional nutrient starvation of intracellular viral invaders (Nandy et al. 2009). This could effectively reduce the cytoplasmic concentrations of free ribonucleotides, especially C, G and U. It is known that interferon-gamma results in high expression rates of the *MVP* gene, both at the transcript and protein level. Apart from the effect on vault, interferon-gamma based upregulation of tryptophanyl-tRNA synthetase and indole-2,3-dioxygenase could also trigger additional nutrient starvation responses involving antimicrobial toxins derived from Trp (Schroecksnadel et al. 2012; Wood et al. 2004; Narui et al. 2009). Interferon-gamma inducible Nitric Oxide Synthases (iNOS) offer further antimicrobial effects from NO (nitric oxide) derived from Arg (Bronte and Zanovello 2005). Vaults being enriched in Arg could very well serve as a dependable reservoir for the NOS systems. *MVP*-knockout mice suffer from poor survival rate due to increased bacterial load and decreased ability to clear *Pseudomonas aeruginosa* during lung infection (Kowalski et al. 2007). Apart from regulating the immune response, sequestering of amino acids into vaults may simply help in maintaining low intracellular amino acid and nucleotide concentrations that could effectively control the growth of invaders. Hence vault may play an important role in innate immunity through intracellular pathogen starvation.

4.4 A RELIABLE STORE OF AMINO ACID-BASED NEUROTRANSMITTERS

The abundance of vault in neuronal tissue, including the glial cells, its anterograde and retrograde axonal transport and its distribution proximity to neurosecretory organelles also fit the proposed novel function for vaults(Li et al. 1999; Herrmann et al. 1996; Chugani et al. 1991). Vaults have been found to move along individual microtubules and are also highly expressed in presynaptic and postsynaptic structures (Paspalas et al. 2009). It has been speculated that more than half of the vaults that accumulate within nerve terminals are degraded (Li et al. 1999). It may be prudent to consider that amino acids, Glu, Asp and Gln, could be delivered to axon terminals via vault particles to stock or replenish vesicle bound neurotransmitter stores while other amino acids like Val may help power remote mitochondria. A clear mechanism detailing the bulk accumulation or movement of amino acids within cells has never been put forward. The movement of vault particles by motorized transport along microtubules effectively explains the transport of amino acids to distal parts of cells without the need for coordinated mechanisms involving amino acid gradients or exchangers. These aggregation and transport mechanisms would become very important in conditions of prolonged starvation, which may explain the lack of observed vault phenotypes in consistently fed animal models. Morphological differences in size of cell bodies and axon length are noted in edycosozoan protostomes lacking vault, as well as the lack of centralized brain tissues, which are found in lophotrocozoan protostomes.

4.5 AN ELUSIVE RESERVE OF ENERGY AND BUILDING BLOCKS

Oocytes or embryos at various stages of development have consistently showed accumulation of vaults around lipid inclusions or membrane vesicles in cytoplasm (Sutovsky et al. 2005). An ubiquitin-proteasome dependent turnover for vaults has been suggested and vaults are found to accumulate in poor quality oocytes or embryos (Sutovsky et al. 2005). Collectively, this evidence suggests that vaults could succumb to significant turnover by the proteasome machinery and possibly, also through autophagy, and provide amino acids to cells that most need them. The developing or regenerating cells are in a constant demand for nutrients that serve as building blocks for cell proliferation, cell specialization and overall development. Not only can vaults provide an amino acid reserve but also serve to provide important building blocks for nucleotide or new protein synthesis. This can explain why vaults are present in copy numbers as high as 10^7 in embryos (Hamill and Suprenant 1997).

Vaults have been found within the nucleus in numerous occasions. It has also been reported that the expression of *vRNA* increases about 1000 fold during infections. With a conserved compositional bias, the *vRNA* can indeed function as a nucleotide trapping mechanism to polymerize U, G and C nucleotides. Since ATP represents the energy unit of the cell, adenine nucleotides are probably not captured and polymerized as *vRNA.* Also, with vault amino acids providing sufficient starting material to drive *de novo* synthesis of nucleotides, it is reasonable to speculate that vaults move into the nucleus to possibly provide for nucleotide biosynthesis. All these mechanisms

may become more pronounced during high metabolic activity or during conditions of stress or starvation. The reported accumulation of vaults within the nucleus during embryonic development could be explained in light of this proposed function (Hamill and Suprenant 1997).

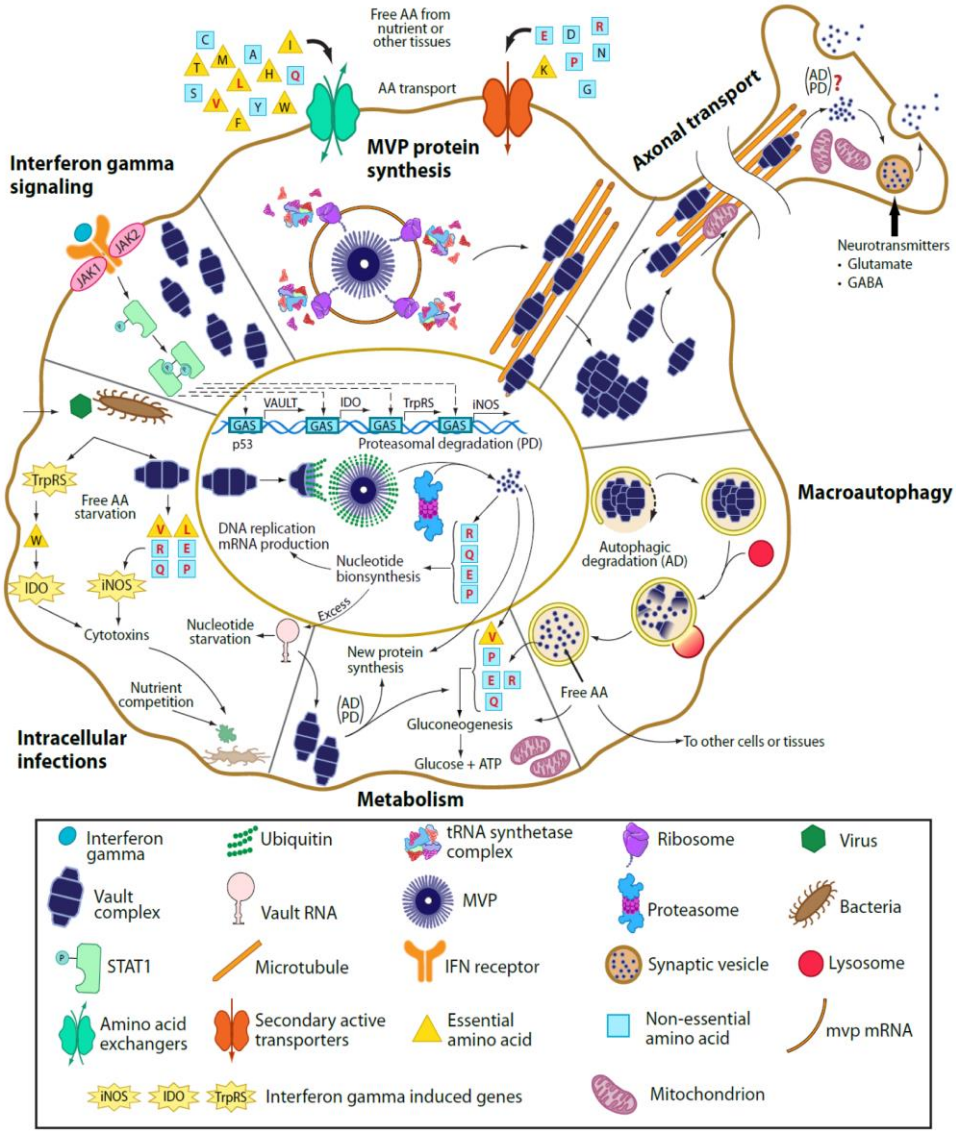


Figure 4.1 Proposed schematic model explaining role of vaults in a generalized cell
Various reported vault functions including axonal transport along microtubules (Eichenmüller et al. 2003), vaults in the nucleus, interferon-gamma induced *MVP* expression(Steiner et al. 2006), vaults in *Dictyostelium*(Vasu and Rome 1995) and sea urchin oocytes(Hamill and Suprenant 1997) (macroautophagy), high expression of *MVP* and *vRNA* during infection(Berger et al. 2008) and proteasomal degradation of *MVP*(Sutovsky et al. 2005) are explained in the context of the nutrient particle hypothesis for the vault complex. Based on our analysis of metabolic turnover, the glucose, ATP, nucleotides and new proteins possibly made from recycled vault particle have also been represented. Each functional segment in the model is based on experiments and observations from the literature. The vault enriched amino acids Val (V), Leu (L), Glu (E), Gln (Q), Pro (P) and Arg (R) are indicated in red and AA denotes amino acid.

Chapter 5

Evaluating Nutrient-Related Phenotype in Ancient Single-Celled Eukaryote

5.1 INTRODUCTION

The protozoan parasite *Trypanosoma brucei* is the causative agent of African trypanosomiasis that causes sleeping sickness in humans and Nagana in cattle. Two subspecies of this parasite, the *T. brucei gambience* and the *T. brucei rhodesiense*, are morphologically indistinguishable parasitic hemoflagellates,and are known to be responsible for chronic and acute infections in humans, respectively. The disease is characterized by two stages, a haemolymphatic stage associated with swelling of lymph nodes and a neurological stage that develops when the parasite crosses the blood-brain barrier. The disease is extremely fatal without medical intervention and the development of an efficient vaccine is challenged by antigenic variation of the surface protein exhibited by this parasite.

The parasite displays a complex life cycle that alternates between the insect (tsetse fly) and the mammalian hosts. The infection in mammals begins when the tsetse fly makes a blood meal and injects metacyclic trypomastigotes into the blood stream. The parasite undergoes transformation in the mammalian host into a long slender bloodstream form that starts to multiply in various body fluids including blood and lymph nodes. The proliferative form undergoes a density dependent differentiation into a non-proliferative short stumpy bloodstream form that has the kinetoplast, the mitochondrial genome holding organelle, in the terminal position. When the insect makes a blood meal, these short-stumpy bloodstream forms are ingested and differentiate into procyclic trypomastigotes in the midgut of the insect and start to divide. The procyclic form is morphologically different from the bloodstream form and has an extended and elaborated mitochondrion. In about two weeks, some of these proliferative forms leave the midgut and reach the salivary gland, where they become the epimastigotes and remain attached to the salivary gland. The epimastigotes later multiply and again convert to the metacyclic form that infects the mammalian host, thus completing one cycle of infection.

During its complex life cycle through the mammalian and insect hosts, the extracellular parasite undergoes marked morphological changes and displays different metabolic requirements. The cell typically alternates between stages of proliferation and stages of differentiation as it adapts to the varied environment. The most fascinating feature of these parasites is their ability to alter their surface proteins on differentiation and during infection. In the insect host, the procyclic forms expresses procyclins and as it differentiates into the bloodstream form it expresses variable surface glycoproteins (VSG)

that helps evade host immune responses. The parasite has the ability to express a new set of surface glycoproteins in approximately 1 out of every 100 cell divisions, thus challenging the host immune system.

T. brucei contains several single copy organelles including a motile flagellum characterized by an axoneme that is closely associated with a single pair basal body situated at the base of the flagellum responsible for axoneme duplication, a single Golgi and associated ER exit and a kinetoplastid representing the compacted genome of the elongated mitochondrion. Though both the procyclic and bloodstream forms have been studied in vitro, a majority of cytological studies in understanding the parasite have been carried out in the procyclic form (McKean 2003).

5.1.1 Cellular Architecture

The complex cellular construction retains the shape and form of the cell relatively unchanged during the course of cell-cycle events. Cytokinesis in *Trypanosoma* is a highly regulated process that initiates at the anterior end of the cell and continues longitudinally along the posterior end of the cell. The duplication and segregation of organelles during cell division are intricately orchestrated in a spatial and temporal manner, complimented by a well-defined cytoskeletal structure that maintains the shape of the cell (Woodward and Gull 1990; Gull 1999). In this section the architecture of the parasite in light of cell division is described.

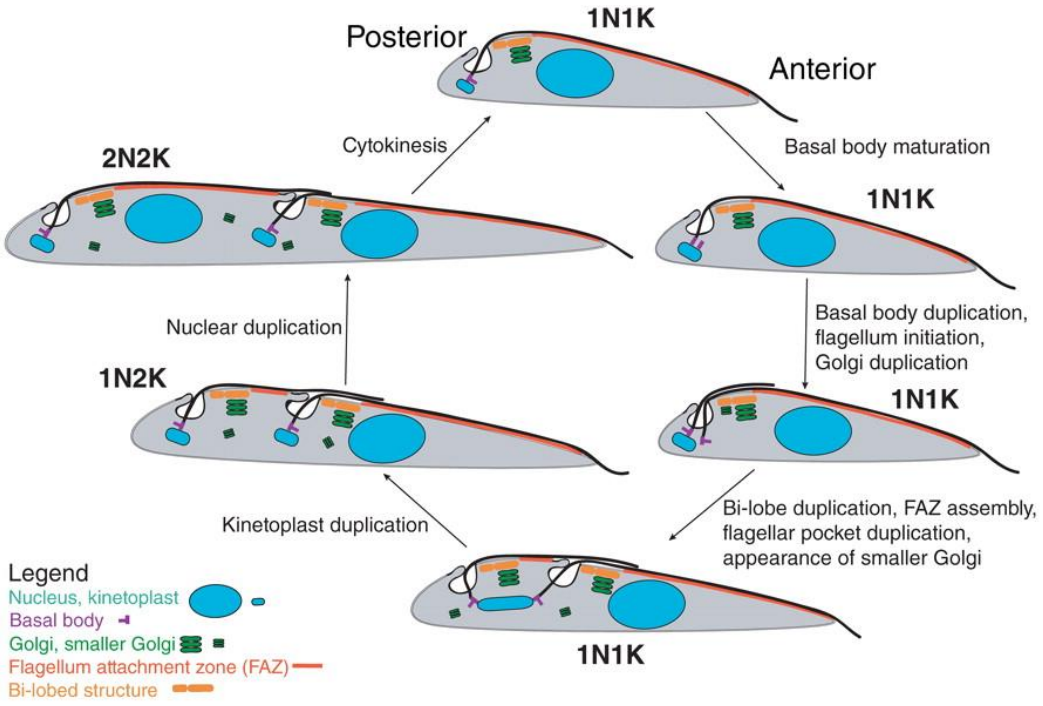


Figure 5.1 Schema showing *Trypanosoma brucei* cell cycle events
The various morphological changes that occur during the course of the cell division and cytokinesis events are depicted. Figure reprinted with permission from (de Graffenried et al. 2008), © de Graffenried et al. 2008. Originally published in Journal of Cell biology, doi 10.1083/jcb.200708082

5.1.1.1 Flagellum

The most prominent feature of *Trypanosoma* is the single flagellum that plays a significant role in cell morphogenesis events. The flagellum is characterized by a canonical 9+2 microtubule axenome and an extra-axonemal paraflagellar rod (PFR) that can function like a biomechanical spring by transmitting energy derived from axonemal beating (Hughes et al. 2012). The flagellum emerges from the flagellar pocket at the posterior side of the cell and remains attached through the length of the cell body with an overhang at the anterior side of the cell. The flagellar pocket serves as an important player in the parasites' defense against the host system, serves as the only site of

clathrin-mediated endocytosis and exocytosis and is also involved in the recycling and trafficking of surface proteins of tsetse midgut form, procyclin, and of the mammalian bloodstream form, variable surface glycoprotein (VSG) (Allen et al. 2003; Field and Carrington 2009). Duplication of the basal body and nucleation of the flagellum are the first events that mark the beginning of the cell cycle. As the flagellum exits via the flagellar pocket, events leading to assembly of external PFR and cytoplasmic FAZ components initiate and elongate over the course of the cell division (Kohl et al. 1999). Resistance builds up when the flagellar connector, that marks the distal tip of the new flagellum, reaches the old flagellum and a series of segregation events follow (Davidge et al. 2006). As the organelles duplicate they retain their close association with the new flagellum. It is interesting to note that the flagellum also displays remarkable control over cytoskeletal structures as non-flagellated cells are short, devoid of polarity and do not undergo cytokinesis (Kohl et al. 2003).

5.1.1.2 Flagellar Attachment Zone

In *T. brucei* the flagellum exits from the flagellar pocket and remains attached through the length of the cell body by a structure termed the Flagellar Attachment Zone (FAZ). It forms a link between the cell body and the flagellum by defining discrete structures between the cell and flagellar membrane (Lacomble et al. 2009). The structure is characterized by a quartet of specialized microtubules (MtQ) and an electron dense protein filament, FAZ filament. The FAZ which originates at the basal bodies plays a very important role in setting the initiation site of the cell cleavage furrow. The axis of cytokinesis cleavage begins at the anterior end of the new FAZ (Robinson

et al. 1995). Like in most eukaryotes, the cytokinesis initiation is modulated by Polo-like kinases that associate with the FAZ region (Li et al. 2010). It is interesting to note that that attachment of flagellum to the cell body and its growth positively regulates assembly of FAZ and a shorter FAZ results in shorter daughter cells (Vaughan et al. 2008; Vaughan 2010). The FAZ along with the growing new flagellum co-ordinates early events including basal body segregation and subsequent events pertaining to other organelles and structures (Absalon et al. 2007).

5.1.1.3 Cytoskeletal Structure

The cytoskeleton in *Trypanosoma* is defined primarily by microtubules representing the flagellar axenome and those of the subpellicular microtubule corset. Other microtubule based structures include the basal bodies and the intranuclear mitotic spindle. The subpellicular microtubule corset, consisting of up to 100 microtubules, forms a layer of connected microtubules under the cell membrane and helps the parasite retain its shape. It is also worth noting that the inter-microtubule distance remains constant even when the cell volume increases, due to the presence of specific linkers that cross link them with each other and with the plasma membrane (Lacomble et al. 2009; Hemphill et al. 1991). The plus end of the microtubule is placed at the posterior end of the cell and new microtubules intercalate and insert into existing microtubules as cell cycle proceeds (Robinson et al. 1995; Sherwin and Gull 1989). In addition, *Trypanosoma* also have a set of four microtubules, called the "subpellicular microtubule quartet" (MtQ) that assembles early during the cell division near the FAZ region (Gallo and Precigout 1988). Extending between the basal body, the MtQ traverses around

the flagellar pocket and runs beneath the length of the flagellum, parallel to the FAZ filament. Displaying a polarity opposite to that of the rest of the subpellicular microtubules, with their plus ends at the anterior side of the cell, the MtQ is found in close association with the endoplasmic reticulum (Robinson et al. 1995).

5.1.1.4 Other Organelles

Apart from functioning as a mere motility organelle, the flagellum initiates a series of cell cycle events through its association with various other cell structures. An organelle that highlights its role very early in the cell cycle is the basal body that functions as the microtubule organizing center (MTOC). The maturation of probasal body, situated anterior to the mature basal body, is one of the first signs for a new round of cytokinesis. The basal body defines the proximal end of the flagellum and plays a significant role in the segregation of flagella and is physically linked to the kinetoplast. Microtubule-mediated separation of basal bodies during cell cycle directly controls segregation of kinetoplast (Robinson and Gull 1991). The newly matured basal body also undergoes a rotational movement that mediates the formation of new flagellar pocket (Lacomble et al. 2010).

The *Trypanosoma* is marked by a single Golgi and ER exit site. It is known that duplication of both the Golgi and the ER export site almost occurs at the same time and that the new Golgi appears with some material transferred from the old Golgi (He et al. 2004). The duplication and segregation of the ER exit site and the Golgi complex is mediated by a bi-lobed structure, found near the flagellar pocket juxtaposed to the Golgi (He et al. 2005). The bi-lobed structure predominantly contains a class of calcium

binding proteins called centrins, which are a known component of centrosomes.

In brief, the cell cycle initiates with the duplication of the basal body and nucleation of the flagellum, followed by the flagellar pocket nucleation. The bi-lobe in close association with the FAZ later duplicates as the FAZ extension begins. The elongation of the FAZ and the flagellum closely follows segregation of the bi-lobe. As elongation proceeds, the kinetoplast and the single nucleus divide and segregate, followed by remodeling of the cell membrane and cytokinesis. While the subpellicular microtubules, basal bodies and Golgi are inherited in a semi-conservative fashion, the microtubules comprising the MtQ, FAZ and the flagellum are synthesized *de novo* in the daughter cell (Farr and Gull 2012; He et al. 2004).

5.1.2 A Single-Celled Eukaryote

Trypanosoma belongs to an evolutionarily distant group of microorganisms called the Kinetoplastida, which represent a deep rooting eukaryote. Apart from the *Trypanosoma*tid parasites, Kinetoplastids also include organisms belonging to the genus *Leishmania*. These single-celled protists are part of a larger supergroup called the Excavates. Phylogenetic reconstruction of Excavates groups Kinetoplastids and closely related euglenids alongside amoeboflagellate Heterolobosea (*Naegleria gruberi*) and heterotrophic flagellates jakobids forming a distinct clade. This clade is evolutionarily distinct from other clades represented by mitochondrion lacking flagellates diplomonads and parabasalids, or the oxymonads and Trimastix (Simpson et al. 2006). The kinetoplastids are marked by unique traits including extensive mitochondrial RNA editing, elaborated mitochondrial

architecture and their ability to vary their surface protein coat and evade host response. Their genes are also predominantly arranged in giant polycistronic clusters and more often have duplicated genes that point to adaptive evolution. Genes related to surface antigens, amino acid transporters and development are present in multiple paralogous copies and are positively selected favoring the evolution and survival of the parasite (Emes and Yang 2008).

5.1.3 Purpose of this study

The only single-celled eukaryotes in which vaults have been studied to some extent are the slime molds *Dicytostelium discoideum*. In spite of being a single celled eukaryote, these protists display a multicellular like lifestyle. The single celled spores grow into unicellular organisms but differentiate to form multicellular fruiting bodies in response to starvation to re-form a smaller population of spores. Unlike the higher eukaryotes which harbor only one *MVP* ortholog, the purified vaults in slime molds are composed of two paralogs, *MVPA* and *MVPB*. The characteristic vault like structure is lost when *MVPA* gene is disrupted by homologous recombination (Vasu et al. 1993). Though the two *MVP* genes are not essential for normal cell growth, as the cells remain viable on disrupting both the genes, it is worth noting that the $MVPA^{-/-}/MVPB^{-/-}$ cells reach limited cell density during conditions of nutrient stress (Vasu and Rome 1995).

The kinetoplastids belonging to the Excavates are evolutionarily distant from the slime molds, which are closer to multicellular metazoans. Phylogenetic reconstructions (refer Chapter 2) suggested an under-representation of vault homologs in protists including the Excavates

suggestive of possible acquisition of vault genes horizontally later through evolution. A majority of studies in characterizing vaults have been carried out in multicellular eukaryotes, including the invertebrate echinoderms. There has been no study so far that has been carried specifically to study vaults in other protists including the Kinetoplastids. The current study focuses on characterizing vaults in ancient single-celled eukaryote *Trypanosoma brucei* and also evaluating if the nutrient-related phenotype is a common feature of all vault-containing organisms, including the Kinetoplastids, which may have acquired vault genes later through evolution.

5.2 MATERIALS AND METHODS

5.2.1 Cell lines and cell culture

The procyclic form *Trypanosoma brucei*, YTat1.1 and 29.13 cell lines were used for the study. The YTat1.1 cell line was grown in Cunningham's medium supplemented with 15% heat-inactivated fetal bovine serum (Hyclone) and maintained at 28°C. The 427 cell line 29.13, expressing T7 RNA polymerase and tetracycline repressor was maintained at 28°C in Cunninghams's medium supplemented with 15% heat inactivated, tetracycline-free fetal bovine serum (Clontech) along with 15 µg/ml G418 and 50 µg/ml hygromycin. The cells were diluted and maintained at exponential growth phase with fresh medium every two days.

5.2.2 Plasmid Construction

The GeneDB database accession number for *TrypanosomabruceiTbMVP1* gene is Tb927.5.4460. The full length coding sequence corresponding to 838 amino acids were amplified from *T.brucei* genomic DNA using the primer pairs Hind3*TbMVP1*FP and Nhe1*TbMVP1*RP, digested with HindIII and Nhe1 and cloned into a pXS2 backbone vector that contains YFP coding sequence as a C terminal tag. For inducible overexpression, the full length coding sequence was amplified using Hind3*TbMVP1*FP and Hind3*TbMVP1*RP, single-digested with HindIII and cloned into a modified pLEW100 vector expressing YFP as a C-terminal tag. For expression of truncated proteins, coding regions corresponding to 1-512 amino acids and 513-838 amino acids were amplified with the primer pairs Hind3*TbMVP1*FP/Nter*TbMVP1*Nhe1RP and Cter*TbMVP1*Hind3FP/

Nhe1*TbMVP1*RP cloned into the pXS2-based vector for construction of N- and C-terminal truncated *TbMVP1*, respectively.

Endogenous replacement of one of the *TbMVP1* alleles with *YFP-TbMVP1* encoding allele was obtained using a double homologous recombination procedure. This was accomplished using a pCR4Blunt-Topo that allows cloning of both upstream and encoding genes regions on either side of regions containing blasticidin resistance gene, tubulin intergenic region and YFP-coding genes. The upstream region of *TbMVP1* corresponding to the 5'-untranslated region (5'-UTR) was amplified using 5'utr*TbMVP1*Pac1F and 5'utr*TbMVP1*Hind3R and cloned into the pCR4Blunt-Topo vector ahead of the blastidicn resistance gene. A region corresponding to 1000 bp of *TbMVP1* was amplified using primer pair*TbMVP1*BamH1F and 1000-*TbMVP1*-NsiR and cloned into the same vector downstream of the YFP-coding region. The targeting construct thus established contained both the 5'-UTR and *TbMVP1*-coding regions. For construction of inducible and inheritable RNAi, suitable RNA fragment was selected using RNAit (http://trypanofan.path.cam.ac.uk/software/RNAit.html), amplified using *TbMVP1*RnaiXba1F/*TbMVP1*RnaiXba1R and cloned into the pZJM vector.

