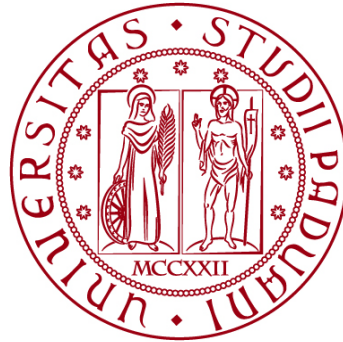


**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI BIOLOGIA**

**Corso di Laurea in Biotecnologie**



**ELABORATO DI LAUREA**

**LONGIN DOMAIN CONTAINING PROTEINS AS  
ESSENTIAL GENES**

**Tutor: Prof. Francesco Filippini  
Dipartimento di Biologia**

**Laureando: Pietro Biasi**

**ANNO ACCADEMICO 2024/2025**

# ***Index***

## ***Abstract***

### ***Chapter 1 – The essential genes***

*1.1 – Essential genes identification*

*1.2 – Gene essentiality is not absolute*

*1.3 – Essential genes applications*

### ***Chapter 2 – Longin SNARE proteins***

*2.1 – VAMP7*

*2.2 – SEC22b*

*2.3 – Ykt6*

### ***Chapter 3 – Bioinformatic analysis of the Longin domain***

## ***Conclusion***

## ***Bibliography***

## ***Abstract***

SNAREs (Soluble NSF Attachment protein REceptors) are a superfamily of proteins whose members are responsible in eukaryotes for almost all events in the extracellular and intracellular secretory pathway. SNARE proteins are especially important for vesicular transport and (in animals) neurotransmitter release at synapses. These proteins are present in the membrane of the vesicle and the target cell and form a complex that promotes membrane fusion.

In 2001, a family of vesicular v/R-SNAREs was identified that shared a conserved N-terminal domain. Due to their comparatively greater length with respect to 'Brevins', such as synaptobrevins, these proteins were thus named "Longins". Consequently, their N-terminal domain was designated as the 'Longin domain'. A comparative analysis of major Eukaryotic taxa has demonstrated that Longins and their longin domains are conserved in all Eukaryotic taxa. In contrast, Brevins are a class of R-SNAREs present only in Opisthokonta (animals and fungi). Furthermore, evidence suggests that each eukaryotic genome contains a minimum of three Longin genes belonging to three subfamilies prototyped by Ykt6, Sec22b and VAMP7.

Since all eukaryotes have Longins, these are likely to be essential genes.

This thesis focuses on highlighting the crucial cellular roles of Longins in subcellular trafficking, which underlies their essential nature.

## ***Chapter 1 – The essential genes***

The term 'essential genes' is employed to describe those genes that are absolutely required for cell survival and reproduction. The disruption of these genes has been demonstrated to induce cell death, thus indicating that these genes perform essential biological functions. The essentiality of a gene is defined as the indispensability of the gene's product to the survival of an organism.

The loss of function of an essential gene compromises the viability of the cell/organism or results in a serious loss of fitness.

### ***Essential genes identification***

In 1951, Horowitz and Leupold advanced a theory proposing that the majority of proteins are likely to be indispensable for viability.

In their experiment, a significant number of temperature-sensitive mutants of *Escherichia coli* and *Neurospora crassa* were isolated and cultivated in minimal medium. It was observed that when subjected to restricted temperature conditions, only 25% of the *E. coli* mutants and 50% of the *N. crassa* mutants did not grow (even in a rich medium).

This finding indicates that the majority of the genes are non-essential for both organisms.

In the decades following Horowitz and Leupold's experiment, random mutagenesis using chemical agents made it possible to estimate the essential genome of *C. elegans* to be approximately 15%.

It is evident that the possession of the complete genomic sequence of an organism is a fundamental prerequisite for the development of strategies aimed at identifying essential genes, whether these strategies are random or targeted.

The first genomes to be completely sequenced were those of *Haemophilus influenzae* and *Mycoplasma genitalium*, and in fact, these were the first organisms to be screened to determine their essential genome.

In the early days of the genomic era, researchers attempted to identify a minimal genome using comparative genomics. They assumed that genes shared between organisms are likely to be essential. This analysis led to the identification of approximately 250 candidates that were deemed essential. This research was further stimulated by the availability of the first genome sequence of *H. Influenzae* and *M. genitalium* [table 1].

As the number of genome sequences available increased, the number of shared genes decreased until the conclusion that the minimal genome consisted of a core of 40 genes. This small number is due to the fact that many essential cellular functions can be encoded by non-orthologous genes.

The first experimental study identified 300 genes that could not be inactivated by transposon mutagenesis in *M. genitalium*.

Organism	Number of protein-coding essential genes	Method of identification
<i>Saccharomyces cerevisiae</i>	878	Single-gene-specific mutagenesis
<i>Haemophilus influenzae</i>	642	Saturation transposon mutagenesis
<i>Acinetobacter baylyi</i>	499	Single-gene-specific mutagenesis
<i>Mycoplasma pulmonis</i>	461	Saturation transposon mutagenesis
<i>Mycoplasma genitalium</i>	381	Saturation transposon mutagenesis
<i>Salmonella enterica</i> Typhi	356	Saturation transposon mutagenesis
<i>Pseudomonas aeruginosa</i>	335	Saturation transposon mutagenesis
<i>Helicobacter pylori</i>	323	Saturation transposon mutagenesis
<i>Escherichia coli</i>	303	Single-gene-specific mutagenesis
<i>Staphylococcus aureus</i>	302	Antisense RNA inhibition
<i>Bacillus subtilis</i>	271	Single-gene-specific mutagenesis
<b>Numbers of genomes compared</b>		
2	256	Comparative genomics
5	180	Comparative genomics
7	156	Comparative genomics
100	63	Comparative genomics
147	35	Comparative genomics

*Table 1 - number of genes constituting the minimal genome in different organisms and the method by which they were identified*

This method involves randomly inserting a transposon (a mobile genetic element) into the genome of interest.

