Hierarchical Evolution of Endosomal Trafficking Proteins Reveals Conserved Catalytic Cores and Lineage-Specific Divergent Regulatory Layers Across Eukaryotes

By

Silmi Suhair

22336010

Marisha Ahmad

22336012

Md Raiyan Kabir

22336008

A thesis submitted to the Department of Biotechnology in partial fulfillment of the requirements for the degree of Bachelor of Science in Biotechnology

Department of Biotechnology
School of Life Sciences
BRAC University
April 2026

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at BRAC University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

_____

Silmi Suhair

22336010

_____

Marisha Ahmad

22336012

_____

Md Raiyan Kabir

22336008

Approval

The thesis/project titled **"**Hierarchical Evolution of Endosomal Trafficking Proteins Reveals Conserved Catalytic Cores and Lineage-Specific Divergent Regulatory Layers Across Eukaryotes**"** submitted by

1. Silmi Suhair (22336010)
2. Marisha Ahmad (22336012)
3. Md Raiyan Kabir (22336008)

of Fall 2025 has been accepted as satisfactory in partial fulfilment of the requirement for the degree of Bachelor of Science in Biotechnology on 7th April, 2026.

Examining Committee:

Supervisor:

Md Tawsif Ur Rashid
Lecturer
Department of Biotechnology
School of Life Sciences
BRAC University

Department Head:
(Chairperson)

Dr. Iftekhar Bin Naser
Associate Professor and Chairperson
Department of Biotechnology
School of Life Sciences
BRAC University

Department Dean:

Dr. Aparna Islam
Professor and Assistant Dean
School of Life Sciences
BRAC University

Abstract

Membrane trafficking proteins are essential for organelle communication. However, whether different functional classes evolve under distinct evolutionary pressures remains unclear. In order to investigate this, eight protein families (ESCRT-0, ESCRT-I, ESCRT-II, ESCRT-III, Rab GTPases, VPS4, SNAREs and tetraspanins) were taken across *Saccharomyces cerevisiae*, *Arabidopsis thaliana* and *Homo sapiens*. Sequence identity, domain architecture, motif conservation, phylogenetic relationship, structural overlays and intrinsic disorder patterns were inspected to find out which features remain conserved and which are not. The investigation showed a clear hierarchy where the catalytic proteins (Rab GTPases, VPS4) showed high sequence identity, near perfect motif conservation along with strong structural similarity. However, structural proteins (ESCRT-I, II, III and SNAREs) showed moderate conservation with preserved folds despite sequence divergence. And finally the regulatory adaptor proteins (ESCRT-0, tetraspanins) showed the greatest divergence, including missing domains, very low sequence identity, and increased disorder. These results from our investigation led us to believe that conservation in the membrane trafficking systems is not uniform. Rather, it follows a hierarchy in which it offers compensation in the structural and adaptor regions while remaining highly conserved for the catalytic roles.

Keywords: Membrane trafficking proteins; conserved; catalytic proteins; structural proteins; regulatory adaptor proteins; hierarchy

Acknowledgement

We would like to extend our sincere gratitude to our supervisor, Md Tawsif Ur Rashid for his constant support and guidance throughout our thesis. His valuable insights and constructive feedback greatly contributed to the successful completion of our work.

We would also like to express deep gratitude towards other faculty members for their valuable lessons throughout our undergraduate life. We are also deeply thankful to our seniors and classmates who have helped us at different times during this academic journey.

Lastly, we would like to thank all our family members and friends who have been a great source of strength and inspiration for us.

Sincerely,

Silmi Suhair (22336010)

Marisha Ahmad (22336012)

Md. Raiyan Kabir (22336008)

Department of Biotechnology

School of Life Sciences

BRAC University

Table of Contents

List of Tables

List of Figures

List of Acronyms

MVB	Multivesicular Bodies
RAB	Ras-associated binding
GTP	Guanosine Triphosphate
GDP	Guanosine Diphosphate
ESCRT	Endosomal Sorting Complex Required for Transport
VPS4	Vascular Protein Sorting-associated protein 4
TOL	TOM1-like
HRS	Hepatocyte growth factor-regulated tyrosine kinase substrate
STAM	Signal Transducing Adaptor Molecule
TSG	Tumor Susceptibility Gene
STP	Suppressor protein
CHMP	Charged multivesicular body protein
VAM	Vacuolar morphogenesis protein
SYP	Syntaxin-related protein
STX	Syntaxin
TET	Tetraspanin
CD	Cluster of Differentiation
SNARE	Soluble N-ethylmaleimide-sensitive factor attachment protein receptors
TEM	Tetraspanin-enriched microdomain
EC	Extracellular loop
DNA	Deoxyribonucleic Acid
SCOP	Structural Classification of Proteins
LUCA	Last Universal Common Ancestor
LECA	Last Eukaryotic COmmon Ancestor
EMBOSS	European Molecular Biology Open Software Suite
MAFFT	Multiple Alignment using Fast Fourier Transform
MEGA	Molecular Evolutionary Genetics ANalysis
TM-score	Template Modeling score
RMSD	Root Mean Square Deviation
VHS	Vps27, Hrs, STAM domain

UIM	Ubiquitin Interacting Motif
UEV	Ubiquitin Enzyme Variant
SH3	Src Homology 3
TM	Transmembrane
SB	Steadiness Box
GLUE	GRAM-like ubiquitin binding in EAP45
MIM	MIT-Interacting Motif
MIT	Microtubule Interacting and Transport

Chapter 1

Introduction:

1.1 Membrane trafficking and cellular organization

Membrane trafficking is a fundamental feature of eukaryotic cells that maintains organelle identity, mediates cargo exchange and supports cellular homeostasis **[1-2]**. Eukaryotic cells contain an interconnected network of membrane bound organelles that exchange proteins, lipids and other cargos through coordinated vesicle mediated transport. The membrane trafficking system supports a variety of cellular processes including the export of newly synthesized proteins, uptake of extracellular material, intracellular targeting of proteins and lipids and maintenance of compartmentalization within the cell. The broad conservation of endomembrane compartments and trafficking pathways across eukaryotes indicates an ancient evolutionary origin of this system and the persistence of shared core mechanisms **[1].**

The endomembrane system consists of several interconnected compartments that exchange materials through vesicle mediated transport in several trafficking pathways **[1, 3-4]**. Major components include the endoplasmic reticulum, the Golgi apparatus, early and recycling endosomes, late endosomes or multivesicular bodies (MVBs), lysosomes or vacuoles and the plasma membrane **[5-6]**. Transport between these compartments is mostly mediated by vesicles that bud from a donor membrane and subsequently fuse with a target acceptor organelle **[7-9]**. Protein trafficking through this system begins with the synthesis of membrane and secretory proteins. In eukaryotic cells, many newly synthesized proteins are first translocated into the endoplasmic reticulum during translation. From there, proteins are transported to the Golgi apparatus where they undergo further processing and are sorted into different trafficking pathways that determine the final cellular destination. Proteins destined for secretion or plasma membrane localization enter the secretory pathway and proteins enter the degradative compartments like vacuoles by the vacuolar protein sorting pathways **[10]**.

In addition to biosynthetic trafficking, eukaryotic cells internalize plasma membrane proteins and extracellular materials through endocytosis. Cargo proteins are delivered to endosomes after internalization which functions as the central sorting compartment within the endosomal system. At this stage, proteins may be recycled back to the plasma membrane or transported to other intracellular compartments. Some proteins are returned to the Golgi apparatus and others are

targeted for degradation through the maturation of early endosomes into late endosomes and multivesicular bodies **[3, 10-11]**. The multivesicular body pathway represents an important degradative route within the endosomal system. Multivesicular bodies are specialized late endosomal compartments characterized by the presence of intraluminal vesicles formed through inward budding of the endosomal membrane **[12-14]**. During this process, membrane proteins destined for degradation are incorporated into these vesicles. The multivesicular bodies fuse with lysosomes to form endolysosomes or with vacuoles and release the vesicles along with their cargos into the lumen for degradation **[7, 15-17]**.

Cargo sorting into the multivesicular body pathway commonly depends on ubiquitination which is a conserved post translational modification that acts as a signal for degradation **[3, 10, 13, 18]**. This process ensures that membrane proteins which must be removed or downregulated are efficiently delivered to lysosomes or vacuoles. The inward budding of endosomal membranes during vesicle formation represents an important form of membrane remodelling that contributes to intracellular trafficking and the regulation of protein turnover **[12].**

1.2 Overview of trafficking protein families

Every eukaryotic cell faces a fundamental challenge on how to selectively transport proteins and lipids between membrane-bound organelles while maintaining the distinct identity of each compartment. This challenge is met by a complex membrane trafficking system that relies on the precise cooperation and coordination of multiple protein families. At the molecular level, Rab GTPases establish compartment identity and trafficking specificity by cycling between active GTP-bound and their inactive GDP-bound states. This allows them to recruit specific effector proteins to the right membranes at the right time **[19-21]**. Among these effectors are the endosomal sorting complexes required for transport (ESCRT), a multi-subunit machinery composed of ESCRT-0, -I, -II and -III that drives inward membrane remodelling and packages cargo into intraluminal vesicles within multivesicular bodies (MVBs) **[22-25]**. This ESCRT driven budding generates intraluminal vesicles **[26-27]** and has been seen to have direct implications for exosome biogenesis **[28-30]**. To sustain this process, the ESCRT machinery must be recycled after each round. This is where the VPS4 ATPase (vacuolar protein sorting associated protein 4) performs its function. It assembles into hexameric rings that disassembles ESCRT III filaments and recycle the system for subsequent sorting **[31-34]**.

The ESCRT budding cycle does not happen randomly, rather it takes place within specialised microdomains organised by tetraspanins. These proteins cluster through lateral interactions to form tetraspanin-enriched microdomain (TEMs) platforms for protein sorting and vesicle formation **[35-37]**. Tetraspanins have conserved extracellular loops (EC1 and EC2) containing disulfide bonds and a signature CCG motif which helps them to organise the endosomal membrane by concentrating proteins into areas primed for budding **[38-40]**. CD63, CD81 and CD9 are common standard biomarkers found on exosomes which reflect their enrichment in MVB membranes and packaging into intraluminal vesicles **[35, 41-43]**. Once the vesicles form, they must deliver cargo to the right destination. SNARE proteins handle this final step. SNAREs (Soluble N-ethylmaleimide-sensitive factor attachment protein receptors) are not involved in vesicle formation but are essential for ensuring vesicle fuse with correct target membranes **[44-46]**. Vesicle anchored SNARE (v-SNAREs) on vesicles paired with complementary target membrane anchored (t-SNAREs) on target membranes, zippering into a four-helix bundle that pulls membranes close enough to fuse **[47-49]**. Different SNARE combinations create different specificity, which can ensure a vesicle headed for the lysosome does not accidentally fuse with the golgi **[44, 50].** Thus, these interconnected protein trafficking families work as an integrated network to guide cargo from initial sorting to their respective destinations.

1.3 Evolutionary conservation of trafficking systems

The membrane trafficking system of eukaryotes composed of interconnected cell organelles involves a sequential process of vesicle formation, transport and membrane fusion. These processes are carried out by conserved protein families like the small GTPases, vesicle coat complexes, cargo adaptors, tethering complexes and SNARE proteins **[51]**. From a comparative genomic and phylogenetic analysis, studies show that the membrane trafficking system is highly conserved across eukaryotic lineages. The core components of the trafficking system are present in organisms ranging from fungi and plants to protists and animals, suggesting that the membrane trafficking system emerged early in eukaryotic evolution and was present in the last eukaryotic common ancestor (LECA) **[1, 52]**. Orthologous proteins in distantly related organisms often have similar functions and cellular localization, indicating that the basic framework of vesicular transport is conserved across eukaryotes **[1,53]**.

Despite the conservation of core trafficking machinery, evolutionary divergence has occurred through gene duplication and specialization of protein families. Such paralogs function in distinct trafficking pathways or organelles, adding to the complexity of the endomembrane system while still maintaining a conserved core machinery **[1, 51]**. A key feature of such conservation is the presence of conserved catalytic cores within trafficking protein families. For example, Rab GTPases contain highly conserved GTP-binding domains responsible for regulating vesicle targeting and trafficking through GTP hydrolysis **[19**, **54]**. Additionally, SNARE proteins are core molecular machinery with stable helical bundles which drive membrane fusion. The domains and motifs of the protein families are conserved as the structural and functional unit. For instance, the SNARE protein families contain a conserved coiled-coil SNARE domain involved in the formation of the SNARE complex during membrane fusion **[55].** The preservation of these domains and motifs across eukaryotic species reflects the evolutionary pressure to maintain the structural and functional roles in the membrane trafficking system **[56]**. Moreover, studies on a wide range of eukaryotic genomes demonstrate that most components of ESCRT-I, ESCRT-II, ESCRT-III and ESCRT-III associated factors are conserved throughout the eukaryotic lineage. Among these components, ESCRT-III subunits and the AAA-ATPase Vps4 constitute the conserved functional core of the system **[57]**.

Moreover, divergence also exists in regulatory components involved mainly in cargo recognition and recruitment of ESCRT components to endosomal membranes **[58]**. An example of this is the ESCRT-0 complex which is responsible for the recognition of ubiquitinated cargo and the recruitment of downstream ESCRT complexes. Comparative genomic analyses show that the ESCRT-0 complex is restricted primarily to fungi and animals, suggesting that other eukaryotic lineages use alternative mechanisms for the initial recognition of ubiquitinated cargo and recruitment of the ESCRT machinery **[57-58]**. Besides, because of gene duplication and expansion of protein families like Rab GTPases, they have generated lineage-specific regulatory networks causing an increased complexity in membrane trafficking **[56, 59]**. Despite sequence divergence across species, the conserved cores maintain similar mechanisms and structural features, reflecting strong evolutionary constraints in their functions **[56]**.

1.4 Model organisms used for evolutionary comparison

The budding yeast, *Saccharomyces cerevisiae* is the first eukaryote with full genome sequenced and is widely used as a model organism for studying eukaryotic cells **[60-61]**. Its high experimental tractability and conserved cellular features makes it an ideal fungal model system. As a unicellular fungus, it shares the complex intracellular organization of higher eukaryotes including a nucleus, endoplasmic reticulum, the Golgi apparatus, vacuole etc. **[10]**. Yet, it has a short generation time, rapid growth and can be easily genetically manipulated. Furthermore, various fundamental cellular pathways are also conserved between yeast and higher eukaryotes, making it an effective model organism for studying evolutionary conservation across different lineages **[62-64]**. The highly efficient homologous recombination machinery enables precise genome editing in yeast cells, allowing researchers to create mutants for studying its cellular mechanisms **[10]**. Research in yeasts played a crucial role in studying the molecular basis of vesicular transport. Through the isolation of mutants defective in secretion (sec), vacuolar protein sorting (vps) and endocytosis (end), researchers were able to identify key components involved in intracellular trafficking. Research in yeasts has also improved the understanding of the ESCRT machinery responsible for MVB formation and vacuolar degradation. As key trafficking proteins are highly conserved between yeast and higher eukaryotes, discoveries in yeast provide insights into similar processes in more complex organisms **[10]**.

Arabidopsis thaliana is an established model organism for plant biology due to its practical advantages and genetic study **[65]**. It has a short generation time, less requirements for growth and prolific seed production through self-pollination, enabling large scale mutant screens and genetic analysis for research **[66-67]**. Moreover, the membrane trafficking system in *Arabidopsis thaliana* comprises a network of similar cell organelles like ER, Golgi apparatus, endosome and vacuoles that communicate through vesicle-mediated transport **[68]**. Most of its core components identified in sorting, packaging and fusion of vesicles have clear orthologs in yeast, mammals and other eukaryotes, reflecting a common ancestral trafficking system **[69-70]**. This conservation validates its use as a model organism to explore fundamental trafficking mechanisms across the eukaryotic lineages.

Homo sapiens serves as a representative model organism for studying metazoan genomics due to its high conservation of core genetic and cellular machinery across the animal kingdom. Approximately 55-60% of human protein-coding genes are descended from the last common

ancestor of all animals (Urmetazoan) **[71]**. The membrane trafficking system in humans is highly conserved across eukaryotes, sharing core organelles and molecular machinery **[1]**. It comprises the ER, Golgi apparatus, endosomes and lysosomes that communicate through vesicle-mediated transport **[72]**. Besides, the fundamental components such as Rab GTPases, SNARE proteins, ESCRT and coat complexes involved in vesicle sorting share strong evolutionary conservation with homologous proteins in yeasts and plants. However, in metazoans, these systems have undergone further diversification to support specialized cellular functions **[73]**. Thus, human cells serve as representative systems for studying membrane trafficking in metazoa as they contain full components of endomembrane organelles and regulatory machineries of complex multicellular eukaryotes.

1.5 Evolutionary principles relevant to protein conservation

Evolutionary principles provide an important framework for understanding patterns of protein conservation across species. Proteins evolve under the influence of natural selection where mutations that reduce protein function are typically removed from populations through purifying selection. The rate ratios of synonymous (silent) and nonsynonymous (amino acid changing) mutations can provide insight about DNA sequence evolution **[74]**. At the molecular level, this form of selection can be detected when the rate of synonymous substitutions (dS) exceeds the rate of nonsynonymous substitutions (dN) which indicates that amino acid changing mutations are selectively eliminated to ensure protein function or stability. As a result, most protein coding genes experience strong purifying selection **[75].**

Different residues evolve at different rates within proteins depending on their structural and functional roles. Residues that contribute directly to protein stability or biochemical activity are subject to stronger selective pressure and therefore evolve more slowly than less critical regions. Catalytic sites in enzymes represent particularly conserved positions as they are directly responsible for substrate binding and strong selective pressure acts to maintain their function across evolutionary time **[76-78]**. Studies have shown that residues located close to catalytic sites tend to display a gradient of conservation and evolutionary conservation decreases with increasing distance from the catalytic center in the three dimensional structure. This pattern supports the fact that enzyme catalysis requires coordinated contributions from multiple residues along with the key catalytic amino acids and this imposes evolutionary constraints on a larger portion of the protein structure **[76]**. Another key principle of protein evolution is that structural

constraints strongly influence evolutionary rates at individual sites. Residues located in the protein core are generally more conserved than surface exposed residues with more variation **[79]**. Similar conservation patterns are observed across the evolutionary family of enzymes where enzymes with more similar sequences typically diverge more recently from a common ancestor. In some cases, the structural folds and active sites of enzymes are so strongly conserved that they can be traced back to the Last Universal Common Ancestor (LUCA) **[77]**. Thus, across evolutionary time, structural constraints often preserve functionally important folds even when amino acid sequences diverge substantially. This principle is especially relevant for trafficking proteins where catalytic motifs and structured cores may remain conserved despite divergence in flexible terminal or regulatory regions.

Protein structure is widely recognized to be more conserved than amino acid sequences across evolutionary time **[80]**. Analysis of Structural Classification of Proteins (SCOP) superfamilies show that homologous proteins often retain similar structural folds even when their sequences diverge substantially. Study showed that distant homologs can share comparable three dimensional structures despite low sequence identity and structurally conserved regions remain detectable within highly divergent protein families **[81].** Similarly, structural features are considered more resistant to the effects of mutation and therefore preserve evolutionary information over longer timescales than primary sequence data. The primary amino acid sequences of a protein are highly variable while the secondary, tertiary and quaternary structure of proteins tend to be more conserved **[82].**

1.6 Knowledge gap

A major limitation in membrane trafficking research is that core trafficking machineries have rarely been analyzed as an integrated evolutionary system across multiple eukaryotic lineages. The individual families like the Rab GTPases, ESCRT complexes, VPS4, tetraspanins and SNARE proteins have been extensively studied but in isolation. Researchers have clearly classified and explained Rab functions in yeast secretion **[19]** while also establishing SNARE specificity in plant vacuolar trafficking **[83]**. But most of these investigations were limited to their individual protein families within single model organisms. This leaves a clear gap for a systematic comparative analysis that examines multiple trafficking protein families together across major eukaryotic lineages.

Studying this gap remains significant due to the fact that the endomembrane system itself is considered to be ancient and conserved. But the proteins in these systems have undergone lineage specific expansions and sequence divergence **[84]**. This matter can be better understood when they are compared across a multitude of species. Humans have more than 70 different Rab GTPase family members while, in contrast, the yeast *Saccharomyces cerevisiae* has only 11 Rab GTPases (often referred to as Ypt proteins) **[85-86]**. Again, while humans possess 33 tetraspanins compared to 17 in *Arabidopsis thaliana*, the evolutionary implications of such differences can be significant **[38, 87]**. A comparative approach spanning yeast, plant, and human system could be used to distinguish core universally conserved features from lineage-specific adaptations. Our study addresses this gap by systematically comparing membrane trafficking proteins from *Saccharomyces cerevisiae, Arabidopsis thaliana*, and *Homo sapien*s using sequence, structural, and phylogenetic tools to uncover patterns of conservation and divergence across the eukaryotic tree of life.

1.7 Aim of the study

The aim of this study is to analyze the evolutionary conservation and divergence of membrane trafficking proteins across eukaryotic lineages including yeast (*Saccharomyces cerevisiae*), plant (*Arabidopsis thaliana*) and human (*Homo sapiens*). A fundamental feature of these eukaryotic cells is their membrane trafficking system. The core machinery is widely conserved across diverse taxa, indicating their essential roles in eukaryotic cells and ancient origin **[1]**. Despite such conservation, lineage-specific expansion, secondary losses and functional diversification have been observed, reflecting organism-specific adaptation **[88].** Through a comparative sequence, structural and phylogenetic analysis of the key protein families involved in membrane trafficking including ESCRT-0, ESCRT-I, ESCRT-II, ESCRT-III, Rab GTPases, Disassembly, SNAREs and tetraspanins, the study will identify the conservation pattern of these proteins and the evolutionary relationships of the representative organisms.

The study is based on the hypothesis that different trafficking protein families evolve under distinct evolutionary constraints. Catalytic regulators such as Rab GTPases and VPS4 are expected to exhibit high conservation due to strict functional constraints and conserved molecular mechanisms **[89-90]**. Additionally, structural scaffold proteins including ESCRT-I, ESCRT-II, ESCRT-III and SNARE-associated proteins are expected to show moderate conservation **[57,**

91-92]. Lastly, regulatory adaptor proteins like ESCRT-0 and tetraspanins are likely to display less conservation across the species **[40, 57]**. Ultimately, this analysis will provide insights into how the fundamental mechanisms of the membrane trafficking system have been conserved across major eukaryotic lineages.

Protein families	**Protein type**	**Conservation level**
Rab GTPases, VPS4	Catalytic regulators	High conservation
ESCRT-I, ESCRT-II, ESCRT-III, SNAREs	Structural machinery	Moderate conservation
ESCRT-0, tetraspanins	Regulatory adaptor proteins	High divergence

Table 1: Conservation level of different protein families across eukaryotes

Chapter 2

Materials & Methods:

2.1 Sequence retrieval from UniProt database

Protein sequences for yeast (*Saccharomyces cerevisiae*), human (*Homo sapiens)* and plant (*Arabidopsis thaliana)* were retrieved from the UniProt database (https://www.uniprot.org/) for the following protein families: ESCRT-0, ESCRT-I, ESCRT-II, ESCRT-III, Rab GTPases (Rab5/Rab7), VPS4, SNAREs and Tetraspanins. Representative sequences with similar functions were selected from each protein family across the three species based on curated annotations and literature evidence. Amino acid sequences were downloaded in FASTA format and their UniProt accession numbers were recorded for subsequent analysis.

2.2 Multiple sequence alignment using MAFFT

Multiple Sequence Alignment (MSA) for each protein family was performed using the MAFFT alignment program (https://mafft.cbrc.jp/) with default parameters. The final alignments were downloaded in FASTA format for further analysis.

2.3 Pairwise sequence identity and similarity analysis using EMBOSS Needle

Global pairwise sequence alignments were performed using the EMBOSS Needle tool (https://www.ebi.ac.uk/jdispatcher/psa/emboss_needle). Three pairwise comparisons were formed for each protein family: yeast-human, yeast-plant and human-plant. The analysis was conducted separately for full length unaligned sequences and aligned functional domains and motifs. Percent identity, percent similarity and alignment lengths were recorded in comparative tables.

2.4 Domain architecture analysis using UniProt annotations

Domain organization for each protein was determined using annotations under the Family & Domains section of the UniProt database. The position and boundaries of major functional domains were recorded including catalytic cores, regulatory domains, transmembrane helices and low complexity regions. Schematic domain diagrams were generated using BioRender (https://www.biorender.com/) which showed protein length and domain positions. The diagrams

of yeast, human and plant proteins were aligned side by side for each protein family for visual comparison of domain organization across species.

2.5 Motif validation using multiple sequence alignment

Motif validation was performed using the MAFFT sequence alignment files taking the yeast protein sequence as reference. Known functional motifs for each protein family were identified in the yeast protein based on UniProt annotation and other literature evidence. The corresponding regions in plant and human sequences were examined within the alignment to determine whether these motifs were conserved. Conserved residues and conservative substitutions within the motif regions were recorded. Motif conservation was further visualized by generating sequence logos with alignment regions containing motifs of interest using the Weblogo server (https://weblogo.threeplusone.com/).

2.6 Phylogenetic tree construction with bootstrap support using Maximum Likelihood method

Phylogenetic trees were constructed with the Maximum Likelihood method based on MAFFT multiple sequence alignment using MEGA Alignment Explorer (https://www.megasoftware.net/). In addition to yeast, plant and human proteins, sequences from three additional species *(Mus musculus, Drosophila melanogaster* and *Caenorhabditis elegans)* were included to improve phylogenetic resolution. The analysis was performed using protein sequences with automatic model selection (JTT model) and 1000 bootstrap replicates to estimate branch support. The trees were exported in image and Newick format with bootstrap values.

2.7 Structural conservation analysis using AlphaFold structures and TM-align

Three dimensional structures for all protein sequences were obtained from the AlphaFold Protein Structure Database (https://alphafold.ebi.ac.uk/) using UniProt accession numbers. Pairwise structural comparisons (yeast-human, yeast-plant and human-plant) were performed using TM-align web server (https://aideepmed.com/TM-align/). The Root Mean Square Deviation (RMSD) value and template modelling score (TM-align) were recorded. Structural conservation was interpreted based on established thresholds: RMSD <2 Å: very strong structural conservation,

RMSD 2–3 Å: moderate conservation, TM-score between 0.3 and 0.5: partial structural similarity, TM-score > 0.5: same fold and TM-score > 0.7: highly similar.

2.8 Motif mapping onto three-dimensional structures using PyMol

Conserved residues and motifs identified from multiple sequence alignments were mapped and highlighted onto representative protein structures visualized using PyMOL (https://www.pymol.org/).

2.9 Intrinsic disorder prediction using IUPred2A

Intrinsic disorder predictions were performed using the IUPred2A web server (https://iupred2a.elte.hu/). Amino acid sequences from UniProt were submitted to the server with default parameters. Residues with disorder scores greater than 0.5 were considered intrinsically disordered. Disordered probability plots for each protein were then compared.

2.10 Cross-complex evolutionary comparison of protein families

A comparative table was constructed to understand evolutionary patterns by summarizing the key metrics for each protein family: overall sequence identity range, core domain conservation, structural conservation and identification of major divergent regions. A unified evolutionary model was developed to describe the differential conservation patterns across all protein families.

Chapter 3

Results

3.1 Sequence identity comparison across species

Protein families	Comparison	% Identity	% Similarity
ESCRT-0	Yeast (Vps27)-Human (Hrs)	21.1	34.9
	Yeast (Vps27)-Plant (TOL2)	10.6	20.3
	Human (Hrs)- Plant (TOL2)	9.4	16.5
	Yeast (Vps27)-Human (Hrs)	21.1	34.9
	Yeast (Vps27)-Plant (TOL6)	18.0	32.6
	Human (Hrs)- Plant (TOL6)	19.8	31.4
	Yeast (Hse1)- Human (STAM)	22.8	37.6
	Yeast (Hse1)-Plant (TOL2)	14.7	28.4
	Human (STAM)-Plant (TOL2)	15.0	24.1
	Yeast (Hse1)- Human (STAM)	22.8	37.6
	Yeast (Hse1)-Plant (TOL6)	15.5	26.9
	Human (STAM)-Plant (TOL6)	16.4	25.9
ESCRT-I	Yeast (Stp22)- Human (Tsg101):	25.6	45.5
	Yeast (Stp22)-Plant (ELC):	22.4	40.0
	Human (Tsg101)-Plant (ELC):	29.2	47.0
ESCRT-II	Yeast (VPS36)-Human (VPS36)	16	29.7
	Yeast (VPS36)-Plant (VPS36)	17.6	31.2
	Plant (VPS36)-Human (VPS36)	33.3	49.8
ESCRT-III	Yeast (Snf7)-Plant (Snf7)	31.7	47.9
	Yeast (Snf7)-Plant (Snf7.1)	28.8	42.1
	Plant (Snf7.1)-Human (CHMP1A)	22.6	36.6
	Plant (Snf7.1)-Human (CHMP1A)	22.2	36.2

	Yeast (Snf7)-Human (CHMP1A)	19.9	36.7
RAB5	Yeast (Ypt52)-Plant (RAB5A)	41.9	55.6
	Yeast (VPS21)-Plant (RAB5A)	46	63.3
	Yeast (Ypt51)-Human (RAB5)	50.2	63.7
	Yeast (Ypt52)-Human (RAB5)	42.7	54.2
	Plant (RAB5A)-Human (RAB5)	59.4	71.4
RAB7	Yeast (Ypt7)-Human (RAB7)	61.4	74.0
	Yeast (Ypt7)-Plant (RAB7)	35.5	52.7
	Plant (RAB7)-Human (RAB7)	59.4	71.4
Disassembly	Yeast (VPS4)-Human (VPS4)	59.0	73.6
	Yeast (VPS4)-Plant (VPS4)	54.8	68.1
	Plant (VPS4)-Human (VPS4)	53.0	68.8
SNAREs	Yeast (VAM3)-Human (STX10)	16.6	38.0
	Yeast (VAM3)-Plant (SYP121)	18.6	33.2
	Plant (SYP121)-Human (STX10)	13.2	25.7
Tetraspanins	Plant (TET8)-Human (CD63)	21.8	37.3
	Plant (TET8)-Human (CD81)	22.0	33.7
	Plant (TET8)-Yeast (SUR7)	11.2	16.3
	Yeast (SUR7)-Human (CD63)	1.2	2.5
	Yeast (SUR7)-Human (CD81)	2.0	3.5

Table 2: Pairwise sequence identity and similarity values across species. The table shows all the identity and similarity values obtained by pairwise comparison of full length unaligned sequences EMBOSS NEEDLE.

Pairwise sequence comparisons revealed substantial variation in conservation across membrane trafficking protein families indicating that different functional modules evolve under different evolutionary constraints. In ESCRT-0 (Vps27-Hrs-TOL2), yeast Vps27 and human Hrs showed more identity (21.1%) and similarity (34.9%) with each other than with plant TOL2 which

suggests divergence of plant homologs despite conservation of core functional roles. Similar results were observed in ESCRT-0 (Vps27-Hrs-TOL6) but plant TOL6 showed more identity and similarity with Vps27 and Hrs proteins compared to TOL2. Yeast Hse1 and Human STAM showed more identity (22.8%) and similarity (37.6%) with each other than with plant TOL2 and TOL6 in ESCRT-0 (Hse1-STAM-TOL2) and ESCRT-0 (Hse1-STAM-TOL6) further indicating the divergence of plant ESCRT-0 components.

In ESCRT-I (STP22-TSG101-ELC), somewhat increased identity and similarity values were observed with the highest values in human Tsg101 and plant ELC comparison (29.2% identity and 47% similarity) which suggests a stronger conservation of core complex assembly functions. ESCRT-II and ESCRT-III showed further increases in conservation. Plant and human VPS36 showed the most identity (33.3%) and similarity (49.8%) in ESCRT-II (VPS36-VPS36-VPS36). Highest identity (31.7%) and similarity (47.9%) was observed in plant and yeast Snf7 in ESCRT-III (Snf7-Snf7-Snf7.1-CHMP1A).

In contrast to ESCRT-0 and SNARE associated comparisons, Rab GTPases and VPS4 proteins showed markedly higher identity and similarity values across species, consistent with stronger conservation of catalytic trafficking regulators. In RAB5 (YPT52-VPS21-RAB5A-RAB5), the identity values ranged from 41.9-59.4% and the similarity from 54.2-71.4% with the highest values observed in plant RAB5A and human RAB5. The identity values ranged from 35.5-61.4% and the similarity values ranged from 52.7-74% in RAB7 (YPT7-RAB7-RAB7) with the highest values in yeast Ypt7 and human RAB7. In Disassembly (VPS4-VPS4-VPS4), high identity (52-59%) and similarity (68.1-73.6%) in all three pairs.

SNAREs and Tetraspanins protein families showed comparatively lower sequence conservation, suggesting greater divergence in membrane fusion associated or membrane organizing proteins than in catalytic trafficking regulators. In SNAREs (VAM3-SYP121-STX10), identity values ranged from 13.2-18.6% and the similarity from 25.7-33.2% with the highest values in yeast VAM3 and plant SYP121. In case of Tetraspanins (SUR7-TET8-CD63-CD81), plant TET8 showed most identity and similarity with human proteins CD63 (21.8% identity and 37.3% similarity) and CD81 (22.0% identity and 33.7% similarity). Yeast SUR7 showed the lowest identity and similarity value with human proteins CD63 (1.2% identity and 2.5% similarity) and CD81 (2.0% identity and 3.5% similarity).

3.2 Domain architecture comparison

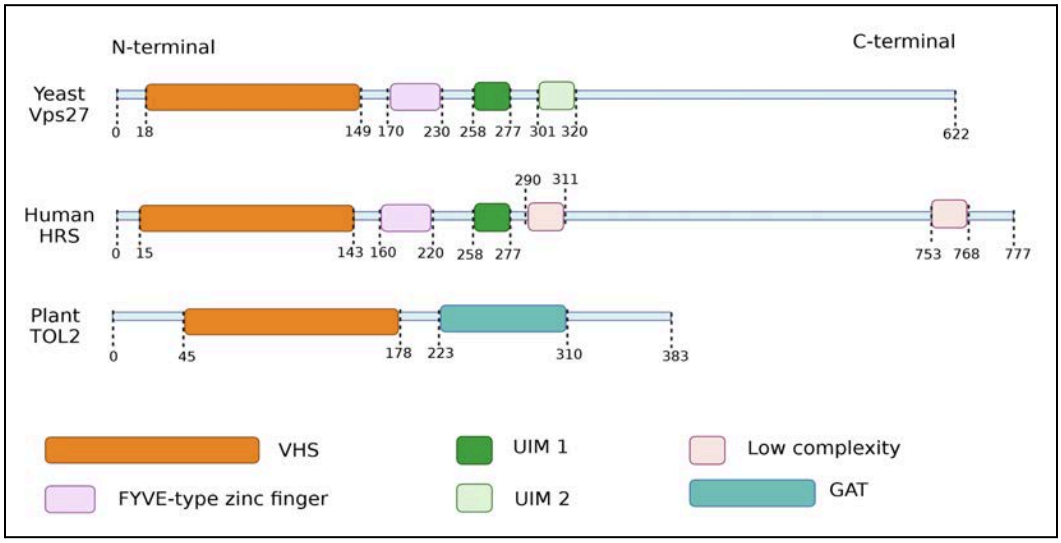


Figure 1.1: ESCRT-0 (Vps27-Hrs-TOL2) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: VHS (orange), FYVE type zinc finger (light purple), UIM motifs (dark and light green), GAT (blue) and low complexity regions (pink) arranged along a linear protein scale from N- to C- terminal.

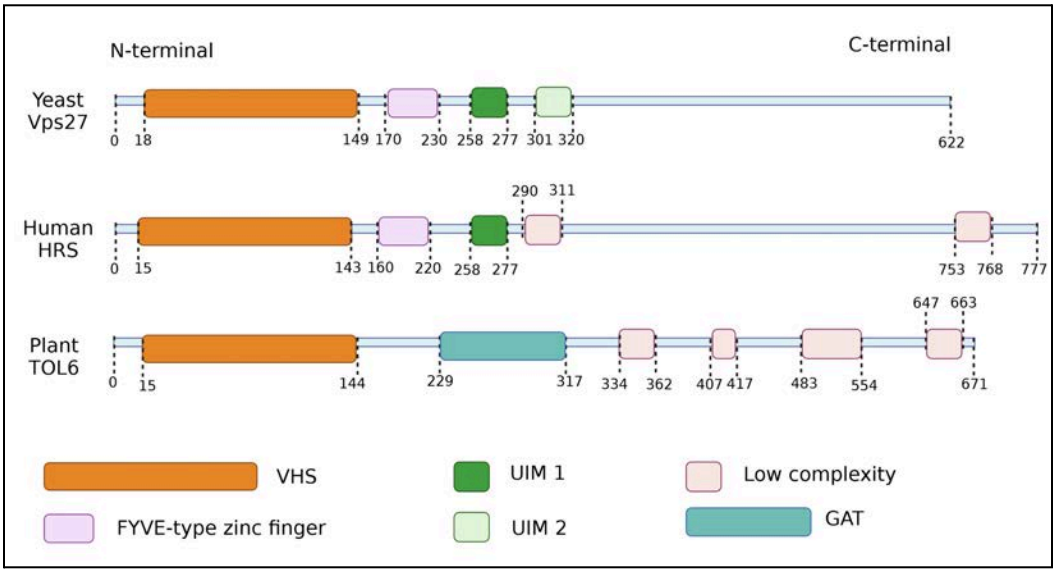


Figure 1.2: ESCRT-0 (Vps27-Hrs-TOL6) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: VHS (orange), FYVE type zinc finger (light purple), UIM motifs (dark and light green), GAT (blue) and low complexity regions (pink) arranged along a linear protein scale from N- to C- terminal.

Vps27 and Hrs retain comparable VHS, UIM and FYVE-type zinc finger domain architecture whereas plant TOL proteins show partial architectural correspondence with marked divergence in UIM and FYVE related features which indicates lineage specific remodelling of regulatory architecture. Low complexity regions are present in human HRS and plant TOL6. The VHS domain is retained as an N-terminal conserved module in all three sequences.

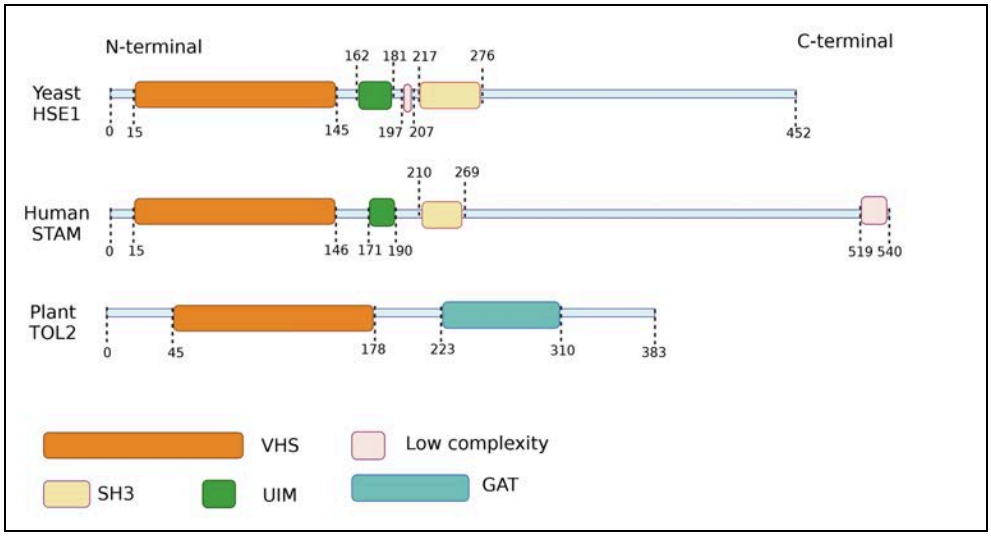


Figure 1.3: ESCRT-0 (Hse1-STAM_TOL2) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: VHS (orange), SH3 (yellow), UIM motif (dark green), GAT (blue) and low complexity regions (pink) arranged along a linear protein scale from N- to C- terminal.

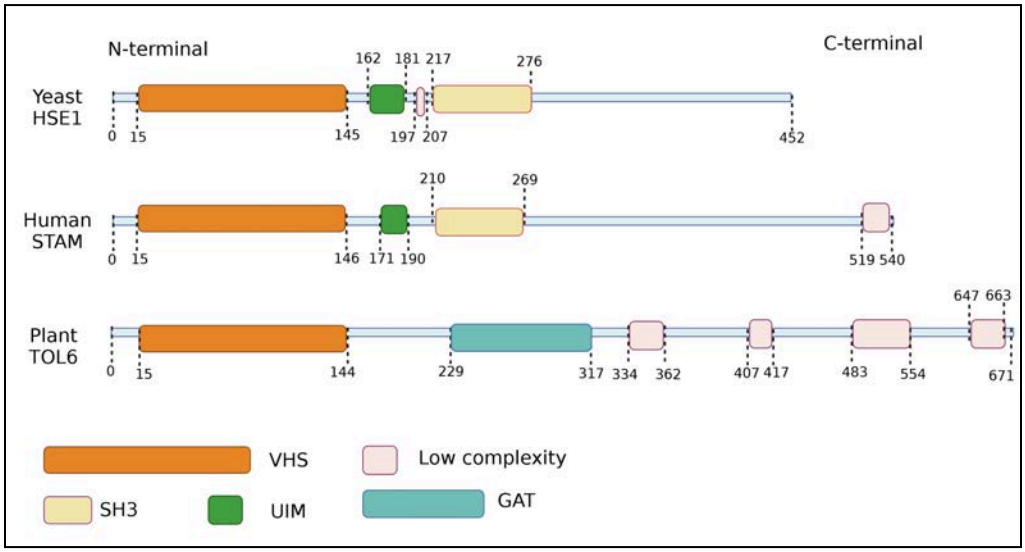


Figure 1.4: ESCRT-0 (Hse1-STAM_TOL6) protein family schematic diagram (Generated using BioRender), Domains are represented as colored boxes: VHS (orange), SH3 (yellow), UIM motif

(dark green), GAT (blue) and low complexity regions (pink) arranged along a linear protein scale from N- to C- terminal.