The GeneDB database accession numbers for *TrypanosomabruceiTbMVP2* and *TbMVP3* genes are Tb927.10.1990 and Tb927.10.6310, respectively. The full-length *TbMVP2* gene encoding 863 amino acids and the *TbMVP3* gene encoding 862 amino acids were amplified using primer pairs Hind3*TbMVP2*FP/Nhe1*TbMVP2*RP and Hind3*TbMVP3*FP/Nhe1*TbMVP3*FP. The PCR fragments were digested using

HindIII and Nhe1 and cloned into the pXS2-based vector for expression of fusion protein with a C terminal YFP tag.

Table 5.1 List of primers used for establishing various constructs

Vector Name	Primer Name	Primer Sequence
*TbMVP1*Topo	*TbMVP1BamH1F*	5'-CG GGA TCC ATG AGT GAT ATC ATA CGA ATT AAA CGT C-3'
	1000-TbMVP1-NsiR	5'-TGC ATG CAT GTA GTG CCT CAT TCT TCC CGA TAG-3'
5'UTR*TbMVP_1*	*5'utrTbMVP1Pac1F*	5'-CCT TAA TTA ATT GAA TCT GAT GAG GTG TAG GGA-3'
Topo	*5'utrTbMVP1Hind3R*	5'-CCC AAG CTT CTT CGA AAC AGT TGG ACA AAA ATG-3'
*TbMVP1_* pXS2	*Hind3TbMVP1FP*	5'-CCC AAG CTT ATG AGT GAT ATC ATA CGA ATT AAA CG-3'
	Nhe1TbMVP1RP	5'-CTA GCT AGC TGA CTT TGT CTC GGT TGT TTG-3'
*TbMVP3_* pXS2	*Hind3TbMVP3FP*	5'-CCC AAG CTT ATG AAT GAT TAT TTA GCG AAT GAG CTG-3'
	Nhe1TbMVP3FP	5'-CTAGCT AGCCTG CTG CAC ATG ACC AGT C-3'
*TbMVP2_* pXS2	*Hind3TbMVP2FP*	5'-CCC AAG CTT ATG GTG GAC AAG GAG AAT CAG GTG A-3'
	Nhe1TbMVP2Rp	5'-CTA GCT AGC TGG CAA CGC GTC GTT CC-3'
*TbMVP1*_plew100	*Hind3TbMVP1FP*	5'-CCC AAG CTTATG AGT GAT ATC ATA CGA ATT AAA CG-3'
	Hind3TbMVP1RP	5'-CCC AAG CTT TGA CTT TGT CTC GGT TGT TTG-3'
*TbMVP1*_pZJM	*TbMVP1RnaiXba1F*	5'-GC TCT AGA TGA ACA CCA CTA CGG GTG AA-3'
	TbMVP1RnaiXba1R	5'-GC TCT AGA AAG AGT CAA AGT CCT CCG CA-3'

Nter-*TbMVP1*	*Hind3TbMVP1FP*	5'-CCC AAG CTTATG AGT GAT ATC ATA CGA ATT AAA CG-3'
	NterTbMVP1Nhe1RP	5'-CTA GCT AGC ACC CAA AAA GAG CTG TAG AGC-3'
Cter-*Tb**MVP1***	*CterTbMVP1Hind3FP*	5'-CCC AAG CTT ATG CCT CGT TTC TCC AGT GAC ACG-3'
	Nhe1TbMVP1RP	5'-CTA GCT AGCTGA CTT TGT CTC GGT TGT TTG

The greyed region refers to inclusion of restriction tag for the purpose of cloning.

5.2.3 Stable and Transient Transfection

A stables line of *TbMVP1* overexpression was established by linearizing pXS2 construct with Nsi1 and transfecting into procyclic cells 29.13. Endogenous replacement of *TbMVP1* was achieved by transfecting Ytat.1 cells using linearized pCR4Blunt-Topo vector with Nsi1. Not1 digested pLEW100 construct was transfected into 29.13 cells for establishing a stable inducible overexpression line. The establishment of inducible *TbMVP1* RNAi involved linearization of the pZJM construct with Not1and transfection into stable line endogenously expressing *YFP-TbMVP1*.

For all stable transfections, cells in exponential growth phase were electroporated with 15 µg of linearized plasmid DNA using Gene Pulser Xcell^{TM} system (Bio-Rad Laboratories) and plated in 96 well plate containing appropriate antibiotics (Blasticidin 10 µg/ml (for pXS2 and pCR4Blunt-Topo vectors) or Phleomycin 5 µg/ml (for pLEW100 vector) or both (for pZJM RNAi). The positive clones were selected and maintained in respective media. For transient transfection 50 µg of plasmid DNA was transfected into 29.13 cells by electroporation.

5.2.4 Immunofluorescence Microscopy

Cells in log phase were spun down and settled onto coverslips for 20 minutes. Intact cells were either fixed and permeabilized in methanol at -20°C for 10 minutes followed by rehydration with PBS for 10 minutes or fixed for 7 minutes with ice cold 4% (w/v) paraformaldehyde in phosphate-buffered saline (PBS) followed by permeabilization for 5 minutes with 1% Triton-X-100 (v/v). For detergent and salt extractions, cells on coverslips were treated with 1% NP-40 (v/v) and 1% NP40 (v/v) + 1M KCl in PBS for 5 minutes followed by fixation with ice cold 4% (w/v) paraformaldehyde in PBS. The coverslips were either directly mounted or prepared for indirect immunofluorescence. For antibody labeling coverslips were blocked with 3% BSA in PBS for 1 hour and then incubated with the indicated primary antibody for 1 hour and washed. Polyclonal anti-GFP (1:200), L3B2 (1:25), anti-CC2D (1:1000), anti-PAR (1:1000) and GRASP (1: 1000) antibodies were used to label YFP, FAZ filament, paraflagellar rod and golgi bodies respectively (Kohl et al. 1999; Zhou et al. 2011; He et al. 2004; Ismach et al. 1989). The fluorescent secondary antibodies (Invitrogen) were then used at 1:2000 dilutions for 1 hour followed by DAP1 staining (20 µg/µl) to label DNA for 15 minutes). The coverslips were washed and mounted before visualization. Fixed cells were observed using Zeiss Axio Observer Z1 fluorescence microscope (Carl Zeiss MicroImaging, Germany) with 63x/1.40 oil DIC objective and image acquisition were performed using a CoolSNAP HQ^2 CCS camera (Photometrics). The images were processed with ImageJ and Adobe Photoshop Elements 9.

5.2.5 Cell Fractionation

Intact vault particles are known to pellet at 100,000g (Kickhoefer et al. 1998). To perform cell fractionation to pellet vaults, 10^9 *TbMVP1* overexpressing cells were used. Cells were pelleted at 3000 RPM for 7 minutes followed by washing twice with cold 10 ml of Buffer A (50 mM Tris-HCl, pH 7.4, 75 mM NaCl, 1.5 mM $MgCl_2$). The cells were then resuspended in 10 ml of Buffer A containing 1% Triton-X-100 (v/v), 1mM DTT and 1X protease inhibitors (cOmplete, EDTA-free Protease Inhibitor Cocktail Tablets, Roche Applied Science, USA) and incubated for 20 minutes at 4°C with intermittent vortexing. All the extraction procedures detailed below was carried out at 4°C .The extracted cells were pelleted at 8000 RPM for 15 minutes to separate the extracted cytoskeleton and soluble cytoplasmic pool. The cytoskeletal and soluble pool were treated independently to obtain the heavy weight pellet. The soluble cytoplasmic pool was further pelleted at 20,000 g for 20 minutes to remove any cytoskeletal debris and the supernatant (soluble cytoplasmic pool) was used for analysis. The extracted cytoskeleton was washed twice with 10 ml of Buffer A containing 1% Triton-X-100 (v/v), 1mM DTT and 1X protease inhibitors and resuspended in 1 ml of the same buffer. The extracted cytoskeleton was subjected to sonication (2 seconds pulse for 3 times with 10 seconds rest) at 80% output and spun down for 15 minutes at 13,000 RPM to obtain a supernatant fraction that is enriched in proteins released from the cytoskeletal extract. The obtained cytoskeletal fraction and soluble pool fraction was rotated at 100,000g for 1 hour using a Sw41 Ti swinging bucket rotor to obtain P100-cytoskeletal and P100-soluble fraction, respectively. The supernatants S100-cytoskeletal and S100-soluble were also processed for immunoblotting analysis.

5.2.6 Purification of Vault Particles from *Trypanosoma* Cells

10^9 cells expressing YFP tagged TbMVP1 were pelleted at 3000 RPM for 7 minutes followed by washing twice with ice cold 10 ml of Buffer A ((50 mM Tris-HCl, pH 7.4, 75 mM NaCl, 1.5 mM $MgCl_2$). The pelleted cells were resupended in 10 ml of Buffer A containing 1% Triton-X100 (v/v), 1 mM dithiothreitol (DTT) and 1X protease inhibitors (cOmplete, EDTA-free Protease Inhibitor Cocktail Tablets, Roche Applied Sciences, USA) and incubated on ice for 20 minutes with intermittent vortexing. Unbroken cells and other heavy organelles were pelleted at 20,000 x g for 20 minutes at 4 °C and the supernatant fraction S20 was used for further analysis. A crude extract containing vaults were obtained by further centrifugation at 100,000 x g for 1 hour at 4 °C. The supernatant was discarded and pellet was resuspended in 1 ml of Buffer A containing 1 mM DTT and 1X protease inhibitors. The resuspended pellet was carefully layered on a sucrose step gradient of 20, 30, 40, 45, 50 and 60% sucrose layers prepared in Buffer A, and centrifuged for 25,000 RPM at 4 °C in a SW41 Ti rotor for 16 hours. The fractions corresponding to each sucrose layer were later collected and diluted 1:10 with Buffer A containing 1 mM DTT and 1X protease inhibitors. The individual fractions were centrifuged for 3 hours at 100, 000 x g at 4 °C to pellet vaults. The pellets were resuspended in 100 µl of Buffer A containing 1 mM DTT and 1X protease inhibitors and analyzed using Western blot with anti-GFP antibody.

5.2.7 Negative Staining and Electron Microscopy

The fractions corresponding to 40 and 45% sucrose layers were adsorbed onto carbon formvar coated nickel grids (EMS) for 4 minutes at room temperature. Following sample adsorption the grids were floated on 20 µl of 1% uranyl acetate for 1 minute and blot dried using filter paper. The grids were stored in dry cabinet prior to viewing in a JEOL JEM 2010F HRTEM microscope.

5.2.8 Immunoelectron microscopy

Immunogold labeling was performed on detergent-extracted cells that maintains the cell shape including the FAZ but removes most of the cytoplasmic pool. Briefly 10^7 cells were harvested, washed twice with PBS and resuspended in 50 µl of PBS prior to being attached into carbon formvar coated nickel grids (EMS). The cells were extracted in freshly prepared PEM buffer containing 1% NP40 for 5 minutes. The extracted cells were fixed with 2.5% glutaraldehyde for 15 minutes followed by washing with PBS for 3 times 5 minutes each. The grids were then neutralized with 100mM glycine for 5 minutes followed by blocking with 1% BSA for at least 30 minutes. The grids were labeled with mouse L3B2 (1:20) and rabbit anti-GFP (1:20) for 1 hour. The labeled grids were washed 5 times 5 minutes each with 1% BSA prior to labeling with Goat anti-mouse IgG conjugated to 40 nm gold (1:100) and Goat anti-rabbit IgG conjugated to 25 nm gold (1:100) for 1 hour. The grids were washed with PBS 3 times 5 minutes each and later fixed with 2.5% glutaraldehyde for 15 minutes. The grids were washed once with PBS and twice with filtered water and briefly stained in a drop of 0.6% Aurothioglucose prepared in distilled deionized water. The grids were air-dried prior to imaging by JEM 2010F HRTEM microscope.

5.2.9 RNAi Induction

The *TbMVP1*-RNAi knockdown cell line was induced with 10 µg/ml of Tetracycline and monitored for cell proliferation in control and starvation media. The 15% tetracycline free 29.13 culture medium was diluted 10 times in HBSS (137 mM NaCl, 5.4 mM KCl, 0.25 mM Na_2HPO_4, 0.44 mM KH_2PO_4, 1.3 mM $CaCl_2$, 1mM MgSO4, 4.2 mM $NaHCO_3$, pH 7.3) with glucose (1 g/l) and used as the nutrient limiting medium. 2 x 10^6 cells/ml were maintained in nutrient rich or nutrient limiting medium with tetracycline and monitored for cell proliferation every 24 hours. Uninduced *TbMVP1*-RNAi cell line was used as a control in both nutrient conditions. Cells in nutrient rich medium were diluted with fresh media when the culture density exceeds 10^7 cells/ml. Only motile cells were counted as viable cells every 24 hours using a hemocytometer.

5.2.10 Immunoblotting Analysis

10% SDS-PAGE gel was used for protein electrophoresis in a Tris-Glycine-SDS running buffer (25mM Tris, 192mM glycine, 0.1% SDS; 1st BASE). A 3x loading buffer (150mM Tris-HCl, 6% SDS, 30% Glycerol, 3% β-Mercaptoethanol, 37.5mM EDTA, 0.06% Bromophenol Blue, pH 6.8) was used for sample preparation. The proteins were electrophoresed at a constant voltage of 100v. Precision Plus Protein Dual Color Standards ((Bio-Rad, USA) was used as a marker. For immunoblotting, the Polyvinylidene difluoride (PVDF) membrane (Bio-Rad, USA) was activated with absolute methanol, rinsed with transfer buffer (0.3% Tris, 1.45% glycine and 20% methanol) and set for transfer at 70v for 1 hour at 4°C. After transfer, the membrane was blocked with 5% skimmed milk in TBST (0.1% Tween-

20,10mM Tris-HCl, 150mM NaCl, pH7.6) for at least 1 hour at room temperature with shaking. The membrane was then incubated with primary antibody (Anti-GFP) or anti-tubulin (mouse kmx1) at a dilution of 1: 500 or 1:3000 respectively, overnight with gentle rocking. After washing for 3 times with 5%skimmed milk, the membrane was incubated with the desired HRP (Horseradish Peroxidase)-conjugated secondary antibody at a dilution of 1:5000 for 1 hour. The membrane was given a final wash with TBST before the substrate solution from SuperSignal® West Dura Extended Duration Substrate Kit (Thermo Scientific, USA) was added. ImageQuant LAS 4000 (GE Healthcare, UK) was used to scan the membrane and acquire the images.

5.3 RESULTS

5.3.1 Identification of three vaults genes in kinetoplastids

Sequence analysis and phylogenetics reveal that Kinetoplastids evolved three different *MVP* homologs. *MVP* originated in the common ancestor of kinetoplastids and underwent duplication events to give rise to differentially diverging paralogs prior to speciation events separating *Leishmania* and *Trypanosomes* (Figure 5.2). The sequence designated as *MVP1,* is closer in terms of sequence similarity to vault protein orthologs from other well-studied organisms compared to *MVP2* and *MVP3.*

In *Trypanosoma brucei* the three homologs are identified as *TbMVP1*, *TbMVP2* and *TbMVP3* (annotated as Tb927.5.4460, Tb927.10.1990 and Tb927.10.6310 respectively, in GeneDB). The *TbMVP1* is an 838 amino-acid protein with a predicted molecular weight of 94.1 kDa and is closer in sequence similarity to *MVP* orthologs described in other eukaryotes. *TbMVP2* is an 863 amino acid protein (95 kDa) that has diverged considerably from the short-branching *TbMVP1* over the years of evolution. *TbMVP3* is an 862 amino acid protein (96.2 kDa), but is significantly different from the 863 amino acid *TbMVP2* protein. An alignment between the three homologs is shown in Figure 5.3. Interestingly, while *TbMVP1* maps to chromosome 5, both *TbMVP2* and *TbMVP3* maps to chromosome 10, hinting at a possible tandem duplication event. It is also worth noting that the shorter homolog in all identified *Leishmania* species also map to chromosome 5. Hence, it could be deduced that the *MVP* gene originated in a kinetoplastid ancestor as a single copy gene and later duplicated to give rise to multiple paralogs. (Refer Appendix for characterization studies of *TbMVP* paralogs)

Vaults purified from multicellular eukaryotes also have minor vault proteins, *VPARP* and *TEP1* in addition to small untranslated *vRNA*. Protein BLAST search based on human *VPARP* sequence failed to identify any orthologs in kinetoplastid genomes. However, potential homologs of *TEP1* proteins from *Trypanosomabrucei* (but not in *Leishmania*) were identified using a human *TEP1*ortholog as query. However, pair wise alignment of the identified *TEP1* homolog with human *TEP1* revealed poor conservation through the length of the sequence (detailed in 2.3.4). It is known that *TEP1*, in addition to interacting with vault complex, also associates with Telomerase RNA complex. In the vault complex, *TEP1* is known to stabilize the associated *vRNA*. However, no homologs of *vRNA* have been identified in kinetoplastids. Hence, it seems unlikely that *TEP1* forms a stable association with vaults in kinetoplastids.

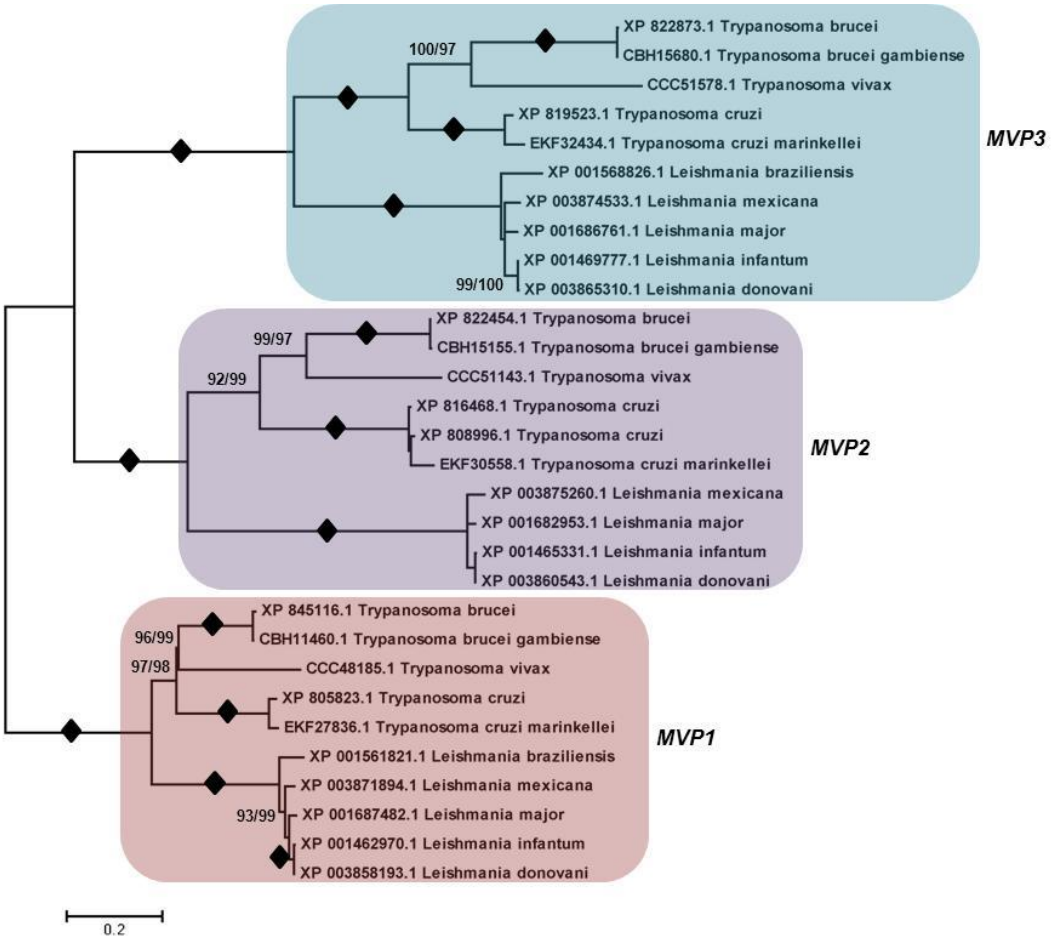


Figure 5.2 Distribution of *MVP* homologs across the Kinetoplastids
Evolutionary relationships among the *MVP* homologues in kinetoplastids. Shown is the ML bootstrap consensus tree obtained based on the JTT+G model of evolution. The tree is robustly resolved and clearly supports the origin of paralogs in an ancestral kinetoplastid that could have likely preceded the divergence of *Trypanosoma* and *Leishmania*. The two *T.cruziMVP2* sequences map to different genomic loci and hence, both were retained for analysis.

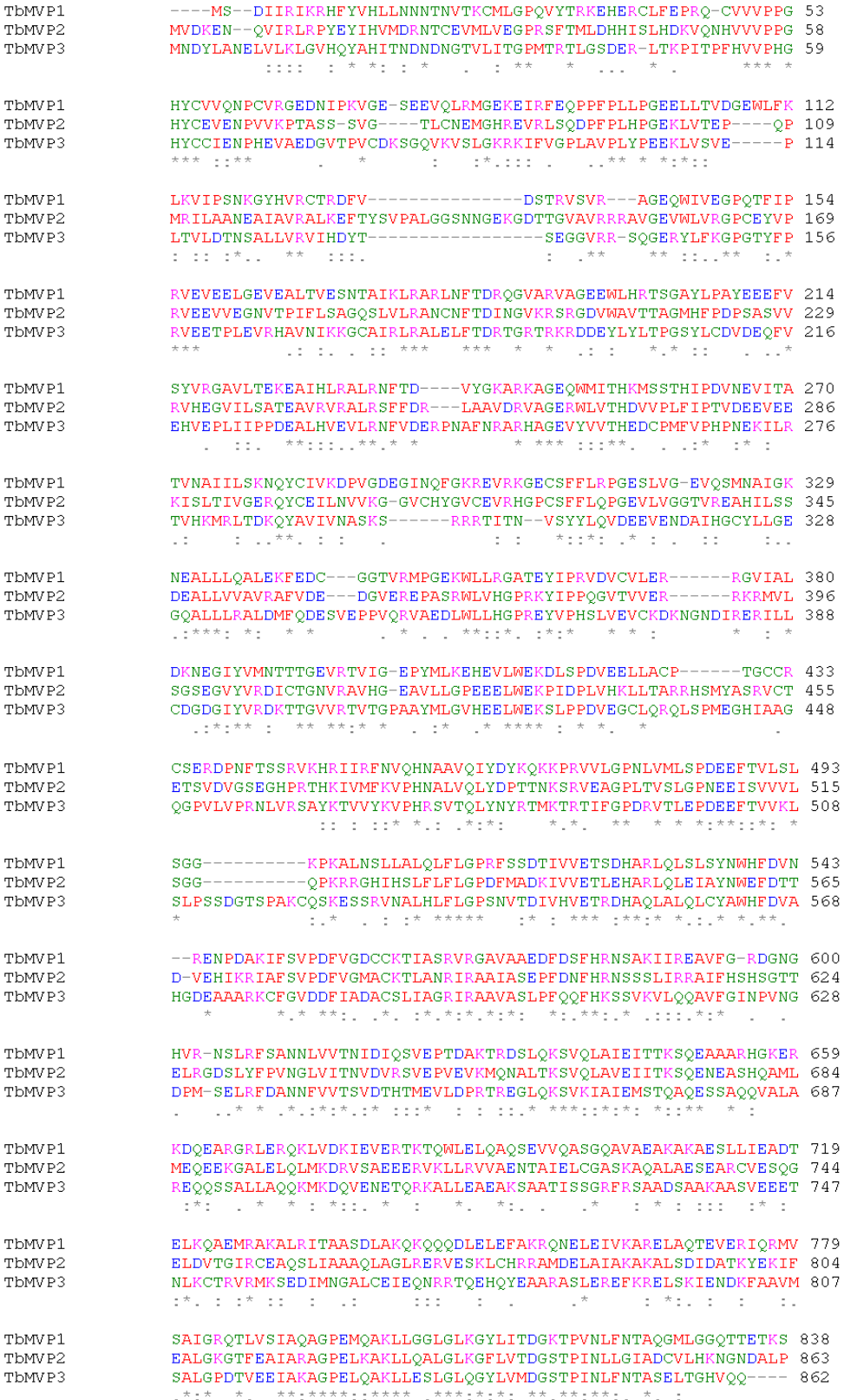


Figure 5.3 Clustal alignment between the *Trypanosoma MVP* paralogs

5.3.2 Endogenous expression of *MVP*1 shows punctate distribution

To identify the subcellular distribution of *TbMVP1*, a stable line expressing YFP tagged *TbMVP1* from one endogenous allele was generated by homologous replacement. The YFP tag was fused at the N terminus and hence, is expected to be enclosed within intact vault complexes. The production of *YFP-TbMVP1* protein of the correct molecular weight was verified by Western blotting. The distribution of epitope tagged *TbMVP1* was followed at all stages of *Trypanosoma* life cycle. It was observed that *TbMVP1* displayed punctate distribution through the cytoplasm but was excluded from the nucleus and kinetoplastid regions (Figure 5.4). The stable cell line was also subjected to indirect immunofluorescence with anti-GFP to confirm the expression of properly folded YFP. Previous studies with GFP tagged vaults in mammalian cells have also shown similar punctate distributions (Kickhoefer 2005).