When a library is created that is large enough so that each different position in the genome has at least one clone with a transposon at that position, this is referred to as saturation mutagenesis.

This technique, when employed in combination with NGS, facilitates the high-throughput identification of essential genes, which are those that, following sequencing, exhibit no insertions (or have very few and always at the margins of the ORF).

This approach has been used in a wide range of organisms, leading to the identification of minimal genomes of varying sizes. This is probably because transposon insertion is not entirely random. In complex eukaryotes, for example, DNA is organised into chromatin, and the accessibility of different regions varies. Another limitation of this method is that not all genes will be disrupted, even in saturating conditions, especially in higher organisms, where the coding portion of the genome is only a small part of the total.

Another important consideration is that not all mutations result in the complete disruption of gene expression, which can lead to misidentification (false negatives). Targeted approaches, in which the entire open reading frame of a specific gene is deleted, are more accurate for identifying essential genes.

The first attempts to move from microorganisms to higher animals were very difficult, given that homologous recombination in somatic tissues is very inefficient, and so the first attempts at gene knockout failed.

The discovery that homologous recombination works surprisingly well in embryonic stem (ES) cells derived from mouse blastocysts led to the development of the first KO mice.

Another step forward in identifying essential genes was the discovery of RNA interference, a phenomenon discovered as part of the plant immune system against pathogens, which allows specific mRNAs to be targeted and degraded through the use of siRNAs that are recognised by the RISC cellular complex and guided to the target transcript.

It is possible to create iRNA libraries that cover the entire genome, or that are specific to certain classes of genes.

Candidates for essential genes are those that lead to cell death when silenced.

Another fundamental discovery for investigating essential genes in higher eukaryotes was the identification of haploid cell lines from diploid organisms, which made it possible to use insertional mutagenesis methods such as gene trapping, which allows a DNA fragment to be inserted into the ORF of a gene to inactivate its function and observe its effect on cell viability. In the case of diploid cell lines, the unmutated copy can compensate for the inactivation of one allele, even if it is an essential gene.

The revolutionary impact of CRISPR-Cas9 technology has also given rise to the identification of essential genes.

In particular, dead Cas9, which has inactivating mutations in the endonuclease domains, is unable to cut DNA but retains the ability to bind with high precision to a specific sequence.

This dCas9 constitutes a potent tool for the regulation of gene expression, given its capacity to be associated with factors that inhibit or induce the expression of the target gene: CRISPRi (interference) and CRISPRa (activation) are powerful genomic screening tools.

CRISPRi can be used to identify essential candidate genes, which, when inactivated, result in cell death.

Gene essentiality in vivo may be assessed through the collection of genome sequencing data at the population level, and as a result, essential genes would be those that are rarely or never disrupted or truncated in the general population.

Genome sizes vary greatly among species, and focusing on the species of which essentialomes have been nearly completely reported, we note that genomes with a larger number of ORFs have a lower percentage of essential genes: the intracellular parasite *M. genitalium* has one of the smallest known genomes (482 genes), and  $\approx 80\%$  of the genes were reported to be essential. It is estimated that

approximately 22% of the 1,600 genes in the H. influenzae genome are essential for viability.

### ***Gene essentiality is not absolute***

The original definition of 'essentiality' was conceived of as a binary and static property. However, it became evident that this classification of genes as either essential or non-essential was overly simplistic. Gene essentiality is often dependent on the environmental context (genes necessary for the synthesis of cellular building blocks are essential only if those building blocks are absent in the growth medium or the natural environment) or the genetic context (often a gene becomes essential when a second gene's function is lost).

A study conducted on the Saccharomyces model has grouped essential genes into four categories:

1. Conditional essential → genes that are essential only under specific growth conditions
2. Essential → genes required for survival under the optimal growth conditions
3. Redundant essential → gene duplication for producing redundant pathways (paralog genes) is a strategy evolved by eukaryotes to improve their fitness. Synthetic lethality is a situation that occurs when a mutation of both genes causes cell death, while the single inactivation is viable.
4. Absolute essential (minimal genes) → genes that are necessary for supporting cellular life in an environment that provides all nutrients and that is free from any form of stress.

The essentiality of a gene is subject to change during the course of evolution. Indeed, as genomes evolve, the genetic background can be altered in such a manner that the essentiality of a gene is modified. This can be due to various reasons: the cellular network during evolution can become more robust by acquiring different pathways for the same function, and genes could be replaced. This suggests that essentiality is more likely to be a property of a functional model rather than of an individual gene.

### ***Essential genes applications***

The identification of the genes that are indispensable for an organism in a given biological context has significant applications across diverse scientific disciplines. The comprehension of the minimal gene sets and the functions performed by the products of these genes facilitates our understanding of the fundamental processes that are vital for a cell.