The domains of Hse1 and STAM (VHS, UIM and SH3) are conserved. Plant TOL proteins show partial architectural similarity with marked divergence in UIM and SH3 related features which indicates lineage specific remodelling of regulatory architecture. Low complexity regions are present in all sequences in various ranges. The VHS domain is retained as an N-terminal conserved module in all three sequences.

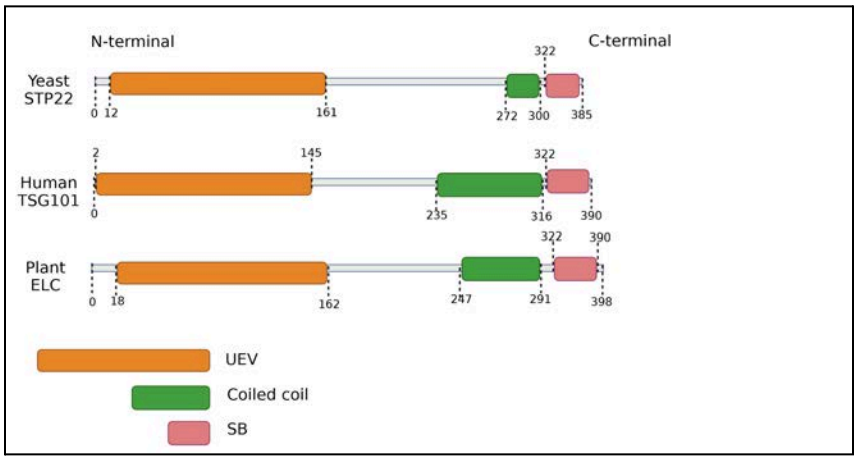


Figure 1.5: ESCRT-I (TSG101-STP22-ELC) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: UEV (orange), Coiled coil (dark green), SB (dark pink) arranged along a linear protein scale from N- to C- terminal.

The ESCRT-I sequences have UEV and SB domains with similar domain positions indicating the preservation of overall domain architecture across species.

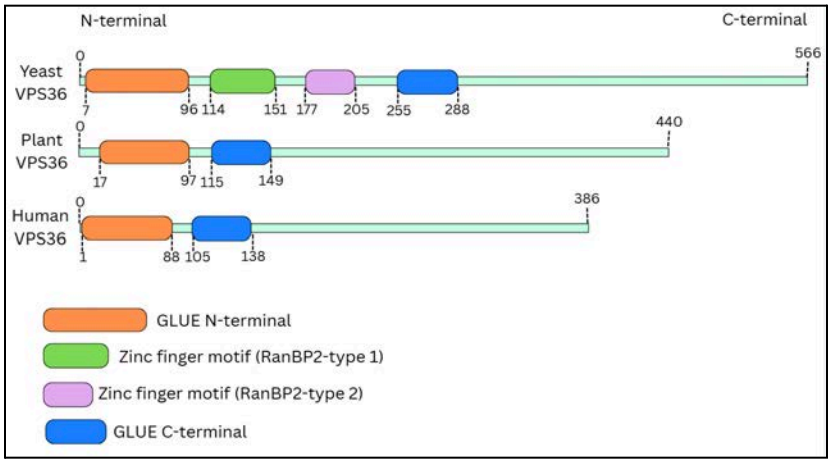


Figure 1.6: ESCRT-II (VPS36-VPS36-VPS36) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: GLUE N-terminal (orange), Zinc

Finger motif RanBP2-type 1(light green), Zinc Finger motif RanBP2-type 2 (purple) and GLUE C-terminal (blue) arranged along a linear protein scale from N- to C- terminal.

The domains GLUE N-terminal and GLUE-C terminal are present in all sequences which indicate conservation of core functional modules. However, yeast VPS36 is the longest sequence and contains two other domains which are absent in plant and human VPS36 suggesting lineage specific domain expansion.

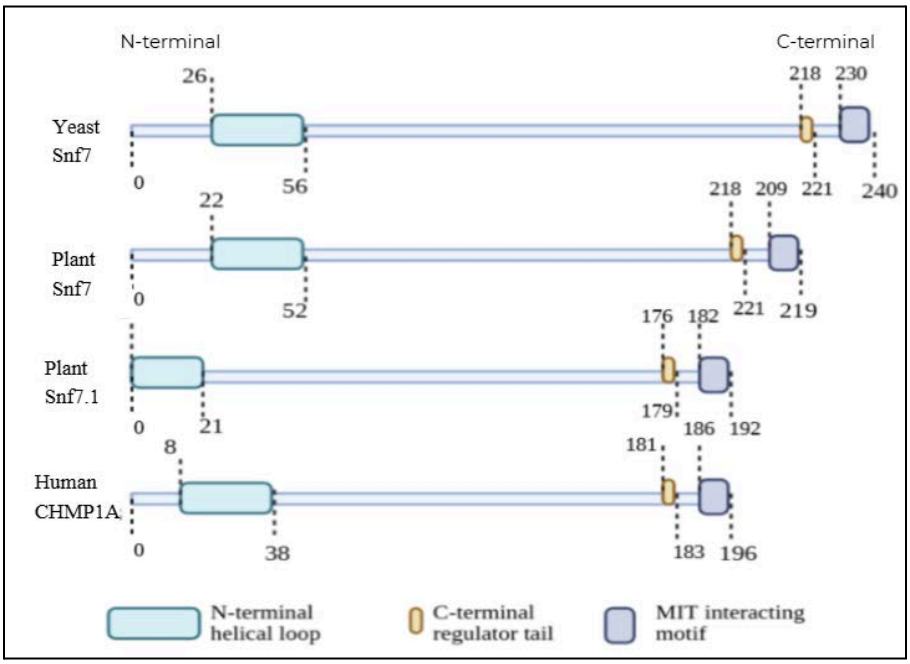


Figure 1.7: ESCRT-III (Snf7-Snf7-Snf7.1-CHMP1A) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: N-terminal helical loop (light blue), C-terminal regulator tail (yellow) and MIT interacting motif (light purple) arranged along a linear protein scale from N- to C- terminal.

In ESCRT-III, the position of the N-terminal helical loop is conserved in yeast and plant while the C-terminal regulator tail and MIT interacting motif show similar positioning in all three sequences which indicates preservation of key functional regions. Variable regions are observed in the centre positions of all the sequences indicating localized divergence.

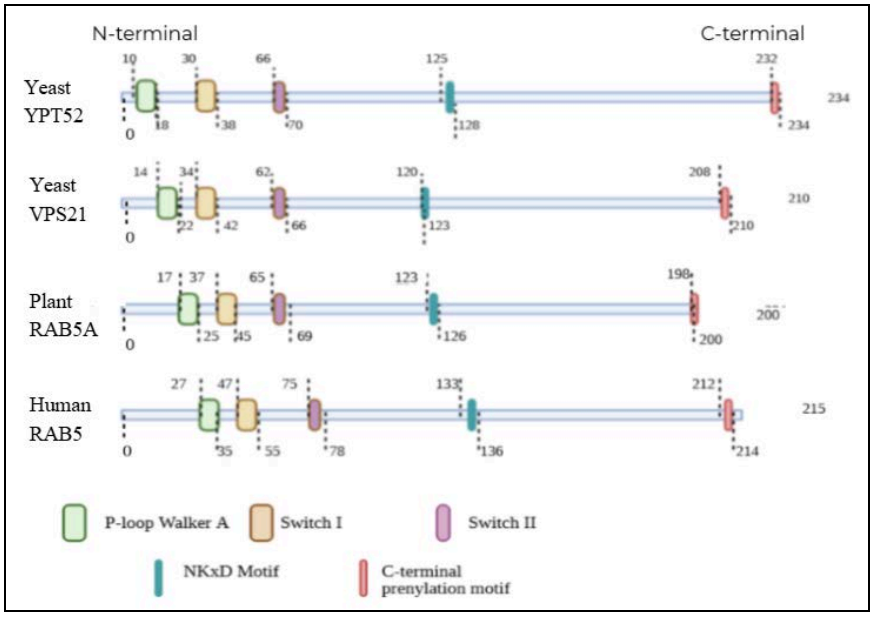


Figure 1.8: Rab GTPases (YPT52-VPS21-RAB5A-RAB5) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: P-loop Walker A (light green), Switch I (light orange), Switch II (light purple), NKxD motif (blue) and C terminal prenylation motif (light red) arranged along a linear protein scale from N- to C- terminal.

The RAB5 sequences have conserved domains and similar domain positions indicating the preservation of overall domain architecture across species.

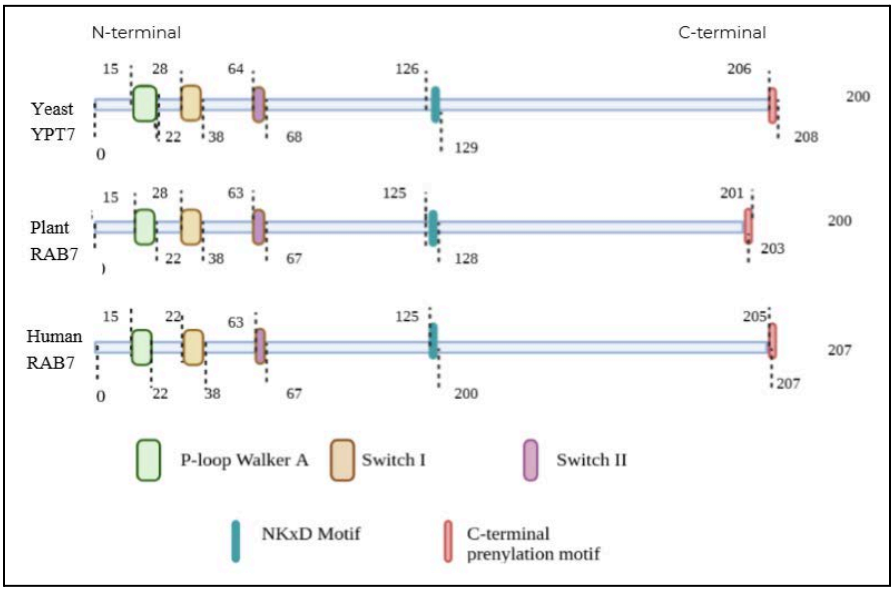


Figure 1.9: Rab GTPases (YPT7-RAB7-RAB7) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: P-loop Walker A (light green),

Switch I (light orange), Switch II (light purple), NKxD motif (blue) and C terminal prenylation motif (light red) arranged along a linear protein scale from N- to C- terminal.

The RAB7 sequences have conserved domains and similar domain positions indicating the preservation of overall domain architecture across species.

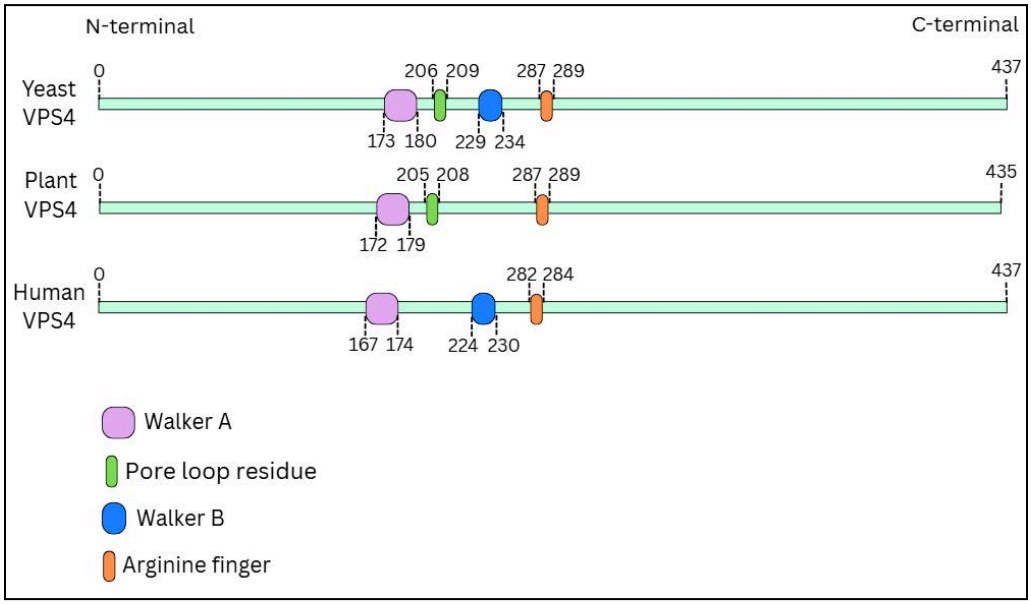


Figure 1.10: Disassembly (VPS4-VPS4-VPS4) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: Walker A (light purple), Pore loop residue (light green), Walker B (blue) and Arginine finger (orange) arranged along a linear protein scale from N- to C- terminal.

For VPS4, all the conserved domains are in the central position and the length of all sequences are also similar which supports strong structural conservation. However, the Pore loop residue is absent in human VPS4 and Walker B region is absent in plant VPS4 which shows lineage specific architectural differences. Variations are also observed in both C terminal and N terminal.

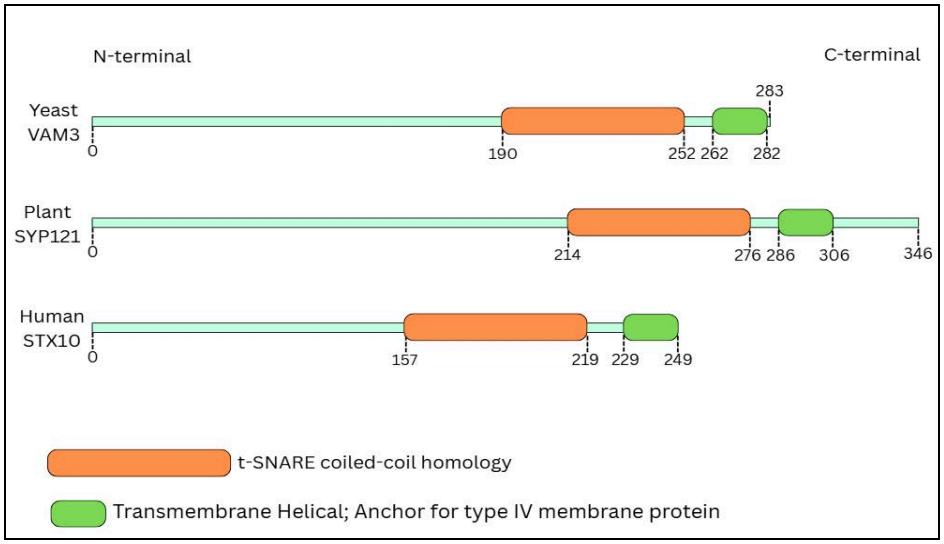


Figure 1.11: SNAREs (VAM3-SYP121-STX10) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: t-SNARE coiled coil homology (orange) and transmembrane helical (light green) arranged along a linear protein scale from N- to C- terminal.

The conserved t-SNARE coiled coil homology and transmembrane helical are positioned at the C terminal in all three sequences which shows preservation of core functional architecture. Variations are observed in the N terminal and the overall sequence lengths differ which suggest divergence in regulatory regions.

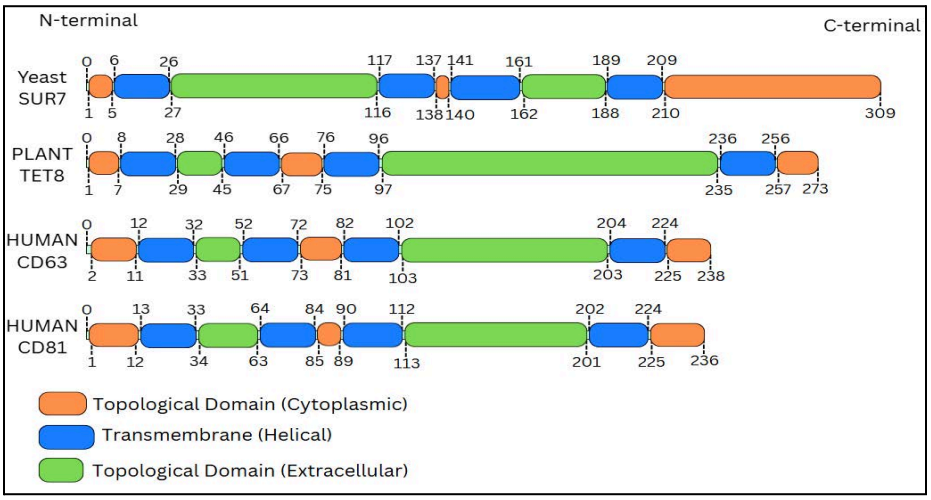


Figure 1.12: Tetraspanins (SUR7-TET8-CD63-CD81) protein family schematic diagram (Generated using BioRender). Domains are represented as colored boxes: Topological domain

(cytoplasmic) (orange), Transmembrane (helical) (blue) and Topological domain (extracellular) (light green) arranged along a linear protein scale from N- to C- terminal.

The same domain is present multiple times in all three sequences. The first topological domain (cytoplasmic) and the last topological domain (extracellular) are both longer in yeast. The second topological domain (extracellular) is longest in plant and shortest in yeast. Human protein sequences are shorter than plant and yeast which shows differences in domain length and organization across species.

Protein family	**Domain**	**Comparison**	**% Identity**	**% Similarity**
ESCRT-0 (VPS27-Hrs-TOL2)	VHS	Yeast-Human	23.7%	38.2%
		Yeast-Plant	16.8%	39.6%
		Plant-Human	27.3%	50.8%
	UIM 1	Yeast-Human	36.8%	57.9%
		Yeast-Plant	20.8%	33.3%
		Plant-Human	13.0%	30.4%
	UIM 2	Yeast-Human	20.0%	50.0%
		Yeast-Plant	14.3%	47.6%
		Plant-Human	22.7%	50.0%
	FYVE-type zinc finger	Yeast-Human	42.6%	65.6%
		Yeast-Plant	14.9%	23.0%
		Plant-Human	1.0%	1.9%
	Tail region (C terminal)	Yeast-Human	16.6%	29.5%
		Yeast-Plant	1.9%	4.3%
		Plant-Human	1.6%	2.2%
ESCRT-0 (VPS27-Hrs-TOL6)	VHS	Yeast-Human	23.7%	38.2%
		Yeast-Plant	20.4%	42.3%

		Plant-Human	36.7%	59.4%
	UIM 1	Yeast-Human	36.8%	57.9%
		Yeast-Plant	15.0%	35.0%
		Plant-Human	3.2%	12.9%
	UIM 2	Yeast-Human	20.0%	50.0%
		Yeast-Plant	5.9%	5.9%
		Plant-Human	3.2%	12.9%
	FYVE-type zinc finger	Yeast-Human	42.6%	65.6%
		Yeast-Plant	0.8%	1.5%
		Plant-Human	5.0%	5.8%
	Tail region (C terminal)	Yeast-Human	16.6%	29.5%
		Yeast-Plant	5.3%	10.5%
		Plant-Human	9.9%	17.8%
ESCRT-0 (Hse1-STAM-TOL2)	VHS	Yeast-Human	25.7%	45.1%
		Yeast-Plant	21.2%	38.6%
		Plant-Human	23.5%	40.9%
	UIM	Yeast-Human	40.0%	75.0%
		Yeast-Plant	6.9%	20.7%
		Plant-Human	25.0%	30.0%
	SH3	Yeast-Human	41.7%	63.3%
		Yeast-Plant	4.9%	11.1%
		Plant-Human	18.8%	29.0%
ESCRT-0 (Hse1-STAM-TOL6)	VHS	Yeast-Human	25.7%	45.1%
		Yeast-Plant	25.9%	43.5%
		Plant-Human	30.4%	45.2%

	UIM	Yeast-Human	40.0%	75.0%
		Yeast-Plant	5.0%	10.0%
		Plant-Human	5.0%	10.0%
	SH3	Yeast-Human	41.7%	63.3%
		Yeast-Plant	3.2%	4.3%
		Plant-Human	2.2%	5.6%
ESCRT-I (STP22- ELC-TSG101)	UEV	Yeast-Human	25.8%	49.0%
		Yeast-Plant	24.0%	40.4%
		Plant-Human	37.9%	60.0%
	Coiled Coil	Yeast-Human	13.8%	48.3%
		Yeast-Plant	20.0%	40.0%
		Plant-Human	26.7%	50.0%
	SB	Yeast-Human	40.6%	55.1%
		Yeast-Plant	23.1%	39.7%
		Plant-Human	35.4%	51.9%
ESCRT II	GLUE N-terminal domain	Yeast-Human	18.9%	38.9%
		Yeast-Plant	21.9%	42.7%
		Plant-Human	32.5%	44.2%
	GLUE C-terminal domain	Yeast-Human	20.5%	48.7%
		Yeast-Plant	31.7%	41.5%
		Plant-Human	27.8%	44.4%
ESCRT-III	N-terminal helical core	Yeast-Human	9.5%	26.2%
		Yeast-Plant1	22.6%	48.4%
		Yeast-plant2	16.1%	38.7%
		Plant1-Human	35.5%	45.2%

		Plant2-Human	29.0%	35.5%
	MIM	Yeast-Human	11.8%	11.8%
		Yeast-Plant1	70.0%	70.0%
		Yeast-Plant2	70.0%	70.0%
		Plant1-Human	28.6%	35.7%
		Plant2-Human	28.6%	35.7%
	C-terminal regulatory tail	Yeast-Human	0.0%	25.0%
		Yeast-Plant1	14.3%	14.3%
		Yeast-Plant2	14.3%	14.3%
		Plant1-Human	0.0%	50.0%
		Plant2-Human	0.0%	50.0%
RAB5	P-loop (Walker A)	Yeast-Human	77.8%	100%
		VPS21-Human	88.9%	100%
		Yeast-Plant	66.7%	66.7%
		VPS21-Plant	55.6%	66.7%
		Plant-Human	55.5%	66.7%
	Switch I	Yeast-Human	77.8%	88.9%
		VPS21-Human	66.7%	77.8%
		Yeast-Plant	77.8%	88.9%
		VPS21-Plant	66.7%	77.8%
		Plant-Human	100%	100%
	Switch II	Yeast-Human	100%	100%
		VPS21-Human	100%	100%
		Yeast-Plant	100%	100%
		VPS21-Plant	100%	100%

		Plant-Human	100%	100%
	NKxD motif	Yeast-Human	75.0%	75.0%
		VPS21-Human	75.0%	75.0%
		Yeast-Plant	75.0%	75.0%
		VPS21-Plant	75.0%	75.0%
		Plant-Human	100%	100%
	C-terminal prenylation motif	Yeast-Human	100%	100%
		VPS21-Human	50.0%	50.0%
		Yeast-Plant	66.6%	66.6%
		VPS21-Plant	25.0%	50.0%
		Plant-Human	66.6%	66.6%
RAB7	P-loop (Walker A**)**	Yeast-Human	100%	100%
		Yeast-Plant	100%	100%
		Plant-Human	100%	100%
	Switch I	Yeast-Human	72.7%	81.8%
		Yeast-Plant	81.0%	100%
		Plant-Human	81.8%	81.8%
	Switch II	Yeast-Human	100%	100%
		Yeast-Plant	100%	100%
		Plant-Human	100%	100%
	NKxD motif	Yeast-Human	75.0%	75.0%
		Yeast-Plant	75.0%	75.0%
		Plant-Human	100%	100%
	C-terminal	Yeast-Human	66.6%	66.6%
		Yeast-Plant	66.6%	66.6%

		Plant-Human	66.6%	66.6%
Disassembly	Walker A	Yeast-Human	100%	100%
		Yeast-Plant	100%	100%
		Plant-Human	100%	100%
	Walker B	Yeast-Human	100%	100%
		Yeast-Plant	71.4%	100%
		Plant-Human	71.4%	100%
	Arginine finger	Yeast-Human	100%	100%
		Yeast-Plant	100%	100%
		Plant-Human	100%	100%
	Pore loop residues	Yeast-Human	100%	100%
		Yeast-Plant	75.0%	75.0%
		Plant-Human	75.0%	75.0%
SNAREs	t-SNARE coiled coil homology	Yeast-Human	19%	46%
		Yeast-Plant	23.8%	46%
		Plant-Human	14.5%	38.7%
	TM Helix	Yeast-Human	13.6%	59.1%
		Yeast-Plant	43.5%	69.6%
		Plant-Human	23.8%	71.4%
Tetraspanins	Four transmembrane helices	Yeast-Human CD63	19.6%	32.6%
		Yeast-Human CD81	16.7%	28.6%
		Yeast-Plant	15.5%	27.2%
		Plant-Human CD63	20.9%	49.3%
		Plant-Human	20%	40%

		CD81		
	Topological Cytoplasmic tails	Yeast-Human CD63	1.4%	2.8%
		Yeast-Human CD81	1.9%	3.3%
		Yeast-Plant	1.7%	3.4%
		Plant-Human CD63	21.2%	31.2%
		Plant-Human CD81	22.2%	31.6%
	Extracellular Topological domains	Yeast-Human CD63	0%	0.6%
		Yeast-Human CD81	20.7%	24.8%
		Yeast-Plant	1.2%	1.8%
		Plant-Human CD63	23.8%	38.1%
		Plant-Human CD81	22.1%	35.3%

Table 3: Domain-wise conservation comparison for each protein family. The table shows all the identity and similarity values obtained by pairwise comparison of aligned functional domains and motifs using EMBOSS NEEDLE.

In ESCRT-0 (VPS27- Hrs-TOL2 and VPS27-Hrs-TOL6), the UIM and FYVE-type zinc finger domains of yeast showed more conservation with human Hrs and the VHS domain showed more conservation with plant TOL2 and TOL6. This suggests that ubiquitin binding and membrane targeting functions are more conserved among yeast and humans while cargo recognition by the VHS domain may have diverged differently across lineages. For ESCRT-0 (Hse1-STAM-TOL2), all three domains (VHS, UIM and SH3) of yeast showed more conservation with human STAM indicating closer functional conservation. In ESCRT-0 (Hse1-STAM-TOL6), more conservation of yeast VHS was seen with plant TOL6 suggesting partial functional adaptation in plants.

In ESCRT-I (STP22-ELC-TSG101), the UEV domain of yeast is more conserved with plants and the SB domain shows more conservation with humans. The sequences corresponding to the GLUE N-terminal domain of yeast showed more conservation among plant and human and C-terminal domain of yeast showed more conservation with plant in ESCRT-II (VPS36-VPS36-VPS36). This suggests that the membrane interaction functions are broadly conserved while other regions may be lineage specific. In ESCRT-III (Snf7-Snf7-Snf7.1-CHMP1A), the sequences corresponding to the N terminal helical core of yeast was most conserved among plant Snf7 and human CHMP1A and the MIM domain of yeast showed conservation with both plant proteins.

In RAB5 (YPT52-VPS21-RAB5A-RAB5), all domains show very high conservation (50-100% identity and 50-100% similarity) with Switch II showing 100% conservation in all sequences and C terminal prenylation motif showing 100% conservation among yeast and human. RAB7 (YPT7-RAB7-RAB7) domains show even higher conservation (66.6-100% identity and 66.6-100% similarity) where P loop (Walker A) and Switch II show 100% conservation in all sequences. This reflects strong evolutionary constraints required for GTP binding and hydrolysis. Highest conservation (71.4-100% identity and 75-100% similarity is observed in Disassembly (VPS4-VPS4-VPS4). Walker A and Arginine finger of VPS4 show 100% conservation in all species while Walker B and P loop residues show 100% conservation among human and yeast. This shows the conservation of VPS4 sequence for ATPase activity across all three species.

TM helix and t-SNARE coiled homology domains of yeast show more conservation with plants in SNAREs (VAM3-SYP121-STX10) which suggest some conservation of membrane fusion machinery despite sequence divergence. In all domains of Tetraspanins (SUR7-TET8-CD63-CD81), more conservation is observed among plant and human proteins. Four transmembrane helices domains of yeast show some conservation with human proteins. However, the lower identity and similarity values indicate significant divergence in tetraspanin domains despite retention of functional roles.

3.3 Motif conservation analysis

VHS:

```
sp|P40343| -------------------MSV----------STPSELDALIEQATSESIPNGDLDLPIA
BAA23366.1 -------------------MGR----------GSGTFERLLDKATSQLLLETDWE--SI
sp|Q9LNC6| MDKLKIAEWGEKLKTGGAQMSRMVSEKVKDMLQAPTLESKMVDEATLETLEEPNWG--MN
                            *.            .   . ::::** : : : :

sp|P40343| LEISDVLRSRRVNPKDSMRCIKKRILNTADNPNTQLSSWKLTNICVKNGGTPFIKEICSR
BAA23366.1 LQICDLIRQGDTQAKYAVNSIKKKV--NDKNPHVALYALEVMESVVKNCGQTVHDEVANK
sp|Q9LNC6| MRICAQINNDEFNGTEIVRAIKRKI--SGKSPVSQRLSLELLEACAMNC-EKVFSEVASE
           :.*.  :..   : .  :..**:::  . ..*     : :: :  . *    . .*:...

sp|P40343| EFMDTMEHVILR-EDSNEELSELVKTILYELYVAFKNDSQLNYVAKVYDKLISRGIKFPE
BAA23366.1 QTMEELKDLLKR-QVEVNVRNKILYLIQAWAH-AFRNEPKYKVVQDTYQIMKVEGHVFPE
sp|Q9LNC6| KVLDEMVWLIKNGEADSENRKRAFQLIRAWGQ--SQDLTYLPVFHQTYMSLEGEN-----
           : :: :  :: . : . :  .. .  *        :: .    . ..*  :  ..
```

UIM 1:

```
sp|P40343| DYSTPEDEEELIRKAIELSLKESRN--------------------SA
BAA23366.1 DETALQEEEEL-QLALALSQSEAEEKERLRQKSTYTSYPKAEPMPSA
sp|Q9LNC6| SMLNTEGKPNHTEDDLTVSLMEK----------------------CK
           .    : : :  .  : :*  *                       .
```

UIM 2:

```
sp|P40343| NEVKRQEIEEEEDPDLKAAIQESLREAEEAKLR
BAA23366.1 PVNSSAPLAEDIDPELARYLNRNYWEKKQEEAR
sp|Q9LNC6| TDDEGVLFEALHLNDELQQVLSSYKKPDETEKK
              .   :       :    :   .  : .: : :
```

FYVE:

```
sp|P40343| KLTLSNSPTAMFDSKTPADWIDSDACMICSKKFSLLNRKHHCRSCGGVFCQEHSSNSIPL
BAA23366.1 ----FKESDAMFAAERAPDWVDAEECHRCRVQFGVMTRKHHCRACGQIFCGKCSSKYSTI
sp|Q9LNC6| ---------------------------------GLHARGEENSMPGQSSLESLMQRPVPV
                                           .:  * ..    *     .  ..  .:

sp|P40343| PDLGIYEPVRVCDSCFEDYD--------LKRHDDSKKSKK-----------HRHKRKKDR
BAA23366.1 PKFGIEKEVRVCEPCYEQLN--------RKAEGKATSTTELPPEYLTSPLSQQSQLPPKR
sp|Q9LNC6| PPPGSY-PVPNQEQALGDDDGLDYNFGNLSIKDKKEQIEI-----------TRNSLELLS
           *  *    *   : .  : :         . ...  .              : .
```

Figure 2.1: MSA showing conserved motifs of ESCRT-0 (VPS27-Hrs-TOL2). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

Some consecutive residues show partial conservation in the VHS domain (LEIS, MRCIKKRI, SWKLTN, KEICSRE). Other domains have no consecutive conserved residues. Overall, the VHS

domain retained several conserved hydrophobic and basic residues across species, the UIM motifs were only partially conserved and FYVE-type zinc binding domains were strongly absent in plant TOL2.

VHS:

```
sp|P40343| MSVSTPSELDALIEQATSESIPNGDLDLPIALEISDVLRSRRVNPKDSMRCIKKRILNTA
BAA23366.1 MGRGSGT-FERLLDKATSQLLLETDWE--SILQICDLIRQGDTQAKYAVNSIKKKV--ND
sp|O80910| MASSSAS-ATVAVDKATSDLLLGPDWT--TNMEICDSVNSLHWQAKDVVKAVKKRL--QH
           *. .: :     :::***: :   *      ::*.* :..   :.*  :..:**::

sp|P40343| DNPNTQLSSWKLTNICVKNGGTPFIKEICSREFMDTMEHVILREDSNEELSELVKTILYE
BAA23366.1 KNPHVALYALEVMESVVKNCGQTVHDEVANKQTMEELKDLLKRQVEVNVRNKILYLIQAW
sp|O80910| KSSRVQLLALTLLETLVKNCGDYLHHQVAEKNILGEMVKIVKKKADMQVRDKILVMVDSW
           ..... * :  : :  *** *  . .::..:: :  : .:: :: . :  .:::  :

sp|P40343| LYVAFKN-DSQLNYVAKVYDKLISRGIKFPEKLTLSNSPTAMFDSKTPADW-------ID
BAA23366.1 AH-AFRN-EPKYKVVQDTYQIMKVEGHVFPE----FKESDAMFAAERAPDW-------VD
sp|O80910| QQ-AFGGPEGKYPQYYWAYDELRRSGVEFPR---------------RSPDASPIITPPVS
              ** . : :      .*: :   *  **.               ..*        :.
```

UIM 1:

```
sp|P40343| DSCFEDYDLKRHDDSKKSKK----------HRHKRKKDRDYSTPEDEEELIRKAIELSL
BAA23366.1 EPCYEQLNRKAEGKATSTTELPPEYLTSPLSQQSQLPPKRDETALQEEEEL-QLALALSQ
sp|O80910| RRLDEAMATEVEGLSLSSIESMRDVM-DLLGDMLQAVDPSDREAVKDEVIV-DLVERCRS
              *    : .. : .: :          .  :     *  : ::*  :   .

sp|P40343| KESRN-----------------------------------------SASSEPIVPVVESKN
BAA23366.1 SEAEEKERLRQK-------------------STYTSYPKAEPMPSASSAPPASSLYSSP
sp|O80910| NQKKLMQMLTSTGDDELLGRGLDLNDSLQILLAKHDAIASGSPLPVQASGSPLSVQASKP
           .: .                                    :* .  .   *.
```

UIM 2:

```
sp|P40343| EVKRQEIEEEEDPDL---------------KAAIQESLREAEEA
BAA23366.1 VNSSAPLAEDIDPEL---------------ARYLNRNYWEKKQE
sp|O80910| ADSSPKSSEAKDSSSIAGSSSPIPATVSTGKSPIDEEYEEEEDE
             .     *  *..                    ::..  * ::
```

FYVE:

```
sp|P40343| LYVAFKN-DSQLNYVAKVYDKLISRGIKFPEKLTLSNSPTAMFDSKTPADW-------ID
BAA23366.1 AH-AFRN-EPKYKVVQDTYQIMKVEGHVFPE----FKESDAMFAAERAPDW-------VD
sp|O80910| QQ-AFGGPEGKYPQYYWAYDELRRSGVEFPR---------------RSPDASPIITPPVS
              ** . : :      .*: :   *  **.               ..*        :.

sp|P40343| SDACMICSKKFSLLNRKHHCRSCG----GVFCQEHSSNSIP------LPDLGIYEPVRVC
BAA23366.1 AEECHRCRVQFGVMTRKHHCRACG----QIFCGKCSSKYST------IPKFGIEKEVRVC
sp|O80910| HPPLRQPQGGYGVPPAGYGVHQAGYGVPQAGYGIPQAGYGVPQAGYGIPQVGYGMPSGSS
                    :.:    :  : .*          .:          :*..*       .

sp|P40343| DSCFEDYDLKRHDDSKKSKK----------HRHKRKKDRDYSTPEDEEELIRKAIELSL
BAA23366.1 EPCYEQLNRKAEGKATSTTELPPEYLTSPLSQQSQLPPKRDETALQEEEEL-QLALALSQ
sp|O80910| RRLDEAMATEVEGLSLSSIESMRDVM-DLLGDMLQAVDPSDREAVKDEVIV-DLVERCRS
              *    : .. : .: :          .  :     *  : ::*  :   .
```

Figure 2.2: MSA showing conserved motifs of ESCRT-0 (VPS27-Hrs-TOL6). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

Some consecutive residues show partial conservation in the VHS domain (ATSE,LEIS, IKKRI VKNGG, EICSRE, KIL, AL, FP). UIM residues DE show some conservation in human and yeast and other domains have no consecutive conserved residues. Overall, the VHS domain retained several conserved hydrophobic and basic residues across species, the UIM motifs were only partially conserved and FYVE-type zinc binding domain were strongly absent in plant TOL.

Figure 2.3: MSA showing conserved motifs of ESCRT-0 (Hse1-STAM-TOL2). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

Partial motif conservation is observed in some consecutive residues in the VHS domain (KAT, IEKRL, LIES, EISS). UIM domain (DDEEL, KMS) and SH3 domain (YDL, ELS, DVIT, NYV)

show very less conservation. Overall, low sequence conservation is observed as only the VHS domain retained several conserved hydrophobic and basic residues across species.

VHS:

```
sp|P38753| MSSSAIK-IRNALLKATDPKLRSDNWQYILDVCDLVKEDPEDNGQEVMSLIEKRLEQQDA
NP_003464. MPLFATNPFDQDVEKATSEMNTAEDWGLILDICDKVGQS-RTGPKDCLRSIMRRVNHKDP
sp|O80910| MASSSAS-ATVAVDKATSDLLLGPDWTTNMEICDSVNSL-HWQAKDVVKAVKKRLQHKSS
           *.  : .     : ***.    . :*   :::** * .  .   :: :  : :*::::..

sp|P38753| NVILRTLSLTVSLAENCGSRLRQEISSKNFTSLLYALIESHSVHITLKKAVTDVVKQLSD
NP_003464. HVAMQALTLLGACVSNCGKIFHLEVCSRDFASEVSNVLNKG--HPKVCEKLKALMVEWTD
sp|O80910| RVQLLALTLLETLVKNCGDYLHHQVAEKNILGEMVKIVKKKA-DMQVRDKILVMVDSWQQ
           .* : :*:*  : ..***. :: ::..::: . :  :::.   .  : . :  :: .  :

sp|P38753| SFKD-DPSLRAMGDLYDKIKRKAPYL--VQPNVPEKHNMS---------TQADNSDDEEL
NP_003464. EFKN-DPQLSLISAMIKNLKEQGVTFPAIGSQAAEQAKASPALVAKDPGTVANKKEEEDL
sp|O80910| AFGGPEGKYPQYYWAYDELRRSGVEFP---------------------------------
            * . : .        .::::..  :
```

UIM:

```
sp|P38753| SFKD-DPSLRAMGDLYDKIKRKAPYL--VQPNVPEKHNMS---------TQADNSDDEEL
NP_003464. EFKN-DPQLSLISAMIKNLKEQGVTFPAIGSQAAEQAKASPALVAKDPGTVANKKEEEDL
sp|O80910| AFGGPEGKYPQYYWAYDELRRSGVEFP---------------------------------
            * . : .        .::::..  :

sp|P38753| QKALKMSLFEYEKQKKLQEQEKESAEVLPQQQQQHQQQNQAPAHKIPAQTVVRRVRALYD
NP_003464. AKAIELSLKEQRQQSTTLSTLYPSTSSLLTNHQHEG----------------RKVRAIYD
sp|O80910| -----------RRSPDASPIITPPVSHPPLRQPQGG-------YGVPPA----------G
                    .:.         ...    .: :                        .
```

SH3:

```
sp|P38753| QKALKMSLFEYEKQKKLQEQEKESAEVLPQQQQQHQQQNQAPAHKIPAQTVVRRVRALYD
NP_003464. AKAIELSLKEQRQQSTTLSTLYPSTSSLLTNHQHEG----------------RKVRAIYD
sp|O80910| -----------RRSPDASPIITPPVSHPPLRQPQGG-------YGVPPA----------G
                    .:.         ...    .: :                        .

sp|P38753| LTTNEPDELSFRKGDVITVLEQVYRDWWKGALRGNMGIFPLNYVTPIVEPSKEEIEKEK-
NP_003464. FEAAEDNELTFKAGEIITVLDDSDPNWWKGETHQGIGLFPSNFVTADLTAEPEMIKTEKK
sp|O80910| YGVHQAGYGVPQAGYGIPQAGYGVPQAGYGIPQVGYGMPSGSSRRLD-----EAMATEVE
             . : .    : *  *.       :   *  : . *: . .          * : .*
```

Figure 2.4: MSA showing conserved motifs of ESCRT-0 (Hse1-STAM-TOL6). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

VHS domain shows partial conservation (KAT, LDVCDLV, TLSL, NCG, LRQEISSKNF, LIES, DKIKRKA). No conserved motifs are observed in other domains. Overall, low sequence conservation is observed as only the VHS domain retained several conserved hydrophobic and basic residues across species.

UEV domain (N-terminal):

```
sp|P25604| MSANGKISVP-------------EAVVNWLFKVIQPIYNDGRTTFHDSLALLDNFHSLRP
NP_006283. M----------------------AVSESQLKKMVSKYKYRDLTVRETVNVITLYKDLKP
sp|Q9LHG8| M-------VPPPSNPQQVQQFLSSALSQRGPSSVPYEESNKWLIRQHLLNLISSYPSLEP
           *                      *: :   . :    .      :. : ::  : .*.*

sp|P25604| RTRVFTHSDGTPQLLLSIYGTISTGEDGSSPHSIPVIMWVPSMYPVKPPFISINLENFDM
NP_006283. VLDSYVFNDGSSRELMNLTGTIPVPYRGNT-YNIPICLWLLDTYPYNPPICFVK------
sp|Q9LHG8| KTASFMHNDGRSVNLLQADGTIPMPFHGVT-YNIPVIIWLLESYPRHPPCVYVN------
               : ..** .  *:.  ***.    * : :.**: :*: . ** :**   ::

sp|P25604| NTISSSLP-IQEYIDSNGWIALPILHCWDPAAMNLIMVVQELMSLL-HEPPQDQAPSLPP
NP_006283. PTSSMTIK-TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSRPISAS
sp|Q9LHG8| PTADMIIKRPHAHVTPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDPPLYSRRRPQP
            * .  :     :: ..* : ** *: *  .  :*: ::. :   :  :**  .     .
```

Coiled Coil:

```
sp|P25604| IIAIDKNHLRAVEQAIEQTMHSLNAQIDVLTANRAKV
NP_006283. DLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEEL
sp|Q9LHG8| ELNIGLKEMVEEKETLEQQLQIISMNTDILDSWVREN
              :   . :.:       ::* :   :. : ::*     :
```

C-terminal Vps23 core / SB domain:

```
sp|P25604| AKTDGLNQLYNLVAQDYALTDTIECLSRMLHRGTIPLDTFVKQGRELARQQFLVRWHIQ-
NP_006283. PTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQK
sp|Q9LHG8| CGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQFFHRATGSK
                 :*: :  * : *: *:*  *.. :: *.: :* :::: * *:*:**  *   .

sp|P25604| -RIT------SPLS-----
NP_006283. ARKT------AGLSDL--Y
sp|Q9LHG8| VRAAQMEVQVAAIAGRLHS
            * :      : ::
```

Figure 2.5: MSA showing conserved motifs of ESCRT-I (STP22-ELC-TSG101). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

UEV domain (HDSLALL, FTHSDG, GTI, HSIPVIMWV, YPVKPP, LPILH, NLIMVV, EPP) and SB domain (NQL, DTI, LDTFVKQ, LARQQF) show some partial motif conservation. Overall, moderate sequence conservation is observed.