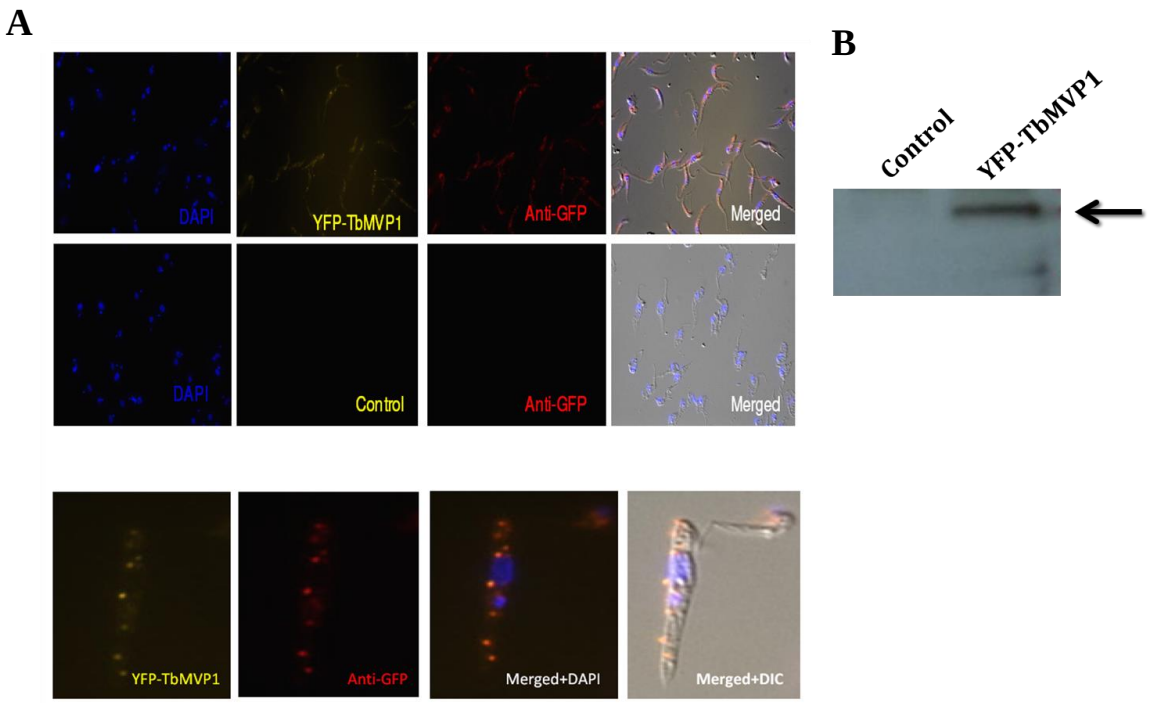


Figure 5.4 Endogenous expression of *YFP-TbMVP1*
(A)Cells stably expressing *TbMVP1* tagged with YFP were fixed with methanol and viewed under fluorescence microscope, either directly or indirectly using anti-GFP. *TbMVP1* shows a clear punctate patterning throughout the cytoplasm and is excluded from the nucleus. Scale bar, 2µm. (B) Western blotting that confirms the expression of YFP tagged *TbMVP1*.

5.3.3 *TbMVP1* can assemble to form intact vault particles

It is known that 78 copies of *MVP* monomers assemble to form an intact vault complex that exhibits dihedral symmetry (Tanaka et al. 2009). Though *TbMVP1* can be visualized as bright puncta, similar to those observed in multi-cellular metazoans, whether it assembles to form intact vault complexes in single celled kinetoplastids remains unknown. It has been previously established that vaults isolated from another single celled protist, *Dictyostelium*, forms intact vault complexes composed of at least two *MVP* paralogs, *MVPA* and *MVPB*. Intact vault complexes are known to pellet at speed as high as 100,000g and this method has been used traditionally to make crude vault preparations (Kickhoefer et al. 1998). It is known that monomeric *MVP* remains in the soluble S20 fraction and it takes about 4 hours for

individual monomers to assemble and form intact vault complexes that is found in the P100 pellet fraction (Zheng et al. 2005). Since, *TbMVP1* is found both in the cytoplasmic and the cytoskeletal pool, it was hypothesized that intact vault complexes can freely float around in the cytoplasm and also associate with cytoskeleton. In that regard, the cells were first detergent extracted to segregate the cytoplasmic and cytoskeletal fractions (detergent-extracted cell that retains the cytoskeleton) and the two samples were treated independently to retrieve the P100 pellet. The extracted cells were sonicated, cell debris removed by centrifugation and the supernatant representing the proteins released off the cytoskeleton was used for further analysis. The supernatants from both the cytoskeletal fraction and soluble fraction were subjected to high-speed centrifugation at 100,000g for 1 hour at 4 °C. The pellets thus obtained, P100-cytoskeletal and P100-soluble, and the supernatants S100-cytoskeletal and S100-soluble were analyzed by immunoblotting using anti-GFP (Figure 5.5). Interestingly P100 fraction was enriched in the soluble cytoplasmic pool than the extracted cytoskeletal pool. The cytoskeletal pool was subjected to more fractionation steps prior to high-speed centrifugation and hence, some amount of sample loss is possible.

In mammalian crude vault extracts, the S100 fraction does not show a positive band as a majority of *MVP* monomers assemble into intact vault particles. However, it was found that the S100 fraction from both cytoplasmic and cytoskeletal extracts showed strong bands for *TbMVP1-YFP*, pointing out to free *MVP* monomers that do not assemble into intact vault particles. Even in insect cells in which vaults are routinely overexpressed using a baculovirus system, the S100 fraction shows a positive band (Stephen et al. 2001). It has

been established previously that *MVP* molecules prefer to stay in complex than as free floating monomers (Zheng et al. 2005). Hence, the high concentration of monomers found in S100 extract could be attributed to the effect of fusion protein overexpression. It should also be noted that the fusion tag is C terminally located and hence, extends out from the cap region outside the vault complex. Ideally, if vaults were assembled purely with *TbMVP1-YFP* fusion proteins, 39 molecules of highly structured YFP extending out from the cap region might pose significant steric hindrance and hence could affect particle stability. A previous study established that when a highly structured 55 amino acid EGF tag was added to the C terminal region of *MVP* and co-expressed with N terminally tagged VSGV-*MVP*, the recombinant vaults only contained 6-8 copies of C-terminally *MVP*-EGF incorporated into an intact particle, while the rest was composed of N-terminally tagged fusion *MVP* (Kickhoefer et al. 2009). This may be due to the spacious packing within the waist region with no potential steric hindrance effect. The positive signal obtained in the S100 fraction could be attributed to the same phenomenon. The presence of *MVP* in the cytoplasmic and the cytoskeletal P100 fractions may imply the assembly of intact vault complexes that comprise YFP tagged *TbMVP1* along with endogenous *TbMVP1*.

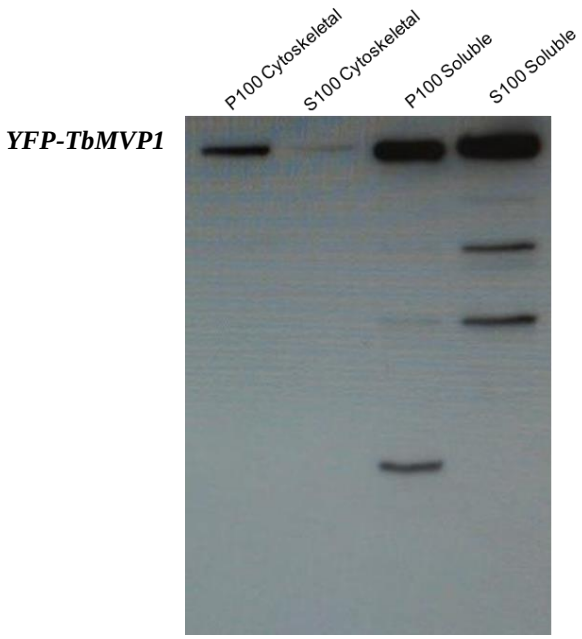


Figure 5.5 *TbMVP1* is retained within heavy weight pellet P100
The high speed supernatant (S100) and high speed pellet (P100) from both the detergent resistant cytoskeletal extract and the soluble cytoplasmic pool were analyzed by immunoblotting with anti-GFP antibody. Intact vault particles pellet at 100,000g and are recovered as crude vault extracts. Western blotting shows positive bands for both cytoskeletal and soluble cytoplasmic pool after high-speed centrifugation. Not all *YFP-TbMVP1* can assemble into intact particles possibly due to crowding from C-terminal YFP at the vault cap region.

Though YFP-tagged *TbMVP1* pelleting at P100 fraction could possibly point to intact vault assembly, such a fractionation may also result from expressed fusion protein forming aggregates within the cells. Apart from pelleting at 100,000 x g fraction, vault purification from various cell types and organisms has revealed that intact vault particles predominantly purifies along the 40 and 45% sucrose layers. A few of the expressed *MVP* proteins also assemble into structures that fractionate at 50 and 60% sucrose layers. Accordingly, *T.brucei* cells expressed YFP tagged *TbMVP1* were subjected to the vault purification on sucrose gradients (20%, 30%, 40%, 45%, 50% and 60% sucrose layers) and analyzed using Western blot (Figure 5.6 A). Similar to what has been observed in other multicellular organisms and *Dictyostelium*, vaults in *T.brucei* also display similar fractionation along the sucrose gradient,

with a majority of the expressed protein found along the 40 and 45% sucrose layers. While additional material was found between 50 and 60% sucrose layers, there was no reactivity at lower sucrose fractions 20 and 30%. This suggests that *TbMVP1* indeed forms vault-like structures and does not assemble into random aggregates that spread across the entire sucrose gradient.

To further ascertain the formation of intact vault complexes, the fraction from the 45% sucrose layer was analyzed using negative stain electron microscopy. Electron micrographs revealed particles displaying characteristic barrel-shaped structures with protruding caps and invaginating waist regions that resemble vault like particles (Figure 5.6 B). The structures were also comparable in size and shape to vaults purified from other multicellular organisms. This suggests that *TbMVP1* sequence from *T.brucei,* in spite of being evolutionarily distant, carries all the necessary information that drives intact vault assembly from individual monomers. It should be noted that *TbMVP1* is also closer in sequence similarity to *MVP* sequences from other well-characterized organisms. Hence, it is suggested that *TbMVP1* likely represents the functional ortholog of vault gene in *T.brucei.*

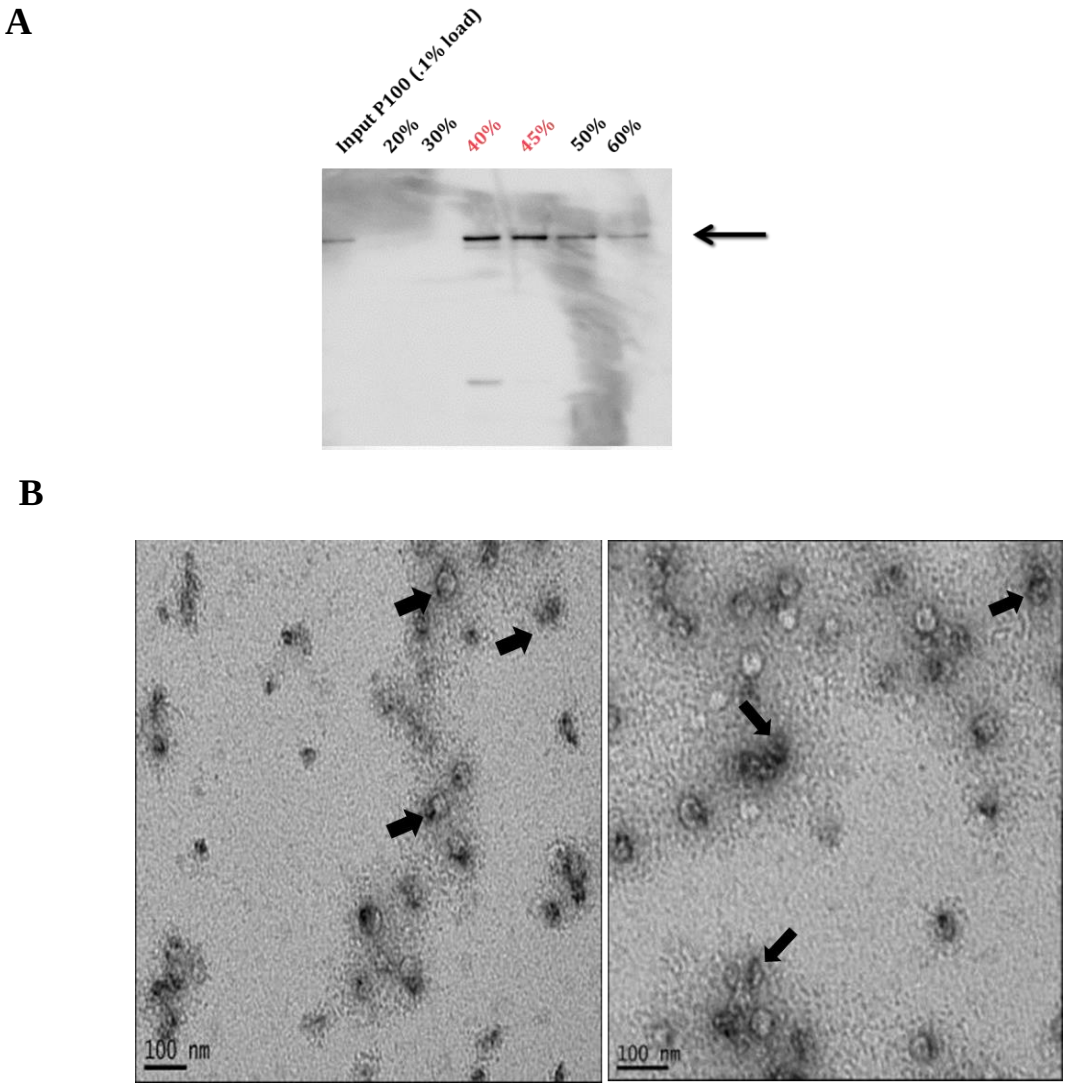


Figure 5.6 *TbMVP1* assembles into vault-like particles
(A)YFP tagged *TbMVP1* was expressed in *Trypanosoma brucei* and the protein lysate obtained was fractionated at 100,000 x g to obtain crude vault containing extract (P100). The P100 fraction was layered on a discontinuous step gradient and the individual fractions analyzed by Western blotting using anti-GFP antibody. The arrow represents full-length *TbMVP1* tagged with YFP fractionating at specific sucrose layers suggestive of assemblage of large protein complexes. (B) Fraction from the 45% sucrose layer viewed under negative staining electron microscopy. Barrel-shaped particles resembling vault are marked with black arrows.

5.3.4 *MVP*1 is not essential for cell proliferation under normal conditions

With the aid of gene specific RNA interference, a stable line in which expression of *TbMVP1* can be knocked-down with the addition of tetracycline was established. Western blots confirmed that expression of endogenous *YFP-TbMVP1* was down-regulated after addition of tetracycline (Figure 5.7 C). The

effect of a specific gene knockdown on cell-division events can be assessed efficiently by a simple cell proliferation assay. In that regard, the cell number for uninduced control and induced knockdown cell lines was assessed in normal growth media supplemented with 15% serum for 10 days. The knockdown cells displayed similar growth patterns to those of control uninduced cells (Figure 5.7 A). This suggests that the knockdown did not affect any changes that come along during *Trypanosoma* cell-cycle events including flagellum duplication, flagellar pocket duplication, FAZ elongation, kinetoplastid duplication and cytokinesis. Thus it is clear that *TbMVP1* is not essential for cell proliferation under normal growth conditions.

5.3.5 Mild nutrient phenotype on *TbMVP1* knockdown at limited nutrient condition

The *Trypanosoma* undergoes complex changes in life-style when acclimatizing to its insect and mammalian host. While in the mammalian host, the parasite is enriched with nutrients, in the insect host there is paucity of nutrients and the parasite has to overcome the nutritional stress. The procyclic form mimics the cell stage that is adapted to the insect host and hence a limited nutrient environment was created to see if knocking down *TbMVP1* in procyclic cells has any potential effect on cell proliferation. The control and induced cells were grown in 10% diluted growth media in HBSS containing 1g/l glucose. This reduces the total serum concentration to about 1.5% and dilutes other nutrients and other growth factors tenfold. On induction of tetracycline-inducible RNAi, *TbMVP1*knockdown cells displayed reduced growth rates in limited-nutrient medium compared to uninduced cells. However, the limiting cell density achieved by the knockdown cells was lower compared to those of the uninduced cells as can be seen in Figure 5.7 B. This

suggests that the role of *TbMVP1* is pronounced only during conditions of nutrient stress or starvation. This corroborates with previous observations in which *MVP* knockout in *Dictyostelium* or mouse embryonic fibroblasts (MEF) display a nutrient phenotype only on starvation but display normal growth otherwise (Vasu and Rome 1995; Kolli et al. 2004).

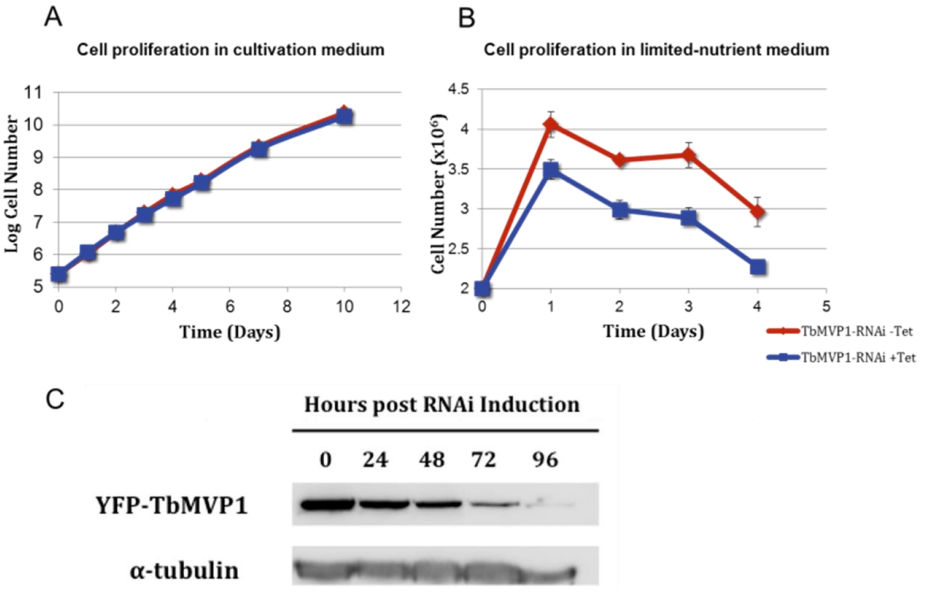


Figure 5.7 *TbMVP1* is important for cell survival at limited-nutrient condition
Depletion of *TbMVP1* did not affect the cell growth under normal conditions but altered the growth rates under limited-nutrient conditions as shown in A and B. RNAi was induced by addition of tetracycline and the cell proliferation rates were monitored over the indicated time points. Viable parasites (judged by motility and cell appearance) were counted with a hemocytometer. The results are presented as mean +/- SD of three independent experiments (C) The efficiency of RNAi in knocking down *YFP-TbMVP1* expression was determined using immunoblotting analysis with anti-GFP. '-'indicates no tetracycline added and '+' indicated addition of tetracycline.

5.3.6 *TbMVP1* knockdown interferes with nutrient-stress related cell adhesion

It has been reported previously that *MVP* knockout MEF cells display a nutrient phenotype only when the serum was completely removed as even cells growing in 2.5% serum had growth rate comparable to those of control cells (Kolli et al. 2004). To see if *TbMVP1* knockdown shows a pronounced effect on lowering the concentration of serum and nutrients further, the media was diluted 20x in HBSS with 1g/l glucose reducing the total serum concentration to 1%. Another set of experiments without glucose was also performed. The uninduced control and tetracycline induced test cells were monitored every 24 hours for 5 days (Figure 5.8 A). Though, the cell numbers were reduced to more than half in both flasks on 24 hours, more freely floating cells were observed in the induced flask than in the control flask. On closer examination it was found that, most of the control cells attached to the walls of the flask, while still being alive, which was evident from their flagellar beating. However, the *TbMVP1* knockdown cells showed no signs of attachment. Since the exact number of attached cells could not be quantified, only the freely floating non-adherent cells were counted. The freely floating cells in the knockdown flask continued to divide and cell numbers increased until day 3 after tetracycline induction. Even when cell numbers began to drop the cells never attached. The control cells, on the other hand, remained attached through the course of the experiment and hence their cell proliferation could not be monitored assessed. To the attached control cells, tetracycline was added to check for any detachment following knockdown of *TbMVP1*. It was found that the floating cell population increased by as much

as tenfold on comparison to other flasks that remained uninduced (data not shown).

It is interesting to note that the induced cells subjected to limited nutrient media without glucose, seemed to exhibit higher proliferation until day 3, but the cell numbers dropped faster compared to other induced cells that were subjected to nutrient limitation with glucose (Figure 5.8 B). Addition of glucose did not have any effect on the uninduced cells. This experiment suggests that on knocking down *TbMVP1*, the parasite loses its ability to initiate a nutrient stress induced attachment phenomenon, which may help the parasite stay in a dormant state until the nutrient conditions are revived.

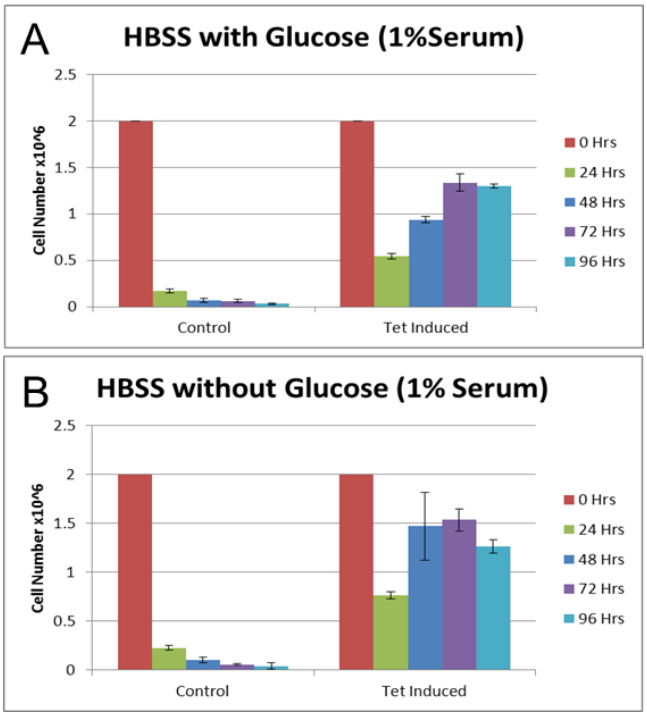


Figure 5.8 *TbMVP1* interference with nutrient-stress mediated cell-adhesion
A severe limited-nutrient condition (1% Serum + 20 times lowered amino acid levels to that of normal cultivation medium) triggered a cell adhesion response whereby cells attached to the walls of the flask while still being alive. While the uninduced control cells remained attached through the indicated time points in both A and B, induction by tetracycline retained more free floating cells and no cell adhesion to the flask was observed. The cell number in the y-axis corresponds to free floating cells. The results are presented as mean +/- SD of three independent experiments.

5.3.7 *TbMVP1* overexpression improves cell survival at limited-nutrient condition

MVP gene is overexpressed during many cellular conditions including cancerous progression and regeneration. To ascertain if *TbMVP1* overexpression displays any effect over the cell proliferation rates in *Trypanosoma,* a stable line constitutively overexpressing YFP-tagged *TbMVP1* was established. Immunofluorescence (Refer Appendix for characterization of overexpression line) and Western blot was performed to confirm the presence of the tagged fusion protein. When the cells were cultured in normal growth media supplemented with 15% serum, the growth pattern of the *TbMVP1* overexpressing cell line was similar to that of control 29.13 cells (Figure 5.9 A). No obvious improvement or retardation of cell growth was observed. However, when the cells were subjected to limited-nutrient environment (10% diluted growth media in HBSS containing 1 g/l glucose) containing only 1.5% serum and tenfold reduced nutrient, the *TbMVP1* overexpression cells displayed a higher limiting cell density compared to those of the control 29.13 cells (Figure 5.9 B). On the second day of starvation, the cells overexpressing *TbMVP1* reached cell densities almost 1.5 times to those of control 29.13 cells. When equal numbers of the cells were analyzed by Western blotting it was observed that in spite of the constitutive overexpression, protein expression of *TbMVP1* progressively went down as cells continued to starve (Figure 5.9 C). This suggests the possible turnover of vaults to overcome the cell starvation stress. However, during states of prolonged starvation, the constitutive overexpression of non-housekeeping gene may add to the cellular burden on the cells. Accordingly, the cell number of the overexpressing cells began to drop into the third day of

continuous starvation. Thus, it is clear that *TbMVP1* overexpression improves cell survival during conditions of nutrient starvation.

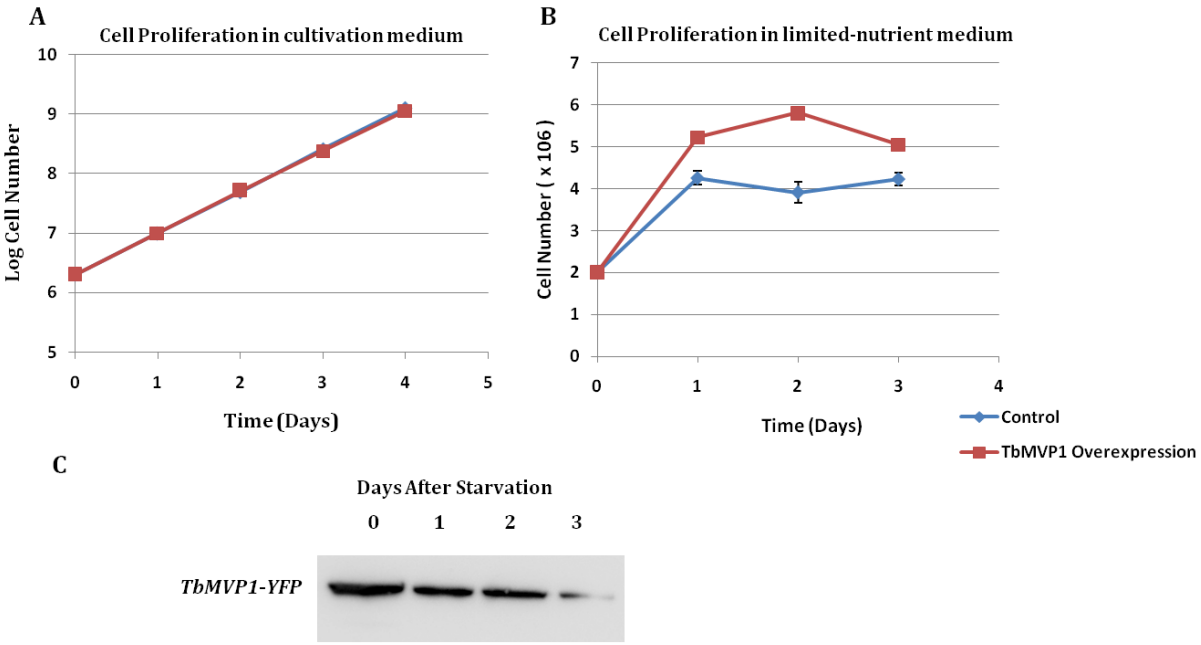


Figure 5.9 Effect of *TbMVP1* overexpression on cell proliferation rates
Overexpression of *TbMVP1* did not affect the cell growth under normal conditions but altered the growth rates under limited-nutrient conditions as shown in A and B. Viable parasites (judged by motility and cell appearance) were counted with a hemocytometer. The results are presented as mean +/- SD of three independent experiments (C) The protein levels of YFP tagged *TbMVP1* continue to drop as the overexpression *Trypanosoma* cells are subjected to limited-nutrient conditions.