In order to reduce the genome to its most basic elements, two approaches have been proposed: the top-down and the bottom-up approach.

The first one proposes deleting random or selected non-essential genomic segments.

The advantage of this method is that, starting from a known genome, it allows any problems that may arise to be remedied by returning to the previous step. On the other hand, this method is very time-consuming.

The bottom-up approach, on the other hand, involves creating a minimal artificial cell by integrating all the functions that are considered essential. The greatest difficulty encountered with this method is that it is extremely complicated to clarify gene interactions without a complete genetic background.

In the pharmaceutical field, the search for essential genes in major human pathogens is a fundamental aspect in the development of new antimicrobial drugs. In fact, identifying possible targets is the first step in the discovery of new drugs and vaccines.

Furthermore, in the field of cancer research, the search for genes that are essential in tumour cell lines but not essential in healthy cells can reveal oncogenic factors.

## ***Chapter 2 - Longin SNARE proteins***

All eukaryotic cells are equipped with a series of membrane-bound organelles whose functions are vital for the survival of the cell. Such organelles include the nucleus, the endoplasmic reticulum, and the Golgi complex, amongst others.

The process of communication between these organelles takes place via transport vesicles that are formed from the membrane of a donor cell and then undergo fusion with the membrane of the target cell.

The process is characterised by the involvement of proteins belonging to the SNARE (Soluble N-ethylmaleimide-sensitive factor Attachment Protein Receptors) family.

As a result of research into the architecture of protein domains, four common functional entities have been identified among the proteins involved in membrane trafficking. One of these is the Longin domain.

Initially, the Longin domain was identified as a regulatory module of proteins belonging to the SNARE family, although more recent studies have found this domain in other vesicular trafficking families as well. The Longin domain has been identified in seven vesicular trafficking protein families, including longin-SNAREs.

Proteins containing the Longin domain have been found to be involved in vesicular budding, vesicle anchoring and membrane fusion reactions.

Among the factors involved in membrane fusion, SNARE proteins are the best characterised molecular factors.

SNARE proteins interact with numerous regulatory proteins that modulate the assembly of the trans-SNARE complex (a complex that joins two membranes), allowing fusion to be inhibited or activated depending on the appropriate biological context.

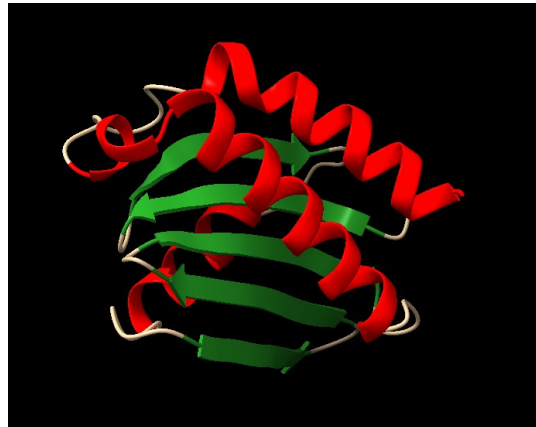
Longin SNAREs possess an N-terminal domain (the Longin domain) that has been demonstrated to regulate the assembly of the trans-SNARE complex and thus membrane fusion.

In addition, longin domains have been identified in non-SNARE protein families that play a crucial role in trafficking processes, including vesicular budding.

It can be concluded that longin domains are indispensable for the regulation of SNARE localisation at their site of action, in addition to their function as a regulatory switch for their fusion activity.

It has been determined that the Longin domain has a globular fold of approximately 120 amino acids, organised in an alpha-beta-alpha sandwich structure (one of the most common) [1].

The topology has been demonstrated to be shared among all seven protein families containing the Longin domain, with five antiparallel beta sheets sandwiched by an alpha helix on one side and two alpha helices on the other side.



*Image 1 – 3D structure of the Longin domain of human VAMP7, with alpha helices highlighted in red and beta sheets in green.*

## **VAMP7**

The VAMP7 protein has been demonstrated to facilitate the fusion of vesicles from the Golgi apparatus, late endosomal compartments and lysosomes with the plasma membrane. This process is therefore involved in numerous cellular functions, including phagocytosis, mitosis and cell migration, amongst others.

VAMP7 was initially identified as being essential to neuronal development, a process involving membrane expansion supported by VAMP7, which provides proteins, lipids and growth factors through membrane fusion.

VAMP7 is also a key protein in the biogenesis of the autophagosome and the subsequent degradation of its contents.

In addition, VAMP7, like other SNAREs, is essential for unconventional protein transport (UPS), an alternative pathway that bypasses the Golgi apparatus.

The classification of UPS is divided into four subclasses: types I-IV. SNARE proteins are vital for type III and IV unconventional protein transport, which are SNARE-dependent vesicular transport.

The majority of UPS pathways are known to be activated by a number of stress conditions, which include, but are not limited to, nutrient deprivation, endoplasmic reticulum stress, mechanical stress and inflammation. These conditions compromise the normal function of the ER–Golgi pathway, increasing the need for alternative secretory pathways.

Furthermore, UPS is necessary for proteins, such as FGF2, which would be inactivated by Golgi glycosylation.

VAMP7 is principally located on late endosomes and the Golgi apparatus. VAMP8, which belongs to the same protein family, has the same subcellular localisation, but its expression is limited to epithelial and immune system cells, while VAMP7 is expressed in all cell types.