GLUE N-terminal domain

```
sp|Q06696| MEYW----------HYVETTSSGQPLLREGEKDIFIDQSVGLYHGKSKILQRQ--RGRIF
NP_057159. MDRF--------------VWTSG--LLEINETLVIQQRGVRIYDGEEKIKFDA---GTLL
sp|Q9FF81| MASGSSSIAIGGLFENAEVTTSGRPVLRRNEVECFLLSSIDIDSEDDPPRFTALRSGNLI
           *                 . :**  :*. .*   :   .: :   ..         * ::

sp|Q06696| LTSQRIIYIDDAKPTQNSLGLELDDLAYVNYSSGFLTRSPRLILFFKDPSSKDELGKSAE
NP_057159. LSTHRLIW-RDQKNHECCMAILLSQIVFIEEQA------------------AGIGKSAK
sp|Q9FF81| LTTHRLIWIPSQSNESVPSSIPLSAVTHIYSHK----------------------KSIK
           *:::*:*:  . .  .   .: *. :..:                         ** :
```

GLUE C-terminal domain

```
sp|Q06696| EKNSLARNKSSHSALSSSSSTGSSTEFVQLSFRKSDGVL-F-----SQATERALENILTE
NP_057159. EPGPF---------------QSSKNSYIKLSFKEHGQIE-FYRRLSEEMTQRRWENMPVS
sp|Q9FF81| DPGSI---------------------VVTIVFRGKGDFDGFLSKLWECWRGRAWEEEEKS
           : ..:                     : : *:  . .  *     .    *  *:   .

sp|Q06696| KNKHIFNQNVVSVNGVDMRKGAS---SHEYNNEVPFIETKLSRIGISSLEKSRENQLLNN
NP_057159. QS-------------LQTNRGPQ-------------PGRIRAVGIVGIERKLEEKRKET
sp|Q9FF81| ES-------------ETSKSGSGTVAQGLYGN-----DGTVRMVGLAGILRKEQEQWEST
           :.               . *.                 :  :*: .: :. :::  ..
```

Zinc finger motif (RanBP2-type 1)

```
sp|Q06696| TASADVVSTWVCPICMVSNETQGEFTKDTLPTPICINCGVPAD
NP_057159. IV-----------------------------------------
sp|Q9FF81| SM-----------------------------------------
```

Zinc finger motif (RanBP2-type 2)

```
sp|Q06696| ANPQNQFGVNSENICPACTFANHPQIGNCEICGHRLP
NP_057159. ------------------------------------
sp|Q9FF81| ------------------------------------
```

Figure 2.6: MSA showing conserved motifs of ESCRT-II (VPS36-VPS36-VPS36). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

A very small number of residues are conserved in the GLUE-N terminal domain (SSG, IFLTSQRIIY) and GLUE C-terminal domain (FR, EN). No conservation is noticed in the Zinc finger motifs. Overall, low sequence conservation is observed.

N-terminal helical core:

```
sp|P39929| MWSSLFGWTSSNAKNKESPTKAIVRLREHINLLSKKQSHLRTQITNQENEARIFLT
sp|P39929| MWSSLFGWTSSNAKNKESPTKAIVRLREHINLLSKKQSHLRTQITNQENEARIFLT
sp|Q9SZE4| MMNRLFG----KPKQEANALQTLDKLNETLEMLEKKEKVLLKKAGAEVEKAKEYSR
sp|Q9SZE4- ------------------------------MLEKKEKVLLKKAGAEVEKAKEYSR
NP_002759. MDDTLF------------------QLKFTAKQLEKLAKKAEKDSKAEQAKVKKALL
                                        *.*  .   ..   :  :.:
```

C-terminal regulatory tail:

```
sp|P39929| AQENANQETSKIVNNNVNAAPISENKVSLPSVPSNKIKQSENSVKDGEEEEDEEDEDEKA
sp|P39929| AQENANQETSKIVNNNVNAAPISENKVSLPSVPSNKIKQSENSVKDGEEEEDEEDEDEKA
sp|Q9SZE4| ESE---ELESQLLQPATTAPP-------LPSVPVPAGRQPARPVPQKRTAEEEE------
sp|Q9SZE4- ESE---ELESQLLQPATTAPP-------LPSVPVPAGRQPARPVPQKRTAEEEE------
NP_002759. AEENGLEVLDQ-----------------LSQLPEGASAVGESSV---RSQEDQ-------
            .*   :  .:                 *..:*         .*   .  *::

sp|P39929| LRELQAEMGL
sp|P39929| LRELQAEMGL
sp|Q9SZE4| LAALQAEMAL
sp|Q9SZE4- LAALQAEMAL
NP_002759. LSRRLAALRN
           *    * :
```

MIM (MIT-interacting motif):

```
sp|P39929| LRELQAEMGL
sp|P39929| LRELQAEMGL
sp|Q9SZE4| LAALQAEMAL
sp|Q9SZE4- LAALQAEMAL
NP_002759. LSRRLAALRN
           *    * :
```

Figure 2.7: MSA showing conserved motifs of ESCRT-III (Snf7-Snf7-Snf7.1-CHMP1A). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

In ESCRT-III, motif pattern based analysis was done. A similar motif pattern was observed in N-terminal helical core (conserved alpha helical region with sequences LREH), C-terminal regulatory tail (variable acidic region with sequences EEEE) and MIM (MIT- interacting motif with sequences LRELQA). All these motifs were present but showed substitutions across species reflecting low sequence conservation.

P-loop (Walker A):

```
sp|P36018|FKLVLLGDSSVGKSS
sp|P36017|IKLVLLGEAAVGKSS
sp|P31582|AKLVLLGDVGAGKSS
NP_004153.FKLVLLGESAVGKSS
          ******: ..****
```

Switch I region:

```
sp|P36018|)ELRESTIGAA
sp|P36017|ENKEPTIGAA
sp|P31582|EFQESTIGAA
NP_004153.IEFQESTIGAA
          * :*.*****
```

NKXD motif:

```
sp|P36018|GNKVDL
sp|P36017|GNKIDM
sp|P31582|GNKADL
NP_004153.GNKADL
          *** *:
```

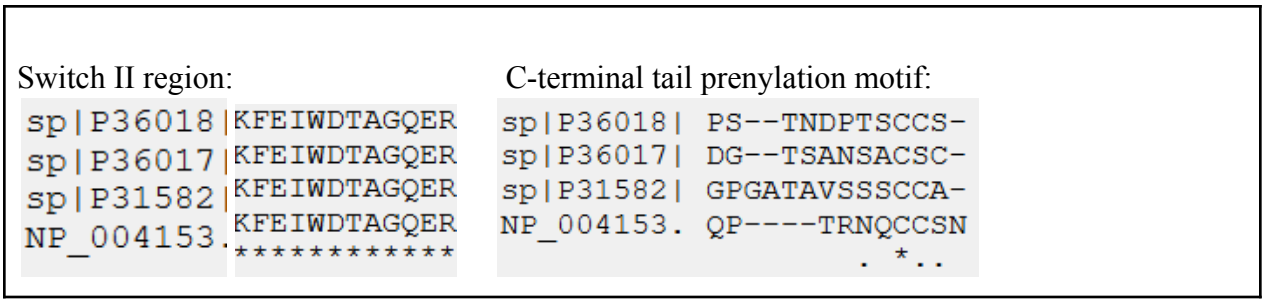


Figure 2.8: MSA showing conserved motifs of RAB5 **(**YPT52-VPS21-RAB5A-RAB5). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

High conservation of motifs was observed in RAB5 as well. P-loop (Walker A) motif with GDSSVG in yeast and C-terminal tail prenylation motif with CCS showed low sequence conservation. Thr/Ser rich Switch I region motif, Switch II region motif (DTAGQ) and NKXD motif with NKVD showed high sequence conservation.

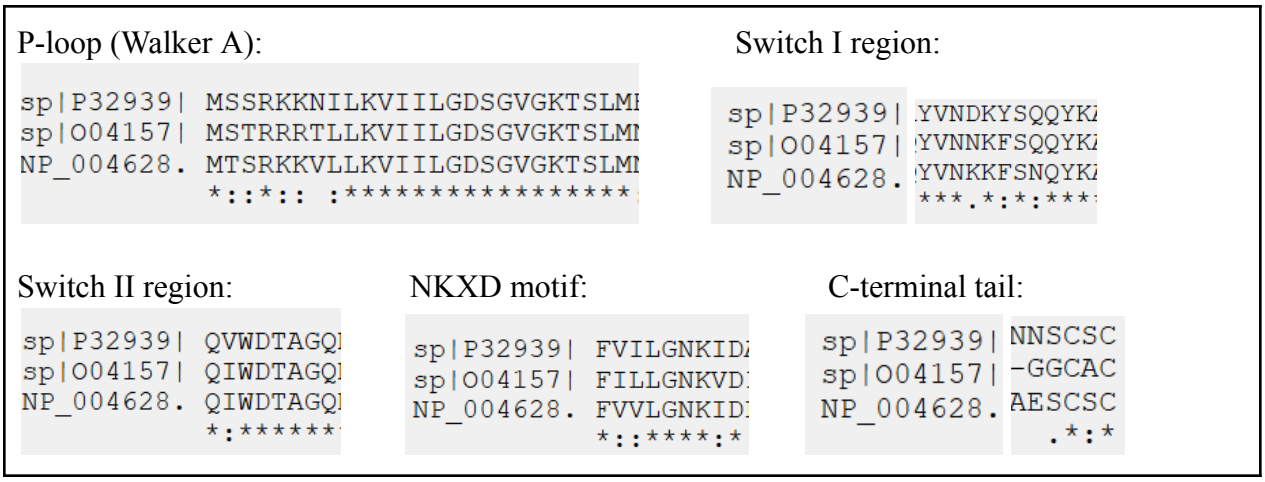


Figure 2.9: MSA showing conserved motifs of RAB7 (YPT7-RAB7-RAB7). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

High conservation of motifs was also observed in RAB7. P-loop (Walker A) motif with GDSGVGKT in yeast, Thr/Ser rich Switch I region motif, Switch II region motif (DTAGQ), NKXD motif with NKVD and C terminal tail with CCS showed high sequence conservation.

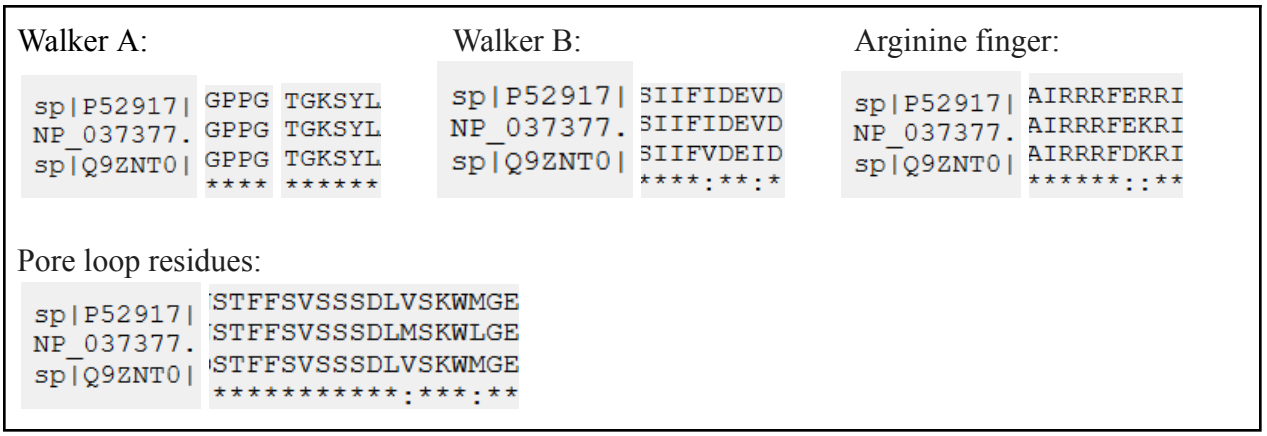


Figure 2.10: MSA showing conserved motifs of Disassembly (VPS4-VPS4-VPS4). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

High motif conservation was observed in VPS4 as well. Walker A motif with GPPGTGKS, Walker B motif with IIFIDE, Arginine finger with RRR and Pore loop with WMGE showed very high sequence conservation across all three species.

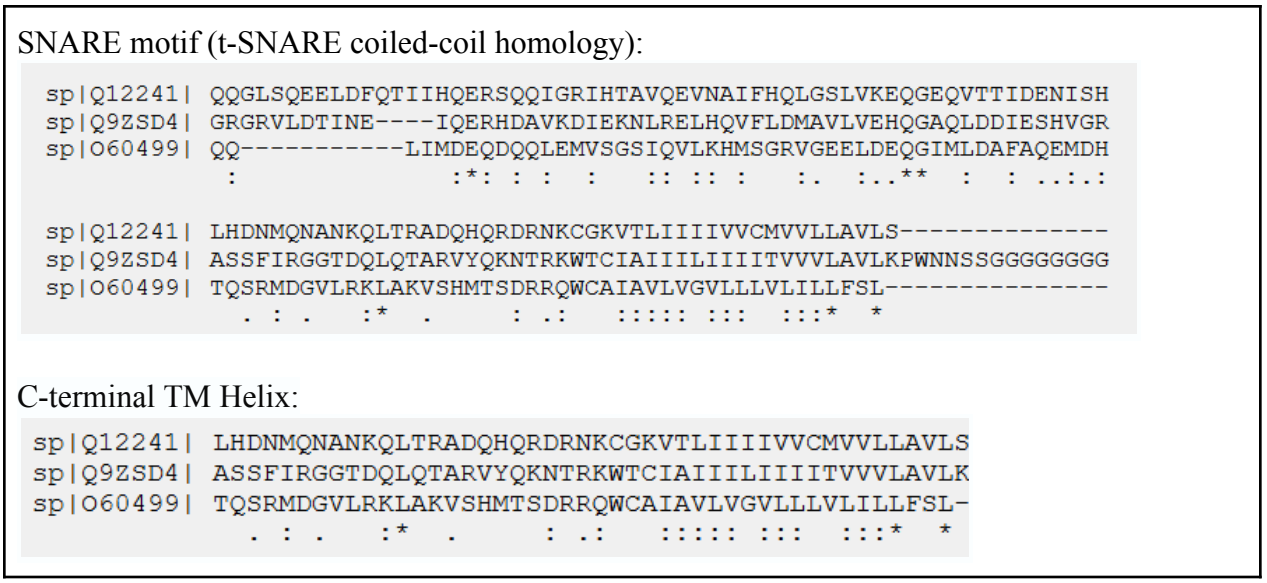


Figure 2.11: MSA showing conserved motifs of SNAREs (VAM3-SYP121-STX10). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

Partial motif conservation was observed in the SNARE motif (QER, VKEQG) and C-terminal TM Helix (VTLII, IVV, VVLL). Overall, moderate sequence conservation was observed.

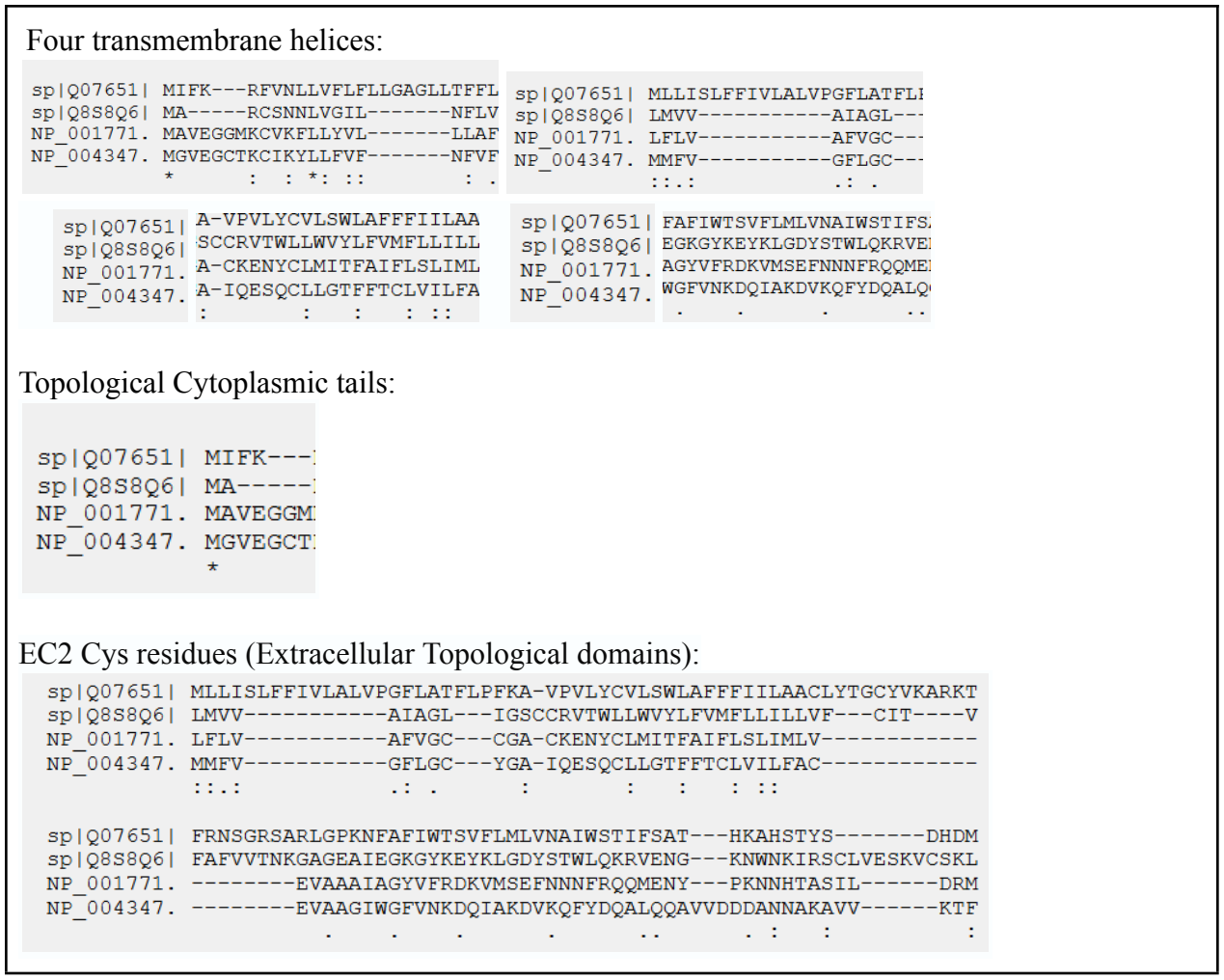
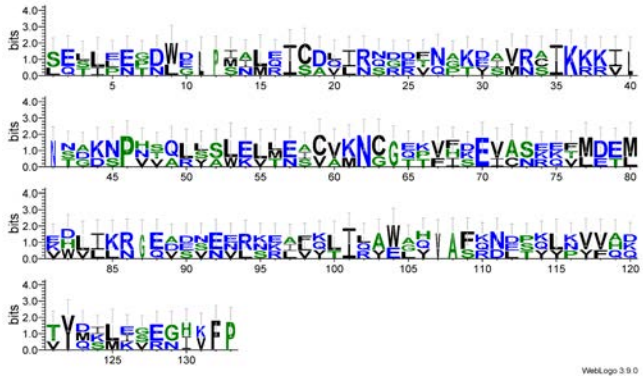


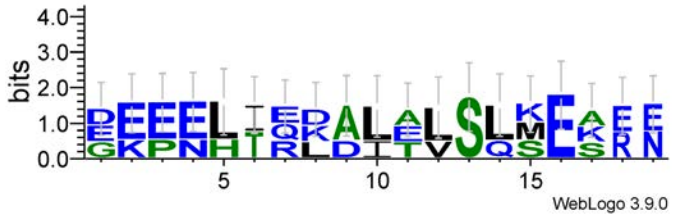
Figure 2.12: MSA showing conserved motifs of Tetraspanins (SUR7-TET8-CD63-CD81). Conserved residues are indicated by asterisks (*), colons (:) and periods (.) represent strongly and weakly similar residues and gaps are shown by dashes (-). MSA regions generated using MAFFT which correspond to the domains are shown.

No significant sequence conservation was observed for tetraspanins.

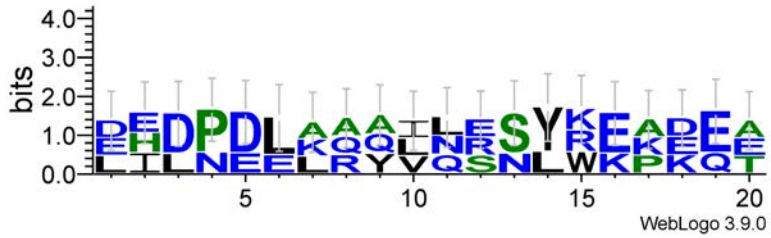
VHS:



UIM 1:



UIM 2:



FYVE:

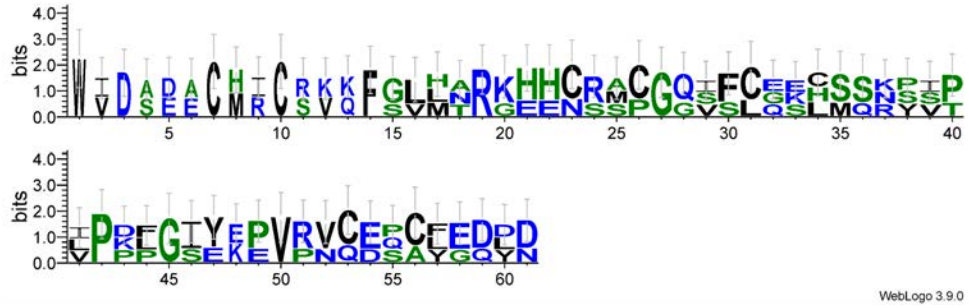


Figure 3.1: Sequence motif logos illustrating conservation of functional motifs of ESCRT-0 (VPS27-Hrs-TOL2). The height of each letter reflects residue conservation at each position. Generated using WebLogo.

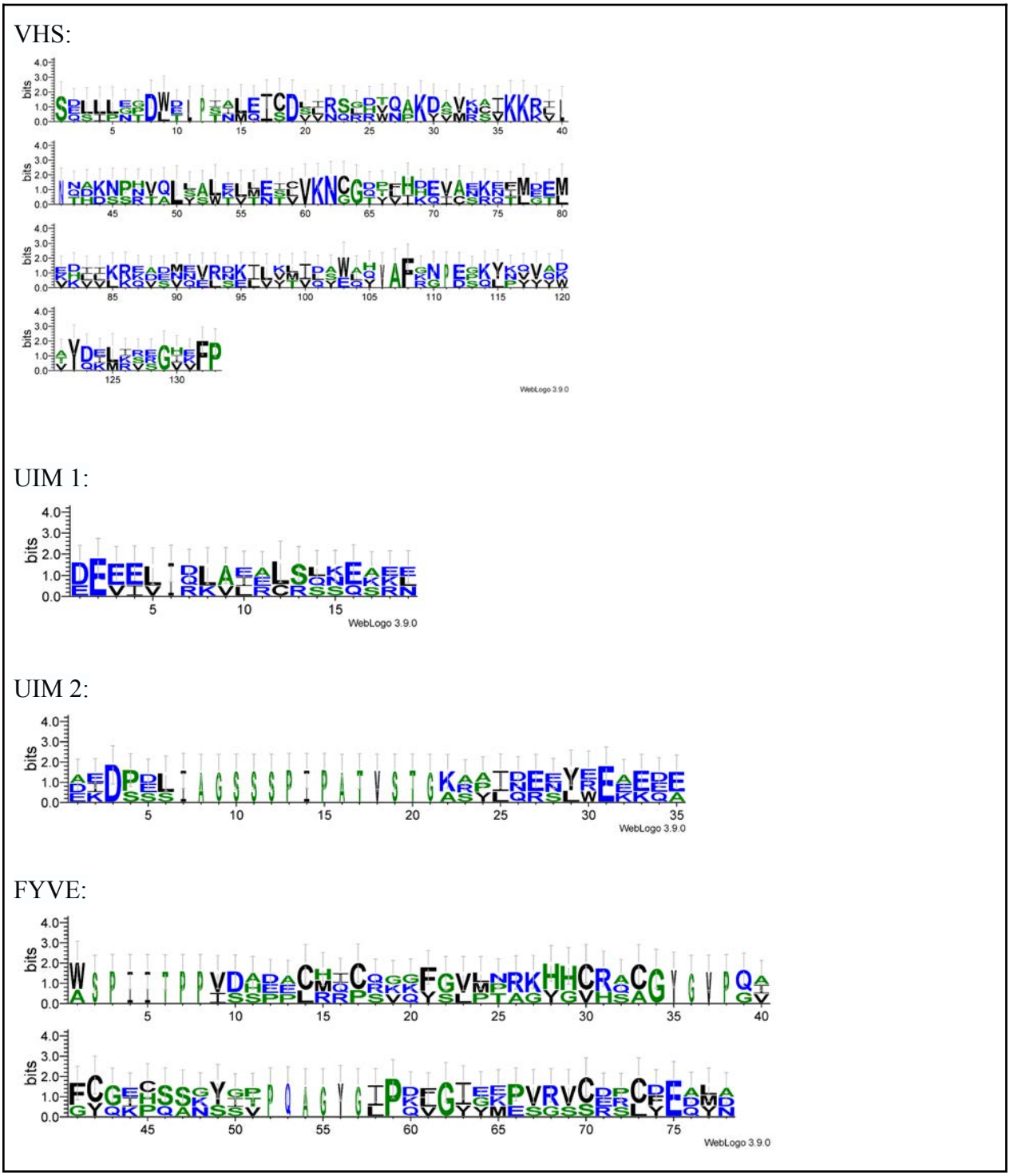


Figure 3.2: Sequence motif logos illustrating conservation of functional motifs of ESCRT-0 (VPS27-Hrs-TOL6). The height of each letter reflects residue conservation at each position. Generated using WebLogo.

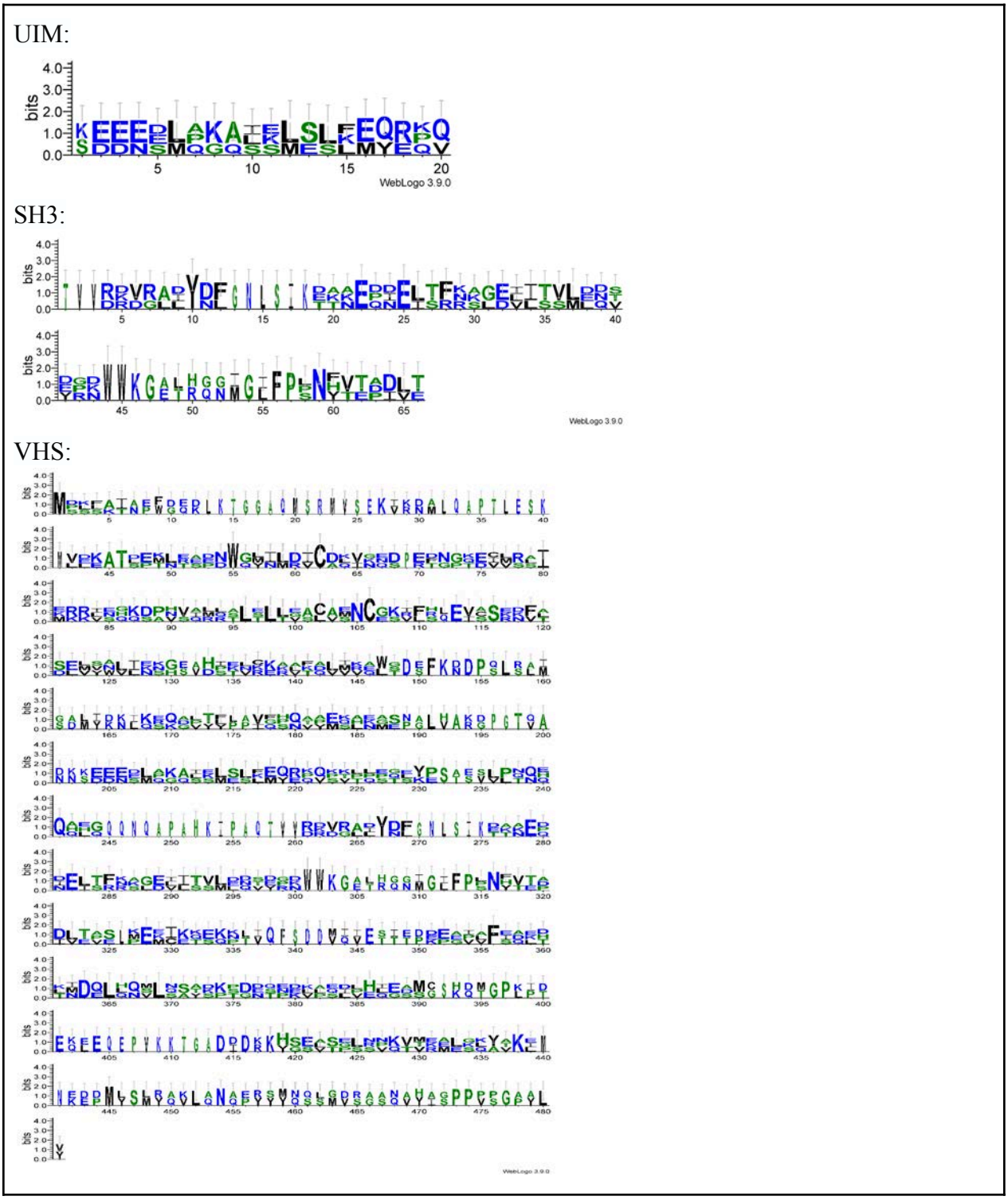
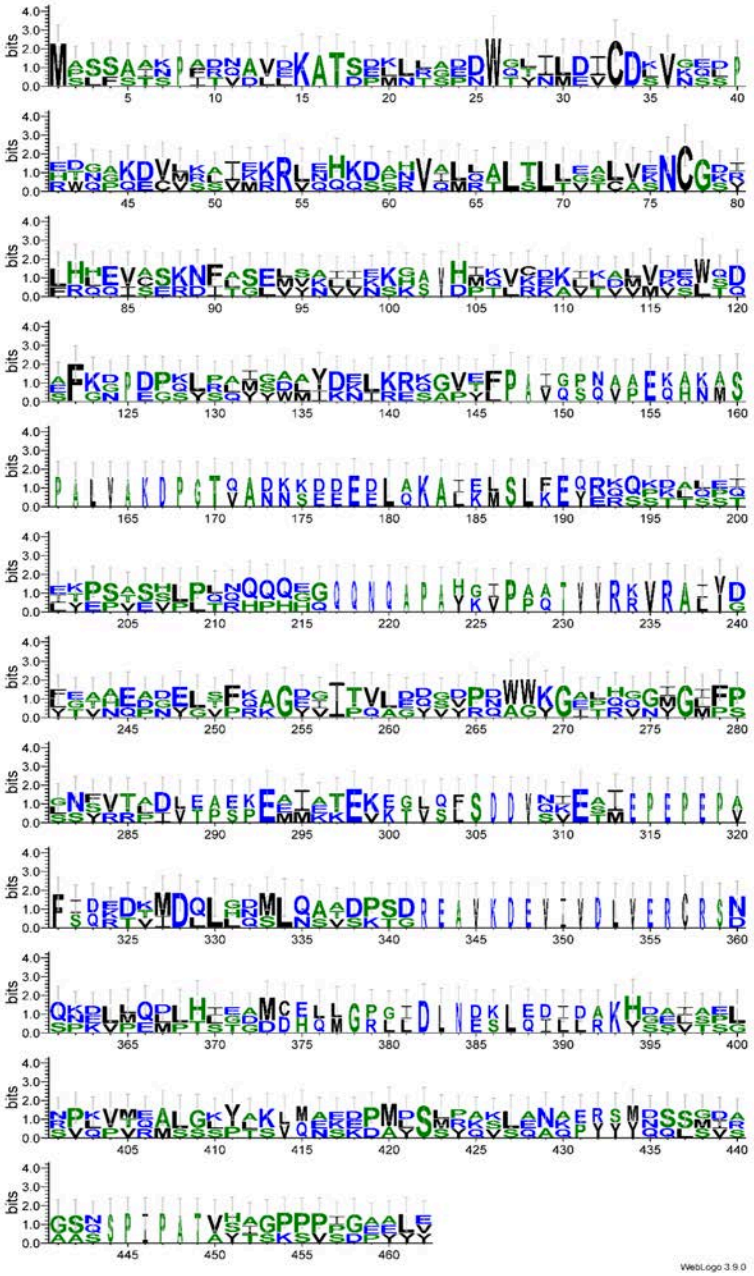
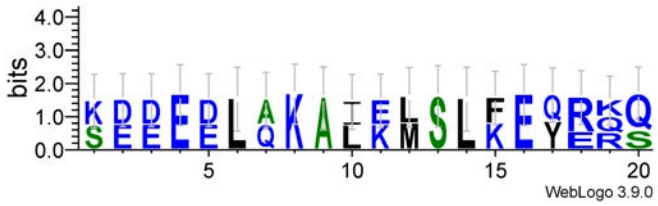


Figure 3.3: Sequence motif logos illustrating conservation of functional motifs of ESCRT-0 (Hse1-STAM-TOL2). The height of each letter reflects residue conservation at each position. Generated using WebLogo.

VHS:



UIM:



SH3:

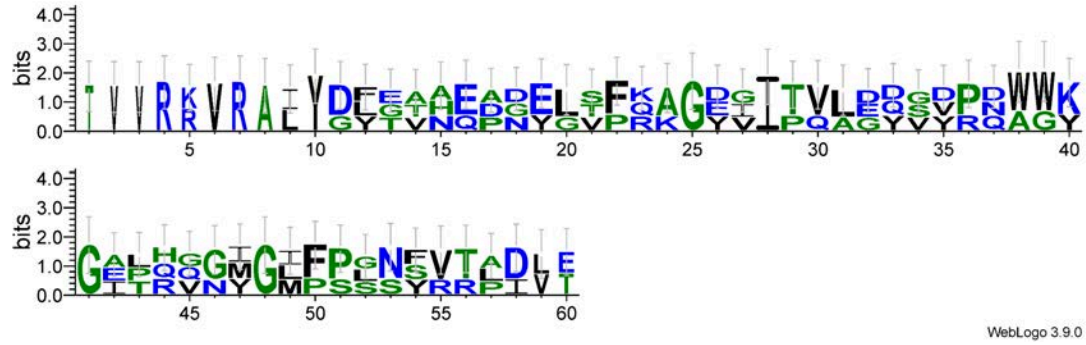
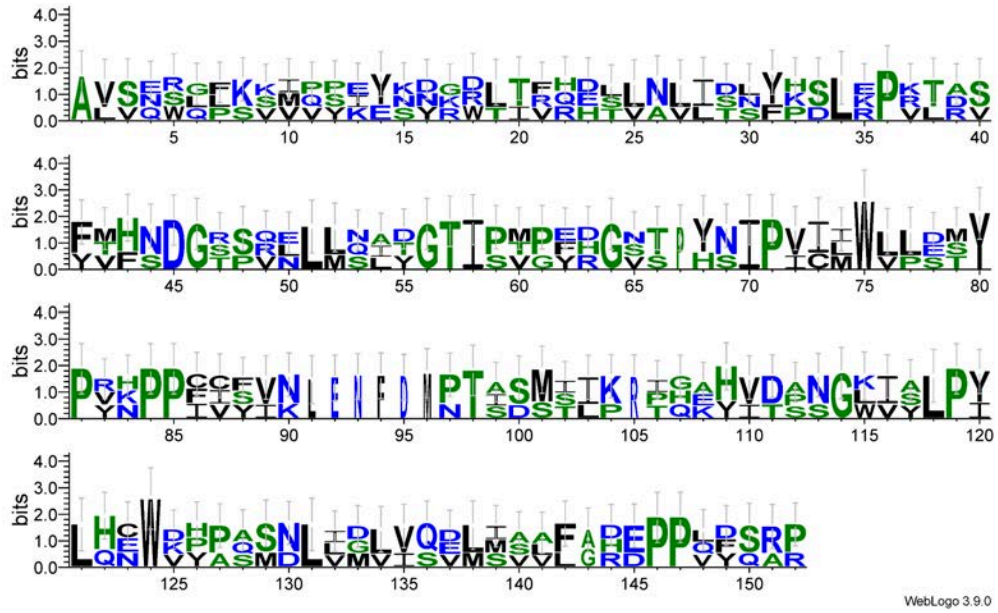
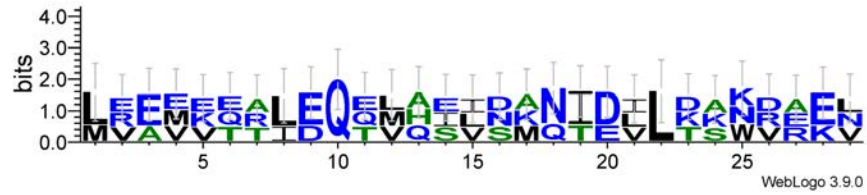


Figure 3.4: Sequence motif logos illustrating conservation of functional motifs of ESCRT-0 (Hse1-STAM-TOL6). The height of each letter reflects residue conservation at each position. Generated using WebLogo.

UEV domain (N-terminal):



Coiled Coil:



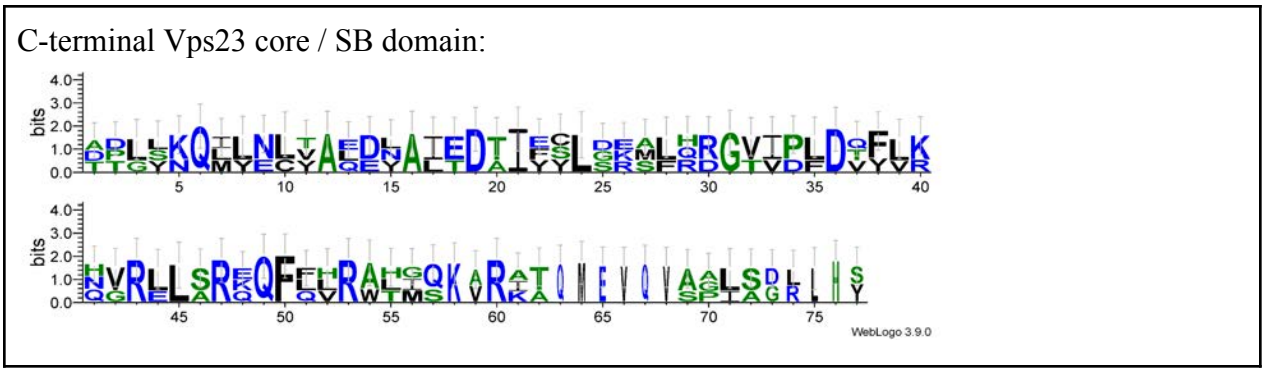


Figure 3.5: Sequence motif logos illustrating conservation of functional motifs of ESCRT-I. The height of each letter reflects residue conservation at each position. Generated using WebLogo.

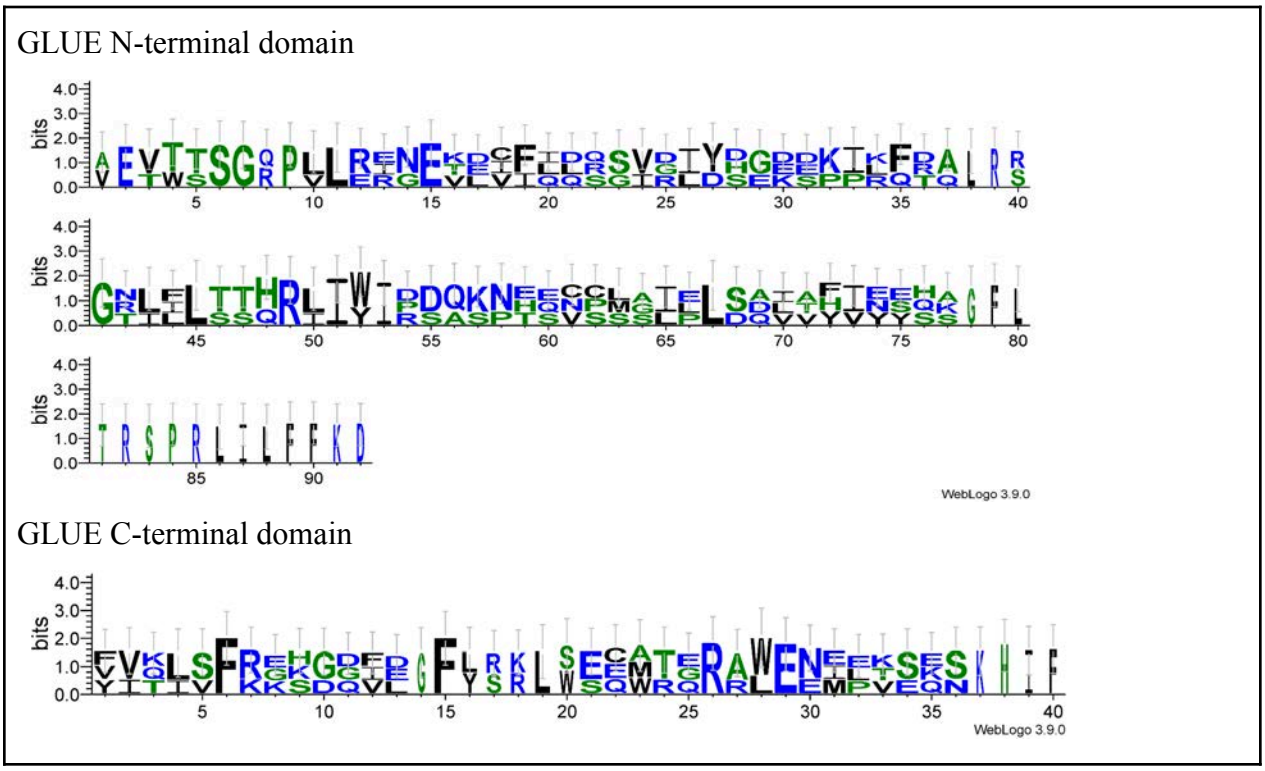


Figure 3.6: Sequence motif logos illustrating conservation of functional motifs of ESCRT-II. The height of each letter reflects residue conservation at each position. Generated using WebLogo.