5.4 DISCUSSION

The kinetoplastids represent a group of flagellated single-celled protists that branch off early in eukaryotic evolution. Vault homologs are widely conserved across the eukaryotes and through sequence homology, kinetoplastids including *Trypanosoma* and *Leishmania*, have been found to harbor vault genes. Kinetoplastids carry multiple copies of the *MVP* gene, each of which maps to different genomic loci. Based on the phylogenetic

reconstruction, *TbMVP1* was found to be more homologous to vault sequences from other multicellular organisms and *Dictyostelium.* The other two duplicated genes have accumulated non-synonymous mutations over the course of evolution to diverge into long-branching paralogs. From the evolutionary tree, it is evident that the ancestral kinetoplast *MVP* gene underwent one round of duplication. The chromosomal proximity of *MVP2* and *MVP3* suggest a second round of duplication, resulting in the three paralogs observed. The tree indicates that all three genes were present in the common ancestor of *Trypanosoma* and *Leishmania*. The *MVP1* gene, in both *Leishmania* and *Trypanosoma,* maps to chromosome 5. *MVP2* and *MVP3* in *Trypanosoma* are located on chromosome 10 while those of *Leishmania* are found in chromosome 21 and 36. In *Trypanosoma*, it is known that the duplicated paralogs are a substrate of positive selection (Emes and Yang 2008). Whether the paralogs perform similar functions in other kinetoplastids or have additional alternative functions is worth investigating.

Based on 3D structure predictions and quaternary structure analysis, a recent study looking at the *MVP* homologs from kinetoplastids revealed that only the short branch length orthologs from kinetoplastids, (*TbMVP1* in the case of *Trypanosoma*), could dock laterally and fold into structures that resemble vault-like structures purified from rats or sea urchins (Daly et al. 2012). Though two other paralogs exist in each of the kinetoplastid species, they have not been shown to form the conventional fold or establish interface interactions to assemble into a vault-like structure. In agreement, *TbMVP1* was found to pellet at similar sucrose gradient fractions (40% -50%) as vaults from other multicellular eukaryotes and also form vault-like structures, suggesting

that *TbMVP1* sequence carries all the inherent information necessary for vault assembly.

A recent study conducted on TOR complexes in *Trypanosoma*, identified *TbMVP1* as a protein that interacts with TbTOR4, one of the four TOR kinase paralogs in *Trypanosoma*, to form a unique TOR complex TbTOR4C (Barquilla et al. 2012). TOR plays an important role in stress resistance, cell growth and also acts as an important sensor for protein synthesis, and ribosome biogenesis. The TbTOR4C is distinct from the TbTORC1 and TbTORC2 complexes that are known to control cell size and cell growth, respectively, in *Trypanosoma*. Triggered by reduced cellular energy state (high AMP:ATP ratio), the TbTOR4C complex is known to negatively regulate the differentiation of the proliferative slender bloodstream form into the G0 state arrested quiescent stumpy form. The quiescent form displays reduced motility and is also pre-adapted to transform into insect procyclic form. This implicates vaults in processes relating to cell-cycle regulation and also cell growth in response to nutrients and energy conditions.

The ubiquitous expression and wide conservation of vaults across many eukaryotes is suggestive of a very fundamental function for vaults. The appearance of vaults in kinetoplastids and their expansion into duplicated paralogs, possibly out of positive selection, is evidence that they evolved for specific functions. Since vault is a huge complex composed of many assembled amino acids, each cell expressing vaults is bound to invest significant amount of cellular energy into translating the protein. So far, the only phenotype to have been reported consistently in any *MVP* knockout studies has been a nutrient phenotype. In MEFs, the nutrient phenotype

became apparent only when serum was withdrawn completely for 24 hours as the *MVP* knockout cells displayed similar proliferation rates as control cells with 10, 5 or 2.5% serum (Kolli et al. 2004). The elevated population of cell arrest at G0 stage and increased cell death was positively correlated with growth factor deprivation. In *Dictyostelium*, the dual *MVPA* and *MVPB* knockout cells that reach only 1/3rd of cell densities when grown in nutrient limited medium do not show any profound effect in normal media (Vasu and Rome 1995). These results are comparable to that observed in *Trypanosoma* on *TbMVP1* knockdown. Cells grown in normal medium supplemented with growth factors and 15% serum showed similar proliferation rates compared to uninduced control cells, while knockdown cells grown in limited nutrient media achieve lesser cell densities than uninduced cells grown in similar conditions. The strategy employed here is an RNAi based knockdown approach, which may not display high efficacy comparable to a full-gene knockout method. The nutrient phenotype may become more obvious if such stringent check on gene expression is employed, and the remaining *MVP2* and *MVP3* genes were to be knocked out as well. On the other hand, overexpression of *TbMVP1* seems to improve cell survival during conditions of nutrient limitation. These observations seem to fit well with the proposed synthesis-turnover based nutrient absorption function for the vault complex. Vaults are possibly involved in mobilizing amino acids within the cells during conditions of amino acid deprivation.

It is also observed that vault containing control cells compromise their motility and attach to the substrate as a nutrient stress response. *MVP*1 knockdown cells with no vault lose the ability to attach to substrate under

these conditions, which may indicate that vaults act as a nutrient sensor, perhaps structurally altering cell motility within the FAZ. In extreme nutrient stress situations, vaults could possibly co-ordinate events that would confer partial resistance to unfavorable conditions in *Trypanosomatids* by maintaining the parasite in a quiescent state. This phenomenon is reminiscent of those observed in *Trypanosoma cruzi*, wherein metacyclogenesis and adhesion to substrate are induced by nutritional stress. Metacyclogenesis occurs when the insect form epimastigotes attach to intestinal walls before and differentiate into infective metacyclic trypomastigotes (Bonaldo et al. 1988). Incubation of the epimastigotes in chemically defined limited nutrient medium TAU3AAG creates a nutrient stress like environment and forces the parasite to attach to the walls of the substrate, before they differentiate (Figueiredo et al. 2000). During this process, T. *cruzi* is known to rely on proteins accumulated in stores called reservosomes as energy source to fuel the attachment and differentiation process, thus establishing an important link between nutrient stress, adhesion and cell differentiation (Soares 1999). Unlike the T. *cruzi*, the epimastigotes of T. *brucei* are known to attach to salivary gland before they differentiate into metacyclic trypomastigotes. The attachment observed in the case of uninduced control procyclic Trypanosomes in limited nutrient conditions may be compared to such a process. Thus, it could be suggested that vaults have a role to play in triggering attachment of cells as response to starvation. The cells probably display a quorum sensing mechanism that forces them into a quiescent non dividing state, thus conferring resistance to unfavorable conditions. A similar response is well known in *Dictyostelium*, whose cells become adherent and non-dividing during starvation-induced

multicellular fruiting body formation. As mentioned earlier, *TbMVP1* is a part of the TbTOR4 complex that regulates differentiation of the proliferative bloodstream form into a quiescent stumpy form. This is suggestive of important roles for vaults in nutrient assisted morphogenetic events widely prevalent in single-celled eukaryotic species. Vaults seem to function both as a nutrient sensor, probably on similar lines of the TOR complex, and may also exert pivotal roles during differentiation process in *Trypanosoma brucei*.

Chapter 6

Conclusions and Future Directions

Unraveling the evolutionary history of the vault genes helped identify unifying traits in all organisms harboring vault genes. The evolutionary analysis also shed light on explaining the puzzling phylogenetic distribution of vaults. Accordingly, it is suggested that *MVP* genes emerged in separate eukaryotic clades and bacterial species by independent horizontal gene transfer events. Vaults are conspicuously missing in organisms that retain essential amino acids biosynthetic capability including plants, algae and fungi. Intriguingly, eukaryotes and bacteria that harbor *MVP* have had ancestral loss of enzymes pertaining to essential amino acid biosynthesis. It appears that organisms with vaults all depend on external sources for essential amino acid needs with the exception of the non-nitrogen-fixing cyanobacteria. However, not all heterotrophic protists have vault genes. The cyanobacterium that carries *MVP* gene also has lost genes pertaining to nitrogen fixation and relies on its internal store of amino acids to derive energy. Given its massive size and amino acid polymerization capability, vaults could very well have originated

in such an organism to compensate for amino acid needs or the loss of nitrogen fixation. The subsequent acquisition of this gene into early single-celled eukaryotes or bacterial heterotrophs may have helped compensate for the loss of essential amino acid synthesis.

The nutrient storage properties of vaults and fate of its protein composition (through catabolism and amino acid recycling into various metabolic routes) revealed interesting results. It was found that the vault amino acid composition is well suited for gluconeogenesis and nucleotide precursor formation, with one vault particle being very similar to a glycogen molecule in terms of carbohydrate equivalents after gluconeogenesis. Also, the proposed synthesis-turnover based nutrient absorption function fits well with various reported vault expression and turnover patterns published in literature and other high throughput expression data. The nutrient absorption and retention function succinctly explains the role of vaults in innate immunity via an intracellular parasite (bacterial or viral) starvation mechanism. A high composition of amino acid precursors for Glutamate explains the axonal transport of vaults as a potential enriched source of neurotransmitter equivalents. These functional roles previously associated with vault that remain unfalsified are, in fact, well explained by considering vault complexes most simply as stable ribonucleoprotein precursor storage particles with regulated synthesis, assembly, disassembly and turnover.

The evidence presented here suggests that the connection between vault and autophagic vesicle activity may have arisen in single-celled heterotrophic eukaryotes through independent lateral transfer events. Vault may have played an important role in facilitating the emergence and radiation

of multicellular heterotrophic animals by first facilitating a flux of amino acid from altruistic storage cells to active germinal tissues as in the case of *Dictyostelium*. This may have led to an evolutionary breakthrough whereby multicellular animals were able to evolve separate specialized digestive cells, protective cells, and glutamate receptor-based nutrient signaling which subsequently evolved into nerve cells and synapses.

If vault had originally contributed a prophylactic advantage against bacterial intracellular parasites via nutrient starvation, loss of vault or down regulation of vault would lead to increased bacterial parasite tolerance. This tolerance could lead to the development of beneficial endosymbiotic relationships so loss of vault in a common ancestor of ecdysozoan protostomes may have been complemented by a general increase in tolerance for endosymbiosis, allowing these creatures to evolve and occupy carbohydrate based nutrient niches (e.g. termites, ants, aphids) that would not support the multicellular phyla with sufficient essential amino acids.

This study extends the characterization of vaults, for the first time, to an ancient single-celled eukaryote *Trypanosoma brucei* in an attempt to track down ancestral roles for vaults and also to evaluate the proposed functional roles for the vault complex. *T. brucei* is evolutionarily distant from the single-celled amoebozoans and other multicellular eukaryotes in which vaults have been studied so far. It is revealed that vault genes have undergone paralogous expansion in a common ancestor of kinetoplastids to give rise to three independently diverging proteins (*TbMVP1-3*).

The knockdown of Tb*MVP*1 intriguingly displayed no altered growth rates under normal cultivation conditions. This suggests that Tb*MVP*1 is

dispensable under normal conditions and may play passive roles during events pertaining to cell division or motility. However under limited-nutrient conditions, knockdown affected the growth pattern resulting in reduced cell densities. The fact that knockdown alters growth patterns during limited-nutrient condition stresses the role of vaults in nutrient-sensing and controlling cell proliferation rates. This becomes important in a parasite that constantly alternates between the mammalian and insect hosts, where it may be subjected to different nutrient environments. The mammalian host provides the parasite with a rich nutrient milieu, while within the insect gut the parasite is subjected to severe nutrient deficit. Thus, Tb*MVP*1 is believed to play important roles during cell survival under nutrient stress or starvation.

Based on the observations made thus far, it is suggested that a limited-nutrient environment pronounces the functions of vaults which otherwise plays subtle roles in the cell. The constantly fed knockout models failed to recapitulate any of the phenotypes established in *in vitro* experiments. This might be because *in vitro* cultures are more often serum-starved to induce cell cycle synchronization, thus triggering *MVP* knockout phenotypes. Future experiments centered on characterizing vault functions in a limited-nutrient environment may provide a more definitive answer with regards to its function.

With respect to *Trypanosoma*, the current study uses the procyclic forms that are adapted to survive in a limited-nutrient environment in the insect hosts. With regards to the proposed function as an ancestral reserve of amino acids, vaults may have a prominent role to play in the bloodstream form in sequestering nutrient amino acids. These sequestered amino acids may help

its survival in the nutrient-limited insect hosts. Future studies focusing on knockdown of Tb*MVP*1 in the bloodstream form of *Trypanosoma* may reveal a stronger phenotype. A triple knockdown may accentuate this phenotype. Autophagy is also more pronounced in the bloodstream form and experiments that look for vault localization within autophagosomes upon induction of macroautophagy may give better insights into the functioning of vaults.

Since vaults are relatively stable and do not interfere with normal cellular functions, the potential of recombinant vaults as a drug delivery vehicle has remained a steady subject of research. With regards to the proposed function of vaults as a nutrient absorption particle, introducing vaults into plants or other cyanobacteria that are routinely used as dietary supplements (like Spirulina), may potentially serve to increase the amino acid content of these organisms. Vaults, being enriched in nutrient amino acids including BCAA, are well suited to be included in dietary supplements. To this end, a fellow student in the Hogue laboratory has pursued the creation of transgenic strains of rice and Arabidopsis with *MVP* genes in the hope of enriching protein content.

It is reasonable to assume that vault structures evolved originally as particles for storing amino acids and later possibly evolved other additional functions with the incorporation of minor vault proteins. In the context of the theory presented here, observed mechanisms for vault expression and degradation may deserve a fresh examination. *VPARP* could represent an internal regulatory sensor molecule that is triggered to disrupt the vault particle via poly-ADP-ribosylation, possibly leading to its recognition by ubiquitin ligases. This would be consistent with the mechanistic role that the

PARP homologues play in disrupting nucleosomal structures during DNA repair (Schreiber et al. 2006). *VPARP* may be sensitive to the oxidative state of the cell or to specific nutrient concentrations. Recruitment of *TEP1* within vaults may have evolved to potentially concentrate valuable ribonucleotides within vault interiors as a potential antiviral mechanism.

In conclusion, if vault is indeed an evolutionarily conserved nutrient sequestering particle, it joins the likes of glycogen, starch, and triglycerides as a metabolically and nutritionally important molecular complex that warrants further study, and there are a large number of experiments that can be done in many organisms to further explore this possibility.

Chapter 7

References

Absalon S, Kohl L, Branche C, Blisnick T, Toutirais G, Rusconi F, Cosson J, Bonhivers M, Robinson D, Bastin P. 2007. Basal body positioning is controlled by flagellum formation in Trypanosoma brucei. PloS one 2(5):e437.

Allen CL, Goulding D, Field MC. 2003. Clathrin-mediated endocytosis is essential in Trypanosoma brucei. EMBO J 22(19):4991-5002.

Amano A, Nakagawa I, Yoshimori T. 2006. Autophagy in innate immunity against intracellular bacteria. J Biochem 140(2):161-166.

An HJ, Ryu SJ, Kim SY, Choi HR, Chung JH, Park SC. 2009. Age associated high level of major vault protein is p53 dependent. Cell Biochem Funct 27(5):289-295.

Anderson DH, Kickhoefer VA, Sievers SA, Rome LH, Eisenberg D. 2007. Draft crystal structure of the vault shell at 9-A resolution. PLoS Biol 5(11):e318.

Andersson JO. 2005. Lateral gene transfer in eukaryotes. Cell Mol Life Sci 62(11):1182-1197.

Andersson JO, Hirt RP, Foster PG, Roger AJ. 2006. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. BMC Evol Biol 6:27.

Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Ségurens B, Daubin V, Anthouard V, Aiach N et al. 2006. Global trends of whole-genome duplications revealed by the ciliate *Paramecium* tetraurelia. Nature 444(7116):171-178.

Barquilla A, Saldivia M, Diaz R, Bart JM, Vidal I, Calvo E, Hall MN, Navarro M. 2012. Third target of rapamycin complex negatively regulates development of quiescence in Trypanosoma brucei. Proc Natl Acad Sci U S A

Basu B, Sandhu SK, de Bono JS. 2012. PARP inhibitors: mechanism of action and their potential role in the prevention and treatment of cancer. Drugs 72(12):1579-1590.

Bateman A, Kickhoefer V. 2003. The TROVE module: a common element in Telomerase, Ro and Vault ribonucleoproteins. BMC Bioinformatics 4:49.

Berger A, Mutch DM, German JB, Roberts MA. 2002. Dietary effects of arachidonate-rich fungal oil and fish oil on murine hepatic and hippocampal gene expression. Lipids Health Dis 1:2.

Berger W, Spiegl-Kreinecker S, Buchroithner J, Elbling L, Pirker C, Fischer J, Micksche M. 2001. Overexpression of the human major vault protein in astrocytic brain tumor cells. Int J Cancer 94(3):377-382.

Berger W, Steiner E, Grusch M, Elbling L, Micksche M. 2008. Vaults and the major vault protein: Novel roles in signal pathway regulation and immunity. Cell Mol Life Sci 66(1):43-61.

Bonaldo MC, Souto-Padron T, de Souza W, Goldenberg S. 1988. Cell-substrate adhesion during Trypanosoma cruzi differentiation. J Cell Biol 106(4):1349-1358.

Brocchieri L, Karlin S. 2005. Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Res 33(10):3390-3400.

Bronte V, Zanovello P. 2005. Regulation of immune responses by L-arginine metabolism. Nat Rev Immunol 5(8):641-654.

Champion CI, Kickhoefer VA, Liu G, Moniz RJ, Freed AS, Bergmann LL, Vaccari D, Raval-Fernandes S, Chan AM, Rome LH et al. 2009. A vault nanoparticle vaccine induces protective mucosal immunity. PloS one 4(4):e5409.

Chen IHB, Huber M, Guan T, Bubeck A, Gerace L. 2006. Nuclear envelope transmembrane proteins (NETs) that are up-regulated during myogenesis. BMC Cell Biol 7:38.

Chugani DC, Kedersha NL, Rome LH. 1991. Vault immunofluorescence in the brain: new insights regarding the origin of microglia. J Neurosci 11(1):256-268.

Chugani DC, Rome LH, Kedersha NL. 1993. Evidence that vault ribonucleoprotein particles localize to the nuclear pore complex. J Cell Sci 106 (Pt 1):23-29.

Citarelli M, Teotia S, Lamb RS. 2010. Evolutionary history of the poly(ADP-ribose) polymerase gene family in eukaryotes. BMC Evol Biol 10:308.

Cmarko D, Smigova J, Minichova L, Popov A. 2008. Nucleolus: the ribosome factory. Histol Histopathol 23(10):1291-1298.

Crepel A, Steyaert J, De la Marche W, De Wolf V, Fryns JP, Noens I, Devriendt K, Peeters H. 2011. Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. Am J Med Genet B Neuropsychiatr Genet 156(2):243-245.

d'Adda di Fagagna F, Hande MP, Tong WM, Lansdorp PM, Wang ZQ, Jackson SP. 1999. Functions of poly(ADP-ribose) polymerase in controlling telomere length and chromosomal stability. Nat Genet 23(1):76-80.

Daly TK, Sutherland-Smith AJ, Penny D. 2012. Beyond BLASTing: Tertiary and quaternary structure analysis helps identify Major Vault Proteins. Genome biology and evolution 5 (1): 217-32, 2013

Davidge JA, Chambers E, Dickinson HA, Towers K, Ginger ML, McKean PG, Gull K. 2006. Trypanosome IFT mutants provide insight into the motor location for mobility of the flagella connector and flagellar membrane formation. J Cell Sci 119(Pt 19):3935-3943.

de Graffenried CL, Ho HH, Warren G. 2008. Polo-like kinase is required for Golgi and bilobe biogenesis in Trypanosoma brucei. J Cell Biol 181(3):431-438.

den Boer ML, Pieters R, Kazemier KM, Rottier MM, Zwaan CM, Kaspers GJ, Janka-Schaub G, Henze G, Creutzig U, Scheper RJ et al. 1998. Relationship between major vault protein/lung resistance protein, multidrug resistance-associated protein, P-glycoprotein expression, and drug resistance in childhood leukemia. Blood 91(6):2092-2098.

Dengjel J, Schoor O, Fischer R, Reich M, Kraus M, Müller M, Kreymborg K, Altenberend F, Brandenburg J, Kalbacher H et al. 2005. Autophagy promotes MHC class II presentation of peptides from intracellular source proteins. Proc Natl Acad Sci U S A 102(22):7922-7927.

Dickenson NE, Moore D, Suprenant KA, Dunn RC. 2007. Vault ribonucleoprotein particles and the central mass of the nuclear pore complex. Photochem Photobiol 83(3):686-691.

DOERING CB, DANNER DJ. 2000. Amino acid deprivation induces translation of branched-chain a-ketoacid dehydrogenase kinase. Am J Physiol Cell Physiol 279:1587-1594.

Dondero F, Dagnino A, Jonsson H, Caprì F, Gastaldi L, Viarengo A. 2006. Assessing the occurrence of a stress syndrome in mussels (Mytilus edulis) using a combined biomarker/gene expression approach. Aquat Toxicol. 1:S13-S24.

Dornan D, Wertz I, Shimizu H, Arnott D, Frantz GD, Dowd P, O'Rourke K, Koeppen H, Dixit VM. 2004. The ubiquitin ligase COP1 is a critical negative regulator of p53. Nature 429(6987):86-92.

Douglas AE. 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria Buchnera. Annu Rev Entomol 43:17-37.

Dumontier M, Yao R, Feldman HJ, Hogue CWV. 2005. Armadillo: domain boundary prediction by amino acid composition. J Mol Biol 350(5):1061-1073.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32(5):1792-1797.

Eichenmüller B, Kedersha N, Solovyeva E, Everley P, Lang J, Himes RH, Suprenant KA. 2003. Vaults bind directly to microtubules via their caps and not their barrels. Cell Motil Cytoskeleton 56(4):225-236.

Eme, Laura, Aurélie Trilles, David Moreira, and Céline Brochier-Armanet. "The phylogenomic analysis of the anaphase promoting complex and its targets points to complex and modern-like control of the cell cycle in the last common ancestor of eukaryotes." *BMC Evol Biol* 11 (2011): 265.

Emes RD, Yang Z. 2008. Duplicated paralogous genes subject to positive selection in the genome of Trypanosoma brucei. PloS one 3(5):e2295.

Engene N, Rottacker EC, Kaštovský J, Byrum T, Choi H, Ellisman MH, Komárek J, Gerwick WH. 2012. Moorea producens gen. nov., sp. nov. and Moorea bouillonii comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. Int J Syst Evol Microbiol 62(Pt 5):1171-1178.

Esfandiary R, Kickhoefer VA, Rome LH, Joshi SB, Middaugh CR. 2008. Structural stability of vault particles. J Pharm Sci 98(4):1376-1386.

Farr H, Gull K. 2012. Cytokinesis in trypanosomes. Cytoskeleton (Hoboken, N.J.) 69(11):931-941.

Field MC, Carrington M. 2009. The trypanosome flagellar pocket. Nat Rev Microbiol 7(11):775-786.

Figueiredo RC, Rosa DS, Soares MJ. 2000. Differentiation of Trypanosoma cruzi epimastigotes: metacyclogenesis and adhesion to substrate are triggered by nutritional stress. J Parasitol 86(6):1213-1218.

Ford LP, Shay JW, Wright WE. 2001. The La antigen associates with the human telomerase ribonucleoprotein and influences telomere length in vivo. RNA 7(8):1068-1075.

Gallo JM, Precigout E. 1988. Tubulin expression in trypanosomes. Biol Cell 64(2):137-143.

Guedes, R L M, F Prosdocimi, G R Fernandes, L K Moura, H A L Ribeiro, and J M Ortega. "Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution." *BMC Genomics* 12 Suppl 4 (2011): S2.

Ge X, Yamamoto S, Tsutsumi S, Midorikawa Y, Ihara S, Wang SM, Aburatani H. 2005. Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. Genomics 86(2):127-141.

Goldsmith LE, Pupols M, Kickhoefer VA, Rome LH, Monbouquette HG. 2009. Utilization of a protein "shuttle" to load vault nanocapsules with gold probes and proteins. ACS nano 3(10):3175-3183.

Goldsmith LE, Yu M, Rome LH, Monbouquette HG. 2007. Vault nanocapsule dissociation into halves triggered at low pH. Biochemistry 46(10):2865-2875.

Gopinath SCB, Matsugami A, Katahira M, Kumar PKR. 2005. Human vault-associated non-coding RNAs bind to mitoxantrone, a chemotherapeutic compound. Nucleic Acids Res 33(15):4874-4881.

Guha S, Rees E, Darvasi A, Ivanov D, Ikeda M, Bergen SE, Magnusson PK, Cormican P, Morris D, Gill M et al. 2013. Implication of a rare deletion at distal 16p11.2 in schizophrenia. JAMA psychiatry (Chicago, Ill.) 70(3):253-260.

Gull K. 1999. The cytoskeleton of trypanosomatid parasites. Annu Rev Microbiol 53:629-655.