VAMP7 has been shown to interact with SNARE partners located on the plasma membrane and with autophagosomes, including syntaxin 1 and 3, SNAP-23 and SNAP-25.

It has been determined that the complexes formed by VAMP7 and these t-SNAREs mediate the fusion of late endosomes with the plasma membrane.

A significant unanswered question regarding UPS is the mechanism by which leaderless proteins are packaged into endosomal vesicles.

VAMP7 has been demonstrated to function as the initiating element for a molecular network that integrates proteins from the predominant classes involved in vesicular trafficking. This system directs VAMP7-containing vesicles along microtubules to the periphery, allowing exocytosis.

In vitro, it has been demonstrated that the expression of the longin domain alone is capable of inhibiting neuronal growth, while its deletion has been shown to stimulate growth.

Research has indicated that the leucine 43 and tyrosine 45 residues of the longin domain are essential for the stabilisation of the closed conformation, due to their interaction with the coiled-coil domain.

The closed conformation of VAMP7 does not necessitate regulatory proteins or post-translational modifications. However, it should be noted that even when in the closed conformation, VAMP7 does not fully prevent the assembly of the SNARE complex. Indeed, VAMP7 can still interact, albeit with reduced affinity, with membrane t-SNARE.

Moreover, the longin domain of VAMP7 has been demonstrated to be indispensable for the sorting and targeting of VAMP7 to the relevant cellular compartments.

This can be attributed to the interaction between the longin domain and the components of the clathrin vesicle coat.

The longin domain has been observed to interact with a component of the AP-3 complex of the vesicle coat, delta-adaptin. This interaction has been hypothesised to direct vamp7 towards synaptic vesicles and endo-lysosomal compartments. Subsequent to vesicle fusion with the plasma membrane, Vamp7 undergoes endocytosis for recycling. The longin domain has also been demonstrated to play a fundamental role in this mechanism, and it has been shown that this mechanism is regulated by an Hrb adapter, whose interaction with the longin domain is essential.

### ***SEC22b***

The protein Sec22b was initially identified as a SNARE protein involved in membrane trafficking between the endoplasmic reticulum (ER) and the Golgi apparatus.

In anterograde transport, from the ER to the Golgi apparatus, sec22 interacts with the COPII complex, formed by sec23A-sec24A, allowing the incorporation of outgoing vesicles.

In the context of retrograde transport, Sec22 has been observed to interact with COPI, facilitating the maintenance of continuous membrane flow.

Sec22 proteins are characterised by the presence of a single transmembrane domain (TMD) at the C-terminus; however, in mammals, Sec22 proteins have been observed to contain multiple TMDs.

This domain is essential for the correct formation of homodimers, which in turn promote the assembly of higher-order SNARE complexes, catalysing membrane fusion. Sec22 has been shown to form efficient homodimers with cell membranes when specific cysteine residues are present at the C-terminus or in the SNARE motif. The presence of cysteine derivatives on both COP-II vesicles (donors) and Golgi membranes (acceptors), in conjunction with the formation of disulphide bridges, provides a clear indicator of the cis and trans arrangement of SNAREs during the fusion event.

A study was conducted in which SEC22 isoforms differing in the number of TMDs were generated. This study demonstrated that the subcellular localisation of Sec22 can vary depending on the number of TMDs, thereby influencing the function of the protein. Indeed, the present study observed that isoforms comprising four TMDs were localised on the Golgi apparatus, while those with fewer than four TMDs were localised on the ER.

Recent studies have demonstrated that Sec22b is implicated in the process of plasma membrane expansion during neuronal development. Indeed, it has been demonstrated that interaction with a plasma membrane t-SNARE, syntaxin 1, Sec22b, induces the approximation of the ER and the plasma membrane.

Inhibition of Sec22 in yeast compromises the modification of plasma membrane phosphoinositides by an ER enzyme. Furthermore, the introduction of a long linker between the coiled-coil domain and the transmembrane domain of Sec22b increases the distance between the endoplasmic reticulum (ER) and the plasma membrane, thereby compromising the growth of axons and dendrites in neurons.

Recent studies have identified SNARE proteins, including SEC22b, as playing a crucial role in various pivotal processes within the immune system. Indeed, a key step in the human immune system is the capacity of cytotoxic lymphocytes (CTLs and natural killer (NK) cells) to identify infected cells and trigger their apoptosis. Following the establishment of contact with the target cell, cytotoxic granules are transported to the immunological synapse, where they release toxic proteins into the synaptic space. Membrane fusion is hypothesised to play a crucial role in the release of these proteins.

Phagocytosis is a process in which SNARE protein-mediated membrane fusion is essential, and SEC22B has been identified on phagosomes. Phagosomes are vesicles formed around particles engulfed by phagocytes, such as macrophages, neutrophils and dendritic cells.

### ***Ykt6***

Ykt6 is a v-SNARE protein that plays a key role in various cellular processes, including transport from the endoplasmic reticulum (ER) to the Golgi apparatus, intra-Golgi transport, endosome-Golgi transitions, and transport to the vacuole.

Ykt6 has been demonstrated to play an important role in the fusion between autophagosomes and lysosomes. This has been shown through immunofluorescence studies carried out on autophagosomes that accumulate in cells in which STX17 has been silenced. This is a Qa-type SNARE protein that is essential for autophagy. It cooperates with the SNAP29 protein to form the SNARE complex, which is necessary for fusion with the lysosome.

