MOLECULAR CHARACTERISATION OF A NOVEL VIPAR INTERACTING PROTEIN IN HEALTH AND DISEASE

BY

BLERIDA BANUSHI

A thesis submitted to
The University of Birmingham
for the degree of
DOCTOR OF PHILOSOPHY

College of Medicine & Dentistry
The University of Birmingham
August 2013

Abstract

Arthrogryposis, Renal dysfunction, and Cholestasis (ARC) syndrome is a multisystem disorder caused by mutations in genes encoding two proteins VPS33B or VIPAR, which appear to be critical regulators of cell polarity. VPS33B and VIPAR may function as part of a multi-protein complex that interacts with an active form of RAB11A and is involved in the transcriptional regulation of E-cadherin.

VPS33B and VIPAR are shown to interact at a protein level forming a stable binary complex. A novel interacting partner of VIPAR was identified, PLOD3, a post-translational modification enzyme with lysyl hydroxylase (LH), collagen galactosyltransferase (GT), and glucosyltransferase (GGT) activities (Wang et al., 2012). In mIMCD-3 polarized cell lines VPS33B and VIPAR are involved in PLOD3 trafficking from the TGN compartment via RAB11A positive vesicles. Study of the topology of this ternary protein complex evidenced that VIPAR is a transmembrane protein with its luminal N-terminal interacting with PLOD3 and its cytosolic C-terminal being involved in the interaction with VPS33B. VPS33B/VIPAR mediates the trafficking of PLOD3 from the TGN to collagen carrier structures where the binding of PLOD3 with collagen takes place. The PLOD3-collagen binding is required for collagen trafficking and an abnormal accumulation of intracellular collagen, associated with failure in PLOD3 delivery to collagen, is observed in Vipar knockdown mIMCD-3 cells.

Abnormal collagen modifications and trafficking in Vps33b/Vipar deficiency can explain the down-regulation of E-cadherin that characterises some polarized cell types in ARC and the cell model for this syndrome. These findings establish a role for VPS33B/VIPAR in the intracellular trafficking of collagen.

Acknowledgments

I am very thankful to all the members of the Medical and Molecular Genetics Department at the University of Birmingham and all members of the MRC-LMCB University College of London for their help and support through the time I spent in each department during my PhD. I would like to sincerely thank my supervisor Prof. Paul Gissen for the continuous support during all the time of my PhD research and writing of this thesis, for his friendly guidance and patience, for his motivation and great enthusiasm.

I would also like also to thank Federico Forneris for his important help through the project, the discussions, the suggestions and his collaborative and enthusiastic support.

Special thanks go to Ania Straatman-Iwanowska and Holly Smith for their precious help throughout the project and for allowing me to work everyday in a very supportive and friendly environment. I would like to thank Holly also for her help in improving my English (a part from the 'biscUit' word ☺) for her firiendship and all the sushi and icecreams I had thanks to her (and all the spinning to fix things back)! I would like to thank Ania also for her friendship and support, her patientce (and sometimes impatience ☺) that definitely helped me in becoming a better lab worker. Now hopefully nothing is between us and our plan to go for walks!

I would like to thank Chris Bruce and Danai Bem for their help during my permanence at the University of Birmingham. I would like to thank Holly Smith, Clare Rogerson and Anne-Marie Lyne for corrections and help with the thesis.

Special thank also to the Biology of Liver and Pancreatic Development and Disease (BOLD) network for finding this project. Finally I would like to thank my family and friends for their continued support and encouragement throughout this project.

Table of contents

List of figures

List of tables

List of Abbreviations

AJ – Adherens Junction

AJC – Apical Junction Complex

AOBS – Acousto Optical Beam Splitter

AP – Adaptor Protein

ARC – Arthrogryposis, Renal dysfunction and Cholestasis (syndrome)