Hakvoort TBM, Moerland PD, Frijters R, Sokolović A, Labruyère WT, Vermeulen JLM, Ver Loren van Themaat E, Breit TM, Wittink FRA, van Kampen AHC et al. 2011. Interorgan coordination of the murine adaptive response to fasting. J Biol Chem 286(18):16332-16343.

Hamill D, Suprenant K. 1997. Characterization of the sea urchin major vault protein: a possible role for vault ribonucleoprotein particles in nucleocytoplasmic transport. (1). Dev Biol 190(1):117-128.

Hamill DR, Suprenant KA. 1997. Characterization of the sea urchin major vault protein: a possible role for vault ribonucleoprotein particles in nucleocytoplasmic transport. Dev Biol 190(1):117-128.

Harrington L, McPhail T, Mar V, Zhou W, Oulton R, Bass MB, Arruda I, Robinson MO. 1997. A mammalian telomerase-associated protein. Science 275(5302):973-977.

Hart SM, Ganeshaguru K, Scheper RJ, Prentice HG, Hoffbrand AV, Mehta AB. 1997. Expression of the human major vault protein LRP in acute myeloid leukemia. Exp Hematol 25(12):1227-1232.

He CY. 2007. Golgi biogenesis in simple eukaryotes. Cell Microbiol 9(3):566-572.

He CY, Ho HH, Malsam J, Chalouni C, West CM, Ullu E, Toomre D, Warren G. 2004. Golgi duplication in Trypanosoma brucei. J Cell Biol 165(3):313-321.

He CY, Pypaert M, Warren G. 2005. Golgi duplication in Trypanosoma brucei requires Centrin2. Science 310(5751):1196-1198.

Hemphill A, Lawson D, Seebeck T. 1991. The cytoskeletal architecture of Trypanosoma brucei. J Parasitol 77(4):603-612.

Henze K, Horner DS, Suguri S, Moore DV, Sánchez LB, Müller M, Embley TM. 2001. Unique phylogenetic relationships of glucokinase and glucosephosphate isomerase of the amitochondriate eukaryotes Giardia intestinalis, Spironucleus barkhanus and Trichomonas vaginalis. Gene 281(1-2):123-131.

Herlevsen M, Oxford G, Owens CR, Conaway M, Theodorescu D. 2007. Depletion of major vault protein increases doxorubicin sensitivity and nuclear accumulation and disrupts its sequestration in lysosomes. Mol Cancer Ther 6(6):1804-1813.

Herrmann C, Golkaramnay E, Inman E, Rome L, Volknandt W. 1999. Recombinant major vault protein is targeted to neuritic tips of PC12 cells. J Cell Biol 144(6):1163-1172.

Herrmann C, Volknandt W, Wittich B, Kellner R, Zimmermann H. 1996. The major vault protein (*MVP*100) is contained in cholinergic nerve terminals of electric ray electric organ. J Biol Chem 271(23):13908-13915.

Holzmann K, Ambrosch I, Elbling L, Micksche M, Berger W. 2001. A small upstream open reading frame causes inhibition of human major vault protein expression from a ubiquitous mRNA splice variant. FEBS Lett 494(1-2):99-104.

Hong YK, Foreman K, Shin JW, Hirakawa S, Curry CL, Sage DR, Libermann T, Dezube BJ, Fingeroth JD, Detmar M. 2004. Lymphatic reprogramming of blood vascular endothelium by Kaposi sarcoma-associated herpesvirus. Nat Genet 36(7):683-685.

Hopkinson BM, Roe KL, Barbeau KA. 2008. Heme uptake by Microscilla marina and evidence for heme uptake systems in the genomes of diverse marine bacteria. Appl Environ Microbiol 74(20):6263-6270.

Hosono K, Sasaki T, Minoshima S, Shimizu N. 2004. Identification and characterization of a novel gene family YPEL in a wide spectrum of eukaryotic species. Gene 340(1):31-43.

Huffman KE, Corey DR. 2005. Major vault protein does not play a role in chemoresistance or drug localization in a non-small cell lung cancer cell line. Biochemistry 44(7):2253-2261.

Hughes LC, Ralston KS, Hill KL, Zhou ZH. 2012. Three-dimensional structure of the Trypanosome flagellum suggests that the paraflagellar rod functions as a biomechanical spring. PloS one 7(1):e25700.

Huntley S, Zhang Y, Treuner-Lange A, Kneip S, Sensen CW, Søgaard-Andersen L. 2012. Complete genome sequence of the fruiting myxobacterium Corallococcus coralloides DSM 2259. J Bacteriol 194(11):3012-3013.

Iizuka T, Jojima Y, Fudou R, Hiraishi A, Ahn JW, Yamanaka S. 2003. Plesiocystis pacifica gen. nov., sp. nov., a marine myxobacterium that contains dihydrogenated menaquinone, isolated from the Pacific coasts of Japan. Int J Syst Evol Microbiol 53(Pt 1):189-195.

Iranfar N, Fuller D, Loomis WF. 2003. Genome-wide expression analyses of gene regulation during early development of Dictyostelium discoideum. Eukaryot Cell 2(4):664-670.

Ismach R, Cianci CM, Caulfield JP, Langer PJ, Hein A, McMahon-Pratt D. 1989. Flagellar membrane and paraxial rod proteins of Leishmania: characterization employing monoclonal antibodies. J Protozool 36(6):617-624.

Izquierdo MA, Scheffer GL, Flens MJ, Giaccone G, Broxterman HJ, Meijer CJ, van der Valk P, Scheper RJ. 1996. Broad distribution of the multidrug resistance-related vault lung resistance protein in normal human tissues and tumors. Am J Pathol 148(3):877-887.

Izquierdo MA, Scheffer GL, Schroeijers AB, de Jong MC, Scheper RJ. 1998. Vault-related resistance to anticancer drugs determined by the expression of the major vault protein LRP. Cytotechnology 27(1-3):137-148.

Jackson AP. 2007. Evolutionary consequences of a large duplication event in Trypanosoma brucei: chromosomes 4 and 8 are partial duplicons. BMC Genomics 8:432.

Jackson AP. 2007. Tandem gene arrays in Trypanosoma brucei: comparative phylogenomic analysis of duplicate sequence variation. BMC Evol Biol 7:54.

Jones AC, Monroe EA, Podell S, Hess WR, Klages S, Esquenazi E, Niessen S, Hoover H, Rothmann M, Lasken RS et al. 2011. Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium Lyngbya majuscula. Proceedings of the National Academy of Sciences;

Jones AC, Monroe EA, Podell S, Hess WR, Klages S, Esquenazi E, Niessen S, Hoover H, Rothmann M, Lasken RS et al. 2011. Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium Lyngbya majuscula. Proc Natl Acad Sci U S A 108(21):8815-8820.

Kane CD, Vena RM, Ouellette SP, Byrne GI. 1999. Intracellular tryptophan pool sizes may account for differences in gamma interferon-mediated inhibition and persistence of chlamydial growth in polarized and nonpolarized cells. Infect Immun 67(4):1666-1671.

Kar UK, Srivastava MK, Andersson A, Baratelli F, Huang M, Kickhoefer VA, Dubinett SM, Rome LH, Sharma S. 2011. Novel CCL21-vault nanocapsule intratumoral delivery inhibits lung cancer growth. PloS one 6(5):e18758.

Kato K, Tanaka H, Sumizawa T, Yoshimura M, Yamashita E, Iwasaki K, Tsukihara T. 2008. A vault ribonucleoprotein particle exhibiting 39-fold dihedral symmetry. Acta Crystallographica Section D Biological Crystallography 64(5):525-531.

Ke Y, Zhang EE, Hagihara K, Wu D, Pang Y, Klein R, Curran T, Ranscht B, Feng GS. 2007. Deletion of Shp2 in the brain leads to defective proliferation and differentiation in neural stem cells and early postnatal lethality. Mol Cell Biol 27(19):6706-6717.

Kedersha NL, Heuser JE, Chugani DC, Rome LH. 1991. Vaults. III. Vault ribonucleoprotein particles open into flower-like structures with octagonal symmetry. J Cell Biol 112(2):225-235.

Kedersha NL, Hill DF, Kronquist KE, Rome LH. 1986. Subpopulations of liver coated vesicles resolved by preparative agarose gel electrophoresis. J Cell Biol 103(1):287-297.

Kedersha NL, Miquel MC, Bittner D, Rome LH. 1990. Vaults. II. Ribonucleoprotein structures are highly conserved among higher and lower eukaryotes. J Cell Biol 110(4):895-901.

Kedersha NL, Rome LH. 1986. Preparative agarose gel electrophoresis for the purification of small organelles and particles. Anal Biochem 156(1):161-170.

Kedersha NL, Rome LH. 1990. Vaults: large cytoplasmic RNP's that associate with cytoskeletal elements. Mol Biol Rep 14(2-3):121-122.

Kedersha NL, Rome LH. 1986. Isolation and Characterization of a Novel Ribonucleoprotein Particle: Large Structures Contain a Single Species of Small RNA. J Cell Biol 69:9-709.

Kemkemer C, Kohn M, Cooper DN, Froenicke L, Högel J, Hameister H, Kehrer-Sawatzki H. 2009. Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. BMC Evol Biol 9:84.

Kennedy AR, Pissios P, Otu H, Roberson R, Xue B, Asakura K, Furukawa N, Marino FE, Liu FF, Kahn BB et al. 2007. A high-fat, ketogenic diet induces a unique metabolic state in mice. Am J Physiol Endocrinol Metab 292(6):E1724-E1739.

Kickhoefer VA, Liu Y, Kong LB, Snow BE, Stewart PL, Harrington L, Rome LH. 2001. The Telomerase/vault-associated protein TEP1 is required for vault RNA stability and its association with the vault particle. J Cell Biol 152(1):157-164.

Kickhoefer VA, Rajavel KS, Scheffer GL, Dalton WS, Scheper RJ, Rome LH. 1998. Vaults are up-regulated in multidrug-resistant cancer cell lines. J Biol Chem 273(15):8971-8974.

Kickhoefer VA, Searles RP, Kedersha NL, Garber ME, Johnson DL, Rome LH. 1993. Vault ribonucleoprotein particles from rat and bullfrog contain a related small RNA that is transcribed by RNA polymerase III. J Biol Chem 268(11):7868-7873.

Kickhoefer VA, Siva AC, Kedersha NL, Inman EM, Ruland C, Streuli M, Rome LH. 1999. The 193-kD vault protein, VPARP, is a novel poly(ADP-ribose) polymerase. J Cell Biol 146(5):917-928.

Kickhoefer VA, Stephen AG, Harrington L, Robinson MO, Rome LH. 1999. Vaults and telomerase share a common subunit, TEP1. J Biol Chem 274(46):32712-32717.

Kickhoefer VA, Vasu SK, Rome LH. 1996. Vaults are the answer, what is the question? Trends Cell Biol 6(5):174-178.

Kickhoefer VA. 2005. Engineering of vault nanocapsules with enzymatic and fluorescent properties. Proceedings of the National Academy of Sciences 102(12):4348-4352.

Kickhoefer V, Han M, Raval-Fernandes S, Poderycki M, Moniz R, Vaccari D, Silvestry M, Stewart P, Kelly K, Rome L. 2009. Targeting vault nanoparticles to specific cell surface receptors. ACS Nano 3(1):27-36.

Kickhoefer VA, Emre N, Stephen AG, Poderycki MJ, Rome LH. 2003. Identification of conserved vault RNA expression elements and a non-expressed mouse vault RNA gene. Gene 309(2):65-70.

Kickhoefer VA, Poderycki MJ, Chan EKL, Rome LH. 2002. The La RNA-binding protein interacts with the vault RNA and is a vault-associated protein. J Biol Chem 277(43):41282-41286.

Kiela PR, Midura AJ, Kuscuoglu N, Jolad SD, Sólyom AM, Besselsen DG, Timmermann BN, Ghishan FK. 2005. Effects of Boswellia serrata in mouse models of chemically induced colitis. Am J Physiol Gastrointest Liver Physiol 288(4):G798-G808.

Kim E, Mian SLF, Yun SU, Song M, Yi KS, Ryu SH, Suh PG. 2006. Crosstalk between Src and major vault protein in epidermal growth factor-dependent cell signalling. FEBS Journal 273(4):793-804.

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, Fairclough S, Hellsten U, Isogai Y, Letunic I et al. 2008. The genome of the choanoflagellate Monosiga brevicollis and the origin of metazoans. Nature 451(7180):783-788.

Kitazono M, Sumizawa T, Takebayashi Y, Chen ZS, Furukawa T, Nagayama S, Tani A, Takao S, Aikou T, Akiyama S. 1999. Multidrug resistance and the lung resistance-related protein in human colon carcinoma SW-620 cells. J Natl Cancer Inst 91(19):1647-1653.

Klionsky DJ, Emr SD. 2000. Autophagy as a regulated pathway of cellular degradation. Science 290(5497):1717-1721.

Kohl L, Sherwin T, Gull K. 1999. Assembly of the paraflagellar rod and the flagellum attachment zone complex during the Trypanosoma brucei cell cycle. J Eukaryot Microbiol 46(2):105-109.

Kohl L, Robinson D, Bastin P. 2003. Novel roles for the flagellum in cell morphogenesis and cytokinesis of trypanosomes. EMBO J 22(20):5336-5346.

Kolli S, Zito CI, Mossink MH, Wiemer EAC, Bennett AM. 2004. The major vault protein is a novel substrate for the tyrosine phosphatase SHP-2 and scaffold protein in epidermal growth factor signaling. J Biol Chem 279(28):29374-29385.

Kong LB, Siva AC, Kickhoefer VA, Rome LH, Stewart PL. 2000. RNA location and modeling of a WD40 repeat domain within the vault. RNA 6(6):890-900.

Kong LB, Siva AC, Rome LH, Stewart PL. 1999. Structure of the vault, a ubiquitous celular component. Structure 7(4):371-379.

Kowalski MP, Dubouix-Bourandy A, Bajmoczi M, Golan DE, Zaidi T, Coutinho-Sledge YS, Gygi MP, Gygi SP, Wiemer EAC, Pier GB. 2007. Host resistance to lung infection mediated by major vault protein in epithelial cells. (1). Science 317(5834):130-132.

Kowalski MP, Dubouix-Bourandy A, Bajmoczi M, Golan DE, Zaidi T, Coutinho-Sledge YS, Gygi MP, Gygi SP, Wiemer EAC, Pier GB. 2007. Host resistance to lung infection mediated by major vault protein in epithelial cells. Science 317(5834):130-132.

Kozlov G, Vavelyuk O, Minailiuc O, Banville D, Gehring K, Ekiel I. 2006. Solution structure of a two-repeat fragment of major vault protein. J Mol Biol 356(2):444-452.

Kraft C, Deplazes A, Sohrmann M, Peter M. 2008. Mature ribosomes are selectively degraded upon starvation by an autophagy pathway requiring the Ubp3p/Bre5p ubiquitin protease. Nat Cell Biol 10(5):602-610.

Kuma A, Hatano M, Matsui M, Yamamoto M, Nakaya H, Yoshimori T, Ohsumi Y, Tokuhisa T, Mizushima T. 2004. The role of autophagy during the early neonatal starvation period. NATURE Letters 432

Lacomble S, Vaughan S, Gadelha C, Morphew MK, Shaw MK, McIntosh JR, Gull K. 2009. Three-dimensional cellular architecture of the flagellar pocket and associated cytoskeleton in trypanosomes revealed by electron microscope tomography. J Cell Sci 122(Pt 8):1081-1090.

Lacomble S, Vaughan S, Gadelha C, Morphew MK, Shaw MK, McIntosh JR, Gull K. 2010. Basal body movements orchestrate membrane organelle division and cell morphogenesis in Trypanosoma brucei. J Cell Sci 123(Pt 17):2884-2891.

Lange C, Walther W, Schwabe H, Stein U. 2000. Cloning and initial analysis of the human multidrug resistance-related *MVP*/LRP gene promoter. Biochem Biophys Res Commun 278(1):125-133.

Lavarone E, Puppin C, Passon N, Filetti S, Russo D, Damante G. 2013. The PARP inhibitor PJ34 modifies proliferation, NIS expression and epigenetic marks in thyroid cancer cell lines. Mol Cell Endocrinol 365(1):1-10.

Leonhardt RM, Lee SJ, Kavathas PB, Cresswell P. 2007. Severe tryptophan starvation blocks onset of conventional persistence and reduces reactivation of Chlamydia trachomatis. Infect Immun 75(11):5105-5117.

Li JY, Volknandt W, Dahlstrom A, Herrmann C, Blasi J, Das B, Zimmermann H. 1999. Axonal transport of ribonucleoprotein particles (vaults). Neuroscience 91(3):1055-1065.

Li X, Madison BB, Zacharias W, Kolterud A, States D, Gumucio DL. 2007. Deconvoluting the intestine: molecular evidence for a major role of the mesenchyme in the modulation of signaling cross talk. Physiol Genomics 29(3):290-301.

Li Z, Umeyama T, Li Z, Wang CC. 2010. Polo-like kinase guides cytokinesis in Trypanosoma brucei through an indirect means. Eukaryot Cell 9(5):705-716.

Liang P, Wan Y, Yan Y, Wang Y, Luo N, Deng Y, Fan X, Zhou J, Li Y, Wang Z et al. 2010. *MVP* interacts with YPEL4 and inhibits YPEL4-mediated activities of the ERK signal pathway. Biochem Cell Biol 88(3):445-450.

Liu B, Wang T, Wang L, Wang C, Zhang H, Gao GD. 2011. Up-regulation of major vault protein in the frontal cortex of patients with intractable frontal lobe epilepsy. J Neurol Sci 308(1-2):88-93.

Liu Y, Snow BE, Hande MP, Baerlocher G, Kickhoefer VA, Yeung D, Wakeham A, Itie A, Siderovski DP, Lansdorp PM et al. 2000. Telomerase-associated protein TEP1 is not essential for telomerase activity or telomere length maintenance in vivo. Mol Cell Biol 20(21):8178-8184.

Liu Y, Snow BE, Kickhoefer VA, Erdmann N, Zhou W, Wakeham A, Gomez M, Rome LH, Harrington L. 2004. Vault poly(ADP-ribose) polymerase is associated with mammalian telomerase and is dispensable for telomerase function and vault structure in vivo. Mol Cell Biol 24(12):5314-5323.

Luo A, Qiao H, Zhang Y, Shi W, Ho SY, Xu W, Zhang A, Zhu C. 2010. Performance of criteria for selecting evolutionary models in phylogenetics: a comprehensive study based on simulated datasets. BMC Evol Biol 10:242.

Maraia RJ, Bayfield MA. 2006. The La protein-RNA complex surfaces. Mol Cell 21(2):149-152.

Marroni M, Agrawal ML, Kight K, Hallene KL, Hossain M, Cucullo L, Signorelli K, Namura S, Bingaman W, Janigro D. 2003. Relationship between expression of multiple drug resistance proteins and p53 tumor suppressor gene proteins in human brain astrocytes. Neuroscience 121(3):605-617.

Mason DX, Autexier C, Greider CW. 2001. Tetrahymena proteins p80 and p95 are not core telomerase components. Proc Natl Acad Sci U S A 98(22):12368-12373.

McKean PG. 2003. Coordination of cell cycle and cytokinesis in Trypanosoma brucei. Curr Opin Microbiol 6(6):600-607.

Mikyas Y, Makabi M, Raval-Fernandes S, Harrington L, Kickhoefer VA, Rome LH, Stewart PL. 2004. Cryoelectron microscopy imaging of recombinant and tissue derived vaults: localization of the *MVP* N termini and VPARP. J Mol Biol 344(1):91-105.

Minaguchi T, Waite KA, Eng C. 2006. Nuclear localization of PTEN is regulated by Ca(2+) through a tyrosil phosphorylation-independent conformational modification in major vault protein. Cancer Res 66(24):11677-11682.

Mizushima N. 2007. The role of mammalian autophagy in protein metabolism. Proc. Jpn. Acad., Ser. B 82:39.

Mizushima N, Klionsky DJ. 2007. Protein turnover via autophagy: implications for metabolism. Annu Rev Nutr 27:19-40.

Mizushima N, Levine B, Cuervo AM, Klionsky DJ. 2008. Autophagy fights disease through cellular self-digestion. Nature 451(7182):1069-1075.

Mortimore GE, Pösö AR. 1987. Intracellular protein catabolism and its control during nutrient deprivation and supply. Annu Rev Nutr 7:539-564.

Mossink MH, de Groot J, van Zon A, Fränzel-Luiten E, Schoester M, Scheffer GL, Sonneveld P, Scheper RJ, Wiemer EAC. 2003. Unimpaired dendritic cell functions in *MVP*/LRP knockout mice. Immunology 110(1):58-65.

Mossink MH, van Zon A, FrÃ¤nzel-Luiten E, Schoester M, Kickhoefer VA, Scheffer GL, Scheper RJ, Sonneveld P, Wiemer EAC. 2002. Disruption of the murine major vault protein (*MVP*/LRP) gene does not induce hypersensitivity to cytostatics. Cancer Res 62(24):7298-7304.

Mossink MH, van Zon A, Scheper RJ, Sonneveld P, Wiemer EAC. 2003. Vaults: a ribonucleoprotein particle involved in drug resistance? Oncogene 22(47):7458-7467.

Mossink M, van Zon A, Fränzel-Luiten E, Schoester M, Scheffer GL, Scheper RJ, Sonneveld P, Wiemer EAC. 2002. The genomic sequence of the murine major vault protein and its promoter. Gene 294(1-2):225-232.

Nandy C, Mrázek J, Stoiber H, Grässer FA, Hüttenhofer A, Polacek N. 2009. Epstein-barr virus-induced expression of a novel human vault RNA. J Mol Biol 388(4):776-784.

Narui K, Noguchi N, Saito A, Kakimi K, Motomura N, Kubo K, Takamoto S, Sasatsu M. 2009. Anti-infectious activity of tryptophan metabolites in the L-tryptophan-L-kynurenine pathway. Biol Pharm Bull 32(1):41-44.

Nguyen H, Sankaran S, Dandekar S. 2006. Hepatitis C virus core protein induces expression of genes regulating immune evasion and anti-apoptosis in hepatocytes. Virology 354(1):58-68.

Nosenko T, Bhattacharya D. 2007. Horizontal gene transfer in chromalveolates. BMC Evol Biol 7:173.

Novarino G, El-Fishawy P, Kayserili H, Meguid NA, Scott EM, Schroth J, Silhavy JL, Kara M, Khalil RO, Ben-Omran T et al. 2012. Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy. Science 338(6105):394-397.

Onodera J, Ohsumi Y. 2005. Autophagy is required for maintenance of amino acid levels and protein synthesis under nitrogen starvation. J Biol Chem 280(36):31582-31586.

Otto GP, Wu MY, Kazgan N, Anderson OR, Kessin RH. 2003. Macroautophagy is required for multicellular development of the social amoeba Dictyostelium discoideum. J Biol Chem 278(20):17636-17645.

Owen OE, Kalhan SC, Hanson RW. 2002. The key role of anaplerosis and cataplerosis for citric acid cycle function. J Biol Chem 277(34):30409-30412.

Pan HC, Lin JF, Ma LP, Shen YQ, Schachner M. 2013. Major vault protein promotes locomotor recovery and regeneration after spinal cord injury in adult zebrafish. Eur J Neurosci 37(2):203-211.

Parsons SJ, Parsons JT. 2004. Src family kinases, key regulators of signal transduction. Oncogene 23(48):7906-7909.

Paspalas CD, Perley CC, Venkitaramani DV, Goebel-Goody SM, Zhang Y, Kurup P, Mattis JH, Lombroso PJ. 2009. Major vault protein is expressed along the nucleus-neurite axis and associates with mRNAs in cortical neurons. Cereb Cortex 19(7):1666-1677.

Poderycki MJ, Kickhoefer VA, Kaddis CS, Raval-Fernandes S, Johansson E, Zink JI, Loo JA, Rome LH. 2006. The vault exterior shell is a dynamic structure that allows incorporation of vault-associated proteins into its interior. Biochemistry 45(39):12184-12193.

Poderycki MJ, Rome LH, Harrington L, Kickhoefer VA. 2005. The p80 homology region of TEP1 is sufficient for its association with the telomerase and vault RNAs, and the vault particle. Nucleic Acids Res 33(3):893-902.

Potireddy S, Vassena R, Patel BG, Latham KE. 2006. Analysis of polysomal mRNA populations of mouse oocytes and zygotes: dynamic changes in maternal mRNA utilization and function. Dev Biol 298(1):155-166.

Qu CK, Yu WM, Azzarelli B, Feng GS. 1999. Genetic evidence that Shp-2 tyrosine phosphatase is a signal enhancer of the epidermal growth factor receptor in mammals. Proc Natl Acad Sci U S A 96(15):8528-8533.

Querol-Audí J, Casañas A, Usón I, Luque D, Castón JR, Fita I, Verdaguer N. 2009. The mechanism of vault opening from the high resolution structure of the N-terminal repeats of *MVP*. EMBO J 28(21):3450-3457.

Querol-Audí J, Perez-Luque R, Fita I, Lopéz-Iglesias C, Castón JR, Carrascosa JL, Verdaguer N. 2005. Preliminary analysis of two and three dimensional crystals of vault ribonucleoprotein particles. J Struct Biol 151(1):111-115.

Rao N, Jhamb D, Milner DJ, Li B, Song F, Wang M, Voss SR, Palakal M, King MW, Saranjami B et al. 2009. Proteomic analysis of blastema formation in regenerating axolotl limbs. BMC Biol 7:83.

Raval-Fernandes S, Kickhoefer VA, Kitchen C, Rome LH. 2005. Increased susceptibility of vault poly(ADP-ribose) polymerase-deficient mice to carcinogen-induced tumorigenesis. Cancer Res 65(19):8846-8852.

Ravi V, Venkatesh B. 2008. Rapidly evolving fish genomes and teleost diversity. Curr Opin Genet Dev 18(6):544-550.

Rimsza LM, Campbell K, Dalton WS, Salmon S, Willcox G, Grogan TM. 1999. The major vault protein (*MVP*), a new multidrug resistance associated protein, is frequently expressed in multiple myeloma. Leuk Lymphoma 34(3-4):315-324.