When STX17 is silenced, this mechanism is significantly diminished, though not fully extinguished. In recent years, the discovery of a second alternative SNARE module, mediated by YKT6, has emerged. This module has been shown to partially replace the classic STX17–SNAP29 complex.

However, if both SNAREs are eliminated, i.e. both STX17 and YKT6, the system no longer has any alternative mechanism for bringing the two membranes together and fusing them. Consequently, under these conditions, the process of fusion between the autophagosome and the lysosome is rendered completely inoperative.

In contrast to other longin-SNAREs, Ykt6 does not possess a transmembrane domain. Rather, it has been demonstrated that association occurs through a lipid

anchor, which facilitates a precarious association with the membrane. This phenomenon elucidates the capacity of Ykt6 to effortlessly insert itself into compromised cellular compartments, such as the membranes of the endoplasmic reticulum in Parkinson's disease, with the objective of restoring functional transport.

The anchoring of Ykt6 to the membrane is contingent upon the presence of a palmitoyl group and a farnesyl group, which are bound to two cysteine residues of the "CCAIM" C-terminal motif.

The addition of the farnesyl group constitutes a post-translational modification that effects the transition of Ykt6 from a semi-closed state to a predominantly closed state, in which its fusion is inactivated. This is due to the formation of a hydrophobic pocket at the interface between the longin domain and the coiled coil domain, which inhibits the insertion of Ykt6 into membranes.

Consequently, the farnesyl group stabilises Ykt6 in the inactive closed conformation, thereby ensuring that it does not associate with non-specific t-SNAREs.

Conversely, palmitoylation is reversible, resulting in the anchoring of Ykt6 to the membrane.

Ykt6 has been shown to play a role in suppressing alpha-synuclein toxicity in models of Parkinson's disease. The pathogenesis of this condition is characterised by the formation of insoluble alpha-synuclein fibrils that inhibit ER-Golgi traffic, causing protein accumulation in the ER. In yeast models, it has been demonstrated that the expression of Ykt6 and other proteins can limit this blockage of vesicular traffic by promoting anterograde ER-Golgi transport.

However, evidence suggests that Ykt6 is overexpressed in metastatic tumours, and its expression in epithelial cell lines has been demonstrated to accelerate the cell cycle.

The process of membrane fusion between cargo-containing transport vesicles and the target membrane is dependent on the precise localisation and activation of v-SNARE and t-SNARE proteins. These proteins must be located at the correct site and activated from an inhibited closed conformational state at the opportune moment, thereby enabling interaction with each other at the optimal time.

The N-terminal Longin domain has been demonstrated to play a crucial role in both regulatory principles. It has been shown to control the intracellular sorting and membrane localisation of SNAREs through interactions with vesicle coat components (in the case of VAMP7 and Sec22b) or through a conformational transition (in the case of Ykt6).

It has been established that both VAMP7 and Ykt6 are capable of assuming a closed and stable conformation independently, thereby inhibiting the assembly of the

SNARE complex. In the case of Ykt6, this process involves a delicate interaction between the farnesyl and palmitoyl groups present at the C-terminus of the protein, while for VAMP7, tyrosine residue 45 of the longin domain is crucial, whose phosphorylation increases VAMP7-mediated exocytosis.

Recent studies have demonstrated the significance of longin-SNAREs in the cell growth processes of various organisms. These proteins have been shown to play an important role in both membrane fusion and docking processes that do not result in fusion. The additional level of regulation provided by the longin domain may be relevant in this context.

In the second part of the project, a comparative analysis of proteins belonging to the longin-SNARE family was conducted. The aim of this analysis was to characterise the longin domain in detail and to evaluate its degree of evolutionary conservation. In order to investigate these aspects in greater depth, an integrated mapping of the sequence and structure of the longin domain was performed. This involved a combination of multiple alignment analysis, residue-specific conservation calculation and projection of scores onto three-dimensional protein models.

This approach allows evolutionary patterns to be linked to function: high conservation tends to mark structurally crucial regions.

### **Chapter 3 - Bioinformatic analysis of the Longin domain**

1. Sequence collection and dataset definition
  - a. Sequence collection: the sequences of VAMP7, Sec22b and Ykt6 in Homo sapiens are obtained from Uniprot (accession IDs: P51809, O75396, O15498, respectively)
  - b. Recovery of orthologues: the Orthodb database is used to recover orthologues for the three genes of interest in the following organisms: <sup>1</sup>Mus musculus, <sup>2</sup>Drosophila melanogaster, <sup>3</sup>Caenorhabditis elegans, <sup>4</sup>Arabidopsis thaliana, <sup>5</sup>Danio rerio, <sup>6</sup>Saccharomyces cerevisiae in order to cover a broad evolutionary spectrum.