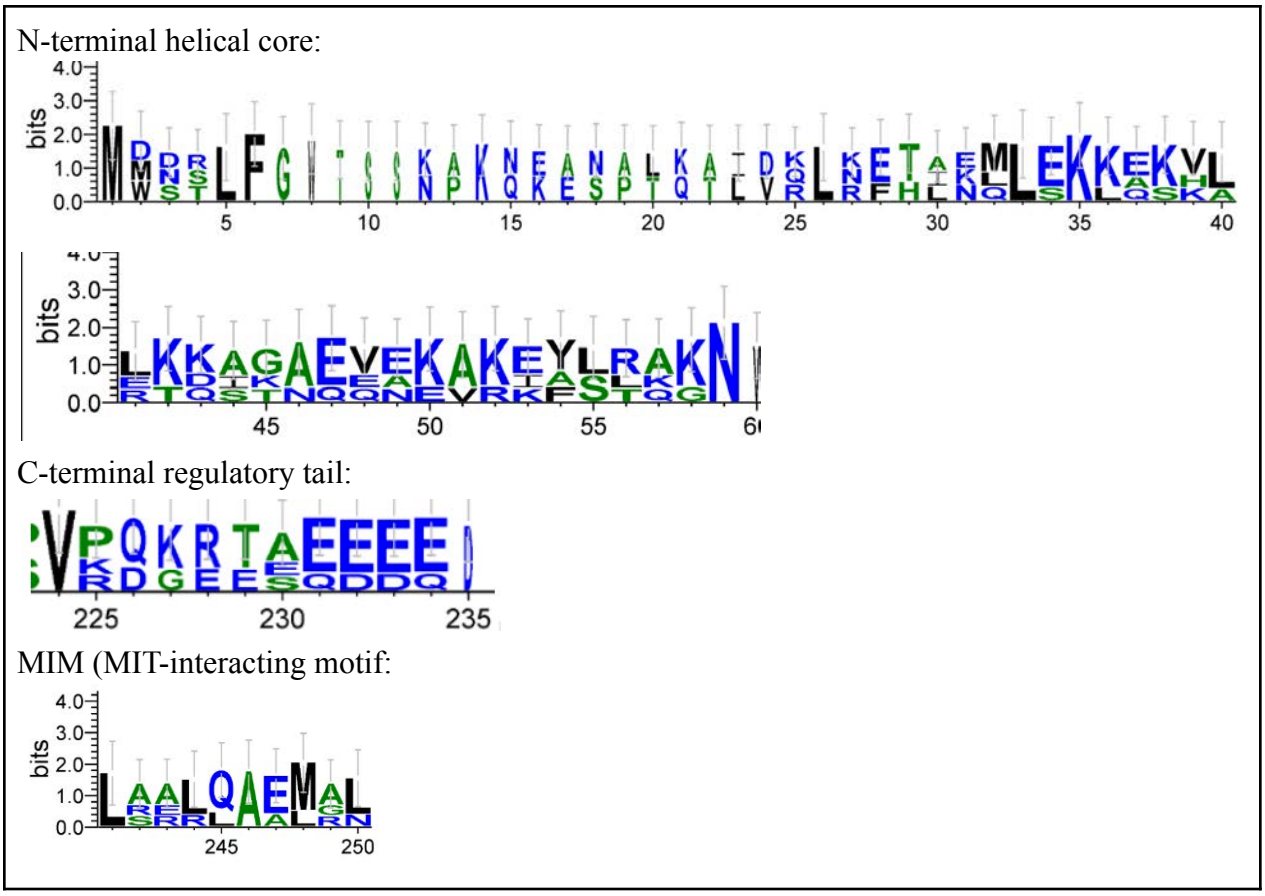
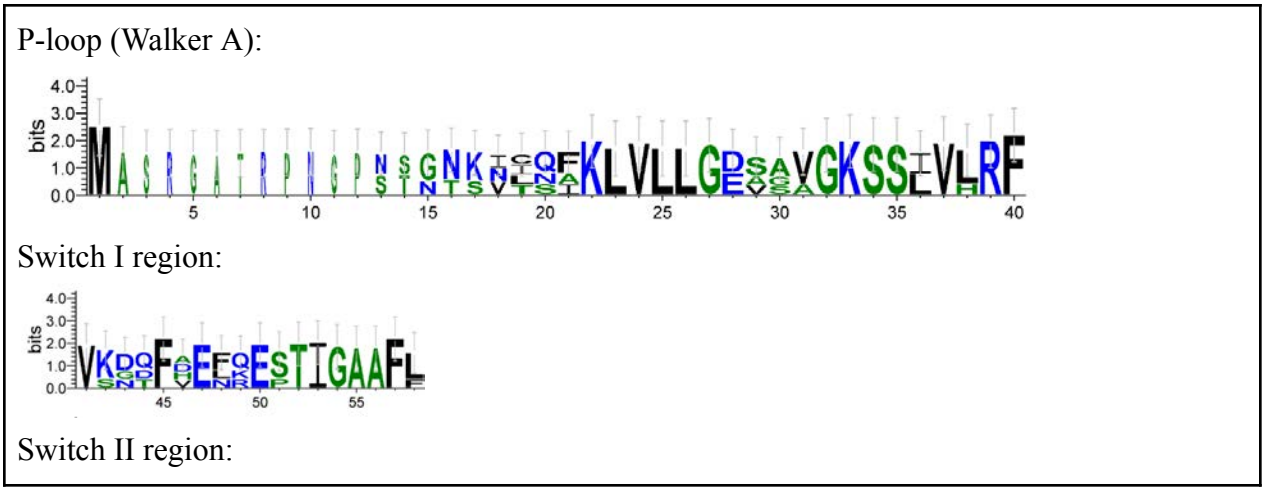


Figure 3.7: Sequence motif logos illustrating conservation of functional motifs of ESCRT-III. The height of each letter reflects residue conservation at each position. Generated using WebLogo.



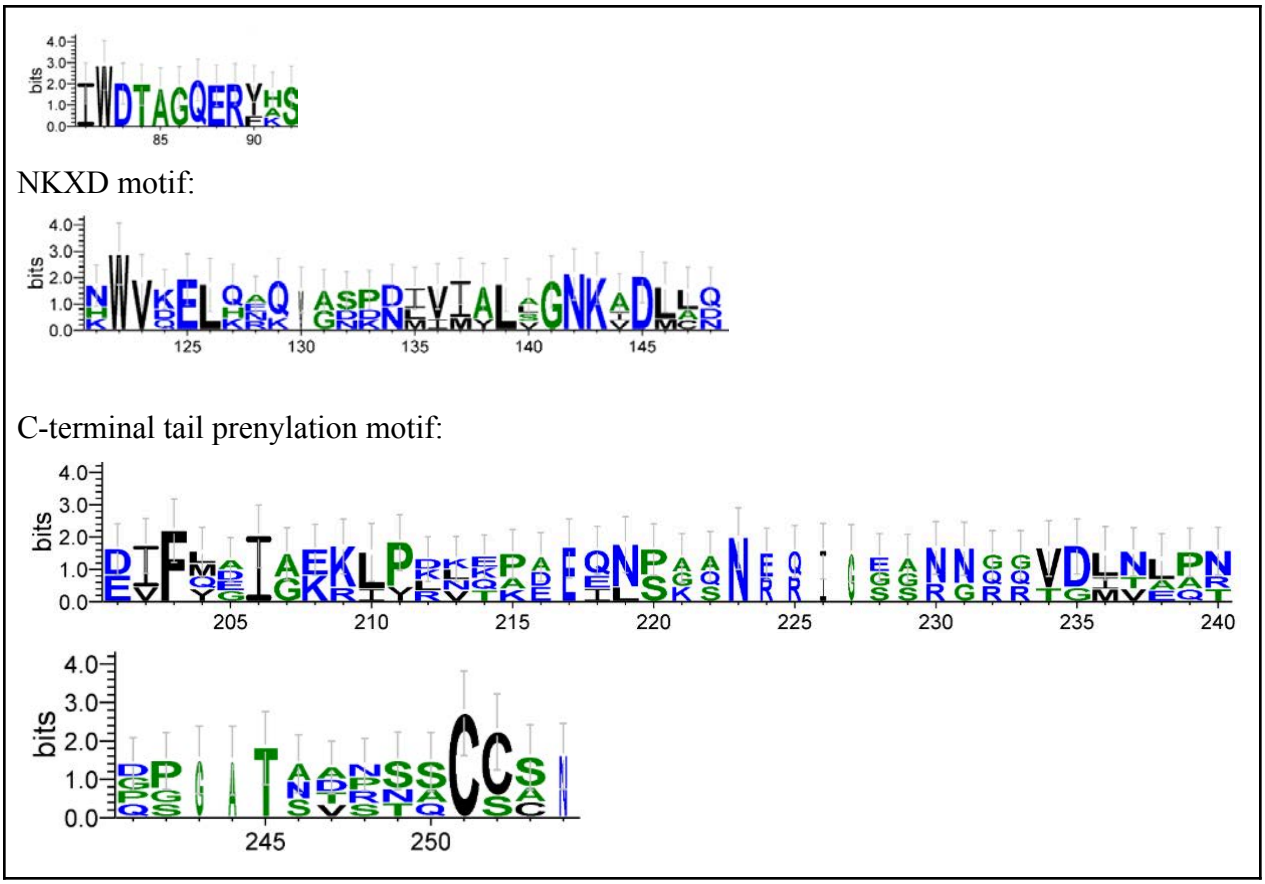
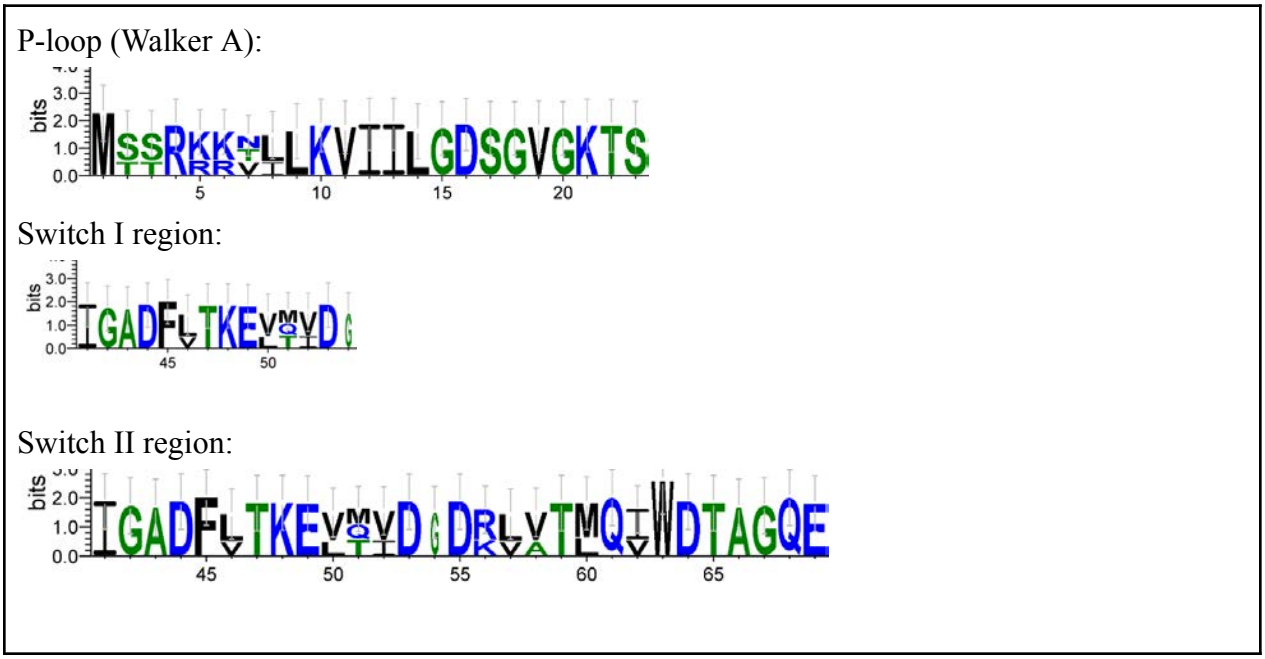


Figure 3.8: Sequence motif logos illustrating conservation of functional motifs of RAB5. The height of each letter reflects residue conservation at each position. Generated using WebLogo.



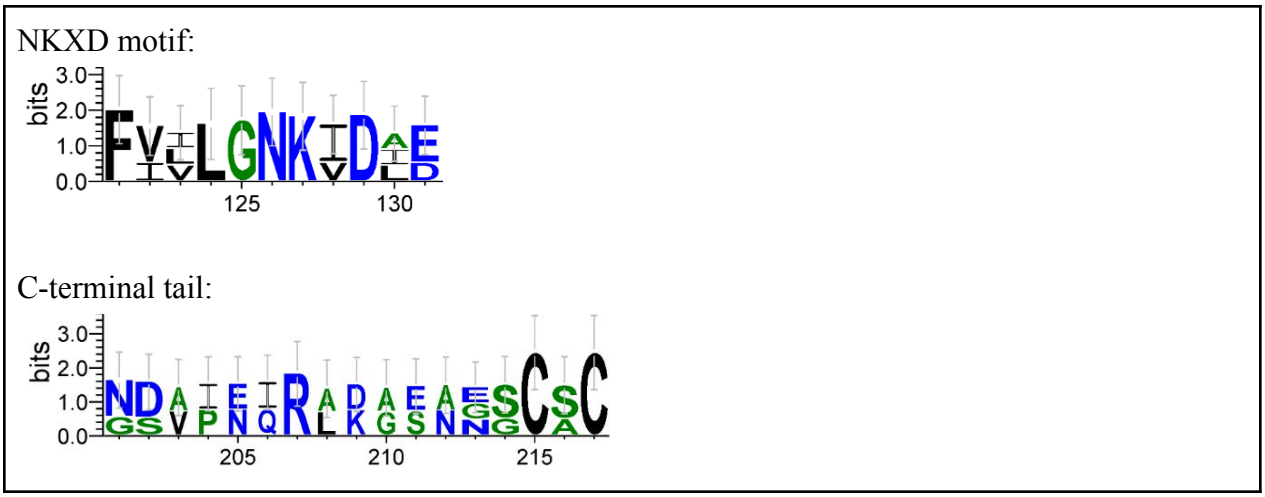


Figure 3.9: Sequence motif logos illustrating conservation of functional motifs of RAB7. The height of each letter reflects residue conservation at each position. Generated using WebLogo.

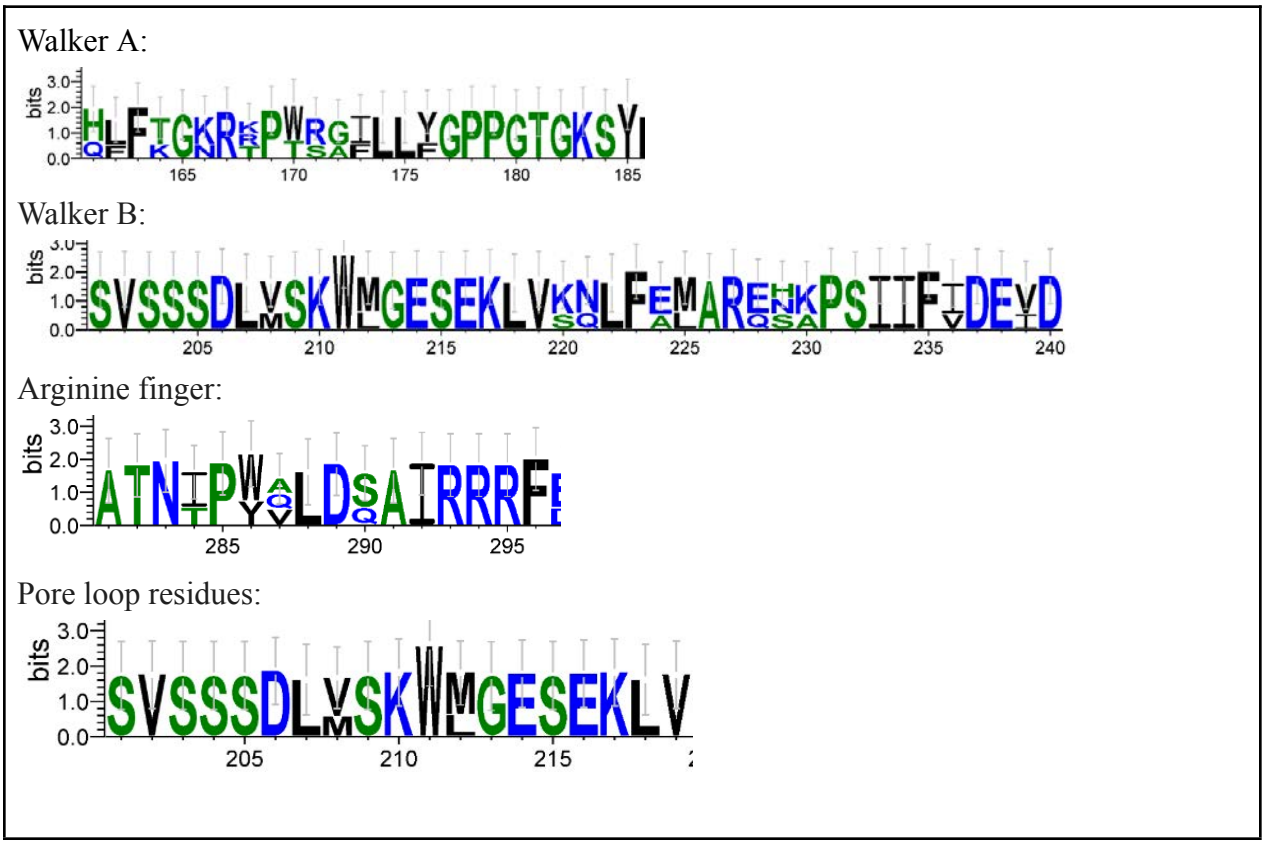
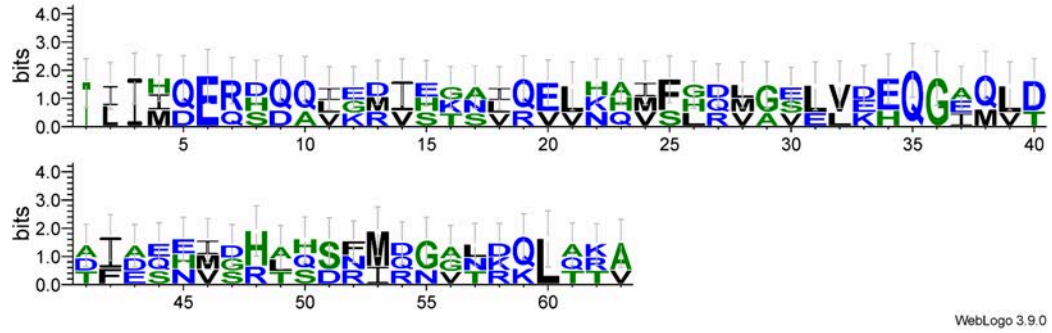


Figure 3.10: Sequence motif logos illustrating conservation of functional motifs of Disassembly. The height of each letter reflects residue conservation at each position. Generated using WebLogo.

SNARE motif (t-SNARE coiled-coil homology):



2C-terminal TM Helix:

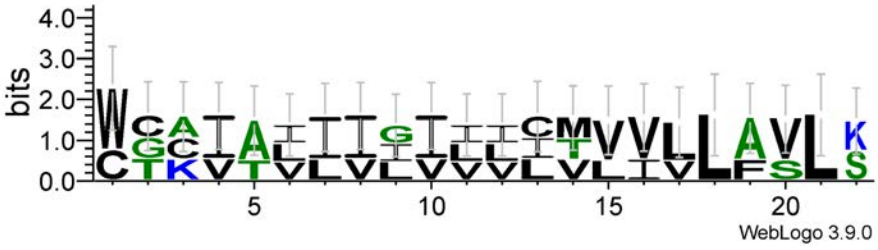
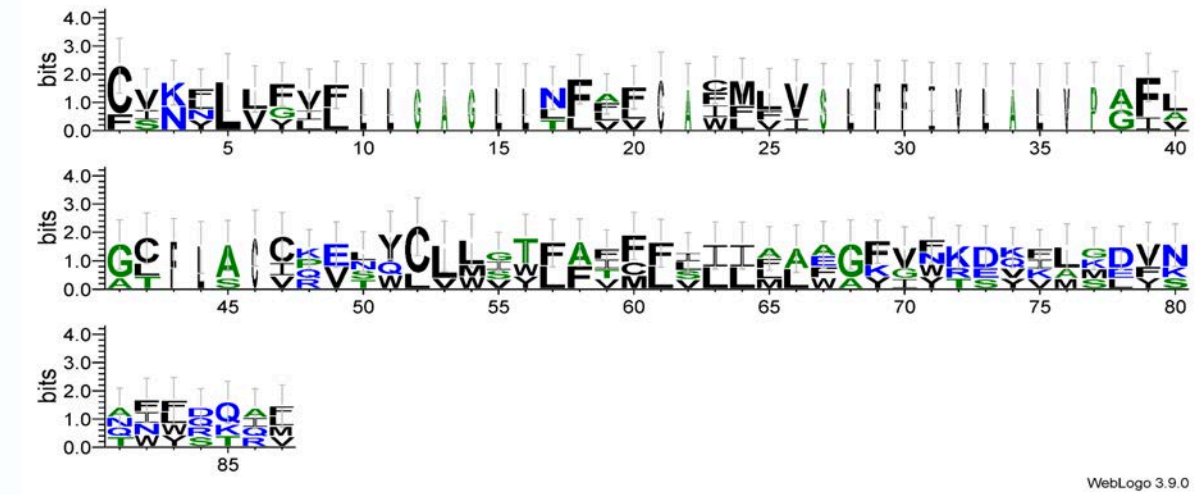


Figure 3.11: Sequence motif logos illustrating conservation of functional motifs of SNAREs. The height of each letter reflects residue conservation at each position. Generated using WebLogo.

Four transmembrane helices:



Topological Cytoplasmic tails:

EC2 Cys residues (Extracellular Topological domains):

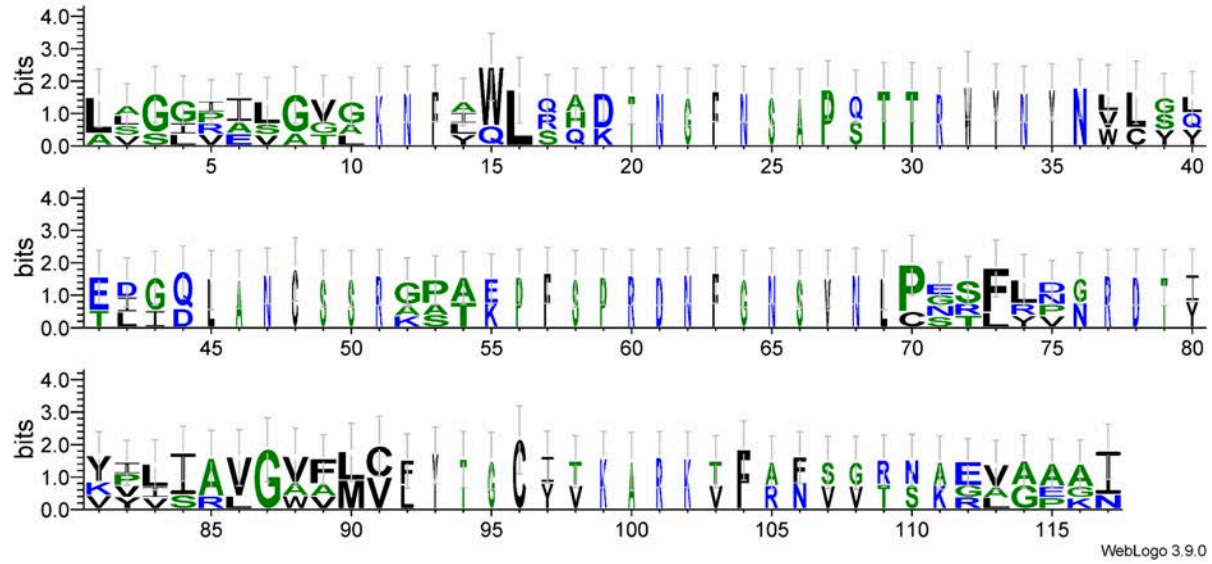


Figure 3.12: Sequence motif logos illustrating conservation of functional motifs of Tetraspanins. The height of each letter reflects residue conservation at each position. Generated using WebLogo.

3.4 Phylogenetic relationships among proteins

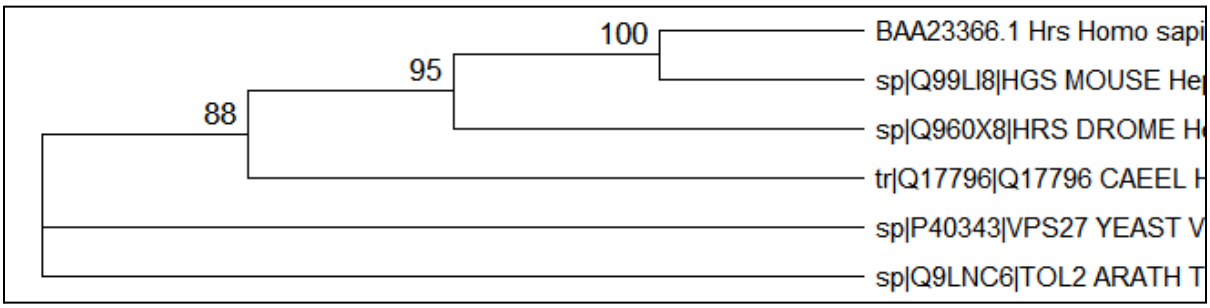


Figure 4.1: Phylogenetic trees with bootstrap support values for ESCRT-0 (Vps27-Hrs-TOL2). Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree shows that human and mouse are the most similar to each other among all the sequences with a bootstrap value of 100. Subsequently drosophila and the nematode cluster with this group with bootstrap values of 95 and 88 respectively. This clustering indicates that these proteins are close orthologs inherited from a shared ancestral protein. Then we find the yeast (Vps27) and plant (TOL2) to be the most divergent from the previous sequences and placed at the basal region of the tree. These sequences are orthologs that have diverged along their respective lineages over time, suggesting that these are more divergent.

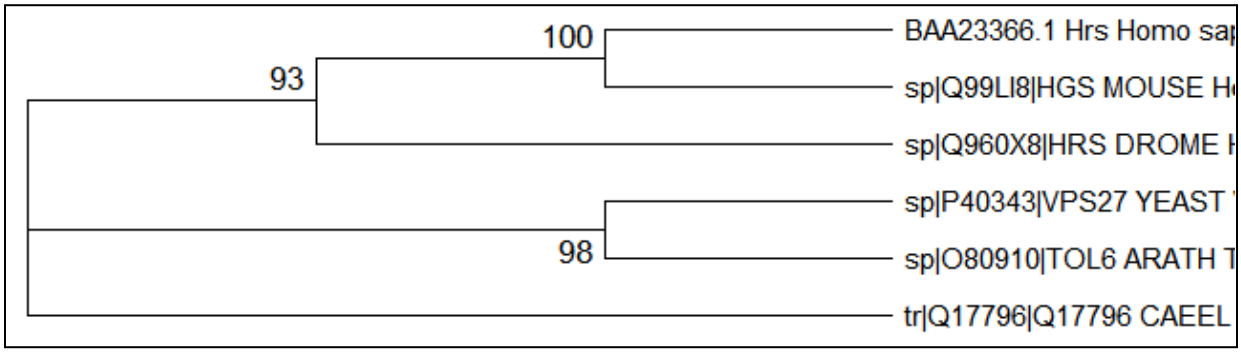


Figure 4.2: Phylogenetic trees with bootstrap support values for ESCRT-0 (Vps27-Hrs-TOL6). Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree shows that human and mouse are most similar with a bootstrap score of 100 indicating as highly reliable. Then *Drosophila* clusters with this clade with a bootstrap value of 93. In the next parallel clade, yeast and plant group together with a bootstrap value of 98, whereas the nematode stays as the most divergent lineage in the base of the tree. The fact that yeast and plant group together while the human, mouse and *Drosophila* sequences form their own cluster is a direct reflection of how these lineages have evolved separately over an evolutionary time period.

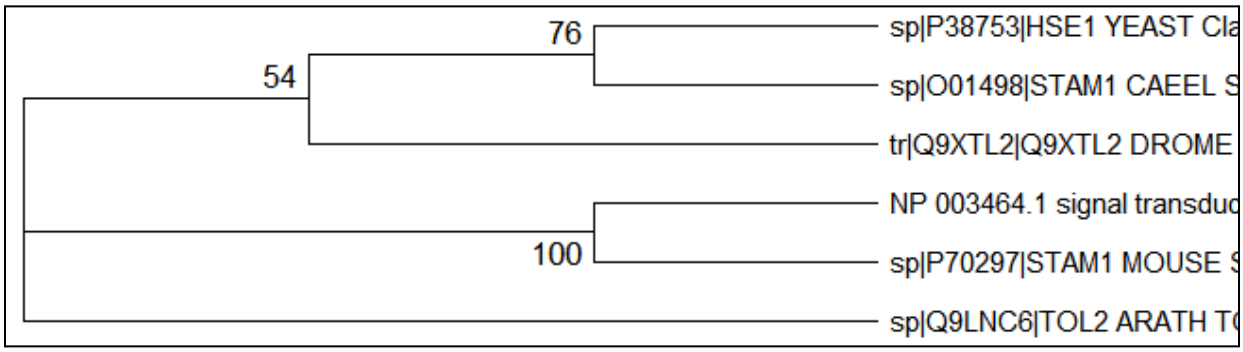


Figure 4.3: Phylogenetic trees with bootstrap support values for ESCRT-0 (Hse1-STAM-TOL2). Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree shows that yeast and nematode share a common evolutionary origin with a bootstrap value of 76. Then drosophila clusters with this clade with a bootstrap value of 54. In another clade, the human and the mouse sequence cluster perfectly with a bootstrap value of 100, where the plant stays as the most divergent lineage. The branching pattern reflects that while the human and mouse sequences have remained the most conserved, the yeast, nematode, and *Drosophila* sequences form a separate evolutionary cluster, reflecting a distinct pattern of lineage evolution and divergence over time.

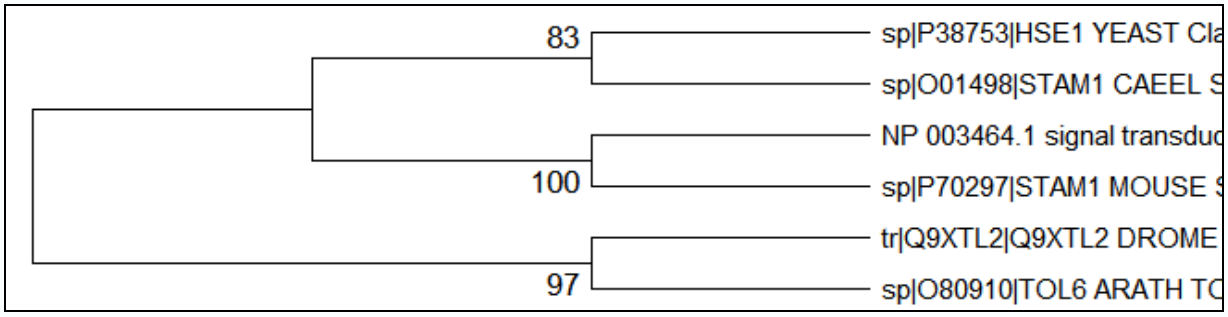


Figure 4.4: Phylogenetic trees with bootstrap support values for ESCRT-0 (Hse1-STAM-TOL6). Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree of ESCRT 0 sequences shows two primary clades where the top one showed close evolutionary relationship of yeast and nematode clustered with human and mouse both at a confidence level of 82% and 100% respectively. This clade was again clustered with the other clade which showed a close evolutionary relationship of drosophila and plant sequences at a bootstrap value of 96. This reflects that the human and mouse orthologs are the most conserved in this dataset, as shown by their perfect 100% bootstrap support. In contrast, the deeper branching of the *Drosophila* and plant lineages reflects a clear pattern of divergence over time.

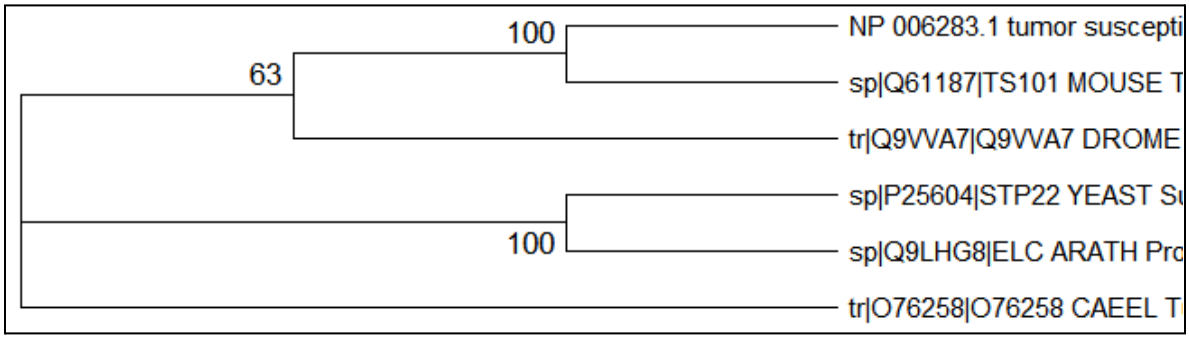


Figure 4.5: Phylogenetic trees with bootstrap support values for ESCRT-I. Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree of ESCRT I sequences shows two primary clades where the top one shows robust evolutionary relationship of human and mouse at bootstrap value of 100, which is clustered with the drosophila sequence at 63% confidence level. The other clades show a near perfect

relationship of the yeast and plant sequence whereas the nematode sequence was left as the basally divergent lineage. This pattern reflects that the human, mouse, yeast, and plant sequences are all highly conserved within their respective clades, as it is seen from their perfect 100% bootstrap support. In contrast, the branching of the *Drosophila* and nematode lineages at lower confidence levels indicates a pattern of divergence, which shows how these orthologs have evolved differently across their lineages over time.

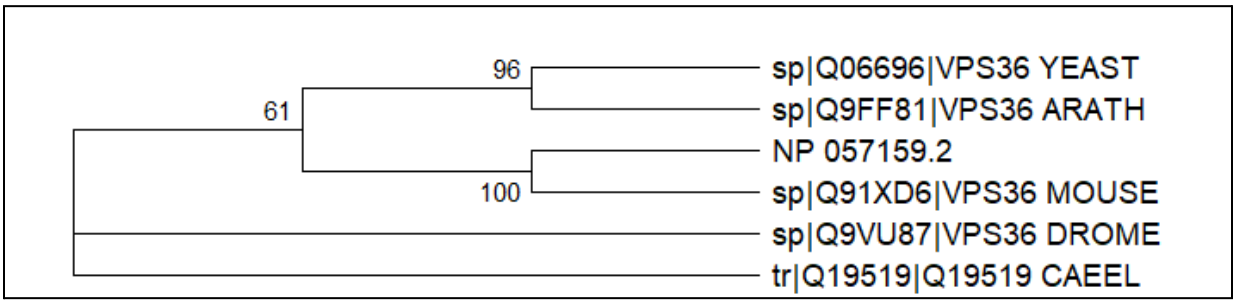


Figure 4.6: Phylogenetic trees with bootstrap support values for ESCRT-II. Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

The phylogenetic tree of ESCRT II (VPS36) sequences shows two primary clades where the top one shows the evolutionary relationship of yeast and plant sequence at 96% confidence level which is clustered with another subgroup of mouse and human. Though mouse and human share a bootstrap value of 100, their cluster shows a confidence level of only 61. Conversely the drosophila and nematode sequences represent the most divergent lineages. This pattern reflects that while the human, mouse, yeast and plant orthologs are highly conserved within their specific clades, they share a more distant common origin. In contrast, the branching of the *Drosophila* and nematode lineages at the base of the tree shows a clear pattern of divergence over time.

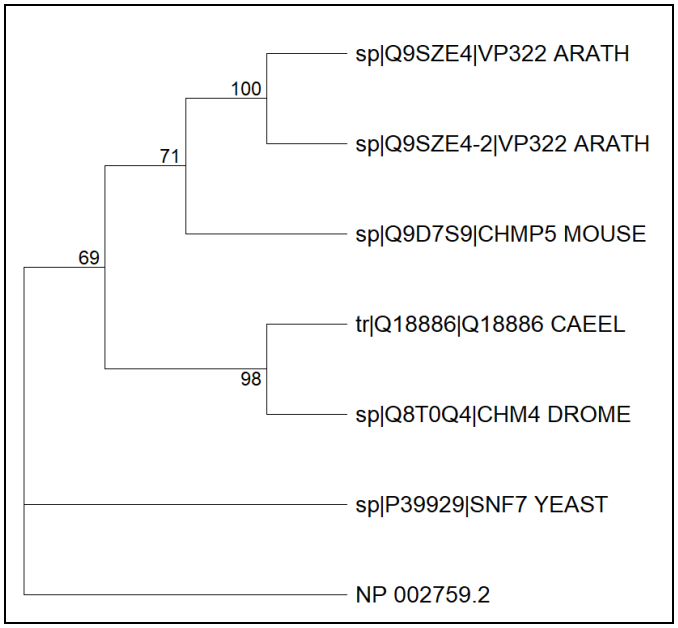


Figure 4.7: Phylogenetic trees with bootstrap support values for ESCRT-III. Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree of ESCRT III sequences shows two primary clades where the top one shows the evolutionary relationship of the two plant sequences at 100% confidence which is clustered with the mouse sequence at 71%. This clade is nested with the other clade of nematode and drosophila at confidence level 69%, even though the bootstrap value within this subgroup was 98. This reveals that the plant,drosophila and nematode subgroups show strong conservation within their lineages. And finally the deeper branching of yeast and human shows significant lineage evolution and divergence over time.

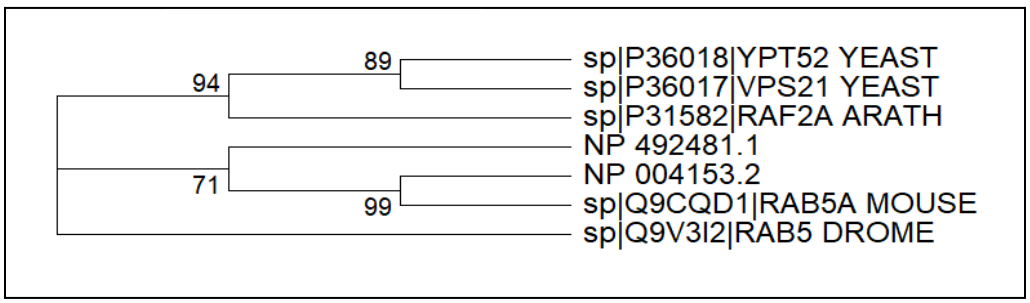


Figure 4.8: Phylogenetic trees with bootstrap support values for RAB5. Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode

sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree of RAB5, YPT52 and VPS21 sequences shows two primary clades where the top one shows the evolutionary relationship of the two yeast sequences at bootstrap value of 89 which is clustered with the plant sequence at 94% confidence. The second clade shows near perfect evolutionary relationship of the mouse and human sequences nested with the nematode sequence at 71% confidence. The tree structure indicates that these sequences may be orthologs that have been conserved across yeast, plant and animals. However, the tight clustering of the human, mouse and insect sequences compared to the yeast and plant clades clearly shows a distinct pattern of lineage evolution and divergence.

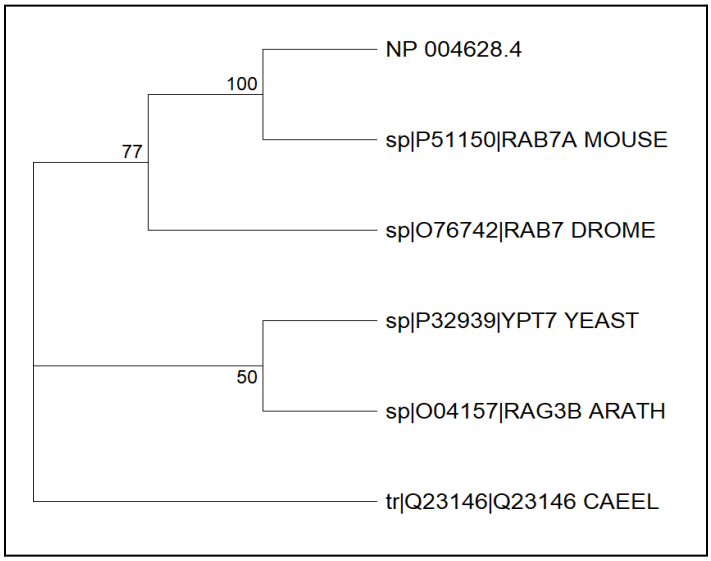


Figure 4.9: Phylogenetic trees with bootstrap support values for RAB7. Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree of RAB7 sequences shows two primary clades where the top one shows the evolutionary relationship of the human and mouse at bootstrap value of 100 which is nested with the drosophila sequence at 77% confidence. In the next clade the yeast and plant sequence showed evolutionary relevance at 50% confidence. This pattern shows the human, mouse and drosophila sequences are highly conserved within their clade. In contrast, the clustering of the

yeast and plant lineage along the distant branch shows a clear pattern of divergence across these orthologs.