Robinson DR, Gull K. 1991. Basal body movements as a mechanism for mitochondrial genome segregation in the trypanosome cell cycle. Nature 352(6337):731-733.

Robinson DR, Sherwin T, Ploubidou A, Byard EH, Gull K. 1995. Microtubule polarity and dynamics in the control of organelle positioning, segregation, and cytokinesis in the trypanosome cell cycle. J Cell Biol 128(6):1163-1172.

Rome L, Kedersha N, Chugani D. 1991. Unlocking vaults: organelles in search of a function. Trends Cell Biol 1(2-3):47-50.

Ruiz-Trillo I, Inagaki Y, Davis LA, Sperstad S, Landfald B, Roger AJ. 2004. Capsaspora owczarzaki is an independent opisthokont lineage. Curr Biol 14(22):R946-R947.

Ruiz-Trillo I, Lane CE, Archibald JM, Roger AJ. 2006. Insights into the evolutionary origin and genome architecture of the unicellular opisthokonts Capsaspora owczarzaki and Sphaeroforma arctica. J Eukaryot Microbiol 53(5):379-384.

Ruiz-Trillo I, Roger AJ, Burger G, Gray MW, Lang BF. 2008. A phylogenomic investigation into the origin of metazoa. Mol Biol Evol 25(4):664-672.

Ryu SJ, An HJ, Oh YS, Choi HR, Ha MK, Park SC. 2008. On the role of major vault protein in the resistance of senescent human diploid fibroblasts to apoptosis. Cell Death Differ 15(11):1673-1680.

Ryu SJ, Park SC. 2009. Targeting major vault protein in senescence-associated apoptosis resistance. Expert Opin Ther Targets 13(4):479-484.

Sanchez S, Hourdez S, Lallier FH. 2007. Identification of proteins involved in the functioning of Riftia pachyptila symbiosis by Subtractive Suppression Hybridization. BMC Genomics

Sanchez S, Hourdez S, Lallier FH. 2007. Identification of proteins involved in the functioning of Riftia pachyptila symbiosis by Subtractive Suppression Hybridization. BMC Genomics 8:337.

Sasaki T, Hankins GR, Helm GA. 2002. Major vault protein/lung resistance-related protein (*MVP*/LRP) expression in nervous system tumors. Brain Tumor Pathol 19(2):59-62.

Saw JHW, Yuryev A, Kanbe M, Hou S, Young AG, Aizawa SI, Alam M. 2012. Complete genome sequencing and analysis of Saprospira grandis str. Lewin, a predatory marine bacterium. Standards in genomic sciences 6(1):84-93.

Scheffer GL, Wijngaard PL, Flens MJ, Izquierdo MA, Slovak ML, Pinedo HM, Meijer CJ, Clevers HC, Scheper RJ. 1995. The drug resistance-related protein LRP is the human major vault protein. Nat Med 1(6):578-582.

Schneiker S, Perlova O, Kaiser O, Gerth K, Alici A, Altmeyer MO, Bartels D, Bekel T, Beyer S, Bode E et al. 2007. Complete genome sequence of the myxobacterium Sorangium cellulosum. Nat Biotechnol 25(11):1281-1289.

Schreiber V, Dantzer F, Ame JC, de Murcia G. 2006. Poly(ADP-ribose): novel functions for an old molecule. Nat Rev Mol Cell Biol 7(7):517-528.

Schroecksnadel S, Kurz K, Weiss G, Fuchs D. 2012. Tryptophan catabolism during intracellular infection. J Infect Dis 205(10):1617-8; author reply 1618.

Schroeijers AB, Reurs AW, Scheffer GL, Stam AGM, de Jong MC, Rustemeyer T, Wiemer EAC, de Gruijl TD, Scheper RJ. 2002. Up-regulation of drug resistance-related vaults during dendritic cell development. J Immunol 168(4):1572-1578.

Schworer CM, Shiffer KA, Mortimore GE. 1981. Quantitative relationship between autophagy and proteolysis during graded amino acid deprivation in perfused rat liver. J Biol Chem 256(14):7652-7658.

Sharp PM, Li WH. 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 15(3):1281-1295.

Sherwin T, Gull K. 1989. Visualization of detyrosination along single microtubules reveals novel mechanisms of assembly during cytoskeletal duplication in trypanosomes. Cell 57(2):211-221.

Shimomura Y, Murakami T, Nakai N, Nagasaki M, Harris RA. 2004. Exercise promotes BCAA catabolism: effects of BCAA supplementation on skeletal muscle during exercise. J Nutr 134(6 Suppl):1583S-1587S.

Shiraishi F, Savageau MA. 1993. The tricarboxylic acid cycle in Dictyostelium discoideum. Systemic effects of including protein turnover in the current model. J Biol Chem 268(23):16917-16928.

Simpson AGB, Inagaki Y, Roger AJ. 2006. Comprehensive multigene phylogenies of excavate protists reveal the evolutionary positions of "primitive" eukaryotes. Mol Biol Evol 23(3):615-625.

Slesina M, Inman EM, Moore AE, Goldhaber JI, Rome LH, Volknandt W. 2006. Movement of vault particles visualized by GFP-tagged major vault protein. Cell Tissue Res 324(3):403-410.

Slesina M, Inman EM, Rome LH, Volknandt W. 2005. Nuclear localization of the major vault protein in U373 cells. Cell Tissue Res 321(1):97-104.

Slovak ML, Ho JP, Cole SP, Deeley RG, Greenberger L, de Vries EG, Broxterman HJ, Scheffer GL, Scheper RJ. 1995. The LRP gene encoding a major vault protein associated with drug resistance maps proximal to MRP on chromosome 16: evidence that chromosome breakage plays a key role in MRP or LRP gene amplification. Cancer Res 55(19):4214-4219.

Soares M. 1999. The reservosome of Trypanosoma cruzi epimastigotes: an organelle of the endocytic pathway with a role on metacyclogenesis. Mem\'o 94:139-141.

Stadler PF, Chen JJL, Hackermüller J, Hoffmann S, Horn F, Khaitovich P, Kretzschmar AK, Mosig A, Prohaska SJ, Qi X et al. 2009. Evolution of vault RNAs. Mol Biol Evol 26(9):1975-1991.

Stein U, Walther W, Laurencot CM, Scheffer GL, Scheper RJ, Shoemaker RH. 1997. Tumor necrosis factor-alpha and expression of the multidrug resistance-associated genes LRP and MRP. J Natl Cancer Inst 89(11):807-813.

Steiner E, Holzmann K, Elbling L, Micksche M, Berger W. 2006. Cellular functions of vaults and their involvement in multidrug resistance. Curr Drug Targets 7(8):923-934.

Steiner E, Holzmann K, Pirker C, Elbling L, Micksche M, Sutterlüty H, Berger W. 2006. The major vault protein is responsive to and interferes with interferon-gamma-mediated STAT1 signals. J Cell Sci. 119(Pt 3):459-469.

Steiner E, Holzmann K, Pirker C, Elbling L, Micksche M, Berger W. 2004. SP-transcription factors are involved in basal *MVP* promoter activity and its stimulation by HDAC inhibitors. Biochem Biophys Res Commun 317(1):235-243.

Steiner E, Holzmann K, Pirker C, Elbling L, Micksche M, SutterlÃ¼ty H, Berger W. 2006. The major vault protein is responsive to and interferes with interferon-gamma-mediated STAT1 signals. J Cell Sci 119(Pt 3):459-469.

Stephen AG, Raval-Fernandes S, Huynh T, Torres M, Kickhoefer VA, Rome LH. 2001. Assembly of vault-like particles in insect cells expressing only the major vault protein. J Biol Chem 276(26):23217-23220.

Stevenson EJ, Koncarevic A, Giresi PG, Jackman RW, Kandarian SC. 2005. Transcriptional profile of a myotube starvation model of atrophy. J Appl Physiol 98(4):1396-1406.

Stewart PL, Makabi M, Lang J, Dickey-Sims C, Robertson AJ, Coffman JA, Suprenant KA. 2005. Sea urchin vault structure, composition, and differential localization during development. BMC Dev Biol 5:3.

Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A et al. 2002. Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A 99(7):4465-4470.

Su YQ, Sugiura K, Woo Y, Wigglesworth K, Kamdar S, Affourtit J, Eppig JJ. 2007. Selective degradation of transcripts during meiotic maturation of mouse oocytes. Dev Biol 302(1):104-117.

Sugawara I, Akiyama S, Scheper RJ, Itoyama S. 1997. Lung resistance protein (LRP) expression in human normal tissues in comparison with that of MDR1 and MRP. Cancer Lett 112(1):23-31.

Sunnaram BL, Gandemer V, Sebillot M, Grandgirard N, Amiot L, Leray E, Goasguen JE. 2003. LRP overexpression in monocytic lineage. Leuk Res 27(8):755-759.

Sutovsky P, Manandhar G, Laurincik J, Letko J, CaamaÃ±o JN, Day BN, Lai L, Prather RS, Sharpe-Timms KL, Zimmer R et al. 2005. Expression and proteasomal degradation of the major vault protein (*MVP*) in mammalian oocytes and zygotes. Reproduction 129(3):269-282.

Sutovsky P, Manandhar G, Laurincik J, Letko J, Caamaño JN, Day BN, Lai L, Prather RS, Sharpe-Timms KL, Zimmer R et al. 2005. Expression and proteasomal degradation of the major vault protein (*MVP*) in mammalian oocytes and zygotes. Reproduction 129(3):269-282.

Szaflarski W, Nowicki M, Zabel M. 2011. [The structure of cellular vaults, their role in the normal cell and in the multidrug resistance of cancer]. Postepy Biochem 57(3):266-273.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28(10):2731-2739.

Tanaka H, Kato K, Yamashita E, Sumizawa T, Zhou Y, Yao M, Iwasaki K, Yoshimura M, Tsukihara T. 2009. The structure of rat liver vault at 3.5 angstrom resolution. Science 323(5912):384-388.

Telford MJ. 2008. Xenoturbellida: the fourth deuterostome phylum and the diet of worms. Genesis 46(11):580-586.

Tian B, Liu J, Liu B, Dong Y, Liu J, Song Y, Sun Z. 2011. p53 Suppresses lung resistance-related protein expression through Y-box binding protein 1 in the MCF-7 breast tumor cell line. J Cell Physiol 226(12):3433-3441.

Tsukamoto S, Kuma A, Mizushima N. 2008. The role of autophagy during the oocyte-to-embryo transition. Autophagy 4(8):1076-1078.

Tsukamoto S, Kuma A, Murakami M, Kishi C, Yamamoto A, Mizushima N. 2008. Autophagy is essential for preimplantation development of mouse embryos. Science 321(5885):117-120.

Tuller T, Waldman YY, Kupiec M, Ruppin E. 2010. Translation efficiency is determined by both codon bias and folding energy. Proc Natl Acad Sci U S A 107(8):3645-3650.

van Zon A, Mossink MH, Schoester M, Scheffer GL, Scheper RJ, Sonneveld P, Wiemer EA. 2001. Multiple human vault RNAs. Expression and association with the vault complex. J Biol Chem 276(40):37715-37721.

van Zon A, Mossink MH, Houtsmuller AB, Schoester M, Scheffer GL, Scheper RJ, Sonneveld P, Wiemer EAC. 2006. Vault mobility depends in part on microtubules and vaults can be recruited to the nuclear envelope. Exp Cell Res 312(3):245-255.

van Zon A, Mossink MH, Schoester M, Houtsmuller AB, Scheffer GL, Scheper RJ, Sonneveld P, Wiemer EAC. 2003. The formation of vault-tubes: a dynamic interaction between vaults and vault PARP. J Cell Sci 116(Pt 21):4391-4400.

van Zon A, Mossink MH, Schoester M, Scheffer GL, Scheper RJ, Sonneveld P, Wiemer EAC. 2002. Structural domains of vault proteins: a role for the coiled coil domain in vault assembly. Biochem Biophys Res Commun 291(3):535-541.

van Zon A, Mossink MH, Schoester M, Scheper RJ, Sonneveld P, Wiemer EAC. 2004. Efflux kinetics and intracellular distribution of daunorubicin are not affected by major vault protein/lung resistance-related protein (vault) expression. Cancer Res 64(14):4887-4892.

Vasu SK, Kedersha NL, Rome LH. 1993. cDNA cloning and disruption of the major vault protein alpha gene (*MVP*A) in Dictyostelium discoideum. J Biol Chem 268(21):15356-15360.

Vasu SK, Rome LH. 1995. Dictyostelium vaults: disruption of the major proteins reveals growth and morphological defects and uncovers a new associated protein. J Biol Chem 270(28):16588-16594.

Vaughan S. 2010. Assembly of the flagellum and its role in cell morphogenesis in Trypanosoma brucei. Curr Opin Microbiol 13(4):453-458.

Vaughan S, Kohl L, Ngai I, Wheeler RJ, Gull K. 2008. A repetitive protein essential for the flagellum attachment zone filament structure and function in Trypanosoma brucei. Protist 159(1):127-136.

Vilalta A, Kickhoefer VA, Rome LH, Johnson DL. 1994. The rat vault RNA gene contains a unique RNA polymerase III promoter composed of both external and internal elements that function synergistically. J Biol Chem 269(47):29752-29759.

Vollmar F, Hacker C, Zahedi RP, Sickmann A, Ewald A, Scheer U, Dabauvalle MC. 2009. Assembly of nuclear pore complexes mediated by major vault protein. J Cell Sci 122(Pt 6):780-786.

Walker J, Su AI, Self DW, Hogenesch JB, Lapp H, Maier R, Hoyer D, Bilbe G. 2004. Applications of a rat multiple tissue gene expression data set. Genome Res 14:742-749.

Walker JR, Su AI, Self DW, Hogenesch JB, Lapp H, Maier R, Hoyer D, Bilbe G. 2004. Applications of a rat multiple tissue gene expression data set. Genome Res 14(4):742-749.

Wang Z, Du J, Lam SH, Mathavan S, Matsudaira P, Gong Z. 2010. Morphological and molecular evidence for functional organization along the rostrocaudal axis of the adult zebrafish intestine. BMC Genomics 11:392.

Wang Z, Wang F, Tang T, Guo C. 2012. The role of PARP1 in the DNA damage response and its application in tumor therapy. Frontiers of medicine 6(2):156-164.

Wernegreen JJ. 2002. Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet 3(11):850-861.

Wood H, Roshick C, McClarty G. 2004. Tryptophan recycling is responsible for the interferon-gamma resistance of Chlamydia psittaci GPIC in indoleamine dioxygenase-expressing host cells. Mol Microbiol 52(3):903-916.

Woodward R, Gull K. 1990. Timing of nuclear and kinetoplast DNA replication and early morphological events in the cell cycle of Trypanosoma brucei. J Cell Sci 95 (Pt 1):49-57.

Yang J, Nagasawa D, Spasic M, Amolis M, Choy W, Garcia H, Prins R, Liau L, Yang I. 2012. Endogenous vaults and bioengineered vault nanoparticles for treatment of glioblastomas: implications for future targeted therapies. Neurosurg Clin N Am. 23(3):451-458.

Yang J, Kickhoefer VA, Ng BC, Gopal A, Bentolila LA, John S, Tolbert SH, Rome LH. 2010. Vaults are dynamically unconstrained cytoplasmic nanoparticles capable of half vault exchange. ACS nano 4(12):7229-7240.

Yi C, Li S, Chen X, Wiemer EAC, Wang J, Wei N, Deng XW. 2005. Major vault protein, in concert with constitutively photomorphogenic 1, negatively regulates c-Jun-mediated activator protein 1 transcription in mammalian cells. Cancer Res 65(13):5835-5840.

Yoshinari N, Ishida T, Kudo A, Kawakami A. 2009. Gene expression and functional analysis of zebrafish larval fin fold regeneration. Dev Biol 325(1):71-81.

Yoshinari N, Kawakami A. 2011. Mature and juvenile tissue models of regeneration in small fish species. Biol Bull 221(1):62-78.

Yu Z, Fotouhi-Ardakani N, Wu L, Maoui M, Wang S, Banville D, Shen SH. 2002. PTEN associates with the vault particles in HeLa cells. J Biol Chem 277(43):40247-40252.

Zheng CL, Sumizawa T, Che XF, Tsuyama S, Furukawa T, Haraguchi M, Gao H, Gotanda T, Jueng HC, Murata F et al. 2005. Characterization of *MVP* and VPARP assembly into vault ribonucleoprotein complexes. Biochem Biophys Res Commun 326(1):100-107.

Zhou Q, Liu B, Sun Y, He CY. 2011. A coiled-coil- and C2-domain-containing protein is required for FAZ assembly and cell morphology in Trypanosoma brucei. J Cell Sci 124(Pt 22):3848-3858.

Zhou Y, Wubneh H, Schwarz C, Landweber LF. 2011. A chimeric chromosome in the ciliate oxytricha resulting from duplication. J Mol Evol 73(3-4):70-73.

Zufferey F, Sherr E, Beckmann N, Hanson E, Maillard A, Hippolyte L. 2012. A 600 kb deletion syndrome at 16p11. 2 leads to energy imbalance and neuropsychiatric disorders. J Med Genet 49(10):660-668.

Chapter 8

Appendices

Table A.1 Routing Vault amino acids through Gluconeogenesis to generate Glucose (Part 1)

Amino Acid	End Products of Amino Acid degradation	Energy/Reducing Equivalents involved in Catabolism				Route to Gluconeogenesis	Amino Acid Type	Energy/Reducing Equivalents Required for Gluconeogenesis			
		ATP	NADH	NADPH	$FADH_2$			ATP	NADH	NADPH	$FADH_2$
A	Pyruvate	0	1	0	0	via pyruvate	Glucogenic	-4	-1	-2	0
C	Pyruvate	0	-1	0	0	via pyruvate	Glucogenic	-4	-1	-2	0
D	Oxaloacetic Acid	0	0	0	0	via oxaloacetate	Glucogenic	-2	-1	-2	0
E	Alpha ketoglutarate	0	0	1	0	via oxalocetate	Glucogenic	-2	1	-1	1
F	Fumaric Acid + Acetoacetic Acid	0	0	-1	0	via oxaloacetate	Gluco/Keto	-2	0	-2	0
G	Pyruvate	0	2	0	0	via pyruvate	Glucogenic	-4	-1	-2	0
H	Glutamic Acid (Alpha Ketoglutarate)	0	0	0	0	via oxaloacetate	Glucogenic	-2	1	-1	1
I	Succinyl CoA + Acetyl CoA	1	2	0	1	via oxaloacetate	Gluco/Keto	-2	0	-1	1
K	Acetoacetyl CoA	0	2	0	2	NA	Ketogenic	NA	NA	NA	NA
L	Acetoacetic Acid + Acetyl CoA	1	1	0	1	NA	Ketogenic	NA	NA	NA	NA
M	Succinyl CoA	-2	1	0	0	via oxaloacetate	Glucogenic	-2	0	-1	1
N	Oxaloacetic Acid	0	0	0	0	via oxaloacetate	Glucogenic	-2	-1	-2	0
P	Glutamic Acid (Alpha Ketoglutarate)	0	0	1	0	via oxaloacetate	Glucogenic	-2	1	-1	1
Q	2 Glutamate (Alpha Ketoglutarate)	0	0	-1	0	via oxalocetate	Glucogenic	-2	1	-1	1
R	Glutamic Acid (Alpha Ketoglutarate)	0	1	0	0	via oxaloacetate	Glucogenic	-2	1	-1	1
S	Pyruvate	0	1	0	0	via pyruvate	Glucogenic	-4	-1	-2	0
T	Acetyl CoA + Glycine (Pyruvate)	0	1	0	0	via pyruvate	Gluco/Keto	-4	-1	-2	0
V	Succinyl CoA	1	3	0	1	via oxaloacetate	Glucogenic	-2	0	-1	1
W	2 Acetyl CoA + Acetoacetyl CoA+ Alanine (Pyruvate)	0	1	-2	0	via pyruvate	Gluco/Keto	-4	-1	-2	0
Y	Fumaric Acid + Acetoacetic Acid	0	0	0	0	via oxaloacetate	Gluco/Keto	-2	0	-2	0

*NA Not Applicable Negative values indicate consumption. Positive values indicate release

Table A.1 Routing Vault amino acids through Gluconeogenesis to generate Glucose (Part 2)

Amino Acid	One Vault	Energy/Reducing Equivalents available from one Vault				Oxaloacetate formed	Pyruvate formed	Energy/Reducing Equivalents Required for Gluconeogenesis/Vault				Glucose Available
		ATP	NADH	NADPH	$FADH_2$			ATP	NADH	NADPH	$FADH_2$	
A	8106	0	8106	0	0	0	8106	-16212	-4053	-8106	0	4053
C	1002	0	-1002	0	0	0	1002	-2004	-501	-1002	0	501
D	5241	0	0	0	0	5241	0	-5241	-2620.5	-5241	0	2620.5
E	8760	0	0	8760	0	8760	0	-8760	4380	-4380	4380	4380
F	3189	0	0	-3189	0	3189	0	-3189	0	-3189	0	1594.5
G	6117	0	12234	0	0	0	6117	-12234	-3058.5	-6117	0	3058.5
H	2208	0	0	0	0	2208	0	-2208	1104	-1104	1104	1104
I	4053	4053	8106	0	4053	4053	0	-4053	0	-2026.5	2026.5	2026.5
K	4992	0	9984	0	9984	NA	NA	NA	NA	NA	NA	NA
L	10542	10542	10542	0	10542	NA	NA	NA	NA	NA	NA	NA
M	1254	-2508	1254	0	0	1254	0	-1254	0	-627	627	627
N	2487	0	0	0	0	2487	0	-2487	-1243.5	-2487	0	1243.5
P	5928	0	0	5928	0	5928	0	-5928	2964	-2964	2964	2964
Q	5808	0	0	-5808	0	5808	0	-11616	5808	-5808	5808	5808
R	6315	0	6315	0	0	6315	0	-6315	3157.5	-3157.5	3157.5	3157.5
S	5571	0	5571	0	0	0	5571	-11142	-2785.5	-5571	0	2785.5
T	4983	0	4983	0	0	0	4983	-9966	-2491.5	-4983	0	2491.5
V	8841	8841	26523	0	8841	8841	0	-8841	0	-4420.5	4420.5	4420.5
W	918	0	918	-1836	0	0	918	-1836	-459	-918	0	459
Y	1908	0	0	0	0	1908	0	-1908	0	-1908	0	954
Total	98223	20928	93534	3855	33420	55992	26697	-115194	201	-64009.5	24487.5	**44248.5**

Table A.1 Routing Vault amino acids through Gluconeogenesis to generate Glucose (Part 3)

Total Energy Available from One Vault Particle			Total Energy required to make ~44,248 molecules of Glucose		
ATP	20928		ATP	115194	
NADH	93534		NADH	201	
NADPH	3855		GTP	64009.5	
$FADH_2$	33420		$FADH_2$	24487.5	
Assuming,			Assuming,		
1 NADH = 3 ATP			1 NADH = 3 ATP		
1 NADPH = 3 ATP			1 NADPH = 3 ATP		
1 $FADH_2$ = 2 ATP			1 $FADH_2$ = 2 ATP		
Total ATP Available from One vault Particle		379,935	Total ATP required to make ~44,248 molecules of Glucose		129,626
Net ATP available = (379,935-129626) = 250309					

Table A.2 Energy available from routing Vault amino acids through Tricarboxylic Acid (TCA) Cycle (Part 1)

Amino Acid	End Products of Amino Acid degradation	Energy Available through TCA Cycle			One Vault	Energy Available through TCA Cycle/Vault		
		NADH	$FADH_2$	GTP		NADH	$FADH_2$	GTP
Alanine	Pyruvate	4	1	1	8106	32424	8106	8106
Cysteine	Pyruvate	4	1	1	1002	4008	1002	1002
Aspartate	Oxaloacetic Acid	3	1	1	5241	15723	5241	5241
Glutamate	Alpha Ketoglutarate	2	1	1	8760	17520	8760	8760
Phenylalanine	Fumaric Acid + Acetoacetic Acid	1	0	0	3189	3189	0	0
Glycine	Pyruvate	4	1	1	6117	24468	6117	6117
Histidine	Glutamic Acid (Alpha Ketoglutarate)	2	1	1	2208	4416	2208	2208
Isoleucine	Succinyl CoA + Acetyl CoA	1	1	1	4053	4053	4053	4053
Lysine	Acetoacetyl CoA	NA	NA	NA	4992	NA	NA	NA
Leucine	Acetoacetic Acid + Acetyl CoA	3	1	1	10542	31626	10542	10542
Methionine	Succinyl CoA	1	1	1	1254	1254	1254	1254
Asparagine	Oxaloacetic Acid	3	1	1	2487	7461	2487	2487
Proline	Glutamic Acid (Alpha Ketoglutarate)	2	1	1	5928	11856	5928	5928
Glutamine	2 Glutamate (Alpha Ketoglutarate)	4	2	2	5808	23232	11616	11616
Arginine	Glutamic Acid (Alpha Ketoglutarate)	2	1	1	6315	12630	6315	6315
Serine	Pyruvate	4	1	1	5571	22284	5571	5571
Threonine	Acetyl CoA + Glycine (Pyruvate)	3	1	1	4983	14949	4983	4983
Valine	Succinyl CoA	1	1	1	8841	8841	8841	8841
Tryptophan	2 Acetyl CoA + Acetoacetyl CoA+ Alanine (Pyruvate)	6	2	2	918	5508	1836	1836
Tyrosine	Fumaric Acid + Acetoacetic Acid	1	0	0	1908	1908	0	0
Total		51	19	19		247350	94860	94860

*NA – Not applicable

Table A.2 Energy available from routing Vault amino acids through Tricarboxylic Acid (TCA) Cycle (Part 2)

Total Energy Available through TCA Cycle		
	NADH	**247350**
	$FADH_2$	**94860**
	GTP	**94860**
Assuming,		
1 NADH = 3 ATP		
1 FADH2 = 2 ATP		
Total ATP Available through TCA Cycle = 1,026,630		

Table A.3 Vault amino acid precursors for *de novo* Nucleotide Biosynthesis (Part 1)