A particular case in the search for orthologs arose during the search for orthologs of VAMP7 in Saccharomyces cerevisiae and C. elegans, as no direct orthologs of VAMP7 were found in these model organisms. VAMP7 is considered a metazoan-specific Longin-SNARE, which appeared relatively late in evolution and is associated with functions typical of complex secretory systems in higher Eukaryotes. Even for A. thaliana, we selected AtVAMP721 as an ortholog, a gene belonging to a subclass called the VAMP-like or VAMP72 family.
  - c. Extraction of the longin: domain using ScanProsite, the sub-sequences related to the Longin domain are extracted in the various orthologs
2. Preparation of multiple sequence alignment (MSA)
  - a. MSA: for each family (VAMP7, Sec22b, Ykt6), a multiple sequence alignment (MSA) was constructed using MAFFT (L-INS-i mode, optimised for accuracy)
  - b. Trimming: the MSAs obtained were subsequently refined with ClipKit to remove columns with poor information content (excessive gaps, ambiguous alignment)
3. Calculation of conservation indices: once trimmed, the MSAs were viewed on Jalview, from which residue-specific conservation indices were extracted. The residue conservation classification was used for subsequent structural mapping.
4. Structural models and 3D conservation mapping
  - a. Predicted models obtained with Phyre2 were used to visualise the three-dimensional structure, in order to obtain a complete three-dimensional representation of the longin domain
  - b. The structures were imported into ChimeraX, and the correct positioning of the residue numbering was verified to ensure correspondence with the multiple alignment
  - c. Mapping conservation scores: the conservation scores obtained from the MSA were associated with the residues of the 3D model, and a colour-coded representation of the protein was created in ChimeraX,

assigning each residue a colour according to its conservation class (blue = variable, green = intermediate, red = highly conserved).

### Multiple alignment and structural conservation analysis

#### VAMP7

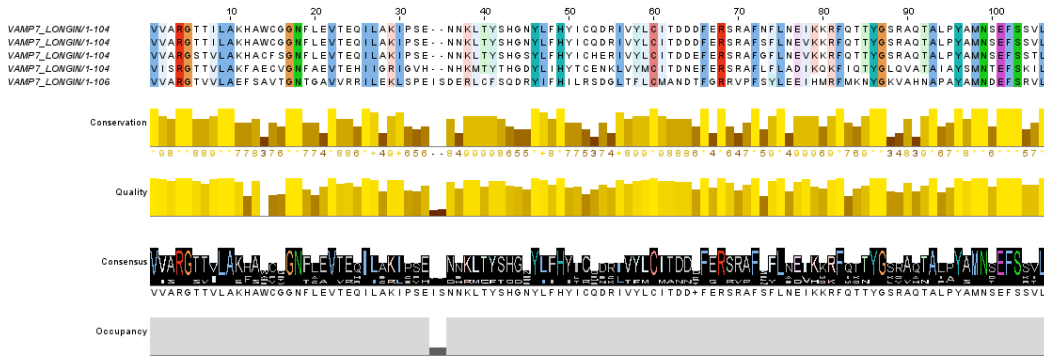


Image 2 – the MSA of the orthologous sequences of VAMP7

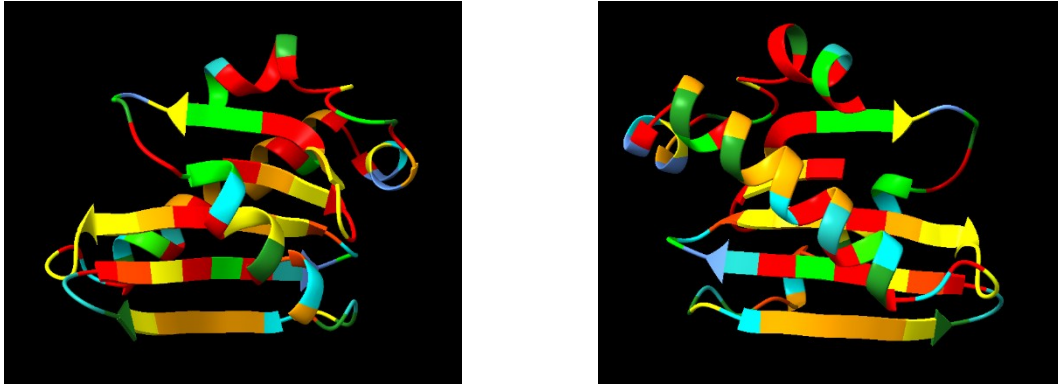
In multiple alignment, as mentioned above, there are no orthologous sequences in *S. cerevisiae* and *C. elegans*.

On Jalview, those residues with a conservation level greater than 80% have been coloured.

Looking at the MSA of VAMP7, we can see that the conservation scores are quite high: most residues are assigned a score higher than 8/9, and residues with a score lower than 3/4, indicative of a medium-low level of conservation, are few and usually well-spaced apart.

We can see that most highly conserved residues are hydrophobic (alanine, valine and others). These amino acids form the inner core of the protein, and their degree of conservation suggests that they are essential for the correct folding of the protein.

Looking at the three-dimensional model of the Longin domain of human VAMP7 displayed on ChimeraX, appropriately coloured with a scale ranging from red to blue for the most and least conserved residues, respectively, we can see how the poorly conserved residues are often concentrated in random-coiled regions. On the other hand, we can see how most of the orange-red coloured residues, and therefore highly conserved ones, are found in regions involved in a very specific conformation: alpha-helices or beta-sheets [3,4].