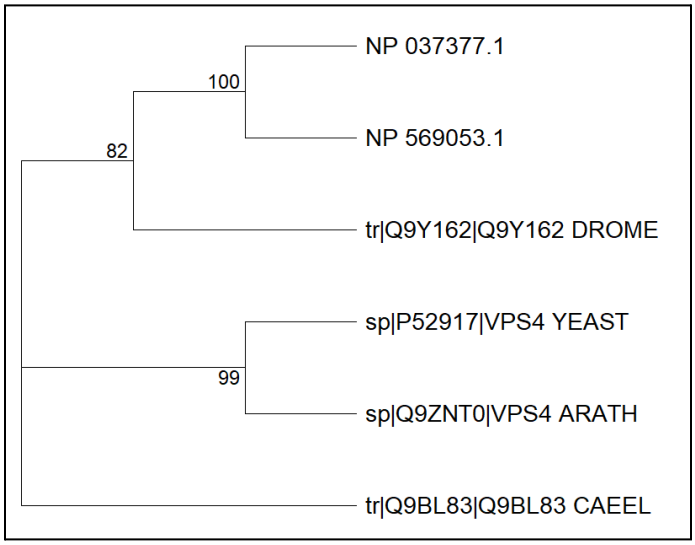


Figure 4.10: Phylogenetic trees with bootstrap support values for Disassembly. Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree of VPS4 sequences shows two primary clades where the top one shows the evolutionary relationship of the human and mouse sequences at bootstrap value of 100 which is nested with the drosophila sequence at 82% confidence. In the next clade the yeast and plant sequence showed evolutionary relevance at 99% confidence. The tight clustering of the human, drosophila, yeast, and plant pairs indicates that these sequences may be orthologs that have remained conserved within their respective clades. But the distinct separation between the clades reflects the divergence over time as these lineages evolved from a common ancestor.

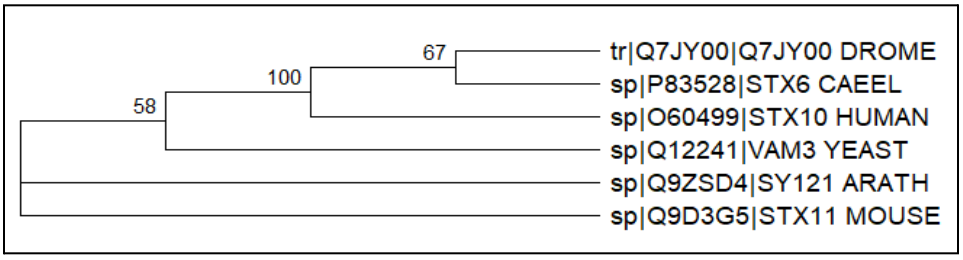


Figure 4.11: Phylogenetic trees with bootstrap support values for SNAREs. Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode

sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree of SNAREs sequences shows two primary clades where the top one shows the evolutionary relationship of the drosophila and nematode sequence at a confidence level of 67% which is nested with the human sequence at bootstrap value of 100. This again was clustered with the yeast sequence at 58% confidence level. These results confirm that these sequences may be orthologs originating from a shared ancestral SNARE gene. However the varying bootstrap values across the tree highlight a pattern of lineage evolution, where the yeast, plant, and mouse sequences show significant divergence compared to the more closely clustered drosophila, nematode, and human group.

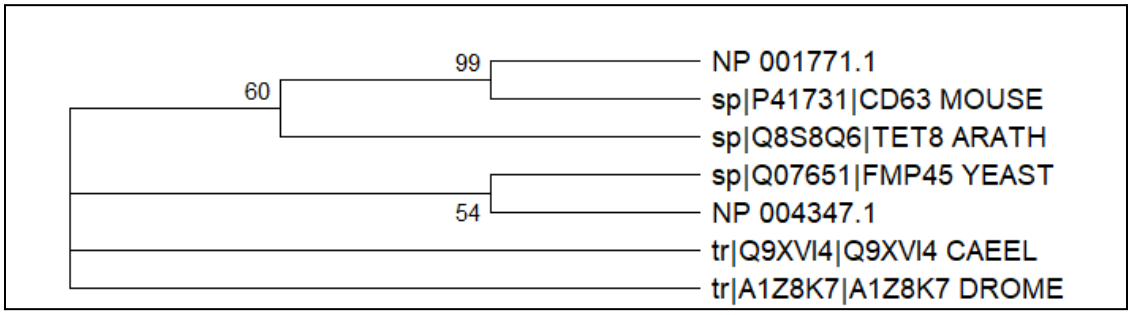


Figure 4.12: Phylogenetic trees with bootstrap support values for Tetraspanins. Branches show evolutionary relationships with bootstrap values at each node. Mouse, *Drosophila* and nematode sequences were included only to meet MEGA's sequence requirements and are not part of study focus.

This phylogenetic tree of Tetraspanin sequences shows a primary clade where the top one shows the evolutionary relationship of the human (CD63) and mouse sequence at a confidence level of 99%. This is clustered with the plant sequence at 60% confidence. The next clade shows the evolutionary relationship with yeast and human (CD81) at bootstrap value of 54. The clustering pattern confirms that these sequences may be orthologs that have diverged from a shared ancestral gene while maintaining their core structural characteristics across species. The higher bootstrap values for the human, mouse, and plant group compared to the yeast and insect branches reflects their distinct lineage evolution and divergence over time.

3.5 Structural conservation across species

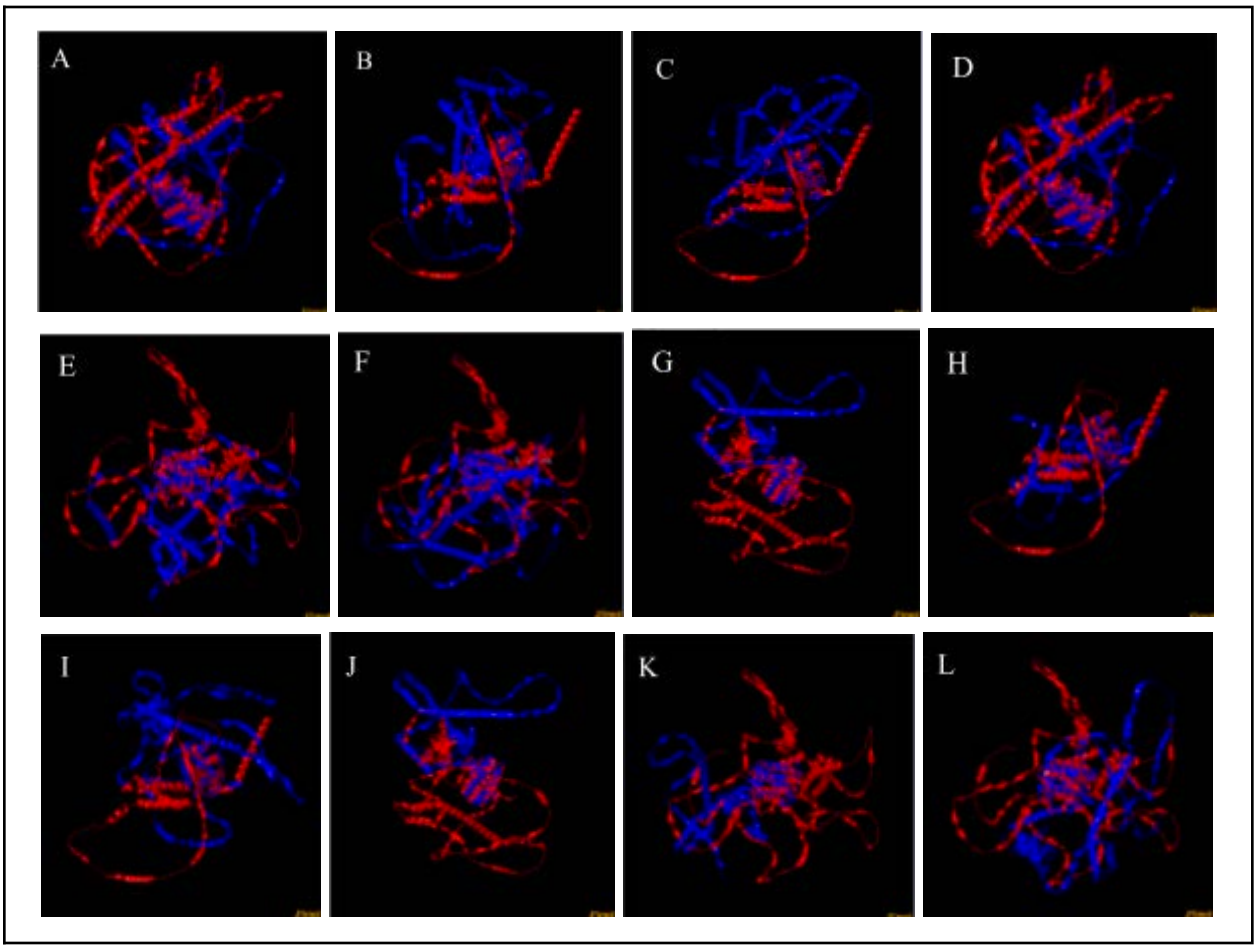


Figure 5.1: Structural overlays of ESCRT-0 proteins (**A**: Yeast_Vps27 vs Human_Hrs; **B**: Yeast_Vps27 vs Plant_TOL2; **C**: Human_Hrs vs Plant_TOL2 ; **D**: Yeast_Vps27 vs Human_Hrs; **E**: Yeast_Vps27 vs Plant_TOL6; **F**: Human_Hrs vs Plant_TOL6; **G**: Yeast_Hse1 vs Human_STAM; **H**: Yeast_Hse1 vs Plant_TOL2; **I**: Human_STAM vs Plant_TOL2; **J**: Yeast_Hse1 vs Human_STAM; **K**: Yeast_Hse1 vs Plant_TOL6; **L**: Human_STAM vs Plant_TOL6). Superimposed structures (blue for protein 1 and red for protein 2) generated using TM-align web server illustrate varying degrees of structural conservation.

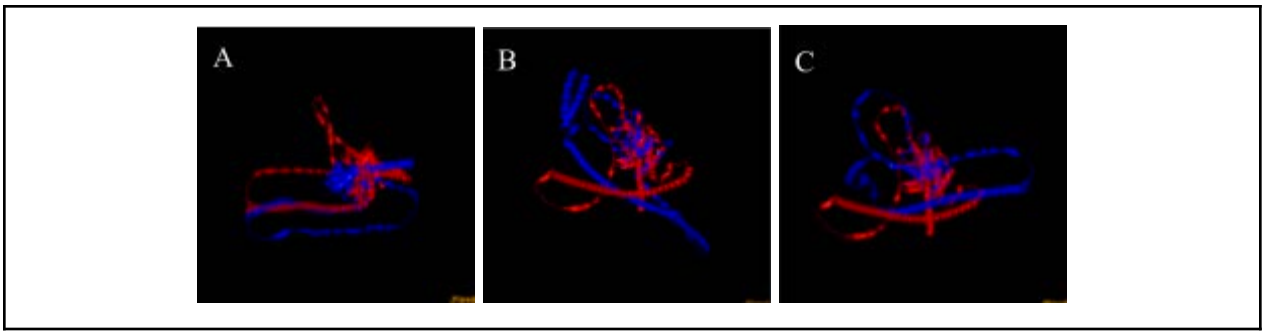


Figure 5.2: Structural overlays of ESCRT-I proteins (**A**: Yeast_Vps23 vs Human_TGS1; **B**: Yeast_Vps23 vs Plant_VPS23; **C**: Human_TGS101 vs Plant_VPS23). Superimposed structures (blue for protein 1 and red for protein 2) generated using TM-align web server illustrate varying degrees of structural conservation.

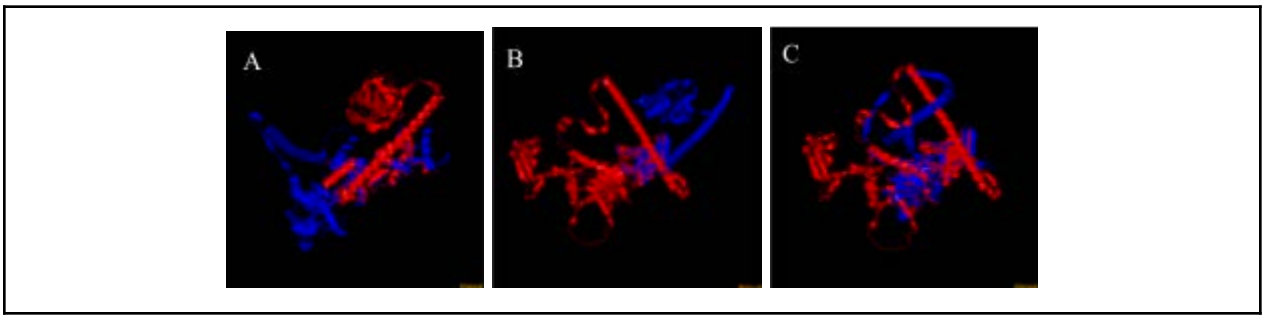


Figure 5.3: Structural overlays of ESCRT-II proteins (**A**: Plant_VPS36 vs Human_VPS36; **B**: Human_VPS36 vs Yeast_Vps36; **C**: Plant_VPS36 vs Yeast_Vps36). Superimposed structures (blue for protein 1 and red for protein 2) generated using TM-align web server illustrate varying degrees of structural conservation.

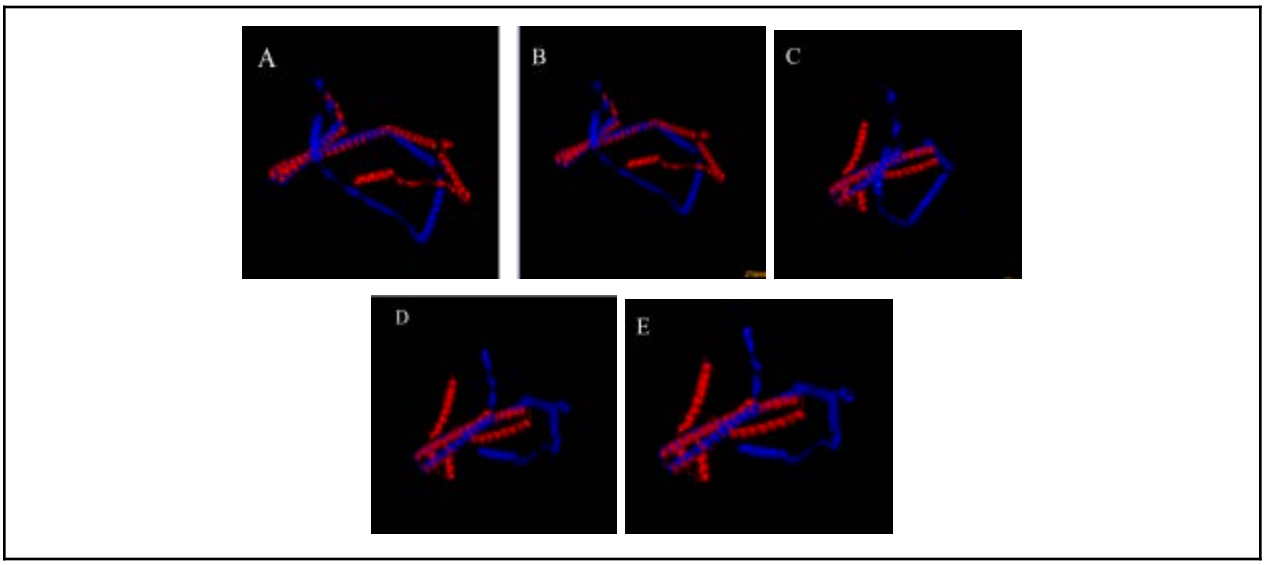


Figure 5.4: Structural overlays of ESCRT-III proteins (**A**: Yeast_Snf7 vs Plant_Snf7; **B**: Yeast_Snf7 vs Plant_Snf7.1; **C**: Yeast_Snf7 vs Human_CHMP1A; **D**: Plant_Snf7 vs Human_CHMP1A; **E**: Plant_Snf7.1 vs Human_CHMP1A). Superimposed structures (blue for protein 1 and red for protein 2) generated using TM-align web server illustrate varying degrees of structural conservation.

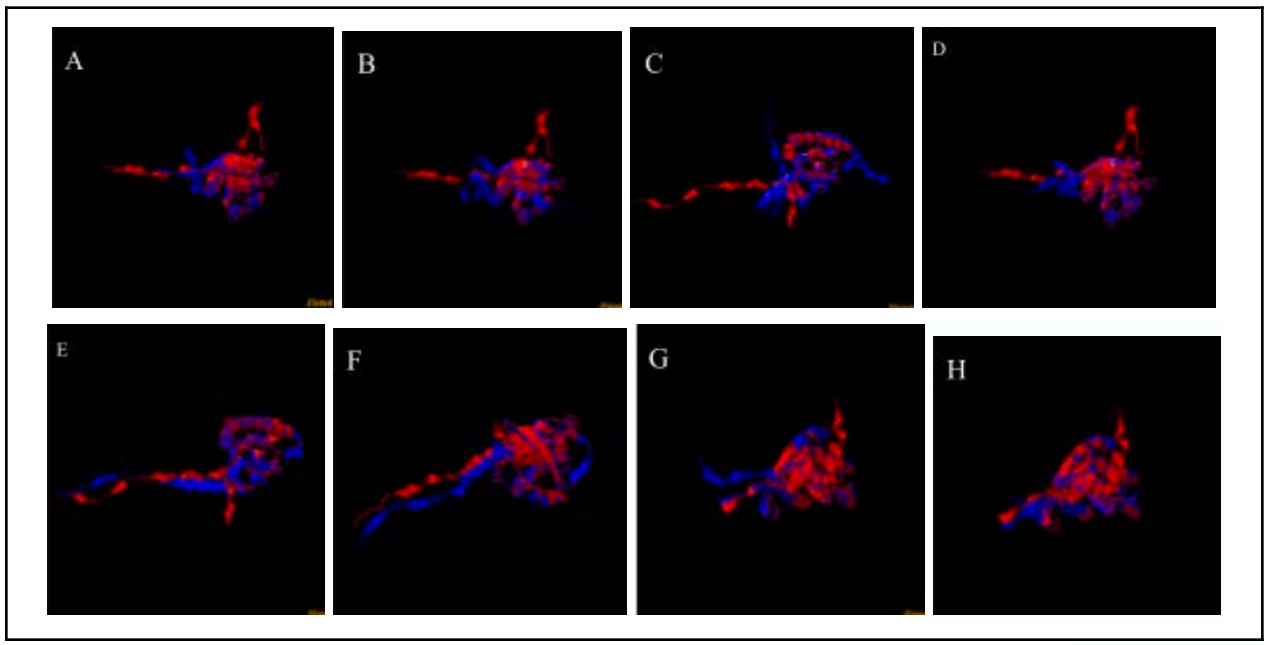


Figure 5.5: Structural overlays of RAB5 (**A**: Human_RAB5 vs Plant_RAB5A; **B**: Human_RAB5 vs Yeast_ypt52; **C**: Yeast_Ypt52 vs Plant_RAB5A; **D**: Yeast_Vps21 vs Human_RAB5; **E**: Yeast_Vps21 vs Plant_RAB5A; **F**: Yeast_Ypt7 vs Plant_RAB7; **G**: Yeast_Ypt7 vs Human_RAB7; **H**: Human_RAB7 vs Plant_RAB7). Superimposed structures (blue for protein 1 and red for protein 2) generated using TM-align web server illustrate varying degrees of structural conservation.

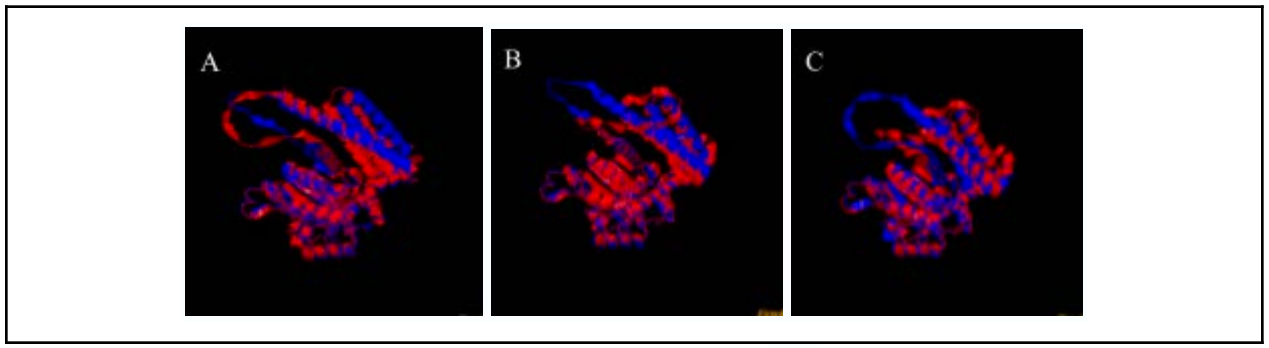


Figure 5.6: Structural overlays of Disassembly proteins (**A**: Yeast_Vps4 vs Human_VPS4l **B**: Yeast_VPS4 vs Plant_VPS4; **C**: Human_VPS4 vs Plant_VPS4). Superimposed structures (blue

for protein 1 and red for protein 2) generated using TM-align web server illustrate varying degrees of structural conservation.

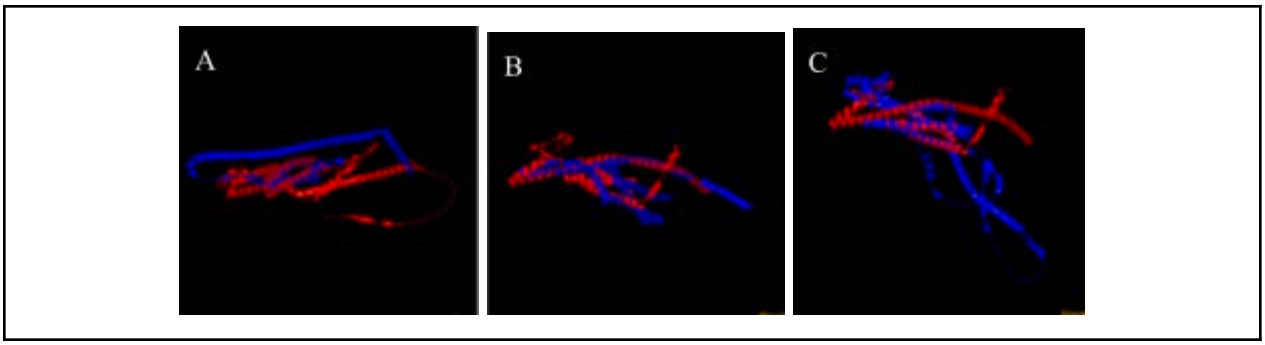


Figure 5.7: Structural overlays of SNAREs proteins (**A**: Human_STX10 vs Plant_SYP121; **B**: Human_STX10 vs Yeast_Vam3; **C**: Yeast_Vam3 vs Plant_SYP121). Superimposed structures (blue for protein 1 and red for protein 2) generated using TM-align web server illustrate varying degrees of structural conservation.

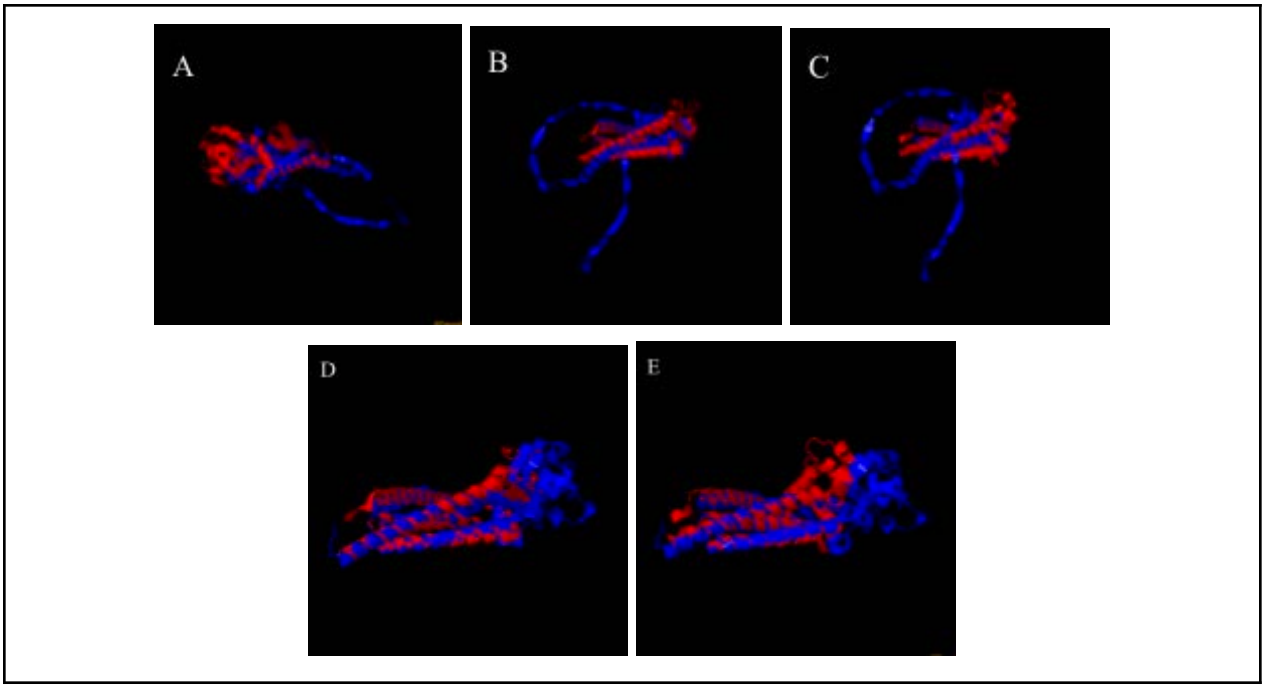


Figure 5.8: Structural overlays of Tetraspanins proteins (**A**: Yeast_SUR7 vs Plant_TET8; **B**: Yeast_SUR7 vs Human_CD63; **C**: Yeast_SUR7 vs Human_CD81; **D**: Plant_TET8 vs Human_CD63; **E**: Plant_TET8 vs Human_CD81). Superimposed structures (blue for protein 1 and red for protein 2) generated using TM-align web server illustrate varying degrees of structural conservation.

Protein family	Species	Aligned length	RMSD	TM-score	Interpretation
ESCRT-0	Yeast (Vps27)-Human (Hrs)	259	6.23	0.296	Low structural similarity
	Yeast (Vps27)-Plant (TOL2)	193	5.26	0.313	Partial structural similarity
	Human (Hrs)-Plant (TOL2)	177	4.38	0.30	Partial structural similarity
	Yeast (Vps27)-Human (Hrs)	259	6.23	0.296	Low structural similarity
	Yeast (Vps27)-Plant (TOL6)	226	5.32	0.283	Low structural similarity
	Human (Hrs)-Plant (TOL6)	225	5.56	0.27	Low structural similarity
	Yeast (Hse1)-Human (STAM)	175	4.80	0.29	Low Structural Similarity
	Yeast (Hse1)-Plant (TOL2)	150	3.69	0.32	Partial Structural Similarity
	Human (STAM)--Plant (TOL2)	179	4.19	0.34	Partial Structural Similarity
	Yeast (Hse1)-Human (STAM)	175	4.80	0.29	Low structural similarity
	Yeast (Hse1)-Plant (TOL6)	203	5.04	0.31	Partial structural similarity
	Human (STAM)--Plant (TOL6)	185	4.60	0.265	Low structural similarity

ESCRT I	Yeast (Stp22)-Human (Tsg101):	200	5.89	0.36	Partial Structural Similarity
	Yeast (Stp22)-Plant (ELC):	158	3.31	0.36	Partial Structural Similarity
	Human (Tsg101)-Plant (ELC):	171	2.89	0.4	Partial Structural Similarity
ESCRT II	Yeast (VPS36)-Human (VPS36)	301	3.81	0.53	Similar overall fold
	Yeast (VPS36)-Plant (VPS36)	167	3.65	0.32	Partial structural similarity
	Plant (VPS36)-Human (VPS36)	150	1.9	0.35	Partial structural similarity
ESCRT III	Yeast (Snf7)-Plant (Snf7)	148	3.81	0.5	Partial structural similarity
	Yeast (Snf7)-Plant (Snf7.1)	148	3.81	0.5	Partial structural similarity
	Plant (Snf7.1)-Human (CHMP1A)	109	2.58	0.46	Partial structural similarity
	Plant (Snf7.1)-Human (CHMP1A)	109	2.58	0.46	Partial structural similarity
	Yeast (Snf7)-Human (CHMP1A)	103	2.11	0.42	Partial structural similarity
RAB5	Yeast (Ypt52)-Plant (Rab5A)	174	1.47	0.77	Similar overall fold and highly conserved.

	Yeast (VPS21)-Plant (RAB5A)	194	1.98	0.875	Similar overall fold and highly conserved.
	Yeast (Ypt51)-Human (RAB5)	177	1.04	0.83	Similar overall fold and highly conserved.
	Yeast (Ypt52)-Human (RAB5)	172	1.35	0.71	Similar overall fold and highly conserved.
	Plant (RAB5A)-Human (RAB5)	176	1.69	0.79	Similar overall fold and highly conserved.
RAB7	Yeast (Ypt7)-Human (RAB7)	202	2.19	0.89-0.89 9	Similar overall fold and can be considered as conserved.
	Yeast (Ypt7)-Plant (RAB7A)	189	1.24	0.89	Similar overall fold and highly conserved.
	Plant (RAB7A)-Human (RAB7)	188	1.61	0.87-0.86	Similar overall fold and highly conserved.
Disassembly	Yeast (VPS4)-Human (VPS4)	410	2.29	0.87	Similar overall fold and partial structural conservation.
	Yeast (VPS4)-Plant (VPS4)	429	3.3	0.87	Similar overall fold and partial structural conservation.
	Plant (VPS4)-Human (VPS4)	410	2.79	0.89	Similar overall fold and partial structural conservation.
SNAREs	Yeast (VAM3)-Human (STX10)	148	4.73	0.36	Partial structural similarity
	Yeast (VAM3)-Plant (SYP121)	129	4.75	0.36	Partial structural similarity

	Plant (SYP121)-Human (STX10)	149	4.5	0.4	Partial structural similarity
Tetraspanin	Plant (TET8)-Human (CD63)	145	4.33	0.38	Partial structural similarity
	Plant (TET8)-Human (CD81)	148	4.96	0.38	Partial structural similarity
	Plant (TET8)-Yeast (SUR7)	144	5.04	0.37	Partial structural similarity
	Yeast (SUR7)-Human (CD63)	194	3.86	0.62	Similar overall fold
	Yeast (SUR7)-Human (CD81)	173	3.54	0.55	Similar overall fold

Table 4: Structural alignment metrics including RMSD and TM-score values. Structural conservation was interpreted based on established thresholds: RMSD <2 Å (very strong structural conservation), RMSD 2–3 Å (moderate conservation), TM-score between 0.3 and 0.5 (partial structural similarity), TM-score > 0.5 (same fold) and TM-score > 0.7: highly similar.

The structural alignment metrics of these proteins reveals a clear spectrum of evolutionary conservation. RAB5, RAB7 and VPS4 families exhibit the highest structural similarity with TM-scores frequently exceeding 0.7 and low RMSD values. This indicates strong preservation of catalytic folds across species. On the contrary, structural alignment revealed limited global structural similarity for the ESCRT-0 and ESCRT-I complexes as reflected by low TM scores (typically below 0.4) and higher RMSD values. This suggests that these proteins show sequence divergence accompanied by substantial fold level divergence. The other families, such as ESCRT-II, ESCRT-III and tetraspanin maintain a similar overall fold with partial structural similarity indicating that the core remains recognisable, but specific functional domains have adapted to their unique requirements.

3.6 Mapping conserved motifs onto protein structures

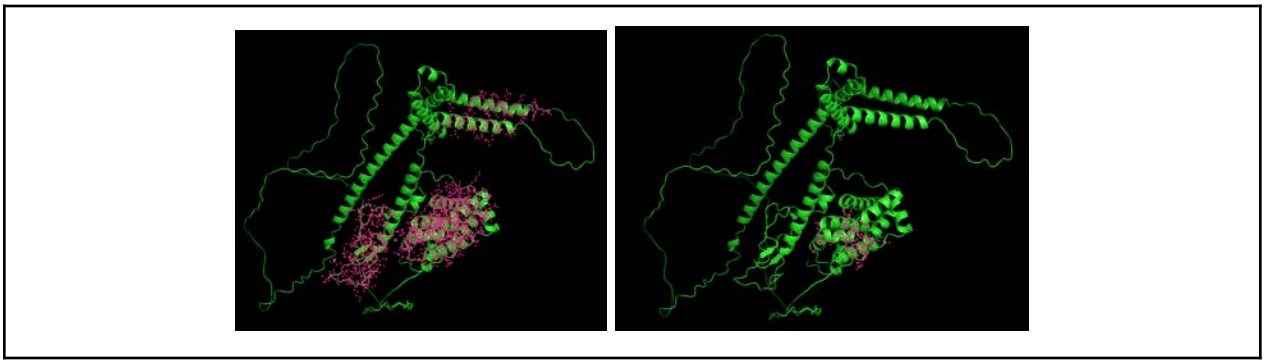


Figure 6.1: Comparative mapping of the conserved domains and motif of ESCRT-0 (Vps27-STAM-TOL2) on representative protein (Vps27). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualisation demonstrates the identification of conserved motifs or residues within the domains of ESCRT-0 proteins Vps27-STAM-TOL2. The initial mapping at the left identifies the primary functional domains of Vps27, characterised by a compact helical VHS domain, the SH3 domain and the FYVE domain. The mapping at the right only shows the residues and motifs that remain highly conserved across all these three sequences. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

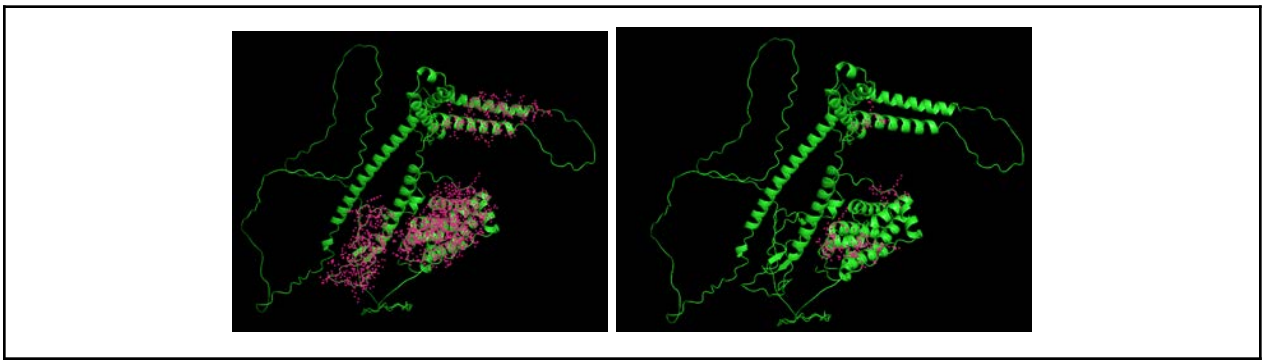


Figure 6.2: Comparative mapping of the conserved domains and motifs of ESCRT-0 (Vps27-STAM-TOL6) on representative structure (Vps27). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which

remains conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualisation demonstrates the identification of conserved structural elements within the domains of ESCRT-0 proteins Vps27-STAM-TOL2. The initial mapping at the left identifies the primary functional domains of Vps27, characterised by a compact helical VHS domain, and SH3 domain and the lower FYVE domain. The mapping at the right only shows the residues and motifs that remain highly conserved across all the sequences. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

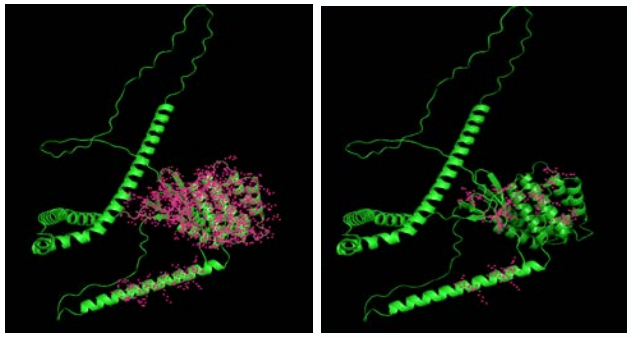


Figure 6.3: Comparative mapping of the conserved domains and motif of ESCRT-0 (Hse1-STAM-TOL2) on representative structure (Hse1). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualisation demonstrates the identification of conserved structural elements within the ESCRT-0 proteins Hse1-STAM-TOL2. The initial mapping at the left identifies the primary functional domains of Hse1, characterised by a compact helical VHS domain, the SH3 domain and the lower helical UIM domain. The mapping at the right only shows the residues and motifs that remain conserved across all the sequences. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

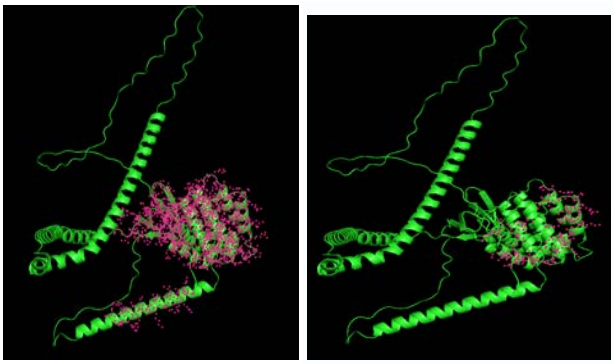


Figure 6.4: Comparative mapping of the conserved domains and motif of ESCRT-0 (Hse1-STAM-TOL6) on representative structure (Hse1). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualisation demonstrates the identification of conserved structural elements within the ESCRT-0 proteins Hse1-STAM-TOL6. The initial mapping at the left identifies the primary functional domains of Hse1, characterised by a compact helical VHS domain, the SH3 domain and the lower helical UIM domain. The mapping at the right only shows the residues and motifs that remain highly conserved across all the sequences. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

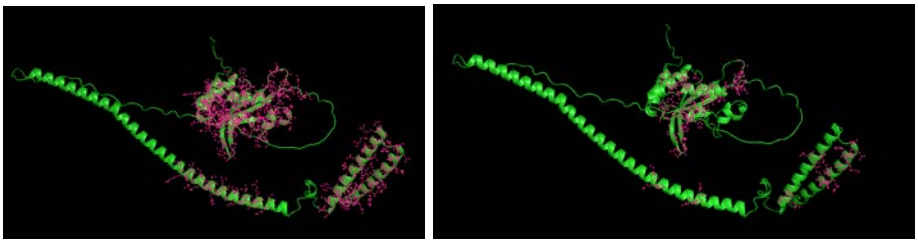


Figure 6.5: Comparative mapping of the conserved domains and motif of ESCRT-I (TSG101-STP22-ELC) on representative structure (TSG101). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which

remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualisation demonstrates the identification of conserved structural elements within the ESCRT-I proteins. The initial mapping (left) identifies the primary functional domains of STP22, characterised by a compact helical UEV domain, the helical structure-coiled coil and helical structure at right-SB. Finally the refined mapping (right) only shows the residues and motifs that remain highly conserved across all the sequences. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

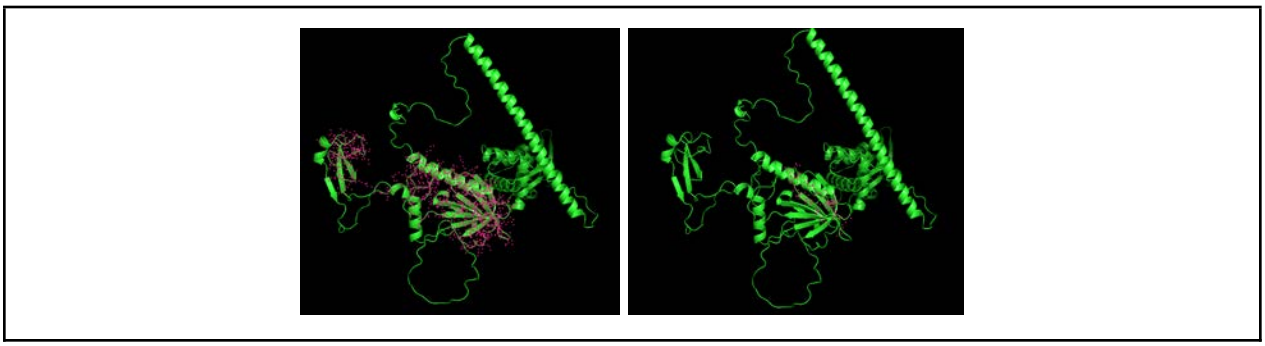


Figure 6.6: Comparative mapping of the conserved domains and motif of ESCRT-II (VPS36-VPS36-VPS36) on representative structure (yeast VPS36). On the left side, the red highlighted region indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualisation illustrates the structural conservation within the ESCRT II proteins. The initial mapping at the left highlighted the functional domains of VPS36, showing the N-terminal GLUE domain and the central VPS36 core domain. At the right, the mapping isolates the specific residues and motifs that are highly conserved across all the sequences. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

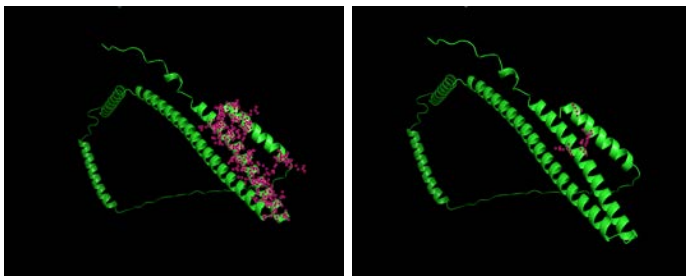


Figure 6.7: Comparative mapping of the conserved domains and motif of ESCRT-III (Snf7-Snf7-Snf7.1-CHMP1A) on representative structure (yeast Snf7). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