Amino Acid Substrates in Nucleotide Biosynthesis		
Route to Glutamine	E	8760
	R	6315
	P	5928
	H	2208
	Q	5808
	D (+N)	**7728**
Total		**36747**
Route to Aspartate	N	2487
	D	5241
	Q (+E+R+P+H)	**29019**
Total		**36747**
Route to Serine	C	1002
	S	5571
	G	6117
Total		**12690**
Route to Glycine	C	1002
	S	5571
	G	6117
Total		**12690**

Note:
Glutamine<---->Glutamate<---->Aspartate Interconvertable
Serine<---->Glycine Interconvertable

Table A.3 Vault amino acid precursors for *de novo* Nucleotide Biosynthesis (Part 2)

Amino Acids to Form PRPP		(-5)	(-1)	(+2)	1	1
		ATP	NADH	NADPH	Ribose5P	PRPP
Leucine	L	nil	nil	nil	nil	nil
Lysine	K	nil	nil	nil	nil	nil
Tryptophan	W	nil	nil	nil	nil	nil
Alanine	A	-20265	-4053	8106	4053	4053
Threonine	T	-12455	-2491	4982	2491	2491
Phenylalanine	F	-7970	-1594	3188	1594	1594
Tyrosine	Y	-4770	-954	1908	954	954
methioinine	M	-3135	-627	1254	627	627
Valine	V	-22100	-4420	8840	4420	4420
Isoleucine	I	-10130	-2026	4052	2026	2026
	Total	-80825	-16165	32330	16165	16165

Negative values indicate consumption. Positive values indicate release

Table A.3 Vault amino acid precursors for *de novo* Nucleotide Biosynthesis (Part 3)

STOICHIOMETRY OF AMINO ACIDS IN NUCLEOTIDE BIOSYNTHESIS (De novo)												
Glutamine	2		Glutamine	3		Glutamine	1		Glutamine	2		Glutamine	1
Aspartate	2		Aspartate	1		Aspartate	1		Aspartate	1		Aspartate	1
Glycine	1		Glycine	1		ATP	2		ATP	5		Serine	1
ATP	4		ATP	5								ATP	2
GTP	1											NADPH	1
NADH	1		NADH	1		NADH	1		NADH	1		NADH	1
NUCLEOTIDES AVAILABLE FROM ONE VAULT PARTICLE (one pathway/vault)												
Glutamine	18372		Glutamine	27558		Glutamine	18373		Glutamine	24498		Glutamine	12690
Aspartate	18372		Aspartate	9186		Aspartate	18373		Aspartate	12249		Aspartate	12690
Glycine	9186		Glycine	9186		ATP	36746		ATP	61245		Serine	12690
ATP	36744		ATP	45930								ATP	25380
GTP	9186											NADPH	12690
NADH	9186		NADH	9186		NADH	18373		NADH	12249		NADH	12690
AMP	**9186**		**GMP**	**9186**		**UMP**	**18373**		**CTP**	**12249**		**dTMP**	**12690**

Key	
	Consumed
	Released

Table A.4 Recycling Vault amino acids to assemble average sized proteins (Part 1)

Amino Acid	Symbol	1 MVP	1 VPARP	1 TEP1	1 Vault Particle		Average Folded Domain		Vault/Avg Protein
					78 MVP + 12 VPARP +3 TEP1	Composition	Average Protein	~360 aa Rounded Off	
Alanine	A	80	101	218	8106	8.53	30.79	31	261.48
Cysteine	C	5	37	56	1002	1.25	4.51	5	200.40
Aspartate	D	49	89	117	5241	5.85	21.12	21	249.57
Glutamate	E	86	134	148	8760	6.64	23.97	24	365.00
Phenylalanine	F	26	78	75	3189	3.98	14.37	14	227.79
Glycine	G	58	89	175	6117	7.52	27.15	27	226.56
Histidine	H	18	46	84	2208	2.29	8.27	8	276.00
Isoleucine	I	35	90	81	4053	5.72	20.65	21	193.00
Lysine	K	43	109	110	4992	5.81	20.97	21	237.71
Leucine	L	92	187	374	10542	9.12	32.92	33	319.45
Methionine	M	9	36	40	1254	2.17	7.83	8	156.75
Asparagine	N	21	53	71	2487	4.22	15.23	15	165.80
Proline	P	54	103	160	5928	4.51	16.28	16	370.50
Glutamine	Q	55	91	142	5808	3.71	13.39	13	446.77
Arginine	R	65	62	167	6315	5.02	18.12	18	350.83
Serine	S	38	160	229	5571	5.69	20.54	21	265.29
Threonine	T	44	98	125	4983	5.58	20.14	20	249.15
Valine	V	91	106	157	8841	7.19	25.96	26	340.04
Tryptophan	W	7	18	52	918	1.42	5.13	5	183.60
Tyrosine	Y	17	37	46	1908	3.57	12.89	13	**146.77**
Total					98223			360	

Limiting amino acid is tyrosine as indicated in bold in last column. Hence, number of average proteins available from one protein is 146. Amino acids used for protein synthesis will be 146 x 360 = 52560. Remaining amino acids catabolized for energy is 98223-58560 = 45663. The calculations for the energy derived out of catabolism are shown below.

End Products of Amino Acid degradation	Energy/Reducing Equivalents involved in Catabolism				Remaining Vault Amino Acids	Energy/Reducing Equivalents involved in Catabolism/Vault			
	ATP	NADH	NADPH	$FADH_2$		ATP	NADH	NADPH	$FADH_2$
Pyruvate	0	1	0	0	3580	0	3580	0	0
Pyruvate	0	-1	0	0	272	0	-272	0	0
Oxaloacetic Acid	0	0	0	0	2175	0	0	0	0
Alpha Ketoglutarate	0	0	1	0	5256	0	0	5256	0
Fumaric Acid + Acetoacetic Acid	0	0	-1	0	1145	0	0	-1145	0
Pyruvate	0	2	0	0	2175	0	4350	0	0
Glutamic Acid (Alpha Ketoglutarate)	0	0	0	0	1040	0	0	0	0
Succinyl CoA + Acetyl CoA	1	2	0	1	987	987	1974	0	987
Acetoacetyl CoA	0	2	0	2	1926	0	3852	0	3852
Acetoacetic Acid + Acetyl CoA	1	1	0	1	5724	5724	5724	0	5724
Succinyl CoA	-2	1	0	0	86	-172	86	0	0
Oxaloacetic Acid	0	0	0	0	297	0	0	0	0
Glutamic Acid (Alpha Ketoglutarate)	0	0	1	0	3592	0	0	3592	0
2 Glutamate (Alpha Ketoglutarate)	0	0	-1	0	3910	0	0	-3910	0
Glutamic Acid (Alpha Ketoglutarate)	0	1	0	0	3687	0	3687	0	0
Pyruvate	0	1	0	0	2505	0	2505	0	0
Acetyl CoA + Glycine (Pyruvate)	0	1	0	0	2063	0	2063	0	0
Succinyl CoA	1	3	0	1	5045	5045	15135	0	5045
2 Acetyl CoA + Acetoacetyl CoA+ Alanine (Pyruvate)	0	1	-2	0	188	0	188	-376	0
Fumaric Acid + Acetoacetic Acid	0	0	0	0	10	0	0	0	0
					45663	11584	42872	3417	15608

Table A.4 Recycling Vault amino acids to assemble average sized proteins (Part 2)

Energy required to make 146 Average Folded Proteins from 52560 amino acids			Total Energy available from Catabolism of Remaining Vault Amino Acids	
Activation of Amino Acid	1 ATP/Amino Acid	52560	ATP	11584
Initiation of Polypeptide Chain	1 GTP/Protein	146	NADH	42872
Elongation	1 GTP/Amino Acid	52560	NADPH	3417
Translocation	1 GTP/Amino Acid	52560	$FADH_2$	15608
Termination	1 GTP/Protein	146		
			Assuming , 1 NADH = 3 ATP; 1 NADPH = 3 ATP; 1 $FADH_2$ = 2 ATP	
Total Energy Required to make 146 Average Folded Proteins		157972	Total ATP available from catabolism	181667
NET Energy Available = (181667 - 157972) = 23695				

Section B

Characterization of Vaults in *Trypanosoma brucei*

B.1 Overexpression causes a majority of *TbMVP1* to assemble near the FAZ region

The full length *TbMVP1* coding region was fused upstream of YFP to construct the *TbMVP1-YFP* overexpression vector. The vector was transiently transfected into Ytat cells and observed for immunofluorescence. *TbMVP1* was observed as very bright puncta, preferentially distributed along the FAZ region. While the puncta could also be observed in the cytoplasm, the signal from the FAZ associated *TbMVP1* was very bright that the cytoplasmic puncta could not be observed clearly, except with very high exposure time. The targeting of *TbMVP1* to the FAZ region was confirmed with two different vectors – i) pXS2 based fusion vector with C terminal YFP tag for constitutive overexpression and ii) modified plew100 fusion vector with C terminal YFP fusion tag for tetracycline inducible overexpression. The vectors were individually transfected into procyclic 29.13 cells and two different stable lines were established. Immunofluorescence confirmed that *TbMVP1* localizes close to the FAZ region on overexpression.

To confirm the accumulation of *TbMVP1* along the FAZ region, dual labeling with antibodies that specifically mark the FAZ region was performed. The monoclonal antibody L3B2 specifically targets the FAZ structure in the cell body side by labeling the FAZ filament protein FAZ1 (Kohl et al. 1999). Indirect

immunofluorescence (IF) and immunogold electron microscopy using L3B2 confirmed the presence of *TbMVP1* in close association with FAZ. IF was also conducted with another marker anti-CC2D that marks the FAZ filament and FAZ juxtaposed ER (Zhou et al. 2011). It was observed that *TbMVP1* partially co-localizes with the FAZ region in all stages of cell cycle as can be seen in Figure I and Figure II.

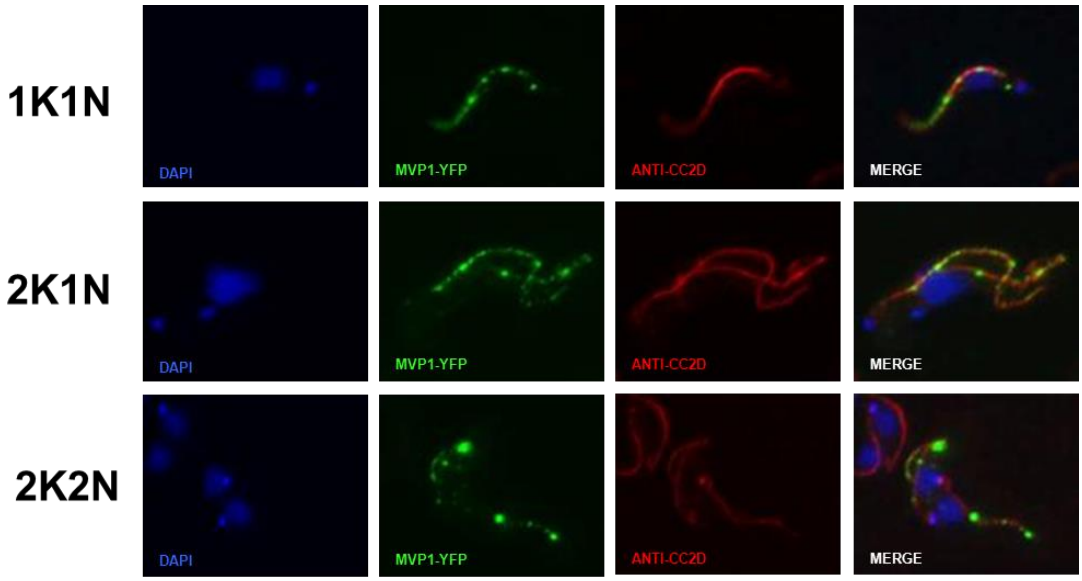


Figure I *TbMVP1* is juxtaposed along the FAZ at all stages of cell cycle
A coiled-coil and C2 domain containing protein (CC2D) that specifically localizes to the FAZ is used as a marker. The DAPI staining clearly marks small structures corresponding to kinetoplasts and larger structures corresponding to nucleus. The kinetoplast divides earlier than the nucleus and hence serves as a reliable marker for cell division.1K1N, 2K1N and 2K2N represents the various stages of cell cycle.

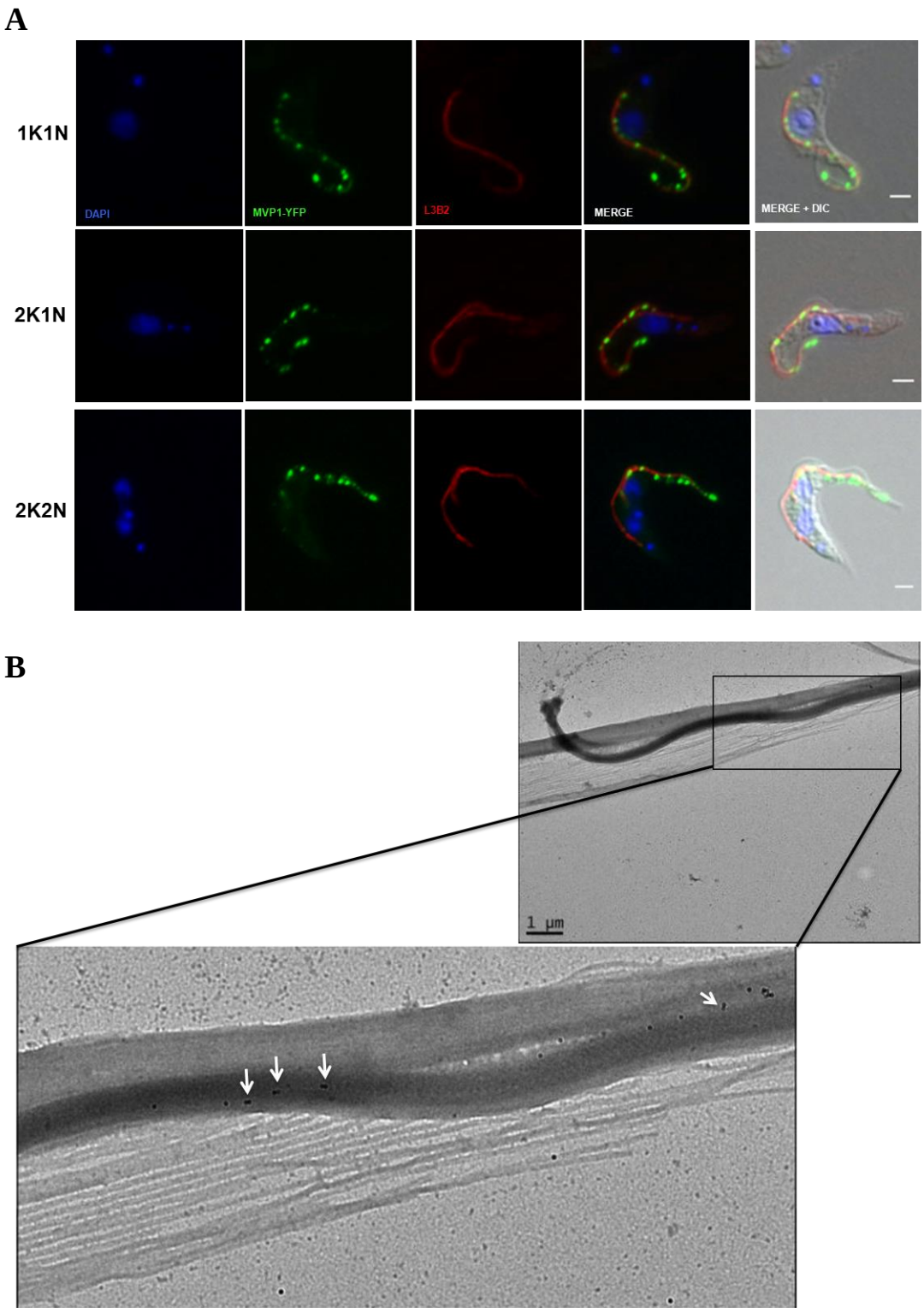


Figure II Partial overlap of *TbMVP1* with another FAZ marker, L3B2
(A)*TbMVP1-YFP* is seen as punctate patterning near the FAZ region. The Merge+DIC image clearly shows the exclusion of *TbMVP1* from the flagella extending from the cell body at the anterior tip. The cytoplasmic pool of *TbMVP1* is visible as faint puncta throughout the cell body. Scale bar, 2 µm (B) Immunogold electron microscopy reveals accumulation of *TbMVP1* along the FAZ region. The white arrows correspond to *TbMVP1*-YFP represented by 25 nm gold and the remaining represent L3B2 labeling FAZ (40 nm)

B.2 *TbMVP1* is excluded from flagella and nucleus

While *TbMVP1* was observed all along the length of the parasite from the anterior to the posterior cell body, whether *TbMVP1* could also associate in parts with the single flagellum of *Trypanosoma* was examined. Indirect immunofluorescence was performed on methanol fixed *TbMVP1-YFP* overexpression stable line using anti-PAR, an antibody that labels the paraflagellar rod and hence, the flagellum. It was observed that *TbMVP1* was specifically excluded from the flagella but was found only within the cell body (Figure III). *TbMVP1* signal was completely missing in the anterior tip detached from the cell body and all through the length of the flagella. This confirms that *TbMVP1* is not a flagellar protein.

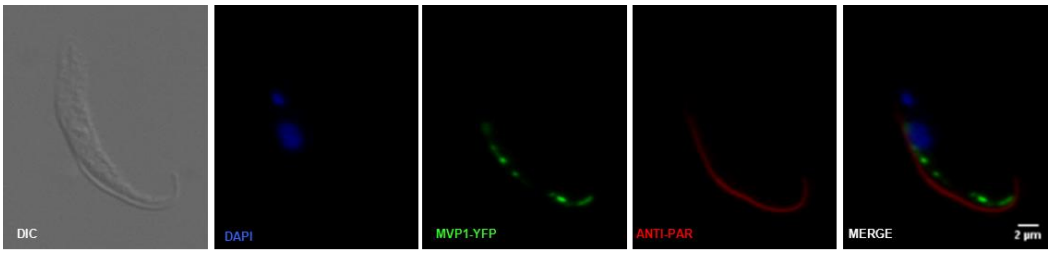


Figure III *TbMVP1* is excluded from the flagella
TbMVP1 overexpressing cells were fixed methanol and labeled for flagella using the anti-PAR antibody that specifically marks the paraflagellar rods. *TbMVP1* clearly runs along the flagella but never overlaps through the length of the cell. The Merge image clearly shows extension of labeling for flagella surpassing the expression of *TbMVP1* towards the anterior end.

B.3 A subset of *TbMVP1* is cytoplasmic

Vaults have been known previously to interact with cytoskeletal elements and sea urchin vaults have been known to co-purify with microtubules. The *Trypanosoma* are characterized by a complex array of cytoplasmic subpellicular microtubule corset under the plasma membranes that helps in retaining the cell shape and remains intact through the complex life cycle. The *Trypanosoma* cytoskeleton being highly cross-

linked can be isolated in an intact fashion by a detergent-based extraction procedure. To examine, if the *MVP* ortholog in *Trypanosoma* can also associate with cytoskeleton, the stable line overexpressing *TbMVP1* was detergent extracted with 1% Triton in PEM to remove the soluble cytoplasmic pool and retain only the detergent resistant cytoskeletal matter comprising subpellicular microtublues, flagellar axenome, microtubule quartet and other microtubules associated structures. The extracted cytoskeleton was fixed and analyzed for immunofluorescence (Figure IV B). It was observed that *TbMVP1* that localized near the FAZ region was still retained in the extracted cytoskeleton; however, the faint puncta observed throughout cytoplasm in intact cells was lost entirely. This confirms that a subset of *TbMVP1* in *Trypanosoma* is cytoskeleton associated while others remain in the soluble pool. Vaults have been observed to move along the microtubules, possibly with the help of molecular motors (van Zon et al. 2006). The strong association of *TbMVP1* with the cytoskeleton might suggest movement of vault complexes along the microtubules in the FAZ region. In *Trypanosoma*, the core cytoskeleton comprising the microtubule quartet and the flagellar axenome remains intact after treatment with 1M NaCl + 1% Triton treatment. To ascertain if *TbMVP1* could represent a part of the core cytoskeleton, the cells were extracted with salt and detergent for 1 minute followed by fixation and viewing under fluorescence microscope. No signal was detected after the salt treatment suggesting that *TbMVP1* does not associate with either the microtubule quartet or the flagellar axenome.

Western blot analysis was also done to confirm the presence of *TbMVP1* in the cytoskeletal fraction (Figure IV A). As mentioned earlier, the fluorescent signal arising from the puncta near the FAZ region was too bright that it occluded potential signal arising from the cytoplasmic pool. However, western blot analysis confirmed

that *TbMVP1* also belongs to the soluble cytoplasmic pool and is not restricted to the FAZ region.

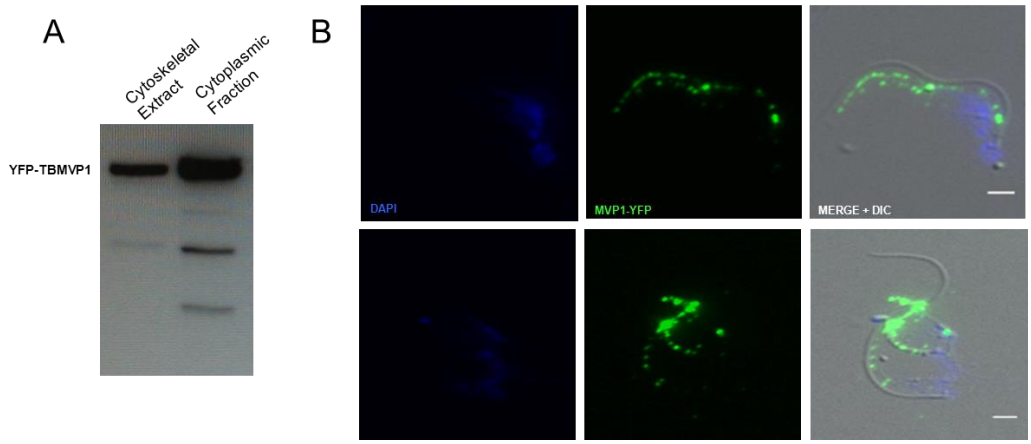


Figure IV: A subset of *TbMVP1* associates with the cytoskeleton
(A) *YFP-TbMVP1* overexpressing cells were fractionated into a soluble cytoplasmic pool and a detergent resistant (1% Triton-X-100 (v/v) treated) cytoskeletal extract. Western blot analysis using anti-GFP antibody reveals that *TbMVP1* is present in both the fractions. Bands of lower sizes correspond to possible degradation products. Vaults are known to associate with cytoskeletal elements, particularly microtubules, in higher eukaryotes. (B)*YFP-TbMVP1* overexpressing cells were detergent extracted for 5 minutes using 1%NP-40 in PBS, fixed with PFA and stained with DAPI. *TbMVP1* can be seen as distinct punta along the FAZ region. The signal from the soluble cytoplasmic pool is lost. Scale bar, 2µm

B.4 Dynamics of *TbMVP1* accumulation along the FAZ region

The vaults have been known to exhibit dynamic behavior in the cytoplasm of mammalian cells. A previous study established that the fluorescence of *MVP-GFP* was recovered within 13 seconds of photobleaching (van Zon et al. 2006). However, when the cells are treated with nocodazole, a drug that inhibits microtubule assembly, a delayed recovery was observed, suggesting that the association of vaults with cytoskeletal elements is responsible for its movement within the cells. Experiments conducted on nerve cells suggested that vaults can be actively transported back and

forth within axons (Li et al. 1999). To trace the dynamic buildup of *TBMVP1-YFP* near the FAZ region an overexpression stable line under the control of tetracycline was used to drive *TbMVP1* expression. After addition of the drug, the cells were tracked for one cell cycle to observe the expression and specific localization of vaults (Figure V). The fluorescence signal was faintly noticeable by 2 hours after addition of tetracycline, however no significant signal was observed. At around 3.5 hours, *TbMVP1-YFP* started to appear as bright punctate structures distributed all through the cell body but excluded from the flagella. Additional faint diffuse patterning was also observed. At each time point, detergent resistant cytoskeletal extracts were also examined. Interestingly, the cytoskeletal extracts retained only the bright punctate structures and the rest of the signal observed earlier was lost. At around 5 hours, a distinct pattern was emerging with more *TbMVP1-YFP* assembling near the FAZ region, but excluded from the flagella. By 6.5 hours, most the cells retained only the punctate distribution along the FAZ and no diffuse patterning within cell body was observed. While the punctate structures were detergent resistant, extraction with salt and detergent, completely removed fluorescent signal arising due to *TbMVP1*, re-emphasizing the association of *TbMVP1* with subpellicular microtubules but not with the core cytoskeleton comprising the microtubule quartet. The dynamic behavior exhibited by *Trypanosoma* vaults is comparable to those observed in other multicellular metazoans where vaults move unconstrained within the cytoplasm along microtubules. The specific accumulation near the FAZ region mid-way through the cell cycle reiterates a plausible unknown role for vaults in these single celled kinetoplastids.