Images 3,4 – 3D structure of the Longin domain of human VAMP7 visualised on ChimeraX, viewed respectively from the side where one and two alpha helices are present

### SEC22b

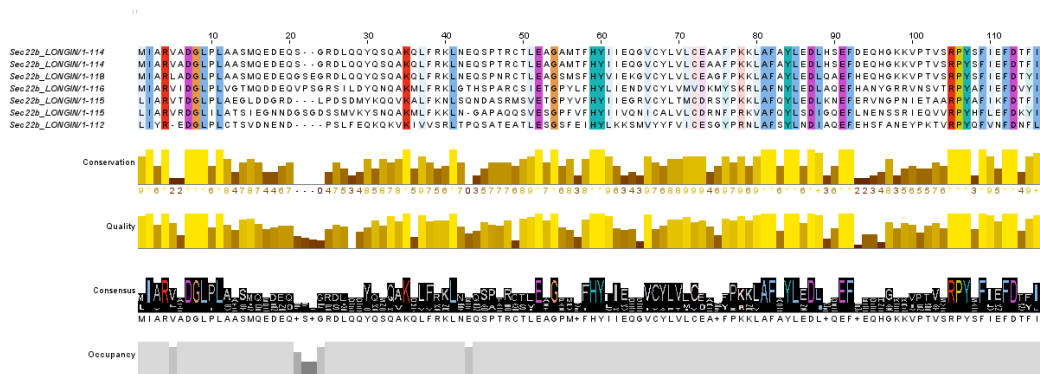


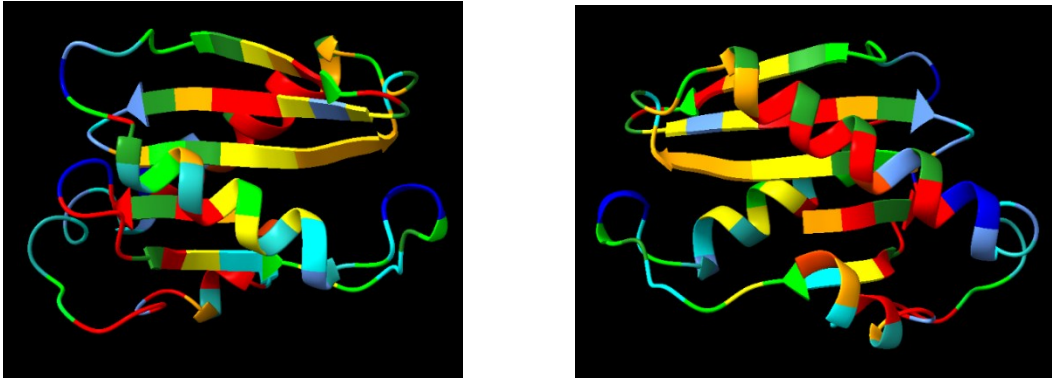
Image 5 – the MSA of the orthologous sequences of SEC22b

Looking at the MSA of SEC22, we can see that, compared to that of VAMP7, there are lower average conservation scores and that these scores are present in larger regions of the alignment, particularly the regions between residues 15-30 and 90-105.

If we look at these regions on the 3D model, we discover that the region between 15-30 there is a short alpha-helix of just 3 amino acids that is not present in the 3D model of VAMP7 but is present, as we will see later, in Ykt6.

If we look at the 3D model in this case, it is even more evident than in the VAMP7 structure how poorly conserved residues are concentrated in random-coiled regions, while alpha-helix and beta-sheet regions are characterised by a higher degree of conservation [6,7].

In this case too, more than half of the invariant residues are hydrophobic and are located in the core of the protein.



Images 6, 7 - 3D structure of the Longin domain of human SEC22b

### Ykt6

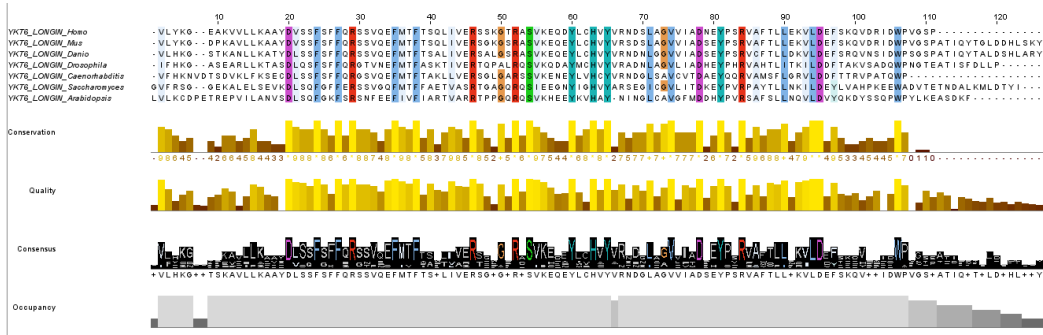
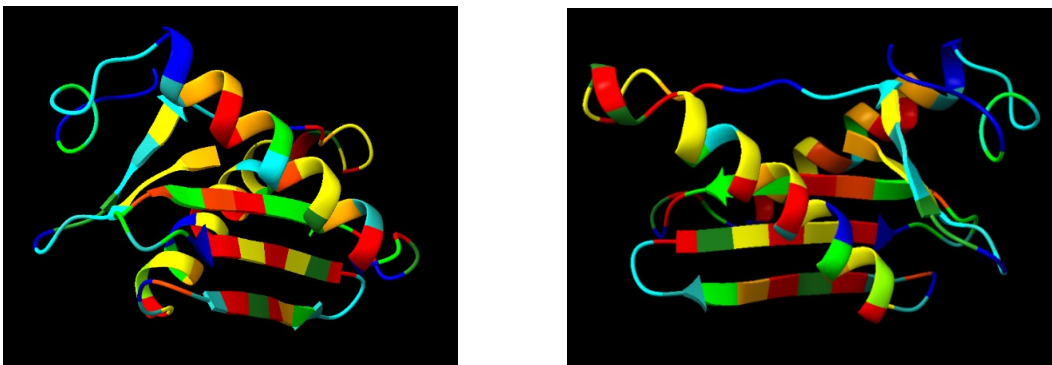