The mapping of the ESCRT-III proteins illustrates the motifs required for membrane remodelling. The left picture identifies the primary helical organisation, including the N-terminal helical core, MIM and the C-terminal regulatory tail. At the right, only the highly conserved MIM (MIT interacting motif) was found across the species. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

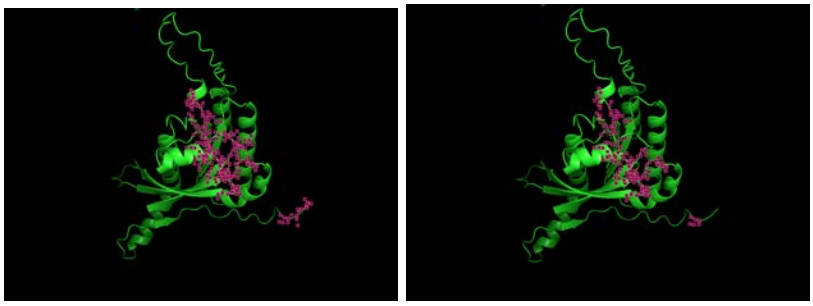


Figure 6.8: Comparative mapping of the conserved domains and motif of Rab GTPases (YPT52-VPS21-RAB5A-RAB5) on representative structure (YPT52). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualization demonstrates the identification of conserved structural elements within the Rab5 family representatives, including VPS21, YPT52, and RAB5A. The initial mapping (top) identifies the primary functional architecture of the G-domain, characterized by the P-loop (Walker A), the NKXD motif, and the regulatory Switch I and II regions. Finally, the refined model (right) isolates only the residues and motifs that remain highly conserved across all sequences. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

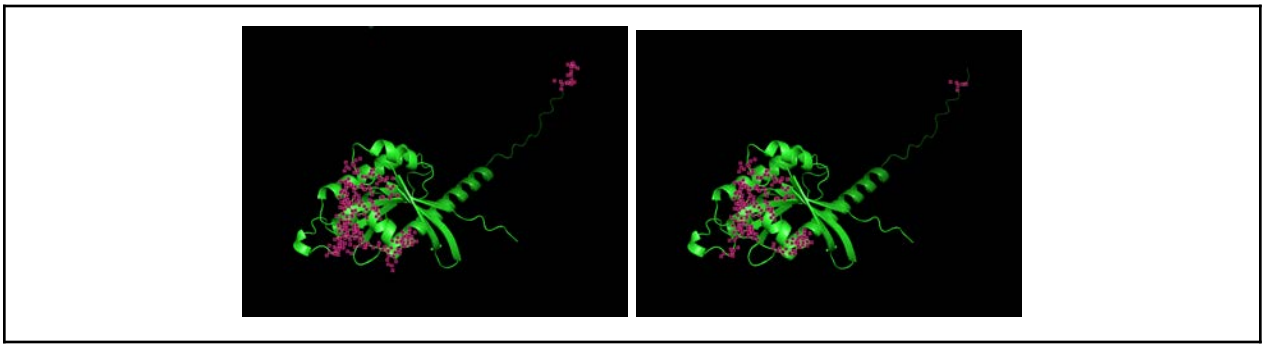


Figure 6.9: Comparative mapping of the conserved domains and motif of Rab GTPases (YPT7-RAB7-RAB7) on representative structure (YPT7). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualization demonstrates the identification of conserved structural elements within the Rab5 family proteins. The initial mapping (top) identifies the primary functional architecture of the G-domain, characterized by the P-loop (Walker A), the NKXD motif, and the regulatory Switch I and II regions. Finally, the refined model (right) isolates only the residues and motifs that remain highly conserved across all sequences. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

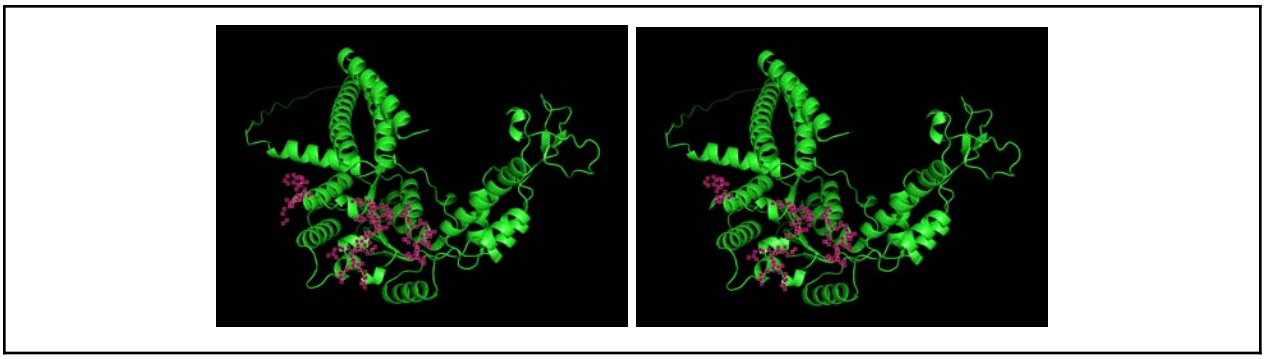


Figure 6.10: Comparative mapping of the conserved domains and motif of VPS4 proteins on representative structure (yeast VPS4). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

The mapping of the VPS4 family identifies the characteristic mechanical organization of a classic AAA ATPase. On the top, the primary domains are identified through the Walker A and Walker B motifs required for ATP binding and hydrolysis, alongside the Arginine finger. Finally, the refined model (right) isolates the highly conserved Pore loop residues, which are evolutionarily maintained to provide the physical force necessary for the disassembly of the ESCRT-III complex. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

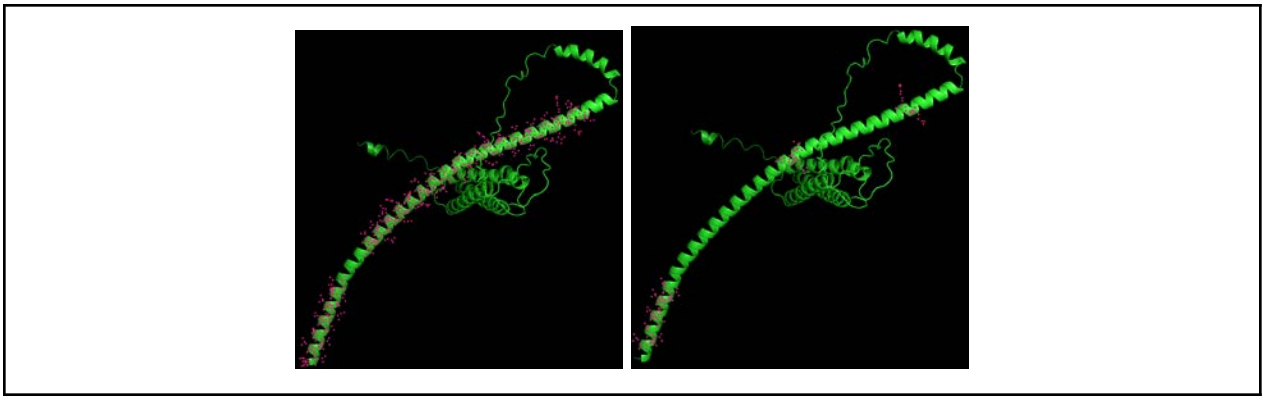


Figure 6.11: Comparative mapping of the conserved domains and motif of SNAREs (VAM3-SYP121-STX10) on representative structure (VAM3). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This visualisation illustrates the structural mapping of the SNARE family. The initial mapping at the top shows the primary functional hallmark of the family, characterised by elongated SNARE motif and the C-terminal TM helix. The mapping on the right only represents those residues which remain strongly conserved across all the species. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

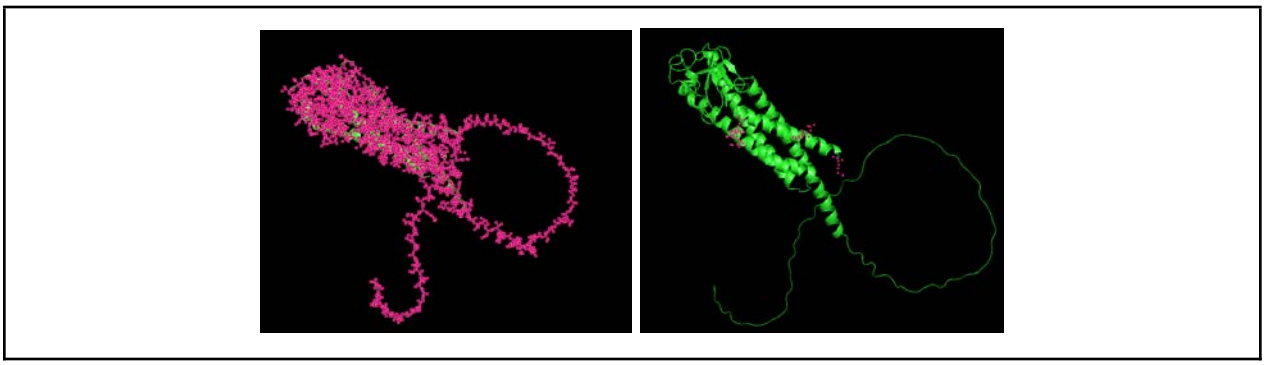


Figure 6.12: Comparative mapping of the conserved domains and motif of Tetraspanins (SUR7-TET8-CD63-CD81) on representative structure (SUR7). On the left side, the red highlighted regions indicate the known functional domains, while on the right are the domain residues which remain conserved. The PDB files of the proteins were visualised using the PYMOL visualisation tool.

This structural mapping of the Tetraspanin family at the top identifies the four transmembrane helices and the cytoplasmic tail. The refined model only shows a few residues which remained conserved across all the species. Mapping of conserved regions onto the representative structure showed that conserved motifs preferentially clustered within compact structured cores whereas more variable residues distributed along peripheral loops, terminal segments or low-complexity regions.

3.7 Intrinsic disorder analysis

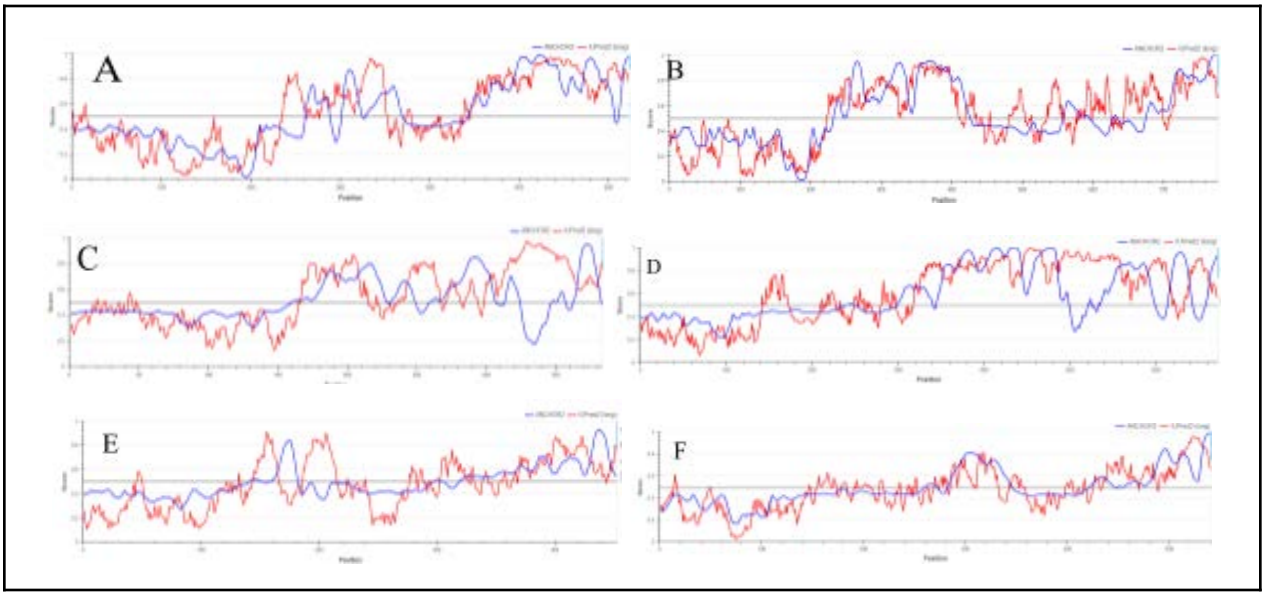


Figure 7.1: Intrinsic disorder prediction of ESCRT-0 proteins (**A**: Yeast_Vps27; **B**:Human_Hrs; **C**: Plant_TOL2; **D**: Plant_TOL6; **E**:Yeast_Hse1; **F**: Human STAM). The X-axis represents the residue positions and the Y-axis denotes the disorder propensity score. The colored curve represents the predicted disorder score along the protein sequence, where the color variations indicate prediction by different models. The disorder prediction graphs were generated using their IUpred2A web server using the respective amino acid sequences.

In case of all the proteins of the three species, the VHS domain tends to be ordered (<0.5 score) while the C-terminal region shows predicted disorder (>0.5 score). Additionally, the FYVE domain is ordered in yeast Vps27 and human Hrs but absent in plant TOL2 and TOL6. The UIM region shows predicted disorder in human Hrs and yeast Vps27 but absent in plant TOL2 and TOL6. Moreover, the UIM and SH3 domains are ordered in yeast Hse1 and human STAM but absent in plant TOL2 and TOL6.

Regions showing lower sequence conservation frequently corresponded to predicted disordered or low-complexity segments whereas the structured catalytic or domain cores tended to remain less disordered across species. Consistently, the MSA results also show higher sequence conservation in the VHS domain than other regions, suggesting that regions with a higher sequence conservation correspond to ordered segments.

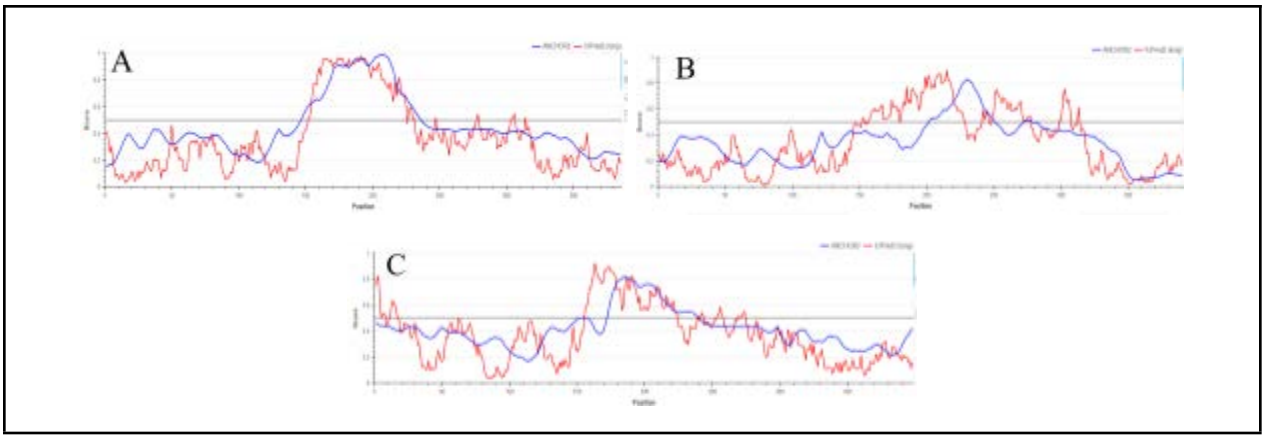
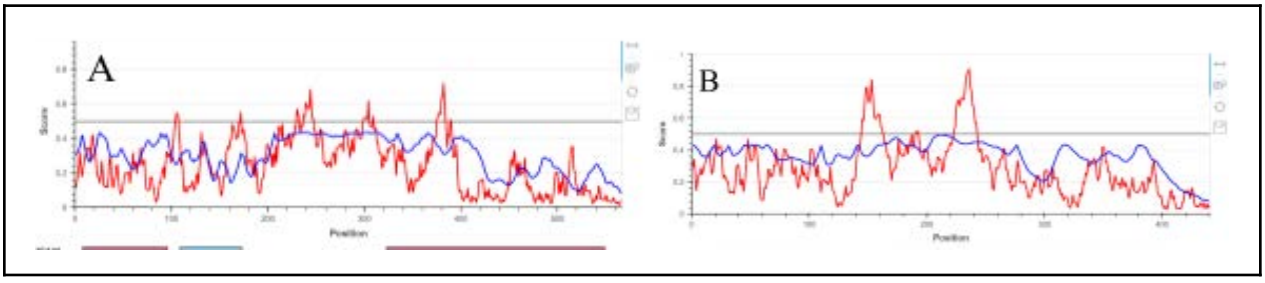


Figure 7.2: Intrinsic disorder prediction of ESCRT-I proteins (**A**:Yeast_STP22; **B**: Human_TSG101; **C**: Plant_ELC). The X-axis represents the residue positions and the Y-axis denotes the disorder propensity score. The colored curve represents the predicted disorder score along the protein sequence, where the color variations indicate prediction by different models. The disorder prediction graphs were generated using the IUpred2A web server using their respective amino acid sequences.

The UEV domain and SB domain / C-terminal region of the three species are ordered (<0.5 score). While the coiled coil domain of yeast is ordered, it is predicted to be somewhat disordered in human and plant (close to 0.5 score). Regions showing lower sequence conservation frequently corresponded to predicted disordered or low-complexity segments whereas the structured catalytic or domain cores tended to remain less disordered across species. Consistently, the MSA results also show that the UEV domain and the SB domain are more conserved across the three species than the coiled coil domain, suggesting that regions with a higher sequence conservation correspond to ordered segments.



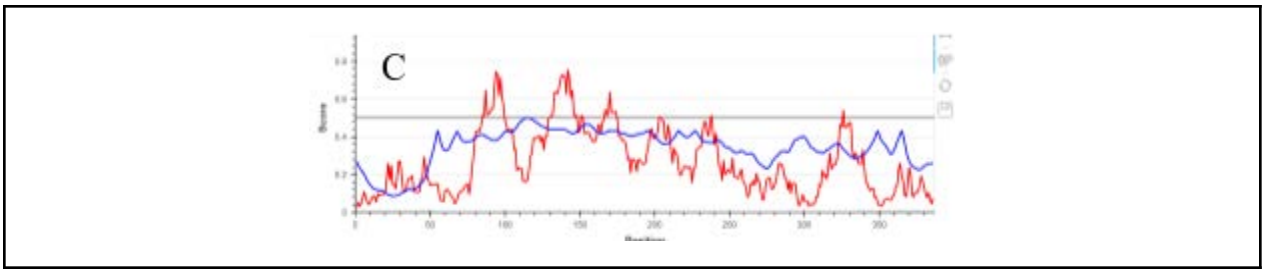


Figure 7.3: Intrinsic disorder prediction of ESCRT-II proteins (**A**:Yeast_VPS36; **B**: Plant_VPS36; **C**: Human_VPS36). The X-axis represents the residue positions and the Y-axis denotes the disorder propensity score. The colored curve represents the predicted disorder score along the protein sequence, where the color variations indicate prediction by different models. The disorder prediction graphs were generated using their IUpred2A web server using the respective amino acid sequences.

The GLUE N-terminal and GLUE C-terminal both are ordered regions (<0.5 score). While the conservation is low, both these regions show more conservation than the other regions (Zinc finger motifs) in the MSA results, suggesting that regions with a higher sequence conservation correspond to ordered segments. In contrast, regions showing lower sequence conservation frequently corresponded to predicted disordered or low-complexity segments. The structured catalytic or domain cores tended to remain less disordered across species.

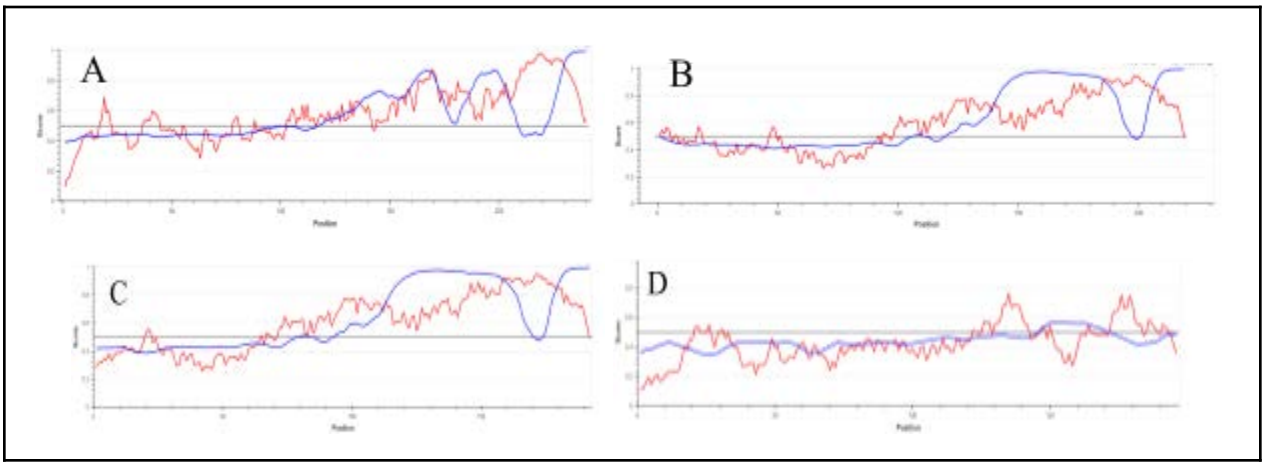


Figure 7.4: Intrinsic disorder prediction of ESCRT-III (**A**: Yeast_Snf7; **B**: Plant_Snf; **C**: Plant_Snf7.1; **D**: Human_CHMP1A). The X-axis represents the residue positions and the Y-axis denotes the disorder propensity score. The colored curve represents the predicted disorder score along the protein sequence, where the color variations indicate prediction by different models.

The disorder prediction graphs were generated using their IUpred2A web server using the respective amino acid sequences.

The N-terminal region is ordered (<0.5 score) across all the three species. The MIT interacting motif and C terminal regions are predicted to be disordered (>0.5 score) in yeast and plant but ordered in human. Regions showing lower sequence conservation frequently corresponded to predicted disordered or low-complexity segments whereas the structured catalytic or domain cores tended to remain less disordered across species. However, the MSA results did not show any significant higher conservation in the N-terminal region than the other regions.

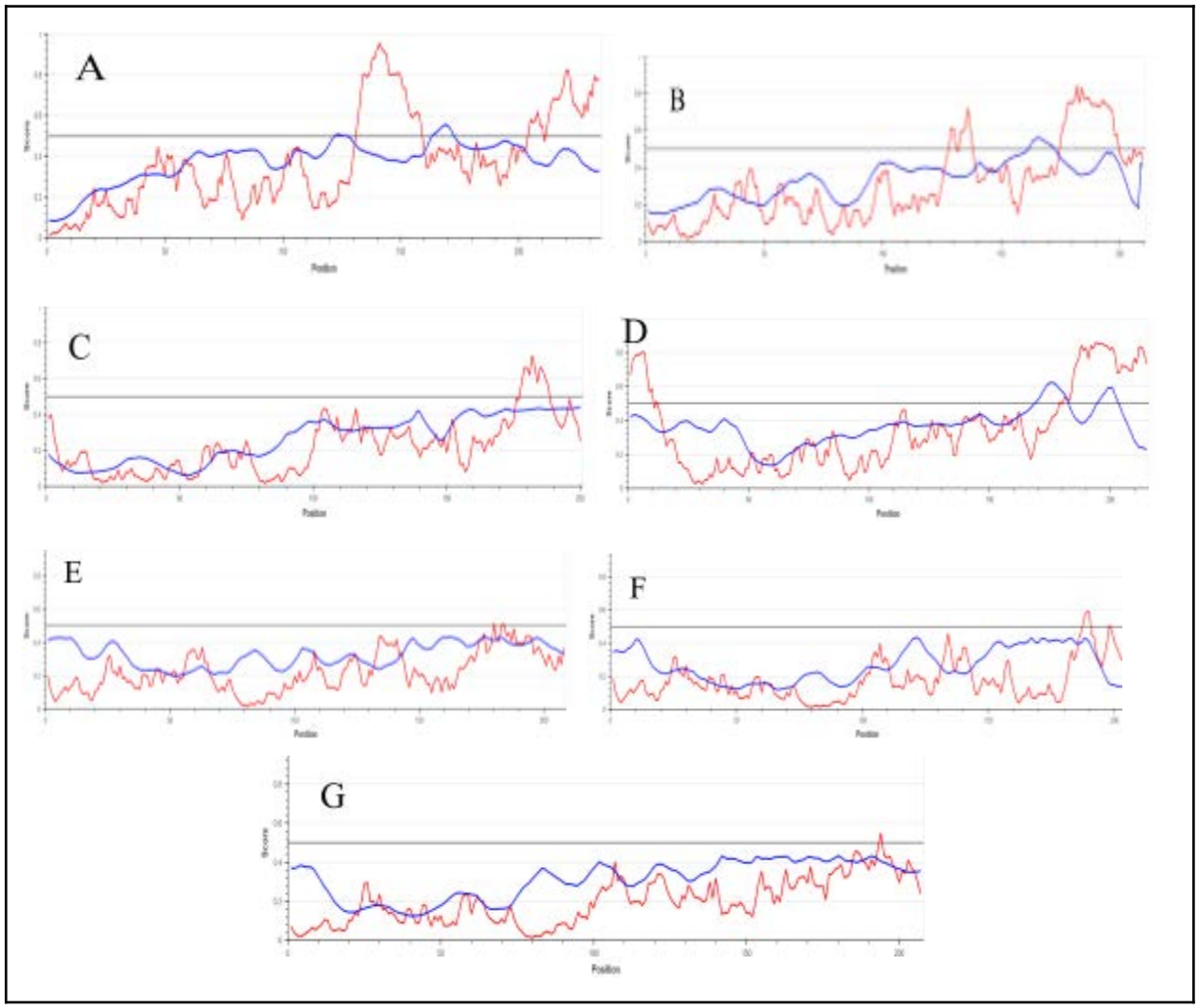


Figure 7.5: Intrinsic disorder prediction of Rab GTPases (**A**: Yeast_Ypt52; **B**: Yeast_VPS21 **C**: Plant_RAB5A; **D**: Human_RAB5; **E**: Yeast_Ypt7; **F**: Plant_RAB7; **G**: Human_RAB7). The X-axis represents the residue positions and the Y-axis denotes the disorder propensity score. The

colored curve represents the predicted disorder score along the protein sequence, where the color variations indicate prediction by different models. The disorder prediction graphs were generated using their IUpred2A web server using the respective amino acid sequences.

The P-loop, Switch I, Switch II and NKxD regions all are ordered (<0.5 score) across the species while the C terminal regions show predicted disorder (>0.5 score). This aligns with the MSA results where higher conservation is noticed in the P-loop, Switch I, Switch II and NKxD domains than the C-terminal region, suggesting that regions with a higher sequence conservation correspond to ordered segments.

Regions showing lower sequence conservation frequently corresponded to predicted disordered or low-complexity segments whereas the structured catalytic or domain cores tended to remain less disordered across species.

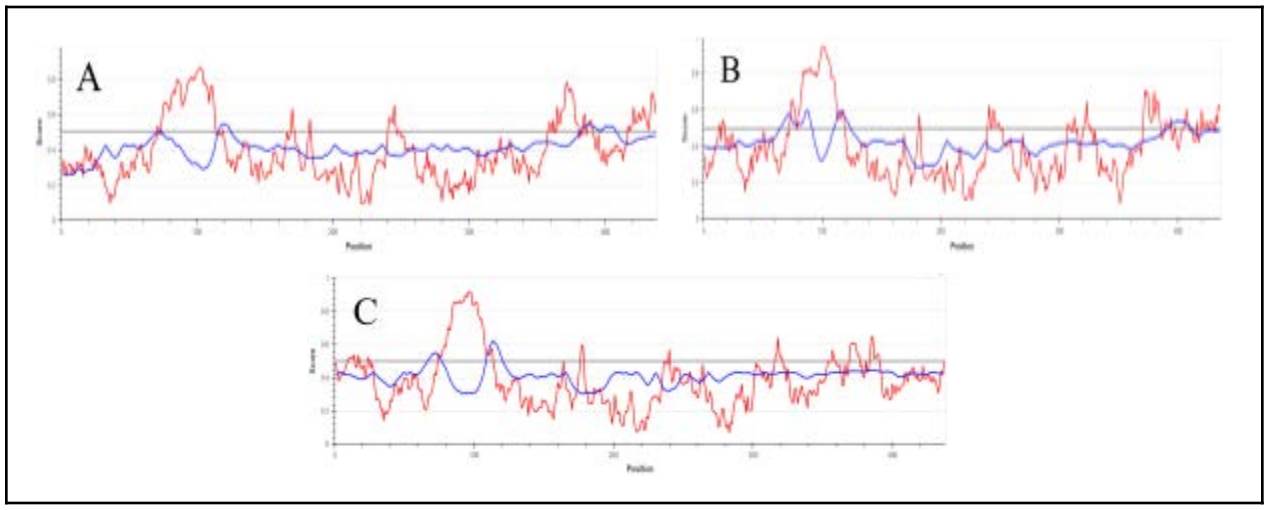


Figure 7.6: Intrinsic disorder prediction of Disassembly proteins (**A**: Yeast_VPS4; **B**: Plant_VPS4; **C**: Human_VPS4). The X-axis represents the residue positions and the Y-axis denotes the disorder propensity score. The colored curve represents the predicted disorder score along the protein sequence, where the color variations indicate prediction by different models. The disorder prediction graphs were generated using their IUpred2A web server using the respective amino acid sequences.

The Walker A, Walker B, Arginine finger and Pore loop residues all are ordered (<0.5 score). This also aligns with the MSA results where high conservation is noticed in all these regions, suggesting that regions with a higher sequence conservation correspond to ordered segments.

Regions showing lower sequence conservation frequently corresponded to predicted disordered or low-complexity segments whereas the structured catalytic or domain cores tended to remain less disordered across species.

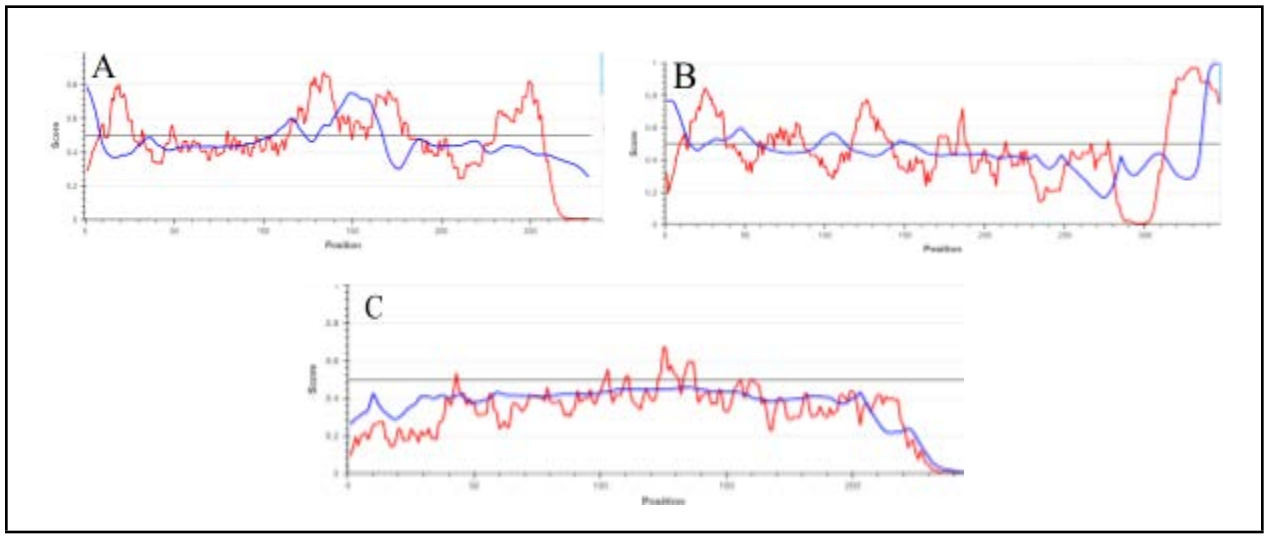


Figure 7.7: Intrinsic disorder prediction of SNAREs proteins (**A**: Yeast_VAM3; **B**: Plant_SYP121; **C**: Human_STX10). The X-axis represents the residue positions and the Y-axis denotes the disorder propensity score. The colored curve represents the predicted disorder score along the protein sequence, where the color variations indicate prediction by different models. The disorder prediction graphs were generated using their IUpred2A web server using the respective amino acid sequences.

The t-SNARE coiled-coil homology is ordered in plant and human (<0.5 score) but shows predicted disorder in yeast (>0.5 score). However, the TM helices are ordered across the three species. This aligns with the MSA results where a comparatively higher conservation is noticed in the TM helix than the t-SNARE coiled-coil homology, suggesting that regions with a higher sequence conservation correspond to ordered segments. Regions showing lower sequence conservation frequently corresponded to predicted disordered or low-complexity segments whereas the structured catalytic or domain cores tended to remain less disordered across species

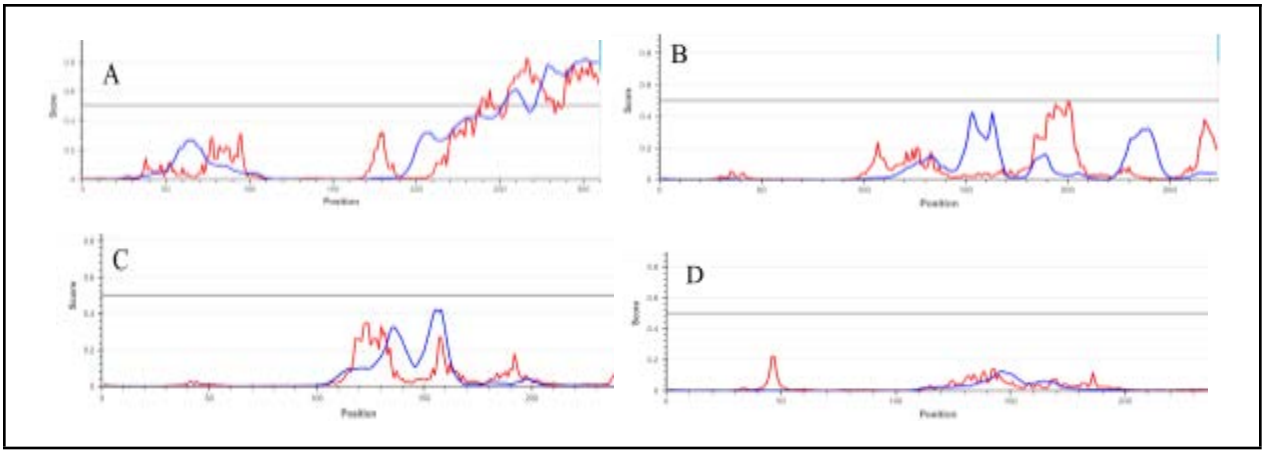


Figure 7.8: Intrinsic disorder prediction of tetraspanins proteins (**A**: Yeast_SUR7; **B**: Plant_TET8; **C**: Human_CD63; **D**: Human_CD81). The X-axis represents the residue positions and the Y-axis denotes the disorder propensity score. The colored curve represents the predicted disorder score along the protein sequence, where the color variations indicate prediction by different models. The disorder prediction graphs were generated using their IUpred2A web server using the respective amino acid sequences.

The Topological domain (Extracellular) and TM helix are ordered (<0.5 score) across all the three species. The Topological domain (Cytoplasmic) is predicted to be disordered in yeast but ordered in plant and human. However, the MSA results show that all the regions exhibit a very low conservation, contradicting that lower sequence conservation corresponds to disordered segments.

3.8 Cross-complex evolutionary synthesis

Protein family	Sequence Conservation	Domain / Motif Conservation	Structural Similarity	Divergent Regions
ESCRT-0	Low sequence conservation	VHS	Low Structural Similarity	FYVE-type zinc finger
				SH3
				UIM
ESCRT-I	Moderate sequence conservation	UEV	Low Structural Similarity	Coiled coil region
		SB		
ESCRT-II	Low sequence conservation	GLUE N-terminal	Partial structural similarity	Zinc finger motif (RanBP2-type 1)
		GLUE C-terminal		Zinc finger motif (RanBP2-type 2)
ESCRT-III	Low sequence conservation	Helical core	Partial structural similarity	N-terminal helical core
		MIM		
		C-terminal		
Rab GTPases	High sequence conservation	Walker A	High structural similarity	Little to no divergence within the motifs
		Switch I		
		Switch II		
		NKxD		
Disassembly	High sequence conservation	Walker A	High structural similarity	Little to no divergence within the motifs
		Walker B		
		Arginine finger		
		Pore loop residues		
SNAREs	Moderate sequence	t-SNARE coiled-coil	Partial structural similarity	Some regions in the SNARE

	conservation	homology		motif
		Transmembrane helices		
Tetraspanins	Very low sequence conservation	Topological domain (Cytoplasmic)	Partial structural similarity	Topological domain (Extracellular)
				TM Helix

Table 5: Cross-complex evolutionary comparison summarizing sequence conservation and structural similarity across protein families

From the table, the findings suggest that the protein families analyzed reveal a hierarchical pattern of evolutionary conservation within the membrane trafficking system. Catalytic proteins such as Rab GTPases and VPS4 reveal the highest level of sequence, motif and structural conservation. In contrast, structural scaffold proteins such as ESCRT-I, ESCRT-II, ESCRT-III and SNAREs display moderate conservation with preservation of core domains despite sequence divergence. At the other end of the hierarchy, regulatory adaptor proteins such as ESCRT-0 and tetraspanins show greater evolutionary divergence in both sequence and domain architecture, often accompanied by an increased disorder or low-complexity regions.

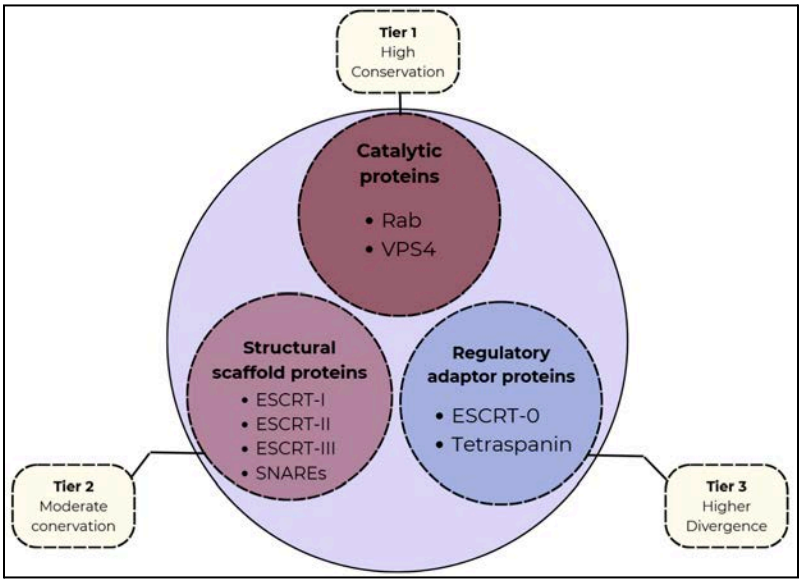


Figure 8: Illustration of conservation tiers across trafficking proteins. Catalytic proteins are shown in dark red (high conservation), structural proteins in light pink (moderate conservation) and regulatory proteins in blue (higher divergence). Generated using BioRender.

Chapter 4

Discussion

4.1 Conservation of catalytic trafficking proteins

The results obtained here clearly demonstrate the strong conservation of catalytic trafficking proteins such as Rab GTPase and VPS4 ATPases across yeast, human and plants. Pairwise comparisons and domain specific alignments indicate that key motifs (P-loop, Switch I and II, NKxD in Rab proteins and Walker A, Walker B, Arginine finger and Pore loop residues in VPS4) are highly conserved. Several of these motifs showed complete conservation across yeast, plant and human in multiple sequence alignments. This supports the claim that strong purifying selection acts on residues involved in catalysis and the regions adjacent to these catalytic cores also show higher conservation **[76-77]**. Phylogenetic analyses highlight the evolutionary divergence of these proteins. In RAB5, yeast sequences cluster together with plants and vertebrate sequences form a separate clade. In RAB7 and VPS4, vertebrate sequences cluster with high bootstrap value and plant and yeast sequences group separately. Here, the key catalytic motifs are conserved despite evolutionary divergence. This indicates that the divergence has occurred in regions outside the functional core and selective pressure has conserved the essential catalytic regions.

Structural alignment further supports the functional importance of this conservation. TM-scores exceeding 0.7 and low RMSD values indicate that the overall three dimensional fold is maintained across species. Additionally, intrinsic disorder analysis show that conserved catalytic motifs correspond to stable ordered regions and less conserved regions are more disordered like the C-terminal tail in Rab proteins. This suggests that structural stability of catalytic regions is critical for maintaining enzymatic activity and other regions may be variable among species. Overall, these findings show that catalytic trafficking proteins Rab GTPase and VPS4 ATPases maintain a highly conserved structural and functional core despite evolutionary divergence.

4.2 Structural conservation despite sequence divergence

The results demonstrate that structural conservation is maintained despite substantial sequence divergence across the protein families. For example, ESCRT complexes exhibit low to moderate sequence identity (ranging from 9 to 33%) but maintain conserved structural elements such as

VHS domain in ESCRT-0, UEV and SB domains in ESCRT-I, GLUE domains in ESCRT-II and helical core and MIT-interacting motif in ESCRT-III. While there are divergent regions observed among these ESCRT complexes, they show partial structural similarity. Similarly, SNARE proteins show moderate sequence conservation while retaining coiled-coil organization and transmembrane helices. Additionally, tetraspanins show higher sequence divergence yet have partial structural similarity.