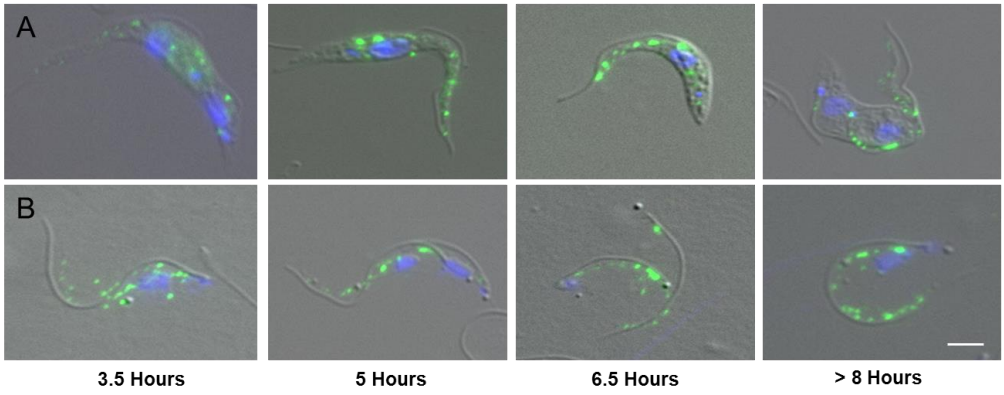


Figure V Dynamic accumulation of *TbMVP1* near the FAZ region
Using a tetracycline-inducible stable line, the expression of *YFP-TbMVP1* was induced by the addition of tetracycline at time = 0 hours. The expression of *TbMVP1* was followed over various time points as indicated. The cells were either methanol fixed (A) or detergent extracted with 1% NP-40 and fixed with PFA (B) before staining with DAPI. At less than 3.5 hours, a diffuse cytoplasmic distribution was observed (data not shown). Punctate structures start to appear at around 3.5 hours and move near the FAZ region over time. Most of the cells show the FAZ localization around 6.5 hours. The punctate structures that appear are detergent resistant. Scale bar, 2 µm.

B.5 Coiled-coil domain in *TbMVP1* is responsible for punctate distribution, but not localization

The crystal structure of *MVP* from rat is composed of defined structural domains that contribute to the unique shape of the vault particle (Tanaka et al. 2009). The interaction between individual *MVP* molecules towards vault assembly is mediated by the C-terminal coiled-coiled domain. The coiled-coil domain marks the region between 648-800 amino acid positions in the human *MVP* protein.

The coiled-coil domain maps to the region termed the cap helix domain. The *TbMVP1* protein was then analyzed using the COILS program (http://embnet.vital-it.ch/software/COILS_form.html) to determine the propensity of *TbMVP1* to form a coiled-coiled domain (Figure VI). The region between 680-800 amino acids favored the formation of coiled-coil domain. To design the truncation constructs, a pairwise

alignment between the rat *MVP* sequence and the *TbMVP1* was performed to establish the conservation among the structural domains (Figure VII). The sequence spanning the nine structural folded domains that builds the barrel constituted the N-terminal truncated *TbMVP1* and the rest of the sequence that included the coiled-coil domain comprised the C-terminal truncated *TbMVP*.

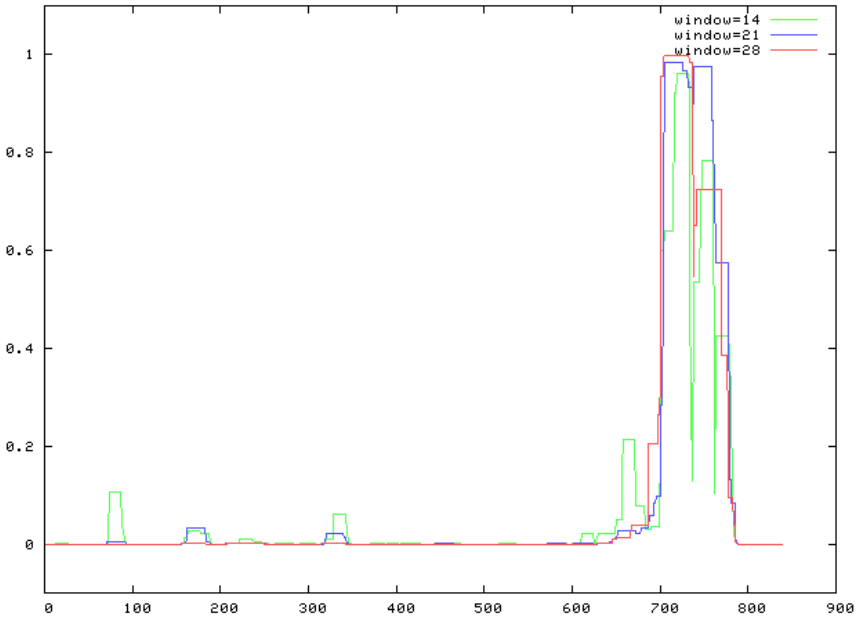


Figure VI *TbMVP1* has a C-terminal coiled-coil region
Vaults are assembled into stable structures by interactions between individual *MVP* chains at the C-terminal coiled-coil region. The propensity of *TbMVP1* to form a coiled-coil domain was predicted using the COILS program. The three-colored lines indicate the various window length in terms of amino acids the program uses to predict the region. The plot shows that a region spanning 680-800 amino acids in *TbMVP1* can potentially form a coiled-coil domain.

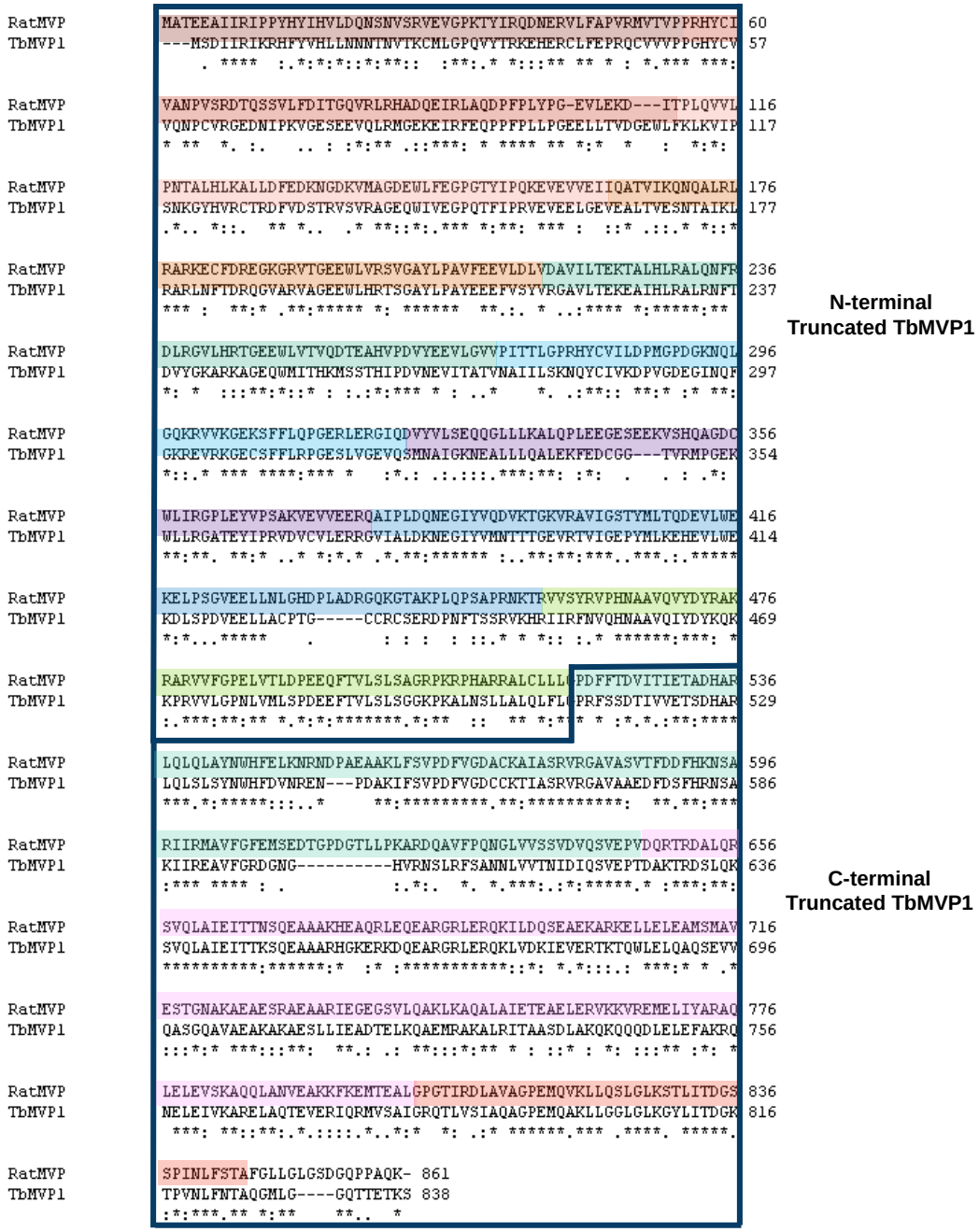


Figure VII Construction of N-terminal or C-terminal truncated proteins
TbMVP1 was aligned with rat *MVP* protein, whose structure has been determined at 3.5 Å, to highlight the various structural regions including the nine structural folded domain, shoulder domain, cap-helix domain and cap-ring domain indicated with different colors in the same order (Tanaka et al. 2009). Based on the alignment, the N-terminal truncated *TbMVP1* was designed to include only the structural folded domains forming the vault barrel and the C-terminal truncated protein included the region from the shoulder till the vault cap.

The sequence encoding regions 1-512 amino acids and 513-838 amino acids were cloned independently into pXS2-YFP vectors resulting in the construction of N- and C- terminal truncated constructs. It was pointed out earlier that punctate distribution of *TbMVP1-YFP* possibly is a result of interaction between individual monomers, resulting in the formation of vault particles. The truncated constructs were designed to check if the coiled-coil domain correlates with the punctate distribution of *TbMVP1-YFP*. In that regard, transient transfection was performed into 29.13 procyclic cells and immunofluorescence was performed. It was observed that, while expression of the C-terminal truncated protein (513-838 aa) exhibited a punctate distribution pattern, the N-terminal truncated protein (1-512 aa) formed a diffuse patterning similar to that observed for *TbMVP3-YFP* (Figure VIII). In spite of the punctate distribution pattern, the specific targeting near the FAZ region was abolished by expressing the C-terminal truncated *TbMVP1 (*513-838 aa). This reiterates that the coiled-coil domain is essential for interaction between *MVP* monomers and the phenomenon seems to be conserved through evolution. The punctate patterning is suggestive of individual monomers interacting and plausible vault particle assembly. Also while C-terminal truncated protein was detergent resistant and adhered strongly to cytoskeletal elements, the N-terminal truncated protein was detergent soluble and hence, remained in the soluble supernatant pool. This also correlates with previous observations that vaults preferentially interact with microtubules via their caps, with their long axis perpendicular to the interacting microtubule axis (Eichenmüller et al. 2003). Hence, the C-terminal truncated proteins are capable of interacting with each other and also assemble along microtubules, however some component necessary for FAZ specific localization appears to reside in the N-terminal portion of the protein.

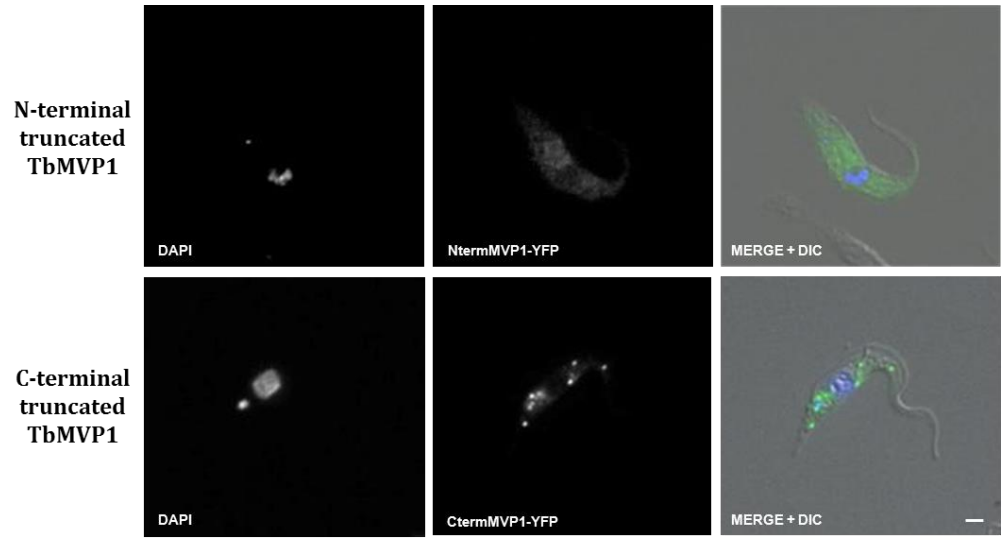


Figure VIII Altered distribution patterns of truncated *TbMVP1*
The preferential FAZ targeting of *TbMVP1* was hampered when the full-length *TbMVP1* protein was truncated C-terminally or N-terminally. The N-terminal *TbMVP1*fragment that retained only the nine structural folded domains showed no puncta but a faint cytoplasmic distribution. The N-terminal truncated protein was detergent-susceptible (data not shown) and does not associate with cytoskeletal elements. The C-terminal protein fragment forms punctate structures throughout the cytoplasm, but shows no accumulation near the FAZ region. The C-terminal protein fragment that retains the coiled-coil region is however detergent resistant (data not shown) and thus capable of interacting with microtubules. Scale bar, 5µm

B.6 Differential subcellular localization of *MVP* paralogs

The paralogs arising out of gene duplication events more often undergo significant evolutionary changes and may have paralogous function. It was hypothesized that if the other two vault paralogs, *TbMVP2* and *TbMVP3*, exhibit similar subcellular localization and punctate distribution akin to *TbMVP1*, the possibility of intact vault structures assembling with all three paralogs are high, like those of *Dictyostelium* composed of two *MVP* paralogs. In that regard, two overexpression vectors with C-terminal YFP tag were created individually from full length *TbMVP2* and *TbMVP3* coding sequences. The vectors were transiently

transfected into 29.13 procyclic cells and subjected to immunofluorescence microscopy.

It was observed that *TbMVP3* was dispersed throughout the entire cell structure with diffuse patterning (Figure IX A). Though an overexpression vector was used, the fluorescence signal was quite faint. There were no punctate structures like those observed for *TbMVP1*. The bright punctate distribution observed in cells expressing *TbMVP1* is suggestive of individual monomers interacting. However, lack of similar patterning in *TbMVP3-YFP* expressing cells raises speculations if such interactions could occur between *TbMVP3* monomers. *TbMVP3-YFP* also seemed to co-localize with the nuclear stain DAPI, that stains the nuclear and kinetoplastid material. The transiently transfected cells were detergent extracted to analyze if *TbMVP3* can associate with cytoskeletal material. Unlike *TbMVP1* which strongly associated with cytoskeleton, *TbMVP3-YFP* showed no fluorescent signal in cytoskeletal preparations, suggesting that it remains in the soluble cytoplasm and does not exhibit associations with cytoskeletal elements.

The distribution pattern of *TbMVP2-YFP* was different from those observed for *TbMVP1* and *TbMVP3*. Transiently transfected procyclic cells expressed *TbMVP2* as one bright puncate spot, more often near the nucleus than the kinetoplastid in 1K1N cell type (Figure IX B). However, in cells in which the kinetoplastid has divided (2K1N) or post mitotic cells (2K2N) more than one bright puncate spot was observed. The patterning was consistent in almost all the cells that expressed *TbMVP2-YFP*. The bright spot was located well within the cell body and it was speculated that they could be closely associated with basal bodies, flagellar pocket or Golgi structures. Also, detergent extracted cytoskeletons retained the punctate

structures intact suggesting that the long branching *TbMVP2*paralogs can still associate with cytoskeletal structures, unlike *TbMVP3*.

Trypanosoma is characterized by a single Golgi structure (He 2007). The protist initially begins with a single Golgi structure in the G1 phase and as the cell proceeds though cell cycle new Golgi start to appear in about 2 hours. However, additional Golgi structures appear towards the end of the cycle (He et al. 2004). As the mother cell undergoes cytokinesis, only the old and new Golgi structures are retained and the additional smaller Golgi disappears. The cells transiently expressing *TbMVP2-YFP* were methanol fixed and indirect immunofluorescence was performed with anti-Tb-GRASP, an antibody that specifically labels the Golgi matrix protein. *TbMVP3-YFP* co-localized entirely with the additional Golgi structures that appeared small and faint compared to either old or new Golgi in most cases. Also, unlike the new or old Golgi that appear close to the kinetoplast, the additional Golgi structures are often found in proximity to the nucleus. The exact functions of these additional Golgi materials and why they appear at this point in the cell cycle is yet unknown. There has been no report of any marker that specifically labels these additional structures and the co-localization experiment suggests a possible role for *TbMVP2* in Golgi duplication.

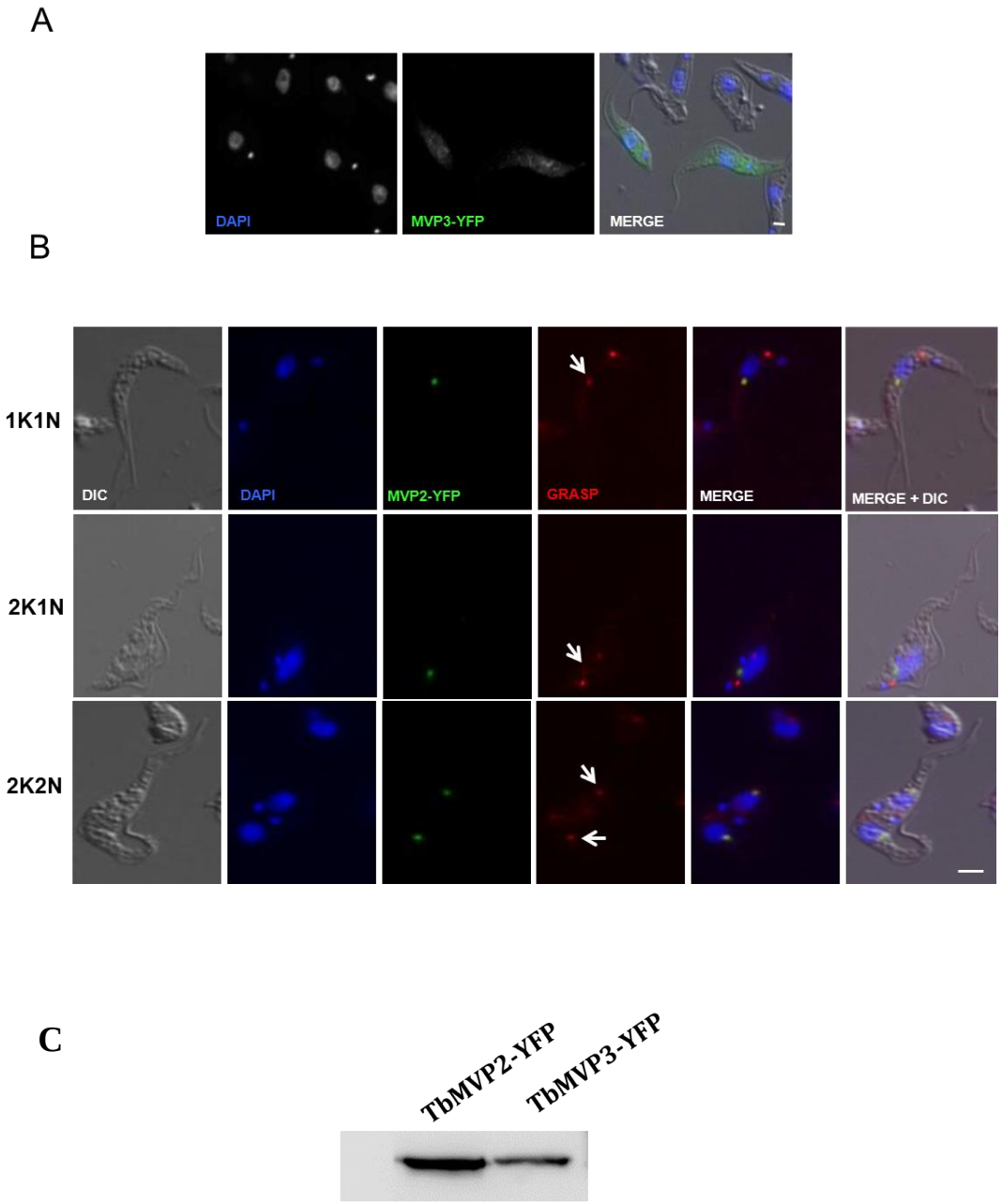


Figure IX Differential localizations of *Trypanosoma MVP* paralogs
(A) Cells overexpressing *TbMVP3-YFP* shows a faint cytoplasmic patterning that spans all along the cell body including the nucleus. No clear punctate structures are formed. (B) Cells overexpressing *TbMVP2-YFP* reveal one or two circular punctate structures that appear close to nucleus. The puncta overlaps with faint structures that correspond to additional Golgi structures (white arrow) that appear during cell division. The pattern is consistent through all stages of cell cycle. Scale bar, 2μm. (C) Western blotting that confirms formation of full-length *TbMVP2* and *TbMVP3* tagged proteins.

B.7 Discussion

It is known that vaults isolated from mammalian cells are entirely composed of one *MVP* ortholog while those from single-celled slime molds are made up of at least two proteins. To deduce if all the three *MVP* paralogs in *Trypanosoma* have similar subcellular localization, hinting at a possible structure composed of all three *MVP* paralogs, immunofluorescence studies were carried out. The subcellular localizations of the paralogs were different, suggestive of a different or evolved additional function for the paralogous genes. While *TbMVP1* had a punctate distribution like those observed for vaults in mammalian cells, *TbMVP3* had a diffuse patterning all over the cell body and *TbMVP2* was expressed as a single bright puncta within the cell body, closer to the posterior end of the cell. About 5% of the total *MVP* expressed in mammalian cells is known to co-localize with the nucleus. Except for *TbMVP3*, which revealed patterns of co-localization with nucleus and kinetoplast, *TbMVP1* and *TbMVP2* remained excluded from the nucleus

While *TbMVP3* forms a diffuse patterning all over the cell, the localization of *TbMVP2* is very specific and co-localizes with the additional Golgi formed during cell cycle. The role of these additional Golgi formed during Golgi duplication and subsequent cell division is unclear, but they are known to disappear shortly before the end of cytokinesis. Their exact route for turnover or degradation is not known. The association of *TbMVP2* along with these additional Golgi inferred by co-localization studies may hint at a possible role for *TbMVP2* as a molecular marker, that helps discriminate the additional Golgi from those that will be inherited by daughter cells. This would allow the cell to maintain tight regulation in ensuring only one Golgi is inherited per daughter cells and any additional material is turned over.

The FAZ region plays a very important role in flagellar assembly, cell flexibility, motility, and is also crucial for cell segregation. In *Trypanosoma*, *TbMVP1* (on overexpression) accumulates near the FAZ region and is expressed all along the cell body but excluded from the flagella. Co-localization studies with FAZ molecular markers, is suggestive of a partial overlap between *TbMVP1* and FAZ filament, with *TbMVP1* preceding the FAZ filament from within the cell body. In *Trypanosoma*, the region around the FAZ, is marked by the MtQ and other microtubule associated structures that regulate cell cycle, and hence it was speculated that knockdown of *TbMVP1* might have adverse effects on cell division. However, the knockdown cells exhibited similar proliferation profiles like those of the control cells and displayed no obvious phenotype. This suggests that *TbMVP1* is a dispensable gene in *Trypanosoma* under normal cellular conditions, as in other eukaryotes.

Various studies have established that vaults preferentially move inside the cell guided by its attachment to cytoskeletal matter. Though vault distribution in mammalian cells exhibits a punctate distribution, the metabolic state of the cell is known to influence their dynamic distribution and cause selective clustering. The fact that they remain enriched within lamellipodia in spreading fibroblasts raised speculations if they associate with actin fibers (Kedersha and Rome 1990). However, co-localization studies confirmed that they do not associate with actin stress fibers; instead they displayed profound co-localization with beta-tubulin molecules (Herrmann et al. 1999). In fact vaults have been observed to arrange in a filamentous pattern along microtubules in a wide range of eukaryotic cells (Hamill and Suprenant 1997). If the movement of the vault across the microtubule is a conserved feature, it should have evolved early in evolution. In *Trypanosoma*, a single-celled eukaryote it was observed that *TbMVP1* is specifically retained in detergent extracted cells,

possibly pointing to its association with the microtubule corset. The subpellicular corset in *Trypanosoma* is arranged as a linear array of microtubules that retain cross-links with each other and with the membranes. The microtubules play a major role in dynamic remodeling of the cell and in maintaining the close tethering between all associated structures. The dynamic study that focused on accumulation of *TbMVP1-YFP* suggested that a subset of vaults is retained after detergent-extraction as observed by bright punctate distribution. Vaults may possibly accumulate near the FAZ region in a microtubule-aided manner. It is worth mentioning that *TbMVP2-YFP* also is cytoskeletal-associated, while *TbMVP3-YFP* has undergone evolutionary changes that somehow abolish its interaction with cytoskeleton, and possibly disrupt vault complex formation.

It is known that the vaults bind to microtubules via their cap region, with their long axis perpendicular to that of microtubules. Truncation studies in *Trypanosoma* suggested that the association of *TbMVP1* with cytoskeletal extract was disrupted when the cap-helix and cap-ring domain were removed in the N-terminal fragment protein. The cap-helix region, which also forms a coiled-coiled domain, is necessary for interaction between *MVP* monomers. The C-terminal fragment protein, which retained the coiled-coiled domain, exhibited association with the cytoskeleton, albeit with no preferential accumulation near the FAZ region. The dynamic behavior of full-length *TbMVP1* suggested that it initially appears as punctate structures spread through the cell body and assembles near the FAZ region over time. It may be reasoned that the C-terminal fragment protein, in spite of forming an association with cytoskeleton, is not specifically bound to the FAZ region as it cannot form complete vault structures. Thus we can conclude that FAZ binding is mediated by some structure in the N-terminal portion of *MVP*1. Fully formed intact complex from

individual monomers may be a prerequisite for movement of vaults across microtubules using molecular motors.

In many eukaryotic cell types studied till date vaults are known to cluster along the growing ends of cell, suggestive of a role for vault in cell growth (Paspalas et al. 2009). As mentioned earlier, a shorter FAZ correlates with shorter daughter cells. The N-terminal primary sequence of *TbMVP1* seems to carry important information concerning this specific targeting, since the truncated proteins failed to exhibit selective distribution. In the case of Tb*MVP*1, the process of accumulation near the FAZ region is almost complete within 6.5 hours. This underscores the importance of cytoskeletal- guided movement of vaults and also points to the involvement of molecular motors to drive the process. The preferential accumulation of *TbMVP1* within this region hints at roles for *MVP* in motility, adhesion and cell-division events.