Image 8 - the MSA of the orthologous sequences of Ykt6

Looking at the multiple alignment of Ykt6, we see that the conservation scores are closer to those of SEC22 than to those of VAMP7. Here too, we have two regions characterised by conservation scores: the first region corresponds to the first 20 residues, while the second is located between residues 95 and 105. These two regions also showed greater variability in SEC22 in correspondence with those regions of the Longin domain, suggesting that greater variability is tolerated in those portions without affecting the functionality of the Longin domain.

As in the two previous cases, observing the 3D model, we see that the most variable regions are not found in regions of the sequence that assume alpha-helix and beta-sheet conformations.



Images 9, 10 - 3D structure of the Longin domain of human Ykt6

## ***Conclusion***

The aim of the study was to evaluate, through a computational and comparative approach, the essentiality of the Longin domain in the VAMP7, Sec22b and Ykt6 proteins through the degree of evolutionary conservation. The analysis integrated multiple sequence alignments and residue-by-residue conservation measures in order to understand whether the Longin domain can be considered an essential module for the function of the proteins in the family.

The results show that the Longin domain is a highly conserved and stable element in the architecture of all three SNAREs analysed, regardless of their taxonomic distribution and differences in biological role. In Sec22b and Ykt6, proteins present in all eukaryotes and involved in ancient and ubiquitous vesicular trafficking pathways, the Longin domain is highly conserved throughout its entire length, reflecting its fundamental role in proper protein folding, conformational control and regulation of fusogenic activity. Its conservation remains high even in phylogenetically distant organisms such as yeasts, plants and nematodes, suggesting a strong selective constraint.

The case of VAMP7 is different, as this protein is absent in yeasts and nematodes and is mainly present in higher Eukaryotes. Despite its more restricted distribution, the Longin domain maintains a significant degree of conservation among the species considered, indicating that, despite appearing later in evolution, the domain has acquired an essential structural and regulatory role. Its function in conformational control and modulation of SNARE activity appears to be indispensable even in more specialised secretory contexts, typical of metazoans. Although VAMP7 is evolutionarily more recent than YKT6 and Sec22b, its MSA shows medium-high levels of conservation. This does not reflect greater antiquity or evolutionary rigidity, but rather the smaller phylogenetic distance between the species included in the VAMP7 dataset, all of which belong to vertebrates.

In contrast, the MSAs of YKT6 and Sec22b include extremely distant species (animals, insects, nematodes, yeasts, and plants), increasing apparent variability and reducing conservation scores.

Overall, the analysis suggests that the essentiality of the Longin domain depends on the very nature of the domain as a conserved structural module.

The observation that all Longin-SNAREs, regardless of their specific role, maintain a highly conserved central hydrophobic core and invariant functional regions supports the idea that the Longin domain represents an evolutionarily robust motif necessary to ensure stability, correct localisation and regulation of this protein family.

## **Bibliography**

1. Rossi V, Banfield DK, Vacca M, Dietrich LEP, Ungermann C, D'Esposito M, Galli T, Filippini F. *Longins and their longin domains: regulated SNAREs and multifunctional SNARE regulators*. Trends Biochem Sci. 2004;29(12):682-688. doi:10.1016/j.tibs.2004.10.002.
2. Daste F, Galli T, Taresté D. *Structure and function of longin SNAREs*. J Cell Sci. 2015;128(23):4263-4272. doi:10.1242/jcs.178574.
3. Vats S, Galli T. *Role of SNAREs in Unconventional Secretion—Focus on the VAMP7-Dependent Secretion*. Front Cell Dev Biol. 2022;10:884020. doi:10.3389/fcell.2022.884020.
4. Adnan M, Islam W, Zhang J, Zheng W, Lu G-D. *Diverse Role of SNARE Protein Sec22 in Vesicle Trafficking, Membrane Fusion, and Autophagy*. Cells. 2019;8(4):337. doi:10.3390/cells8040337.
5. Sun W, Tian B-X. *The function of SEC22B and its role in human diseases*. Cytoskeleton. 2020. doi:10.1002.
6. Rancati G, Moffat J, Typas A, Pavelka N. *Emerging and evolving concepts in gene essentiality*. Nat Rev Genet. 2018;19(1):34-49. doi:10.1038/nrg.2017.74.
7. Kriegenburg F, Bas L, Gao J, Ungermann C, Kraft C. *The multi-functional SNARE protein Ykt6 in autophagosomal fusion processes*. Cell Cycle. 2019;18(6-7):?-?. doi:10.1080/15384101.2019.1580488.
8. Zhang Z, Ren Q. *Why are essential genes essential? – The essentiality of Saccharomyces genes*. Microb Cell. 2015;2(8):280-287. doi:10.15698/mic2015.08.218.
9. Guo Y, Ju Y, Chen D, Wang L. *Research on the Computational Prediction of Essential Genes*. Front Cell Dev Biol. 2021;9:803608. doi:10.3389/fcell.2021.803608.