Structural conservation is observed even when sequence identity is low because proteins can differ extensively in amino acid sequences due to evolution while still maintaining a stable three-dimensional structure and biological functions **[93-94]**. The mentioned protein families of yeast, plant and human have derived from common ancestral genes **[1]**. Though their amino acid sequences have diverged over time due to species-specific adaptation, their overall structure has been maintained to preserve the essential functions of the eukaryotic cells. This indicates the evolutionary pressure to conserve structure rather than retaining the exact amino acid sequences in proteins. As mentioned earlier, in the membrane trafficking system, the core structure of protein families is evolutionarily conserved as it is required for the key processes of vesicle formation, transportation, fusion etc. **[56]**. Despite sequence divergence observed in the protein families, the protein structures are conserved by strong structural and functional constraints required to maintain their stability and core functions **[95-97]**. Thus, even across distant species, proteins can exhibit low sequence identity but retain conserved structural folds.

4.3 Divergence of regulatory and adaptor proteins

The results showed that adapter proteins like ESCRT-0 components and tetraspanins exhibit higher divergence compared to other protein families. Pairwise sequence analysis for ESCRT-0 show yeast and human sequences were more conserved with each other than with plant homologs. In case of tetraspanins, plant TET8 shows moderate similarity with human CD63 and CD81 and yeast SUR7 showed very low identity in pairwise comparison.

Domain specific analysis indicated low sequence conservation of these adaptor proteins across yeast, plant and human as well. The VHS domain of ESCRT-0 shows relatively higher conservation in domain specific analysis and multiple sequence alignment than UIM, SH3 and FYVE-type zinc finger domains. This indicates that the core VHS domain is somewhat conserved in ESCRT-0 but the other regions show sequence variation. The plant sequences did not contain

the other domains except VHS which indicates that these regions may be less critical for maintaining fundamental ESCRT-0 function across all species. In case of tetraspanins, transmembrane helices showed moderate conservation while topological cytoplasmic tails and extracellular topological domains showed more conservation between plant and human among three species in domain specific analysis. However, no significant continuous conserved regions were found in multiple sequence alignment of tetraspanins. Although pairwise comparisons show moderate similarity, the multiple sequence alignment reveals only scattered conserved residues indicating overall low sequence conservation across all species.

Phylogenetic tree also showed widespread evolutionary divergence of ESCRT-0 and tetraspanins sequences which is consistent with lower sequence conservation. Intrinsic disorder prediction showed the VHS domain is more ordered and stable compared to other domains which is supported by its partial sequence conservation. However, TM-scores and RMSD values for ESCRT-0 showed low structural similarity which is consistent with observed high sequence divergence. Overall, it can be interpreted that only the VHS domain may retain some structural and functional features of ESCRT-0 across species. In case of tetraspanins, TM-scores and RMSD value indicated partial structural similarity and similar overall fold across species despite low sequence conservation. Intrinsic disorder pattern also showed Topological domain (Extracellular) and TM helix are ordered in all species indicating stable protein structure despite low sequence similarity. These findings partially support the idea that protein structure is more conserved than sequences **[81-82].** Overall, ESCRT-0 and tetraspanins show high evolutionary divergence compared to catalytic proteins. This may be because catalytic proteins need conservation to maintain enzymatic activity but adapter proteins which primarily mediate interaction with cargos can tolerate sequence changes.

4.4 Evolutionary hierarchy of membrane trafficking proteins

The results collectively suggest an existence of evolutionary hierarchy within these protein families, where the degree of conservation across yeast, plant and human directly reflects their functional roles. At the top tier, RAB GTPases and VPS4 showed the highest sequence identity, structural similarity, motif conservation and consistently ordered regions. In the intermediate tier, ESCRT-I, ESCRT-II, ESCRT-III and SNAREs showed moderate conservation and partial structural similarity. This suggests that they showed signs of maintaining conserved folds and

functional domains while tolerating divergence in regulatory regions. At the base of this hierarchy, ESCRT-0 and tetraspanin showed the greatest evolutionary divergence with low sequence identity, poor structural similarity and minimal conserved residues within their functional domains.

By analyzing the results, it was demonstrated that selective pressure acting on a protein is proportional to the functional consequences of its mutation. In VPS4, Walker A, Walker B, Arginine finger and Pore loop residues were all consistently ordered and highly conserved. In RAB GTPases, although the P-loop, Switch I and Switch II regions were ordered, the NKxD motif and C-terminal regions showed disorder despite retaining relatively high conservation in the sequence analysis. This suggested that conservation and structural order are not always directly coupled. Similarly conservation regions in the UEV domain in ESCRT-I and the GLUE domains in ESCRT-II were consistently ordered, while peripheral regions showed higher disorder scores. This again indicated that only the structurally stable regions are subject to selective conservation. However for tetraspanins, transmembrane helices and extracellular domains were ordered across all species despite very low sequence conservation, which contradicts the expected relationship between disorder and sequence divergence. This indicates that while the disorder analysis generally supports the proposed hierarchy, structural order alone is not a reliable predictor of sequence conservation across all protein families.

For this reason, phylogenetic analysis and PyMOL structural mapping together provide two complementary lines of evidence that further validate this hierarchy. Since the phylogenetic trees used in this study are topology based without branch lengths, they cannot be used to quantify the rate or degree of divergence directly. However, they can establish whether sequence variation is evolutionarily structured or not. If sequences group according to known species relationships, the variation is constrained by lineage rather than being random. For RAB GTPases and VPS4, the trees recover well-resolved clades in which yeast, plant and animal sequences separate into distinct, phylogenetically coherent groupings. This indicates that selective pressure has kept these sequences within predictable evolutionary boundaries. For ESCRT-0 and tetraspanins, the topologies are less resolved, consistent with sequences having diverged more independently across lineages. PyMOL structural mapping then provides the spatial dimension of this argument. For tier 1 proteins, conserved residues are distributed broadly across the entire three-dimensional structure, indicating that constraint is global and that almost no region of the protein can tolerate

substitution without consequence. For tier 3 proteins, conserved residues collapse into a compact, localised core, while peripheral loops, terminal segments and inter-domain regions remain free to vary. This means that a greater fraction of the protein surface can accommodate mutation without disrupting function. The correlation between topological structure in the phylogenetic trees and the spatial distribution of conserved residues in the PyMOL models therefore provides a mutually reinforcing validation of the evolutionary hierarchy.

4.5 Biological implication

The conservation pattern across yeast, plant and human proteins carry significant biological implications which reflects the fundamental nature of vesicular transport in eukaryotic cell. The conservation of catalytic cores suggests that these proteins represent core cellular processes that emerged early in eukaryotic evolution and have remained conserved for a very long evolutionary period **[19]**. This can be seen as basic functions like nutrient uptake, organelle maintenance and intracellular communication depend on intact trafficking machinery and are required by all eukaryotic cells **[98].** In a similar way, the shared conservation of SNARE proteins and ESCRT complexes suggests that the basic mechanisms of vesicle formation and membrane fusion were already established in the last eukaryotic common ancestor (LECA), and they have been retained across diverging lineages **[99]**. So the greater divergence which was seen in the adaptor proteins like ESCRT-0 and tetraspanin could likely reflect how different organisms have adopted tier trafficking systems to meet their specific cellular needs **[100]**.

Thus from an evolutionary perspective, the fact that these core proteins remain conserved across organisms separated by hundreds of millions of years of evolution strongly suggests that mutations disrupting them would cause some form of disturbance which also could be lethal in nature **[19]**.

Chapter 5

Conclusion

This study demonstrates a comparative analysis of different protein families involved in membrane trafficking systems across yeast, plant and human. The studied trafficking proteins exhibit varying degrees of sequence conservation and structural similarity across the eukaryotes. While some proteins exhibit a higher sequence and structural similarity, others demonstrate significant sequence divergence but retain conserved structural features essential for their core function. This highlights that conservation is not uniform across all components of the trafficking system. Instead, it depends on the functional roles of the individual protein families, where structural and catalytic proteins remain more conserved than regulatory adaptor proteins, supporting the evolutionary conservation model. The catalytic proteins like Rab and VPS4 showed higher levels of sequence and structural conservation, reflecting strong evolutionary pressure to retain their essential roles. Therefore, higher conservation of these proteins across diverse eukaryotic lineages highlights their fundamental roles in maintaining key cellular processes. Additionally, some proteins show low sequence identity while retaining their conserved three-dimensional folding structure, suggesting that structural integrity is maintained by evolutionary pressure to sustain their functional roles rather than preserving the exact amino acid sequences. Moreover, the higher level of sequence divergence in regulatory proteins like ESCRT-0 and tetraspanins demonstrates reduced evolutionary constraints compared to catalytic proteins, highlighting their roles in maintaining organism-specific regulatory processes depending on their cellular complexity. As a result, such proteins exhibit greater evolutionary flexibility while still maintaining structural and functional roles.

Altogether, the study supports the existence of an evolutionary hierarchy within the membrane trafficking system where the catalytic proteins are highly conserved, structural scaffold proteins show moderate conservation and regulatory proteins show higher variation. The membrane trafficking system is evolutionarily conserved at the structural and functional level with the presence of sequence variation. Overall, the findings contribute to a broader understanding of how conserved biological systems like membrane trafficking are maintained across diverse eukaryotic lineages.

References

1. More K, Klinger CM, Barlow LD, Dacks JB. Evolution and natural history of membrane trafficking in Eukaryotes. Current Biology [Internet]. 2020 May 1;30(10):R553–64. Available from: https://doi.org/10.1016/j.cub.2020.03.068

2. Kumar A, Ahmad A, Vyawahare A, Khan R. Membrane trafficking and subcellular drug targeting pathways. Frontiers in Pharmacology [Internet]. 2020 May 27;11:629. Available from: https://doi.org/10.3389/fphar.2020.00629

3. Laidlaw KME, MacDonald C. Endosomal trafficking of yeast membrane proteins. Biochemical Society Transactions [Internet]. 2018 Oct 31;46(6):1551–8. Available from: https://doi.org/10.1042/bst20180258

4. Lippincott-Schwartz J, Phair RD. Lipids and cholesterol as regulators of traffic in the endomembrane system. Annual Review of Biophysics [Internet]. 2010 Apr 1;39(1):559–78. Available from: https://doi.org/10.1146/annurev.biophys.093008.131357

5. Aniento F, De Medina Hernández VS, Dagdas Y, Rojas-Pierce M, Russinova E. Molecular mechanisms of endomembrane trafficking in plants. The Plant Cell [Internet]. 2021 Sep 17;34(1):146–73. Available from: https://doi.org/10.1093/plcell/koab235

6. Elias M. Patterns and processes in the evolution of the eukaryotic endomembrane system. Molecular Membrane Biology [Internet]. 2010 Nov 1;27(8):469–89. Available from: https://doi.org/10.3109/09687688.2010.521201

7. Cullen PJ, Steinberg F. To degrade or not to degrade: mechanisms and significance of endocytic recycling. Nature Reviews Molecular Cell Biology [Internet]. 2018 Sep 7;19(11):679–96. Available from: https://doi.org/10.1038/s41580-018-0053-7

8. Barlowe CK, Miller EA. Secretory protein biogenesis and traffic in the early secretory pathway. Genetics [Internet]. 2013 Feb 1;193(2):383–410. Available from: https://doi.org/10.1534/genetics.112.142810

9. Hurley JH, Boura E, Carlson LA, Różycki B. Membrane budding. Cell [Internet]. 2010 Dec 1;143(6):875–87. Available from: https://doi.org/10.1016/j.cell.2010.11.030

10. Feyder S, De Craene JO, Bär S, Bertazzi D, Friant S. Membrane Trafficking in the Yeast Saccharomyces cerevisiae Model. International Journal of Molecular Sciences [Internet]. 2015 Jan 9;16(1):1509–25. Available from: https://doi.org/10.3390/ijms16011509

11. Ng EL, Gan BQ, Ng F, Tang BL. Rab GTPases regulating receptor trafficking at the late endosome–lysosome membranes. Cell Biochemistry and Function [Internet]. 2012 Apr 3;30(6):515–23. Available from: https://doi.org/10.1002/cbf.2827

12. Ahmed I, Akram Z, Iqbal HMN, Munn AL. The regulation of Endosomal Sorting Complex Required for Transport and accessory proteins in multivesicular body sorting and

enveloped viral budding - An overview. International Journal of Biological Macromolecules [Internet]. 2019 Jan 4;127:1–11. Available from: https://doi.org/10.1016/j.ijbiomac.2019.01.015

13. Wang HJ, Hsu YW, Guo CL, Jane WN, Wang H, Jiang L, et al. VPS36-Dependent multivesicular bodies are critical for plasmamembrane protein turnover and vacuolar biogenesis. PLANT PHYSIOLOGY [Internet]. 2016 Nov 22;173(1):566–81. Available from: https://doi.org/10.1104/pp.16.01356

14. Azmi IF, Davies BA, Xiao J, Babst M, Xu Z, Katzmann DJ. ESCRT-III family members stimulate VPS4 ATPASE activity directly or via VTA1. Developmental Cell [Internet]. 2008 Jan 1;14(1):50–61. Available from: https://doi.org/10.1016/j.devcel.2007.10.021

15. Rodrigues J, Lydall D. Cis and trans interactions between genes encoding PAF1 complex and ESCRT machinery components in yeast. Current Genetics [Internet]. 2018 Mar 21;64(5):1105–16. Available from: https://doi.org/10.1007/s00294-018-0828-6

16. Zhu L, Jorgensen JR, Li M, Chuang YS, Emr SD. ESCRTs function directly on the lysosome membrane to downregulate ubiquitinated lysosomal membrane proteins. eLife [Internet]. 2017 Jun 29;6. Available from: https://doi.org/10.7554/elife.26403

17. Huotari J, Helenius A. Endosome maturation. The EMBO Journal [Internet]. 2011 Aug 31;30(17):3481–500. Available from: https://doi.org/10.1038/emboj.2011.286

18. Piper RC, Katzmann DJ. Biogenesis and function of multivesicular bodies. Annual Review of Cell and Developmental Biology [Internet]. 2007 Jul 6;23(1):519–47. Available from: https://doi.org/10.1146/annurev.cellbio.23.090506.123319

19. Stenmark H. Rab GTPases as coordinators of vesicle traffic. Nature Reviews Molecular Cell Biology [Internet]. 2009 Jul 15;10(8):513–25. Available from: https://doi.org/10.1038/nrm2728

20. Li G, Marlin MC. RAB family of GTPases. Methods in Molecular Biology [Internet]. 2015 Jan 1;1298:1–15. Available from: https://doi.org/10.1007/978-1-4939-2569-8_1

21. Kiral FR, Kohrs FE, Jin EJ, Hiesinger PR. RAB GTPASes and membrane trafficking in neurodegeneration. Current Biology [Internet]. 2018 Apr 1;28(8):R471–86. Available from: https://doi.org/10.1016/j.cub.2018.02.010

22. Park J, Kim J, Park H, Kim T, Lee S. ESCRT-III: a versatile membrane remodeling machinery and its implications in cellular processes and diseases. Animal Cells and Systems [Internet]. 2024 Jul 25;28(1):367–80. Available from: https://doi.org/10.1080/19768354.2024.2380294

23. Van Niel G, D'Angelo G, Raposo G. Shedding light on the cell biology of extracellular vesicles. Nature Reviews Molecular Cell Biology [Internet]. 2018 Jan 17;19(4):213–28. Available from: https://doi.org/10.1038/nrm.2017.125

24. Raiborg C, Stenmark H. The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. Nature [Internet]. 2009 Mar 25;458(7237):445–52. Available from: https://doi.org/10.1038/nature07961

25. Gireud-Goss M, Reyes S, Wilson M, Farley M, Memarzadeh K, Srinivasan S, et al. Distinct mechanisms enable inward or outward budding from late endosomes/multivesicular bodies. Experimental Cell Research [Internet]. 2018 Aug 23;372(1):1–15. Available from: https://doi.org/10.1016/j.yexcr.2018.08.027

26. Wenzel EM, Schultz SW, Schink KO, Pedersen NM, Nähse V, Carlson A, et al. Concerted ESCRT and clathrin recruitment waves define the timing and morphology of intraluminal vesicle formation. Nature Communications [Internet]. 2018 Jul 20;9(1):2932. Available from: https://doi.org/10.1038/s41467-018-05345-8

27. Migliano SM, Wenzel EM, Stenmark H. Biophysical and molecular mechanisms of ESCRT functions, and their implications for disease. Current Opinion in Cell Biology [Internet]. 2022 Mar 3;75:102062. Available from: https://doi.org/10.1016/j.ceb.2022.01.007

28. Olmos Y. The ESCRT machinery: remodeling, repairing, and sealing membranes. Membranes [Internet]. 2022 Jun 19;12(6):633. Available from: https://doi.org/10.3390/membranes12060633

29. Vietri M, Radulovic M, Stenmark H. The many functions of ESCRTs. Nature Reviews Molecular Cell Biology [Internet]. 2019 Nov 8;21(1):25–42. Available from: https://doi.org/10.1038/s41580-019-0177-4

30. Mercier V, Larios J, Molinard G, Goujon A, Matile S, Gruenberg J, et al. Endosomal membrane tension regulates ESCRT-III-dependent intra-lumenal vesicle formation. Nature Cell Biology [Internet]. 2020 Aug 1;22(8):947–59. Available from: https://doi.org/10.1038/s41556-020-0546-4

31. Huang LJ, Zhan ST, Pan YQ, Bao W, Yang Y. The role of Vps4 in cancer development. Frontiers in Oncology [Internet]. 2023 Jun 19;13:1203359. Available from: https://doi.org/10.3389/fonc.2023.1203359

32. Obita T, Saksena S, Ghazi-Tabatabai S, Gill DJ, Perisic O, Emr SD, et al. Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4. Nature [Internet]. 2007 Oct 1;449(7163):735–9. Available from: https://doi.org/10.1038/nature06171

33. Hill CP, Babst M. Structure and function of the membrane deformation AAA ATPase Vps4. Biochimica Et Biophysica Acta (BBA) - Molecular Cell Research [Internet]. 2011 Sep 10;1823(1):172–81. Available from: https://doi.org/10.1016/j.bbamcr.2011.08.017

34. Han H, Hill CP. Structure and mechanism of the ESCRT pathway AAA+ ATPase Vps4. Biochemical Society Transactions [Internet]. 2019 Jan 15;47(1):37–45. Available from: https://doi.org/10.1042/bst20180260

35. Mamand DR, Gustafsson O, Sork H, Wiklander RJ, Bazaz S, Liang X, et al. Evaluation of Tetraspanins in Extracellular Vesicle Bioengineering. bioRxiv [Preprint] [Internet]. 2026 Jan 13; Available from: https://doi.org/10.64898/2026.01.13.699196

36. Jiang X, Zhang J, Huang Y. Tetraspanins in cell migration. Cell Adhesion & Migration [Internet]. 2015 Jun 19;9(5):406–15. Available from: https://doi.org/10.1080/19336918.2015.1005465

37. New C, Lee ZY, Tan KS, Wong AHP, Wang DY, Tran T. Tetraspanins: host factors in viral infections. International Journal of Molecular Sciences [Internet]. 2021 Oct 27;22(21):11609. Available from: https://doi.org/10.3390/ijms222111609

38. Charrin S, Jouannet S, Boucheix C, Rubinstein E. Tetraspanins at a glance. Journal of Cell Science [Internet]. 2014 Jan 1;127(Pt 17):3641–8. Available from: https://doi.org/10.1242/jcs.154906

39. Termini CM, Gillette JM. Tetraspanins function as regulators of cellular signaling. Frontiers in Cell and Developmental Biology [Internet]. 2017 Apr 6;5:34. Available from: https://doi.org/10.3389/fcell.2017.00034

40. Toribio V, Yáñez-Mó M. Tetraspanins interweave EV secretion, endosomal network dynamics and cellular metabolism. European Journal of Cell Biology [Internet]. 2022 Apr 27;101(3):151229. Available from: https://doi.org/10.1016/j.ejcb.2022.151229

41. Andreu Z, Yáñez-Mó M. Tetraspanins in extracellular vesicle formation and function. Frontiers in Immunology [Internet]. 2014 Sep 16;5:442. Available from: https://doi.org/10.3389/fimmu.2014.00442

42. Rubinstein E, Théry C, Zimmermann P. Tetraspanins affect membrane structures and the trafficking of molecular partners: what impact on extracellular vesicles? Biochemical Society Transactions [Internet]. 2025 Mar 26;53(02):371–82. Available from: https://doi.org/10.1042/bst20240523

43. Jankovičová J, Sečová P, Michalková K, Antalíková J. Tetraspanins, More than Markers of Extracellular Vesicles in Reproduction. International Journal of Molecular Sciences [Internet]. 2020 Oct 14;21(20):7568. Available from: https://doi.org/10.3390/ijms21207568

44. Spang A. Membrane tethering complexes in the endosomal system. Frontiers in Cell and Developmental Biology [Internet]. 2016 May 8;4:35. Available from: https://doi.org/10.3389/fcell.2016.00035

45. Baker RW, Hughson FM. Chaperoning SNARE assembly and disassembly. Nature Reviews Molecular Cell Biology [Internet]. 2016 Jun 15;17(8):465–79. Available from: https://doi.org/10.1038/nrm.2016.65

46. Yoon TY, Munson M. SNARE complex assembly and disassembly. Current Biology [Internet]. 2018 Apr 1;28(8):R397–401. Available from: https://doi.org/10.1016/j.cub.2018.01.005

47. Jahn R, Cafiso DC, Tamm LK. Mechanisms of SNARE proteins in membrane fusion. Nature Reviews Molecular Cell Biology [Internet]. 2023 Oct 17;25(2):101–18. Available from: https://doi.org/10.1038/s41580-023-00668-x

48. Lou X, Shin YK. SNARE zippering. Bioscience Reports [Internet]. 2016 May 6;36(3). Available from: https://doi.org/10.1042/bsr20160004

49. Liu Y, Wan C, Rathore SS, Stowell MHB, Yu H, Shen J. SNARE Zippering Is Suppressed by a Conformational Constraint that Is Removed by v-SNARE Splitting. Cell Reports [Internet]. 2021 Jan 1;34(2):108611. Available from: https://doi.org/10.1016/j.celrep.2020.108611

50. Duman JG, Forte JG. What is the role of SNARE proteins in membrane fusion? American Journal of Physiology-Cell Physiology [Internet]. 2003 Aug 1;285(2):C237–49. Available from: https://doi.org/10.1152/ajpcell.00091.2003

51. Dacks JB, Field MC. Evolutionary origins and specialisation of membrane transport. Current Opinion in Cell Biology [Internet]. 2018 Jun 18;53:70–6. Available from: https://doi.org/10.1016/j.ceb.2018.06.001

52. Schlacht A, Herman EK, Klute MJ, Field MC, Dacks JB. Missing pieces of an ancient puzzle: evolution of the Eukaryotic Membrane-Trafficking System. Cold Spring Harbor Perspectives in Biology [Internet]. 2014 Oct 1;6(10):a016048. Available from: https://doi.org/10.1101/cshperspect.a016048

53. Thattai M. Molecular and cellular constraints on vesicle traffic evolution. Current Opinion in Cell Biology [Internet]. 2023 Jan 5;80:102151. Available from: https://doi.org/10.1016/j.ceb.2022.102151

54. Pereira-Leal JB, Seabra MC. Evolution of the rab family of small GTP-binding proteins. Journal of Molecular Biology [Internet]. 2001 Nov 1;313(4):889–901. Available from: https://pubmed.ncbi.nlm.nih.gov/11697911/

55. Jahn R, Scheller RH. SNAREs — engines for membrane fusion. Nature Reviews Molecular Cell Biology [Internet]. 2006 Aug 16;7(9):631–43. Available from: https://www.nature.com/articles/nrm2002

56. Gurkan C, Koulov AV, Balch WE. An Evolutionary perspective on eukaryotic membrane trafficking [Internet]. Madame Curie Bioscience Database - NCBI Bookshelf. 2013. Available from: https://www.ncbi.nlm.nih.gov/books/NBK6035

57. Leung KF, Dacks JB, Field MC. Evolution of the multivesicular body ESCRT machinery; retention across the eukaryotic lineage. Traffic [Internet]. 2008 Jul 14;9(10):1698–716. Available from: https://doi.org/10.1111/j.1600-0854.2008.00797.x

58. Winter V, Hauser MT. Exploring the ESCRTing machinery in eukaryotes. Trends in Plant Science [Internet]. 2006 Feb 21;11(3):115–23. Available from: https://doi.org/10.1016/j.tplants.2006.01.008

59. Dunst S, Kazimiers T, Von Zadow F, Jambor H, Sagner A, Brankatschk B, et al. Endogenously tagged RAB proteins: A resource to study membrane trafficking in Drosophila. Developmental Cell [Internet]. 2015 May 1;33(3):351–65. Available from: https://pubmed.ncbi.nlm.nih.gov/25942626/

60. Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, et al. The Reference Genome Sequence ofSaccharomyces cerevisiae: Then and Now. G3 Genes Genomes Genetics [Internet]. 2014 Jan 8;4(3):389–98. Available from: https://doi.org/10.1534/g3.113.008995

61. Kachroo AH, Vandeloo M, Greco BM, Abdullah M. Humanized yeast to model human biology, disease and evolution. Disease Models & Mechanisms [Internet]. 2022 Jun 1;15(6). Available from: https://doi.org/10.1242/dmm.049309

62. Vanderwaeren L, Dok R, Voordeckers K, Nuyts S, Verstrepen KJ. Saccharomyces cerevisiae as a Model System for Eukaryotic Cell Biology, from Cell Cycle Control to DNA Damage Response. International Journal of Molecular Sciences [Internet]. 2022 Oct 1;23(19):11665. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9570374/

63. Zieniuk B, Wierzchowska K, Jasińska K, Kobus J, Piotrowicz A, Uğur Ş, et al. Yeast as a model for human disease. International Journal of Molecular Sciences [Internet]. 2026 Feb 7;27(4):1632. Available from: https://doi.org/10.3390/ijms27041632

64. Nielsen J. Yeast Systems Biology: Model organism and Cell Factory. Biotechnology Journal [Internet]. 2019 Mar 29;14(9):e1800421. Available from: https://doi.org/10.1002/biot.201800421

65. Yaschenko AE, Alonso JM, Stepanova AN. Arabidopsis as a model for translational research. The Plant Cell [Internet]. 2024 Feb 27;37(5). Available from: https://doi.org/10.1093/plcell/koae065

66. Koornneef M, Meinke D. The development of Arabidopsis as a model plant. The Plant Journal [Internet]. 2010 Mar 1;61(6):909–21. Available from: https://doi.org/10.1111/j.1365-313x.2009.04086.x

67. Mathew MM, Prasad K. Model systems for regeneration:Arabidopsis. Development [Internet]. 2021 Mar 15;148(6). Available from: https://doi.org/10.1242/dev.195347

68. Brandizzi F. Transport from the endoplasmic reticulum to the Golgi in plants: Where are we now? Seminars in Cell and Developmental Biology [Internet]. 2017 Jul 5;80:94–105. Available from: https://doi.org/10.1016/j.semcdb.2017.06.024

69. Dahhan DA, Reynolds GD, Cárdenas JJ, Eeckhout D, Johnson A, Yperman K, et al. Proteomic characterization of isolated Arabidopsis clathrin-coated vesicles reveals evolutionarily conserved and plant-specific components. The Plant Cell [Internet]. 2022 Feb 24;34(6):2150–73. Available from: https://doi.org/10.1093/plcell/koac071

70. Bassham DC, Brandizzi F, Otegui MS, Sanderfoot AA. The secretory system of Arabidopsis. The Arabidopsis Book [Internet]. 2008 Jan 1;6:e0116. Available from: https://doi.org/10.1199/tab.0116

71. Paps J, Holland PWH. Reconstruction of the ancestral metazoan genome reveals an increase in genomic novelty. Nature Communications [Internet]. 2018 Apr 24;9(1):1730. Available from: https://doi.org/10.1038/s41467-018-04136-5

72. Tu Y, Zhao L, Billadeau DD, Jia D. Endosome-to-TGN trafficking: Organelle-Vesicle and Organelle-Organelle interactions. Frontiers in Cell and Developmental Biology [Internet]. 2020 Mar 18;8:163. Available from: https://doi.org/10.3389/fcell.2020.00163

73. Adeoye AO. Membrane trafficking mechanisms and their biological relevance. PubMed [Internet]. 2023 Oct 31; Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10998955/

74. Yang Z, Nielsen R. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Molecular Biology and Evolution [Internet]. 2000 Jan 1;17(1):32–43. Available from: https://doi.org/10.1093/oxfordjournals.molbev.a026236

75. Omer S, Harlow TJ, Gogarten JP. Does sequence conservation provide evidence for biological function? Trends in Microbiology [Internet]. 2016 Oct 20;25(1):11–8. Available from: https://doi.org/10.1016/j.tim.2016.09.010

76. Sharir-Ivry A, Xia Y. Quantifying evolutionary importance of protein sites: A Tale of two measures. PLoS Genetics [Internet]. 2021 Apr 7;17(4):e1009476. Available from: https://doi.org/10.1371/journal.pgen.1009476

77. Ribeiro AJM, Riziotis IG, Borkakoti N, Thornton JM. Enzyme function and evolution through the lens of bioinformatics. Biochemical Journal [Internet]. 2023 Nov 22;480(22):1845–63. Available from: https://doi.org/10.1042/bcj20220405

78. Ribeiro AJM, Tyzack JD, Borkakoti N, Holliday GL, Thornton JM. A global analysis of function and conservation of catalytic residues in enzymes. Journal of Biological Chemistry [Internet]. 2019 Dec 3;295(2):314–24. Available from: https://doi.org/10.1074/jbc.rev119.006289

79. Norn C, Oliveira F, André I. Improved prediction of site‐rates from structure with averaging across homologs. Protein Science [Internet]. 2024 Jun 25;33(7):e5086. Available from: https://doi.org/10.1002/pro.5086

80. Illergård K, Ardell DH, Elofsson A. Structure is three to ten times more conserved than sequence—A study of structural response in protein cores. Proteins Structure Function and Bioinformatics [Internet]. 2009 Apr 20;77(3):499–508. Available from: https://doi.org/10.1002/prot.22458

81. Huang IK, Pei J, Grishin NV. Defining and predicting structurally conserved regions in protein superfamilies. Bioinformatics [Internet]. 2012 Nov 28;29(2):175–81. Available from: https://doi.org/10.1093/bioinformatics/bts682

82. Puente-Lelievre C, Malik A, Douglas J. Protein Structural Phylogenetics. Genome Biology and Evolution [Internet]. 2025 Jul 30;17(8). Available from: https://doi.org/10.1093/gbe/evaf139

83. Henne WM, Buchkovich NJ, Emr SD. The ESCRT pathway. Developmental Cell [Internet]. 2011 Jul 1;21(1):77–91. Available from: https://doi.org/10.1016/j.devcel.2011.05.015

84. Chi RJ, Rietzel A, Segarra VA, editors. Endomembrane system diversity and evolution [Internet]. Nature. 2024. Available from: https://www.nature.com/collections/dcaeddgada/guest-editors

85. Reimann R, Kost B, Dettmer J. TETRASPANINs in plants. Frontiers in Plant Science [Internet]. 2017 Apr 18;8:545. Available from: https://doi.org/10.3389/fpls.2017.00545

86. Lipatova Z, Hain AU, Nazarko VY, Segev N. Ypt/Rab GTPases: Principles learned from yeast. Critical Reviews in Biochemistry and Molecular Biology [Internet]. 2015 Feb 23;50(3):203–11. Available from: https://doi.org/10.3109/10409238.2015.1014023

87. Jimenez-Jimenez S, Hashimoto K, Santana O, Aguirre J, Kuchitsu K, Cárdenas L. Emerging roles of tetraspanins in plant inter-cellular and inter-kingdom communication. Plant Signaling & Behavior [Internet]. 2019 Mar 4;14(4):e1581559. Available from: https://doi.org/10.1080/15592324.2019.1581559

88. Brighouse A, Dacks JB, Field MC. Rab protein evolution and the history of the eukaryotic endomembrane system. Cellular and Molecular Life Sciences [Internet]. 2010 Jun 25;67(20):3449–65. Available from: https://doi.org/10.1007/s00018-010-0436-1

89. Porfírio-Sousa AL, Tice AK, Brown MW, Lahr DJG. Phylogenetic reconstruction and evolution of the Rab GTPase gene family in Amoebozoa. Small GTPases [Internet]. 2021 Mar 29;13(1):100–13. Available from: https://doi.org/10.1080/21541248.2021.1903794

90. Nachmias D, Melnikov N, Zorea A, Sharon M, Yemini R, De-Picchoto Y, et al. Asgard ESCRT-III and VPS4 reveal conserved chromatin binding properties of the ESCRT machinery. The ISME Journal [Internet]. 2022 Oct 11;17(1):117–29. Available from: https://pubmed.ncbi.nlm.nih.gov/36221007/

91. Ito E, Uemura T. RAB GTPases and SNAREs at the trans-Golgi network in plants. Journal of Plant Research [Internet]. 2022 Apr 29;135(3):389–403. Available from: https://doi.org/10.1007/s10265-022-01392-x

92. Makarova KS, Tobiasson V, Wolf YI, Lu Z, Liu Y, Zhang S, et al. Diversity, origin, and evolution of the ESCRT systems. mBio [Internet]. 2024 Feb 21;15(3):e0033524. Available from: https://doi.org/10.1128/mbio.00335-24

93. Cocco S, Feinauer C, Figliuzzi M, Monasson R, Weigt M. Inverse statistical physics of protein sequences: a key issues review. Reports on Progress in Physics [Internet]. 2017 Nov 9;81(3):032601. Available from: https://doi.org/10.1088/1361-6633/aa9965

94. Sankar N. Comparative genomic analysis of equilibrative nucleoside transporters suggests conserved protein structure despite limited sequence identity. Nucleic Acids Research [Internet]. 2002 Oct 15;30(20):4339–50. Available from: https://pubmed.ncbi.nlm.nih.gov/12384580/

95. Fay JC, Wu CI. Sequence divergence, functional constraint, and selection in protein evolution. Annual Review of Genomics and Human Genetics [Internet]. 2003 Sep 1;4(1):213–35. Available from: https://doi.org/10.1146/annurev.genom.4.020303.162528

96. Wolf MY, Wolf YI, Koonin EV. Comparable contributions of structural-functional constraints and expression level to the rate of protein sequence evolution. Biology Direct [Internet]. 2008 Oct 7;3(1):40. Available from: https://doi.org/10.1186/1745-6150-3-40

97. Williams SG, Lovell SC. The effect of sequence evolution on protein structural divergence. Molecular Biology and Evolution [Internet]. 2009 Feb 4;26(5):1055–65. Available from: https://doi.org/10.1093/molbev/msp020

98. Bonifacino JS, Glick BS. The mechanisms of vesicle budding and fusion. Cell [Internet]. 2004 Jan 1;116(2):153–66. Available from: https://doi.org/10.1016/s0092-8674(03)01079-1

99. SüDhof TC, Rothman JE. Membrane Fusion: Grappling with SNARE and SM Proteins. Science [Internet]. 2009 Jan 22;323(5913):474–7. Available from: https://doi.org/10.1126/science.1161748

100. Saksena S, Wahlman J, Teis D, Johnson AE, Emr SD. Functional reconstitution of ESCRT-III assembly and disassembly. Cell [Internet]. 2009 Jan 1;136(1):97–109. Available from: https://doi.org/10.1016/j.cell.2008.11.013

Appendix A.

Table S1 Complete pairwise identity matrices

Comparison	% Identity	% Similarity	Alignment Length
Yeast (Vps27)-Human (Hrs)	21.1%	34.9%	848
Yeast (Vps27)-Plant (TOL2)	10.6%	20.3%	714
Human (Hrs)- Plant (TOL2)	9.4%	16.5%	859
Yeast (Vps27)-Human (Hrs)	21.1%	34.9%	848
Yeast (Vps27)-Plant (TOL6)	18.0%	32.6%	752
Human (Hrs)- Plant (TOL6)	19.8%	31.4%	880
Yeast (Hse1)- Human (STAM)	22.8%	37.6%	591
Yeast (Hse1)-Plant (TOL2)	14.7%	28.4%	538
Human (STAM)--Plant (TOL2)	15.0%	24.1%	622
Yeast (Hse1)- Human (STAM)	22.8%	37.6%	591
Yeast (Hse1)-Plant (TOL6)	15.5%	26.9%	728
Human (STAM)--Plant (TOL6)	16.4%	25.9%	793
Yeast (Stp22)- Human (Tsg101):	25.6	45.5	418
Yeast (Stp22)-Plant (ELC):	22.4	40.0	447
Human (Tsg101)-Plant (ELC):	29.2	47.0	17.6
Yeast (VPS36)-Human (VPS36)	16	29.7	607
Yeast (VPS36)-Plant (VPS36)	17.6	31.2	641
Plant (VPS36)-Human (VPS36)	33.3	49.8	454
Yeast (Snf7)-Plant (Snf7)	31.7	47.9	240
Yeast (Snf7)-Plant (Snf7.1)	28.8	42.1	240
Plant (Snf7.1)-Human (CHMP1A)	22.6	36.6	235
Plant (Snf7.1)-Human (CHMP1A)	22.2	36.2	221

Yeast (Snf7)-Human (CHMP1A)	19.9	36.7	267
Yeast (Ypt52)-Plant (Rab5A)	41.9	55.6	241
Yeast (VPS21)-Plant (Rab5A)	46	63.3	215
Yeast (Ypt51)-Human (Rab5)	50.2	63.7	223
Yeast (Ypt52)-Human (Rab5)	42.7	54.2	253
Plant (Rab5A)-Human (Rab5)	59.4	71.4	217
Yeast (Ypt7)-Human (RAB7)	61.4	74.0	215
Yeast (Ypt7)-Plant (RAB7)	35.5	52.7	220
Plant (RAB7)-Human (RAB7)	59.4	71.4	217
Yeast (VPS4)-Human (VPS4)	59.0	73.6	451
Yeast (VPS4)-Plant (VPS4)	54.8	68.1	451
Plant (VPS4)-Human (VPS4)	53.0	68.8	455
Yeast (VAM3)-Human (STX10)	16.6	38.0	295
Yeast (VAM3)-Plant (SYP121)	18.6	33.2	392
Plant (SYP121)-Human (STX10)	13.2	25.7	385
Plant (TET8)-Human (CD63)	21.8	37.3	303
Plant (TET8)-Human (CD81)	22.0	33.7	300
Plant (TET8)-Yeast (SUR7)	11.2	16.3	412
Yeast (SUR7)-Human (CD63)	1.2	2.5	520
Yeast (SUR7)-Human (CD81)	2.0	3.5	509

Table 1: Complete pairwise identity matrices for all protein families

Appendix B.

Supplementary Table S2 Domain / Motif position and residue conservation data

Protein Family	Protein names (Yeast-Plant-Human)	Domain / Motif to Verify	Domain / Motif Position (in yeast protein raw sequence)	Conserved residues / motifs (in yeast)	Motif Pattern / Structural Feature	Functional significance	Conservation Status
ESCRT-0	VPS27-Hrs-TOL2	VHS domain	18-149	LEIS MRCIKKRI SWKLTN KEICSREFMDTM	N-terminal conserved fold	Cargo recognition	Some conserved residues observed across species
		UIM 1	258-227	LS	Short acidic-hydrophobic motifs	Ubiquitin binding	Low conservation across species (no consecutive conserved residues)
		UIM 2	301-320		Short acidic-hydrophobic motifs	Ubiquitin binding	No fully conserved residues
		FYVE domain	170-230		Cys-rich Zn-binding motif	PI3P binding	Low conservation (no consecutive conserved residues)
	VPS27-Hrs-	VHS domain	18-149	IEQATSE	N-terminal	Cargo recognition	Some conserved

	TOL6			LEISDVL VKNGG EICSREGIKFP	conserved fold		residues observed across species
		UIM 1	258-257	DE	Short acidic–hydrophobic motifs	Ubiquitin binding	Low conservation across species
		UIM 2	301-320		Short acidic–hydrophobic motifs	Ubiquitin binding	No conservation observed across species
		FYVE domain	170-230		Cys-rich Zn-binding motif	PI3P binding	No conservation
	Hse1-STAM-TOL2	VHS domain	15-145	KATIEKRLLIESEISS	N-terminal conserved fold	Cargo recognition	Some conserved residues observed across species
		UIM motifs	162-181	DDEELKMS	Short acidic–hydrophobic motifs	Ubiquitin binding	Low conservation across species
		SH3 domain (STAM-like)	217-276	YDL ELS	β-barrel SH3 fold residues	Protein–protein interaction	Low conservation

				DVIT NYV			
	Hse1-STAM-TOL6	VHS domain	15-145	KAT LDVCDL V TLSL NCG LRQEISS KNF LIES DKIKRK A	N-terminal conserved fold	Cargo recognition	Some conserved residues observed across species
		UIM motifs	162-181		Short acidic–hydrophobic motifs	Ubiquitin binding	No consecutive conservation across species
		SH3 domain (STAM-like)	217-276		β-barrel SH3 fold residues	Protein–protein interaction	No consecutive conservation across species
ESCRT-I	STP22-TSG101-ELC	UEV domain (N-terminal)	12-161	HDSLALL FTHSDG GTI HSIPVIMWV YPVKPP LPILH	UEV fold signature residues	Ubiquitin binding	Some conserved residues observed across species

				NLIMVV EPP			
		C-terminal Vps23 core / SB domain	272-300	NQL DTI LDTFVKQ, LARQQF	Conserved core residues	ESCRT-I complex assembly	Some conserved residues observed across species
		Coiled Coil	322- 385	NQL DTI LDTFVKQ LARQQF			Very low conservation
ESCRT-II	VPS36-VPS36-VPS36	GLUE N-terminal domain	7-96	SSG GRIFLTSQRIIY	Phosphoinositide-binding fold	Membrane targeting	Low conservation across the species
		GLUE C-terminal domain	255-288	FR EN	Conserved structured region	ESCRT-III recruitment	Very low conservation across the species
		Zinc finger motif (RanBP2-type 1)	114-151		Cys/His clusters	Structural stabilization	No conservation in plant and human

		Zinc finger motif (RanBP 2-type 2)	177-205		Cys/His clusters	Structural stabilizatio n	No conservation in plant and human
ESCRT-III	Snf7-Snf7-Snf7.1-CHMP1A	N-termi nal helical core	20-120	LREH	Conserv ed alpha-h elical region	Polymeriza tion scaffold	Low conservation across the species
		C-termi nal regulato ry tail	160-210	EEEE	Variable acidic region	VPS4 interaction / regulation	Low conservation across the species
		MIM (MIT-int eracting motif)		LRELQA	LxxLxx L-like motif	VPS4 binding	Low conservation across the species
RAB5	YPT52-VPS21-RAB5A-RAB5	P-loop (Walke r A)	18-25	GDSSVG	Gxxxx GKST	GTP binding	Moderate conservation across the species
		Switch I region	47-55		Conserv ed flexible loop (Thr/Ser -rich)	Effector interaction	High conservation across the species
		Switch II region	60-64	DTAGQ	Conserv ed DxxG motif region	Conformati onal switching	High conservation across the species

		NKXD motif	120-123	NKXD	NKXD	Guanine specificity	High conservation across the species
		C-terminal tail prenylation motif		CCS	CC / CXC / Cys-containing tail	Membrane attachment	Moderate conservation across the species
RAB7	YPT7-RAB7-RAB7	P-loop (Walker A)	10-17	GDSGVGKT	GxxxxGKST	GTP binding	High conservation across the species
		Switch I region	35-42		Conserved flexible loop (Thr/Ser-rich)	Effector interaction	High conservation across the species
		Switch II region	57-61	DTAGQ	Conserved DxxG motif region	Conformational switching	High conservation across the species
		NKXD motif	125-128	NKID	NKXD	Guanine specificity	High conservation across the species

		C-terminal tail prenylation motif		CCS	CC / CXC / Cys-containing tail	Membrane attachment	High conservation across the species
Disassembly	VPS4-VPS4-VPS4	Walker A	173-180	GPPGTGKS	GxxxxGKST	ATP binding	Very high conservation across the species
		Walker B	229-234	IIFIDE	hhhhDE	ATP hydrolysis	Very high conservation across the species
		Arginine finger	287-289 aa	RRR	Conserved Arg residue	ATPase oligomer function	Very high conservation across the species
		Pore loop residues	206-209	WMGE	Conserved aromatic/hydrophobic residues	Substrate translocation	Very high conservation across the species
SNAREs	VAM3-SYP121-STX10	t-SNARE coiled-coil homology	190-252	QER VKEQG	Coiled-coil heptad repeats	Membrane fusion complex	Moderate conservation across the species

		C-terminal TM helix	262-282	VTLII IVV VVLL	Hydrophobic stretch (~20-25 aa)	Membrane anchoring	Moderate conservation across the species
Tetraspanins	SUR7-TET8-CD63-CD81	Four transmembrane helices	TM1: 6-26 TM2: 117-137 TM3: 141-161 TM4: 189-209	LV MLLI	Four hydrophobic segments	Membrane scaffold	Very low conservation across the species
		Extracellular Topological domains	27-116	G	Conserved Cys pairs	Disulfide bond formation	Very low conservation
		Short cytoplasmic tails	N-terminal: 1-5 C-terminal: 210-309	M	Basic charged residues	Signalling regulation	Very low conservation across the species)

Table 2: Domain / Motif position and residue conservation data

Appendix C.

Supplementary Figure S1 Full multiple sequence alignments for each protein family

ESCRT-0 (Vps27-Hrs-TOL2)

Figure 1.1: MAFFT multiple sequence alignment result for ESCRT-0 (Vps27-Hrs-TOL2)

ESCRT-0 (Vps27-Hrs-TOL6)

Figure 1.2: MAFFT multiple sequence alignment result for ESCRT-0 (Vps27-Hrs-TOL6)

ESCRT-0 (Hse1-STAM-TOL2)

```
sp|P38753| MSSSAI-------------------KIRNA----------LLKATDPKLRSDNWQYILD
NP_003464. MPLFATNPFDQD---------------------------VEKATSEMNTAEDWGLILD
sp|Q9LNC6| MDKLKIAEWGEKLKTGGAQMSRMVSEKVKDMLQAPTLESKMVDEATLETLEEPNWGMNMR
           *                                      : :**       :*   :

sp|P38753| VCDLVKEDPEDNGQEVMSLIEKRLEQQDANVILRTLSLTVSLAENCGSRLRQEISSKNFT
NP_003464. ICDKVGQS-RTGPKDCLRSIMRRVNHKDPHVAMQALTLLGACVSNCGKIFHLEVCSRDFA
sp|Q9LNC6| ICAQINND-EFNGTEIVRAIKRKISGKSPVSQRLSLELLEACAMNCEKVFS-EVASEKVL
           :*  : :. . .  : :  * :::. :..     :* *  : . ** . :  *:.*...

sp|P38753| SLLYALIESHSVHITLKKAVTDVVKQLSDSFKDDPSLRAMGDLYDKIKRKAPYL--VQPN
NP_003464. SEVSNVLNKG--HPKVCEKLKALMVEWTDEFKNDPQLSLISAMIKNLKEQGVTFPAIGSQ
sp|Q9LNC6| DEMVWLIKNGEADSENRKRAFQLIRAWG-------------------QSQDLTYLPVFHQ
           . :  :::.   .    :    ::                      : :      :  :

sp|P38753| VPEKHNMS---------TQADNSDDEELQKALKMSLFEYEKQKKLQEQEKESAEVLPQQQ
NP_003464. AAEQAKASPALVAKDPGTVANKKEEEDLAKAIELSLKEQRQQSTTLSTLYPSTSSLLTNH
sp|Q9LNC6| TYMSLEGENGLHARG--------EENSMPGQSSLESLMQRPVPVPPPGSYP----VPNQE
           .  . : .              :::.:    .:.    .               :  :.

sp|P38753| QQHQQQNQAPAHKIPAQTVVRRVRALYDL------TTNEPDELSFRKGDVITVLEQVYRD
NP_003464. QHEG----------------RKVRAIYDF------EAAEDNELTFKAGEIITVLDDSDPN
sp|Q9LNC6| QALG----------------DDDGLDYNFGNLSIKDKKEQIEITRNSLELLSSMLNTEGK
           *                        *::         *  *:: .  :::: : :   .

sp|P38753| WWKGALRGNMGIFPLNYVTPIVEPS-KEEIEKEKN---------------KEAIVFSQKT
NP_003464. WWKGETHQGIGLFPSNFVTADLTAE-PEMIKTEKKTVQFSDDVQVETIEPEPEPAFIDED
sp|Q9LNC6| --------------PNHTEDDLTVSLMEKCKQSQPLIQ----MIIESTTDDEGVLFEALH
                       *..   :  .  *  : .:               .    *

sp|P38753| TIDQLHNSLNAASKTGNSNEVLQDPHIGDM---------------------------YG
NP_003464. KMDQLLQMLQSTDPSDDQPDLPELLHLEAMC-HQMGPLIDEKLE---------DIDRKHS
sp|Q9LNC6| LNDELQQVLSSYKKPDETEKKASIVEQESSGSKDTGPKPTEQEEQEPVKKTGADDDKKHS
             *:* : *.: . ..:  .  .  .                                 :.

sp|P38753| SVTPLRPQVTRMLGKYAKE--KEDMLSLRQVLANAERSYNQLMDRAANAHISPPVPGPAL
NP_003464. ELSELNVKVMEALSLYTKLMNEDPMYSMYAKLQNQPYYMQSSGVSGSQVYAGPPPSGAYL
sp|Q9LNC6| EASGSSNKTVKEEKQAVKI-----------------------------------------
           . :    :. .     .*

sp|P38753| YAG---MTHANNTPVMPPQRQSYQSNEYSP---------------------------YPS
NP_003464. VAGNAQMSHLQSYSLPPEQLSSLSQAVVPPSANPALPSQQTQAAYPNTMVSSVQGNTYPS
sp|Q9LNC6| ------------------------------------------------------------


sp|P38753| NLPIQHP----------------TNSANNTPQY-GYDLGYSVVSQP-----PPGYEQ---
NP_003464. QAPVYSPPPAATAAAATADVTLYQNAGPNMPQVPNYNLTSSTLPQPGGSQQPPQPQQPYS
sp|Q9LNC6| ----------------------------------ELGLS----------SDEDEK---
                                             :*  *          .   ::

sp|P38753| -----
NP_003464. QKALL
sp|Q9LNC6| -----
```

Figure 1.3: MAFFT multiple sequence alignment result for ESCRT-0 (Hse1-STAM-TOL2)

ESCRT-0 (Hse1-STAM-TOL6)

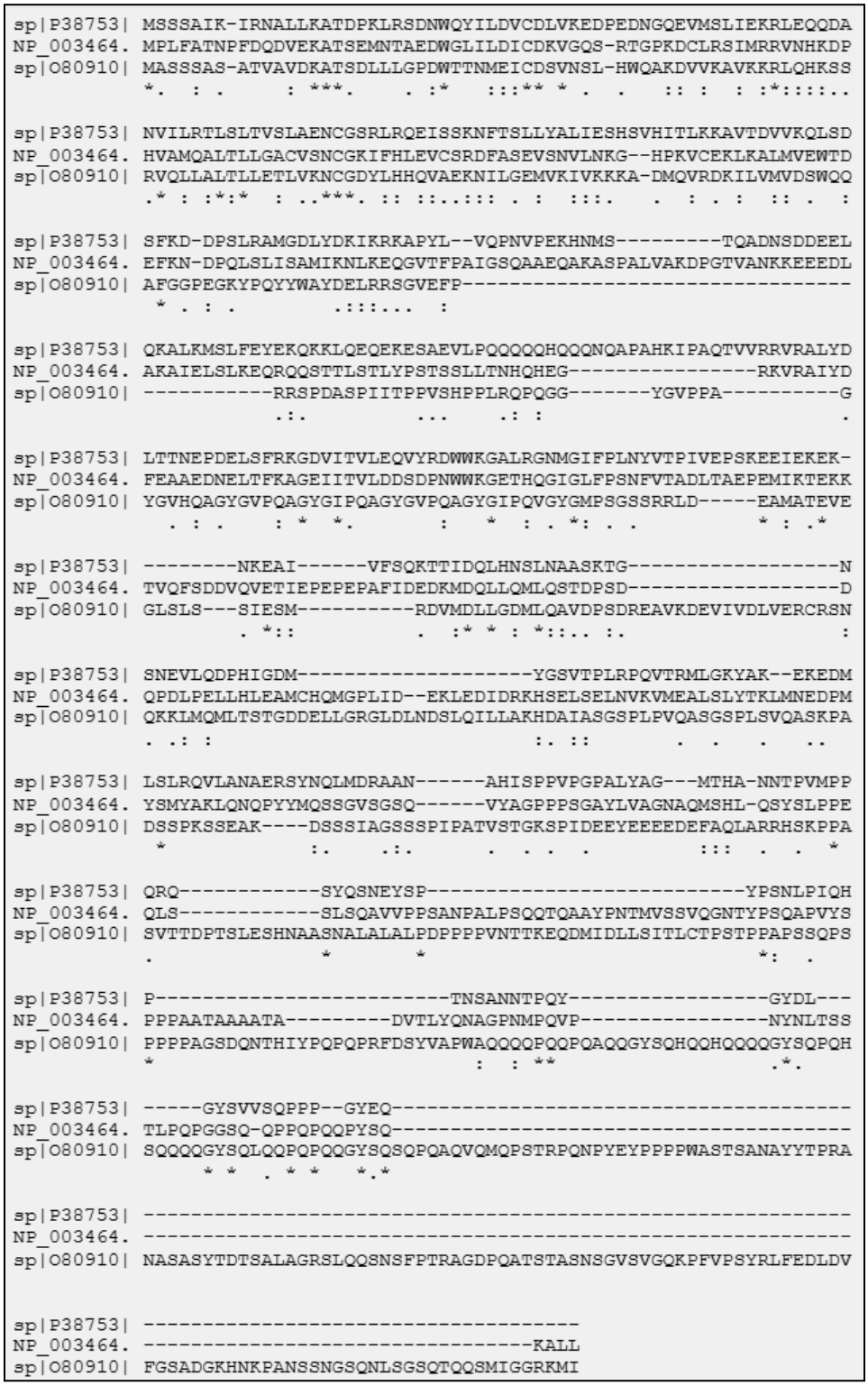


Figure 1.4: MAFFT multiple sequence alignment result for ESCRT-0 (Hse1-STAM-TOL6)

ESCRT-I (STP22-TSG101-ELC)

```
sp|P25604| MSANGKISVP-------------EAVVNWLFKVIQPIYNDGRTTFHDSLALLDNFHSLRP
NP_006283. M----------------------AVSESQLKKMVSKYKYRDLTVRETVNVITLYKDLKP
sp|Q9LHG8| M-------VPPPSNPQQVQQFLSSALSQRGPSSVPYEESNKWLIRQHLLNLISSYPSLEP
           *                      *: :   . :    .      :. : ::  : .*.*

sp|P25604| RTRVFTHSDGTPQLLLSIYGTISTGEDGSSPHSIPVIMWVPSMYPVKPPFISINLENFDM
NP_006283. VLDSYVFNDGSSRELMNLTGTIPVPYRGNT-YNIPICLWLLDTYPYNPPICFVK------
sp|Q9LHG8| KTASFMHNDGRSVNLLQADGTIPMPFHGVT-YNIPVIIWLLESYPRHPPCVYVN------
              : ..** .  *:.  ***.    * : :.**: :*: . ** :**   ::

sp|P25604| NTISSSLP-IQEYIDSNGWIALPILHCWDPAAMNLIMVVQELMSLL-HEPPQDQAPSLPP
NP_006283. PTSSMTIK-TGKHVDANGKIYLPYLHEWKHPQSDLLGLIQVMIVVFGDEPPVFSRPISAS
sp|Q9LHG8| PTADMIIKRPHAHVTPSGLVSLPYLQNWVYPSSNLVDLVSDLSAAFARDPPLYSRRRPQP
            * .  :     :: ..* : ** *: *  .  :*: ::. :   :  :**  .     .

sp|P25604| KPNTQLQQEQNT--PPLPPKPKS----PHLKPPLP--------PPPPPQPASNALDLMDM
NP_006283. YPPYQATGPPNT--SYMPGMPGGISPYPSGYPPNPSGYPGCPYPPGGPYPATTSSQYPSQ
sp|Q9LHG8| PPP----SPPTVYDSSLSRPPSADQSLPRPFPPSPYG--------GGVSRVQVQHVHHQQ
            *        ..  . :.  * .    *   ** *              .       .

sp|P25604| DN-TDISPTNHHEMLQNLQTVVNELYREDVDYVADKILTRQT----VMQE---SIARFHE
NP_006283. PPVTTVGPSRDGTISE--DTIRASL----ISAVSDKLRWRMKEEMDRAQAELNALKRTEE
sp|Q9LHG8| QSDDAAEVFKRNAINKMVEMVHSDL----VSM--RRAREAEAEELLSLQA---GLKRRED
                    .   : :  : :  .*    :.    :            *    .: * .:

sp|P25604| IIAIDKNHLRAVEQAIEQTMHSLNAQIDVLTANRAKVQQFSSTSHVDDE----DVNSIAV
NP_006283. DLKKGHQKLEEMVTRLDQEVAEVDKNIELLKKKDEELS--SALEKMENQSENNDIDEVII
sp|Q9LHG8| ELNIGLKEMVEEKETLEQQLQIISMNTDILDSWVRENQ--GKTKNLVDL----DVDNAFE
            :  . :.:      ::* :  :. : ::*     : .  .  .:: :     *::.

sp|P25604| AKTDGLNQLYNLVAQDYALTDTIECLSRMLHRGTIPLDTFVKQGRELARQQFLVRWHIQ-
NP_006283. PTAPLYKQILNLYAEENAIEDTIFYLGEALRRGVIDLDVFLKHVRLLSRKQFQLRALMQK
sp|Q9LHG8| CGDTLSKQMLECTALDLAIEDAIYSLDKSFQDGVVPFDQYLRNVRLLSREQFFHRATGSK
                 :*: :  * : *: *:*  *.. :: *.: :* :::: * *:*:**  *   .

sp|P25604| -RIT------SPLS-----
NP_006283. ARKT------AGLSDL--Y
sp|Q9LHG8| VRAAQMEVQVAAIAGRLHS
            * :      : ::
```

Figure 1.5: MAFFT multiple sequence alignment result for ESCRT-I (STP22-TSG101-ELC)

ESCRT-II (VPS36-VPS36-VPS36)

Figure 1.6: MAFFT multiple sequence alignment result for ESCRT-II (VPS36-VPS36-VPS36)

ESCRT-III (Snf7-Snf7-Snf7.1-CHMP1A)

```
sp|P39929| MWSSLFGWTSSNAKNKESPTKAIVRLREHINLLSKKQSHLRTQITNQENEARIFLTKGN-
sp|P39929| MWSSLFGWTSSNAKNKESPTKAIVRLREHINLLSKKQSHLRTQITNQENEARIFLTKGN-
sp|Q9SZE4| MMNRLFG----KPKQEANALQTLDKLNETLEMLEKKEKVLLKKAGAEVEKAKEYSRAKN-
sp|Q9SZE4- ------------------------------MLEKKEKVLLKKAGAEVEKAKEYSRAKN-
NP_002759. MDDTLF------------------QLKFTAKQLEKLAKKAEKDSKAEQAKVKKALLQKNV
                                        *.*  .   ..   :  :.:     *

sp|P39929| ---KVMAKNALKKKKTIEQLLSKVEGTMESMEQQLFSIESANLNLETMRAM-QEGAKAMK
sp|P39929| ---KVMAKNALKKKKTIEQLLSKVEGTMESMEQQLFSIESANLNLETMRAM-QEGAKAMK
sp|Q9SZE4| ---KRAAIQCLKRKRLYEGQVEQLGNFQLRIHDQMIMLEGAKATTETVDAL-RSGASAMK
sp|Q9SZE4- ---KRAAIQCLKRKRLYEGQVEQLGNFQLRIHDQMIMLEGAKATTETVDAL-RSGASAMK
NP_002759. ECARVYAENAIRKK--------NEGVNWLRMASRVDAVASKVQTAVTMKGVTKNMAQVTK
              :  * :.:::*        :       : .::  : .   .  *: .: :. *.. *

sp|P39929| TIHSGL---DIDKVDETMDEIREQVELGDEISDAISRPLITGANEVD--EDELDEELDML
sp|P39929| TIHSGL---DIDKVDETMDEIREQVELGDEISDAISRPLITGANEVD--EDELDEELDML
sp|Q9SZE4| AMQKAT---NIDDVDKTMDEINEQTENMKQIQEALATPM-GAAADFD--EDELAAELDEL
sp|Q9SZE4- AMQKAT---NIDDVDKTMDEINEQTENMKQIQEALATPM-GAAADFD--EDELAAELDEL
NP_002759. ALDKALSTMDLQKVSSVMDRFEQQVQNLDVHTSVMEDSM-SSATTLTTPQEQVDSLIMQI
           ::...    :::.*...**.:.:*.:  .   ..:  .:  .*  .   ::::   :  :

sp|P39929| AQENANQETSKIVNNNVNAAPISENKVSLPSVPSNKIKQSENSVKDGEEEEDEEDEDEKA
sp|P39929| AQENANQETSKIVNNNVNAAPISENKVSLPSVPSNKIKQSENSVKDGEEEEDEEDEDEKA
sp|Q9SZE4| ESE---ELESQLLQPATTAPP-------LPSVPVPAGRQPARPVPQKRTAEEEE------
sp|Q9SZE4- ESE---ELESQLLQPATTAPP-------LPSVPVPAGRQPARPVPQKRTAEEEE------
NP_002759. AEENGLEVLDQ-----------------LSQLPEGASAVGESSV---RSQEDQ-------
            .*   :  .:                 *..:*         .*   .  *::

sp|P39929| LRELQAEMGL
sp|P39929| LRELQAEMGL
sp|Q9SZE4| LAALQAEMAL
sp|Q9SZE4- LAALQAEMAL
NP_002759. LSRRLAALRN
           *    * :
```

Figure 1.7: MAFFT multiple sequence alignment result for ESCRT-III (Snf7-Snf7-Snf7.1-CHMP1A)

Rab GTPases (YPT52-VPS21-RAB5A-RAB5)

```
sp|P36018| M-----------------LQFKLVLLGDSSVGKSSIVHRFVKDTFDELRESTIGAAFLSQ
sp|P36017| M-------------NTSVTSIKLVLLGEAAVGKSSIVLRFVSNDFAENKEPTIGAAFLTQ
sp|P31582| MA----------SSGNKNINAKLVLLGDVGAGKSSLVLRFVKDQFVEFQESTIGAAFFSQ
NP_004153. MASRGATRPNGPNTGNKICQFKLVLLGESAVGKSSLVLRFVKGQFHEFQESTIGAAFLTQ
           *                  . ******: ..****:* ***.. * * :*.******::*

sp|P36018| SITIHPNDGNETKDVVIKFEIWDTAGQERYKSLAPMYYRNANAALVVYDITQEDSLQKAR
sp|P36017| RVTIN--------EHTVKFEIWDTAGQERFASLAPMYYRNAQAALVVYDVTKPQSFIKAR
sp|P31582| TLAVN--------DATVKFEIWDTAGQERYHSLAPMYYRGAAAAIIVFDITNQASFERAK
NP_004153. TVCLD--------DTTVKFEIWDTAGQERYHSLAPMYYRGAQAAIVVYDITNEESFARAK
            : :.        : .:************: ********.* **::*:*:*:  *: :*:

sp|P36018| NWVDELKNKVGDDDLVIYLLGNKVDLCQETPSTETSPDSNEGGDEEQKVRAISTEEAKQY
sp|P36017| HWVKELHEQ-ASKDIIIALVGNKIDMLQ-----------EGG-----ERKVAREEGEKL
sp|P31582| KWVQELQAQ-GNPNMVMALAGNKADLLD--------------------ARKVSAEEAEIY
NP_004153. NWVKELQRQ-ASPNIVIALSGNKADLAN--------------------KRAVDFQEAQSY
           :**.**: : .. :::: * *** *: :                     * :  :*.:

sp|P36018| AQEQGLLFREVSAKTGEGVKEIFQDIGEKLYDLKKDEILSKQNRQIGGGNNGQVDINLQR
sp|P36017| AEEKGLLFFETSAKTGENVNDVFLGIGEKIPLKTAEEQNSASNER--ESNNQRVDLNAAN
sp|P31582| AQENSLFFMETSAKTATNVKDIFYEIAKRLPRVQPA-ENP-------------TGMVLPN
NP_004153. ADDNSLLFMETSAKTSMNVNEIFMAIAKKLPKNEP--QNPGAN----SARGRGVDLTEPT
           *:::.*:* *.****. .*:::*  *.:::         .             ..:

sp|P36018| PS--TNDPTSCCS-
sp|P36017| DG--TSANSACSC-
sp|P31582| GPGATAVSSSCCA-
NP_004153. QP----TRNQCCSN
                  . *..
```

Figure 1.8: MAFFT multiple sequence alignment result for Rab GTPases (YPT52-VPS21-RAB5A-RAB5)

Rab GTPases (YPT7-RAB7-RAB7)

```
sp|P32939| MSSRKKNILKVIILGDSGVGKTSLMHRYVNDKYSQQYKATIGADFLTKEVTVDGDKVATM
sp|O04157| MSTRRRTLLKVIILGDSGVGKTSLMNQYVNNKFSQQYKATIGADFVTKELQID-DRLVTL
NP_004628. MTSRKKVLLKVIILGDSGVGKTSLMNQYVNKKFSNQYKATIGADFLTKEVMVD-DRLVTM
           *::*:: :******************::***.*:*:**********:***: :* *::.*:

sp|P32939| QVWDTAGQERFQSLGVAFYRGADCCVLVYDVTNASSFENIKSWRDEFLVHANVNSPETFP
sp|O04157| QIWDTAGQERFQSLGVAFYRGADCCVLVYDVNHLKSFESLDNWHNEFLTRASPRDPMAFP
NP_004628. QIWDTAGQERFQSLGVAFYRGADCCVLVFDVTAPNTFKTLDSWRDEFLIQASPRDPENFP
           *:**************************:**.  .:*:.:..*::*** :*. ..*  **

sp|P32939| FVILGNKIDAE-ESKKIVSEKSAQELAKSLGDIPLFLTSAKNAINVDTAFEEIARSALQQ
sp|O04157| FILLGNKVDIDGGNSRVVSEKKAREWCAEKGNIVYFETSAKEDYNVDDSFLCITKLALAN
NP_004628. FVVLGNKIDLE---NRQVATKRAQAWCYSKNNIPYFETSAKEAINVEQAFQTIARNALKQ
           *::****:* :   .: *: * *:  . . .:*  * ****:  **: :*  *:: ** :

sp|P32939| NQADTEAF-ED-------DYNDAINIRLDGENNSCSC
sp|O04157| -ERDQDIYFQP-------DTGSVPEQR-----GGCAC
NP_004628. -ETEVELY-NEFPEPIKLDKND----RAKASAESCSC
            : : : : :        * ..    *      .*:*
```

Figure 1.9: MAFFT multiple sequence alignment result for Rab GTPases (YPT7-RAB7-RAB7)

Disassembly (VPS4-VPS4-VPS4)

```
sp|P52917| MSTGDFLTKGIELVQKAIDLDTATQYEEAYTAYYNGLDYLMLALKYEKNP-KSKDLIRAK
NP_037377. M-TTSTLQKAIDLVTKATEEDKAKNYEEALRLYQHAVEYFLHAIKYEAHSDKAKESIRAK
sp|Q9ZNT0| M-YSNFKEQAIEYVKQAVHEDNAGNYNKAFPLYMNALEYFKTHLKYEKNP-KIREAITQK
           *   .   :.*: * :* . *.* :*::*   * :.::*:   :*** :. * :: *  *

sp|P52917| FTEYLNRAEQLKKHLESEEANA---AKKSPSAGSGSNGGNKKISQEEGEDNGGEDNKKLR
NP_037377. CVQYLDRAEKLKDYLRSKEKHGKKPVKENQSEGKGSD------SDSEGDN---PEKKKLQ
sp|Q9ZNT0| FTEYLRRAEEIRAVL---DEGGSGPGSNGDAAVATRPKTKPKDGEGGGKDGEDPEQSKLR
            .:** ***:::  *   :  .    .:. :           .:  *.:    ::.**:

sp|P52917| GALSSAILSEKPNVKWEDVAGLEGAKEALKEAVILPVKFPHLFKGNRKPTSGILLYGPPG
NP_037377. EQLMGAVVMEKPNIRWNDVAGLEGAKEALKEAVILPIKFPHLFTGKRTPWRGILLFGPPG
sp|Q9ZNT0| AGLNSAIVREKPNIKWSDVAGLESAKQALQEAVILPVKFPQFFTGKRRPWRAFLLYGPPG
             * .*:: ****::*.******.**:**:******:***::*.*:* *  .:**:****

sp|P52917| TGKSYLAKAVATEA-NSTFFSVSSSDLVSKWMGESEKLVKQLFAMARENKPSIIFIDEVD
NP_037377. TGKSYLAKAVATEANNSTFFSVSSSDLMSKWLGESEKLVKNLFELARQHKPSIIFIDEVD
sp|Q9ZNT0| TGKSYLAKAVATEA-DSTFFSVSSSDLVSKWMGESEKLVSNLFEMARESAPSIIFVDEID
           ************** :***********:***:*******.:** :**:  *****:**:*

sp|P52917| ALTGTRGE-GESEASRRIKTELLVQMNGVGNDSQGVLVLGATNIPWQLDSAIRRRFERRI
NP_037377. SLCGSRNE-NESEAARRIKTEFLVQMQGVGNNNDGTLVLGATNIPWVLDSAIRRRFEKRI
sp|Q9ZNT0| SLCGTRGEGNESEASRRIKTELLVQMQGVGHNDEKVLVLAATNTPYALDQAIRRRFDKRI
           :* *:*.* .****:******:****:***::.: .***.*** *: **.******::**

sp|P52917| YIPLPDLAARTTMFEINVGDTPCVLTKEDYRTLGAMTEGYSGSDIAVVVKDALMQPIRKI
NP_037377. YIPLPEEAARAQMFRLHLGSTPHNLTDANIHELARKTEGYSGADISIIVRDSLMQPVRKV
sp|Q9ZNT0| YIPLPEAKARQHMFKVHLGDTPHNLTEPDFEYLGQKTEGFSGSDVSVCVKDVLFEPVRKT
           *****:  **  **.:::*.**  **. : . *.  ***:**:*::: *:* *::*:**

sp|P52917| QSATHFKDVSTEDDETRK------LTPCSPGDDGAIEMSWTDIE----ADELKEPDLTIK
NP_037377. QSATHFKKVCGPSRTNPSMMIDDLLTPCSPGDPGAMEMTWMDVP----GDKLLEPVVCMS
sp|Q9ZNT0| QDAMFFFK-----------SPDGTWMPCGPRHPGAIQTTMQDLATKGLAEKIIPPPITRT
           *.* .* .                **.* . **:: :  *:     .:::  * :  .

sp|P52917| DFLKAIKSTRPTVNEDDLLKQEQFTRDFGQEGN
NP_037377. DMLRSLATTRPTVNADDLLKVKKFSEDFGQES-
sp|Q9ZNT0| DFEKVLARQRPTVSKSDLDVHERFTQEFGEEG-
           *: : :   ****. .**   ::*:.:**:*.
```

Figure 1.10: MAFFT multiple sequence alignment result for Disassembly (VPS4-VPS4-VPS4)

SNAREs (VAM3-SYP121-STX10)

Figure 1.11: MAFFT multiple sequence alignment result for SNAREs (VAM3-SYP121-STX10)

Tetraspanins (SUR7-TET8-CD63-CD81)

```
sp|Q07651| MIFK---RFVNLLVFLFLLGAGLLTFFL--ILSGGRESGTLKNFYWLQADTNGFNSAPST
sp|Q8S8Q6| MA-----RCSNNLVGIL-------NFLV--FLLSIPILAGG---IWLSQK----------
NP_001771. MAVEGGMKCVKFLLYVL-------LLAFCACAVGLIAVGVG---AQL-------------
NP_004347. MGVEGCTKCIKYLLFVF-------NFVF--WLAGGVILGVA---LWLRHD-------PQT
           *      :  : *: ::        : .     .    .       *

sp|Q07651| TRWYNYNWCGYEDGQLANCSSRAPAKPFSPRDNFGNSVNLPSSFRNNRDTYYYLSRVGWA
sp|Q8S8Q6| ----------------------GSTE--------------CERFLD-----KPVIALGVF
NP_001771. -------VLSQTIIQ-------GAT---------------PGSLLP-----VVIIAVGVF
NP_004347. T-----NLLYLELGD-------KPA---------------PNTFYVG---IYILIAVGAV
                               .:                  :         :  :*

sp|Q07651| MLLISLFFIVLALVPGFLATFLPFKA-VPVLYCVLSWLAFFFIILAACLYTGCYVKARKT
sp|Q8S8Q6| LMVV-----------AIAGL---IGSCCRVTWLLWVYLFVMFLLILLVF---CIT----V
NP_001771. LFLV-----------AFVGC---CGA-CKENYCLMITFAIFLSLIMLV------------
NP_004347. MMFV-----------GFLGC---YGA-IQESQCLLGTFFTCLVILFAC------------
           ::.:           .: .      :       :   :   : ::

sp|Q07651| FRNSGRSARLGPKNFAFIWTSVFLMLVNAIWSTIFSAT---HKAHSTYS-------DHDM
sp|Q8S8Q6| FAFVVTNKGAGEAIEGKGYKEYKLGDYSTWLQKRVENG---KNWNKIRSCLVESKVCSKL
NP_001771. --------EVAAAIAGYVFRDKVMSEFNNNFRQQMENY---PKNNHTASIL------DRM
NP_004347. --------EVAAGIWGFVNKDQIAKDVKQFYDQALQQAVVDDDANNAKAVV------KTF
                     .    .    .      .      ..      . :   :         :

sp|Q07651| YAQYESPSVDTGAQMEKSTYNSGATDGA-----GPITAAPVVGQPQPTTTTTPAGNGKFF
sp|Q8S8Q6| EAKFVNVPVNSFYKEHLTALQSGCCKPSDECGFEYVNPTTWTKNTTGTHTNPDCQTWDNA
NP_001771. QADF-------------------KCCGAANYTDWEKIPSMSKNRVPDSCCINVTVGCGINF
NP_004347. HETL-----------------DCCGSS-----------TLTALTTSVLKNNLCPSGSNI
                               .   :          .    .     .

sp|Q07651| QKLKTRKQVPSAELEPAGDGGLAGPV-------------------------TVRD----
sp|Q8S8Q6| KEKLCF------DCQSCKAGLLDNVKSAWKKVAIVNIVFLVFLIIVYSVGCCAFRNNKRD
NP_001771. NEKAIHKE----GCVEKIGGWLRKNVLVVAAAALGIAFVEVLGIVFACCLVKSIRSG---
NP_004347. ISNL-FKE----DCHQKIDDLFSGKLYLIGIAAIVVAVIMIFEMILSMVLCCGIRNSSV-
            .                 . :                                .*.

sp|Q07651| ------------
sp|Q8S8Q6| DSYSRTYGYKP-
NP_001771. --------YEVM
NP_004347. --------Y---
```

Figure 1.12: MAFFT multiple sequence alignment result for Tetraspanins (SUR7-TET8-CD63-CD81)

Appendix D.

Supplementary Data S1 Phylogenetic tree files (Newick format)

ESCRT-0 (Vps27-Hrs-TOL2):

((((BAA23366.1_Hrs_Homo_sapiens,sp|Q99LI8|HGS_MOUSE_Hepatocyte_growth_factor-regulated_tyrosine_kinase_substrate_OS=Mus_musculus_OX=10090_GN=Hgs_PE=1_SV=2)1.0000,sp|Q960X8|HRS_DROME_Hepatocyte_growth_factor-regulated_tyrosine_kinase_substrate_OS=Drosophila_melanogaster_OX=7227_GN=Hrs_PE=1_SV=1)0.9570,tr|Q17796|Q17796_CAEEL_Hepatocyte_growth_factor-regulated_tyrosine_kinase_substrate_OS=Caenorhabditis_elegans_OX=6239_GN=hgrs-1_PE=1_SV=2)0.8720,'sp|P40343|VPS27 YEAST Vacuolar protein sorting-associated protein 27 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=VPS27 PE=1 SV=3',sp|Q9LNC6|TOL2_ARATH_TOM1-like_protein_2_OS=Arabidopsis_thaliana_OX=3702_GN=TOL2_PE=1_SV=1);

ESCRT-0 (Vps27-Hrs-TOL6)

((((BAA23366.1_Hrs_Homo_sapiens:0.0206656307994185,sp|Q99LI8|HGS_MOUSE_Hepatocyte_growth_factor-regulated_tyrosine_kinase_substrate_OS=Mus_musculus_OX=10090_GN=Hgs_PE=1_SV=2:0.0457361660112471)1.0000[stddev=0.00000000]:0.431796449943375,sp|Q960X8|HRS_DROME_Hepatocyte_growth_factor-regulated_tyrosine_kinase_substrate_OS=Drosophila_melanogaster_OX=7227_GN=Hrs_PE=1_SV=1:0.490138645789766)0.9130[stddev=0.90498619]:0.230843393108645,tr|Q17796|Q17796_CAEEL_Hepatocyte_growth_factor-regulated_tyrosine_kinase_substrate_OS=Caenorhabditis_elegans_OX=6239_GN=hgrs-1_PE=1_SV=2:0.612918824263557)0.9850[stddev=0.44271887]:0.324969331724587,'sp|P40343|VPS27 YEAST Vacuolar protein sorting-associated protein 27 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=VPS27 PE=1 SV=3':1.20608476776199,sp|O80910|TOL6_ARATH_TOM1-like_protein_6_OS=Arabidopsis_thaliana_OX=3702_GN=TOL6_PE=1_SV=1:1.34873230133113);

ESCRT-0 (Hse1-STAM-TOL2)

(('sp|P38753|HSE1 YEAST Class E vacuolar protein-sorting machinery protein HSE1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=HSE1 PE=1 SV=1',sp|O01498|STAM1_CAEEL_Signal_transducing_adapter_molecule_1_OS=Caenorhabditis_elegans_OX=6239_GN=stam-1_PE=1_SV=2)0.7680,tr|Q9XTL2|Q9XTL2_DROME_Jak_pathway_signal_transduction_adaptor_molecule_OS=Drosophila_melanogaster_OX=7227_GN=Stam_PE=1_SV=1,(NP_003464.1_signal_transducing_adapter_molecule_1_isoform_a_Homo_sapiens,sp|P70297|STAM1_MOUSE_Signal_transducing_adapter_molecule_1_OS=Mus_musculus_OX

=10090_GN=Stam_PE=1_SV=3)0.9980,sp|Q9LNC6|TOL2_ARATH_TOM1-like_protein_2_OS=Arabidopsis_thaliana_OX=3702_GN=TOL2_PE=1_SV=1);

ESCRT-0 (Hse1-STAM-TOL6)

(('sp|P38753|HSE1 YEAST Class E vacuolar protein-sorting machinery protein HSE1 OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=HSE1 PE=1 SV=1',sp|O01498|STAM1_CAEEL_Signal_transducing_adapter_molecule_1_OS=Caenorhabditis_elegans_OX=6239_GN=stam-1_PE=1_SV=2)0.8230,(NP_003464.1_signal_transducing_adapter_molecule_1_isoform_a_Homo_sapiens,sp|P70297|STAM1_MOUSE_Signal_transducing_adapter_molecule_1_OS=Mus_musculus_OX=10090_GN=Stam_PE=1_SV=3)1.0000,(tr|Q9XTL2|Q9XTL2_DROME_Jak_pathway_signal_transduction_adaptor_molecule_OS=Drosophila_melanogaster_OX=7227_GN=Stam_PE=1_SV=1,sp|O80910|TOL6_ARATH_TOM1-like_protein_6_OS=Arabidopsis_thaliana_OX=3702_GN=TOL6_PE=1_SV=1)0.9680);

ESCRT-I (STP22-TSG101-ELC)

(((NP_006283.1_tumor_susceptibility_gene_101_protein_Homo_sapiens,sp|Q61187|TS101_MOUSE_Tumor_susceptibility_gene_101_protein_OS=Mus_musculus_OX=10090_GN=Tsg101_PE=1_SV=2)1.0000,tr|Q9VVA7|Q9VVA7_DROME_Tumor_suppressor_protein_101_OS=Drosophila_melanogaster_OX=7227_GN=TSG101_PE=1_SV=2)0.6460,('sp|P25604|STP22 YEAST Suppressor protein STP22 of temperature-sensitive alpha-factor receptor and arginine permease OS=Saccharomyces cerevisiae (strain ATCC 204508 / S288c) OX=559292 GN=STP22 PE=1 SV=3',sp|Q9LHG8|ELC_ARATH_Protein_ELC_OS=Arabidopsis_thaliana_OX=3702_GN=ELC_PE=1_SV=1)0.9990,tr|O76258|O76258_CAEEL_Tumor_susceptibility_gene_101_protein_OS=Caenorhabditis_elegans_OX=6239_GN=tsg-101_PE=1_SV=2);

ESCRT-II (VPS36-VPS36-VPS36)

(((sp|Q06696|VPS36_YEAST,sp|Q9FF81|VPS36_ARATH)0.9600,(NP_057159.2,sp|Q91XD6|VPS36_MOUSE)1.0000)0.6100,sp|Q9VU87|VPS36_DROME,tr|Q19519|Q19519_CAEEL);

ESCRT-III (Snf7-Snf7-Snf7.1-CHMP1A)

((((sp|Q9SZE4|VP322_ARATH,sp|Q9SZE4-2|VP322_ARATH)1.0000,sp|Q9D7S9|CHMP5_MOUSE)0.7110,(tr|Q18886|Q18886_CAEEL,sp|Q8T0Q4|CHM4_DROME)0.9880)0.6950,sp|P39929|SNF7_YEAST,NP_002759.2);

Rab (YPT52-VPS21-RAB5A-RAB5)

(((sp|P36018|YPT52_YEAST,sp|P36017|VPS21_YEAST)0.8950,sp|P31582|RAF2A_ARATH)0.9480,(NP_492481.1,(NP_004153.2,sp|Q9CQD1|RAB5A_MOUSE)0.9990)0.7170,sp|Q9V3I2|RAB5_DROME);

Rab (YPT7-RAB7-RAB7)

(((NP_004628.4,sp|P51150|RAB7A_MOUSE)1.0000,sp|O76742|RAB7_DROME)0.7780,(sp|P32939|YPT7_YEAST,sp|O04157|RAG3B_ARATH)0.5070,tr|Q23146|Q23146_CAEEL);

Disassembly (VPS4-VPS4-VPS4)

(((NP_037377.1,NP_569053.1)1.0000,tr|Q9Y162|Q9Y162_DROME)0.8270,(sp|P52917|VPS4_YEAST,sp|Q9ZNT0|VPS4_ARATH)0.9980,tr|Q9BL83|Q9BL83_CAEEL);

SNAREs (VAM3-SYP121-STX10)

((sp|Q9ZSD4|SY121_ARATH:0.913328808350036,sp|Q9D3G5|STX11_MOUSE:1.209889795038830.5870[stddev=1.56076904]:0.225299212807783,sp|Q12241|VAM3_YEAST:0.574697097327348,(sp|O60499|STX10_HUMAN:0.301482261749452,(tr|Q7JY00|Q7JY00_DROME:0.35754209655391,sp|P83528|STX6_CAEEL:0.626283442354567)0.6700[stddev=1.48694317]:0.206843636291082)1.0000[stddev=0.00]:1.3313145378462);

Tetraspanins (SUR7-TET8-CD63-CD81)

(((NP_001771.1,sp|P41731|CD63_MOUSE)0.9990,sp|Q8S8Q6|TET8_ARATH)0.6090,(sp|Q07651|FMP45_YEAST,NP_004347.1)0.5410tr|Q9XVI4|Q9XVI4_CAEEL,tr|A1Z8K7|A1Z8K7_DROME);