ARE – Apical Recycling Endosome

ATP – Adenosine Tri-Phosphate

BARs – Bin/Amphiphysin/Rvs

BLAST – Basic Local Alignment Search Tool

BM – Basement Membrane

BP – Base Pairs

BSA – Bovine Serum Albumin

BSEP – Bile Salt Export Pump

CAM – Cell Adhesion Molecule

CCV – Clathrin Coated Vesicle

CD26 – dipeptidyl peptidase

CEA – Carcino Embryonic Antigen

CEACAM5 – CarcinEembryonic Antigen related Cell Adhesion Molecule 5

COP – COat Protein

CORVET – class C cORe Vacuole Endosome Tethering

CRE – Common Recycling Endosome

CV – Constitutive secretory Vesicles

DDR – Discoidin Domain Receptors

dHPLC – denaturing High Performance Liquid Chromatography

DIG – DIGoxigenin

DMEM – Dulbecco's Modified Eagle's Medium

DMSO – DiMethyl SulfOxide

DTT – DiThioThreitol

EBNA1 – Epstein Bar virus Nuclear Antigen

EC – Extracellular Cadherin

ECL – Enhance ChemiLuminescence

ECM – ExtraCellular Matrix

EDTA – Ethylene Diamine Tetraacetic Acid

EE – Early Endosome

EEA1 – Early Endosome Antigen 1

EGTA – Ethylene Glycol Tetraacetic Acid

EM – Electron Microscopy

EMT – Epithelial to Mesenchymal Transition

ENTHs – Epsin N Terminal Homology

ER – Endoplasmic Reticulum

ESCRT – Endosomal Sorting Complex Required for Transport

EtBr – Ethidium Bromide

FACIT – Fibril Associated Collagens with Interrupted Triple helices

FBS – Foetal Bovine Serum

FLIM – Fluorescence Life time Imaging

FRET – Förster (or Fluorescence) Resonance Energy Transfer

GAP – Rab GTP Activating Proteins

GBM – Glomerulus Basement Membrane

GDI – Rab GDP Dissociation Inhibitor

GDP – Guanosine Di-Phosphate

GEF – Guanine nucleotide Exchange Factors

GF – Gel Filtration

GFP – Green Fluorescence Protein

GGT – Glucosyltransferase

GM130 – Golgi Marker of 130 kDa

GOLGA5 – Golgin subfamily A member 5

GPI-AP – Glycosyl Phosphatidyl Inositol-Anchored Protein

GT – Galactosyltransferase

GTP – Guanosine Tri-Phosphate

HA – HemAgglutinin

HEK293 – Human Embryonic Kidney 293 (cells)

HEPES – HydroxyEthyl-1-PiperazineEthaneSulfonic acid

HOPS – HOmotypic fusion and vacuole Protein Sorting

HRP – Horse Radish Peroxidase

HSP47 – Heat Shock Protein 47

IgG – Immunoglobulin G

IP – ImmunoPrecipitation

LAMP1 – Lysosome Associated Membrane Protein 1

LB – Luria Broth

LC ESI MS MS – Liquid chromatography ElectroSpray Ionisation tandem Mass Spectrometry

LE – Late Endosome

LH – Lysyl Hydroxylase

LO – Lysyl Oxidase

MALLS – Multi Angle Laser Light Scattering

MDCK – Madin Darby Canine Kidney

mIMCD-3 – Murine Inner Medullary Collecting Duct-3 (cells)

MMPs – Metalloproteinases

MO – Morpholino Oligonucleotide

MS – Mass Spectrometry

MVB – Multi Vesicular Body

MWCO – Molecular Weight CutOff

NC1 – C-terminal non collagenous domain

NLS – Nuclear Localisation Sequence

NPFs – Nucleating Promoting Factors

NSF – N-ethylmaleimide Sensitive Factor

SNAP – Soluble NSF Attachment Protein

OD – Optical Density

pBLAST – protein Basic Local Alignment Search Tool

PBS – Phosphate Buffered Saline

PBST – Phosphate Buffered Saline with 0.1% Tween-20

PCR – Polymerase Chain Reaction

PDI – Protein Disulphide Isomerase

PFA – ParaFormaldehyde

PH – Prolyl Hydroxylase

PLOD – Procollagen Lysine 2 Oxoglutarate 5 Dioxygenase

PM – Plasma Membrane

PPI – Peptidyl Prolyl cis trans Isomerase

PVDF – PolyVinylidene DiFluoride

RabGGT – RAB GeranylGeranyl Transferase

RE – Recycling Endosome

RPM – Revolutions Per Minute

SDS – Sodium Dodecyl Sulfate

SDS PAGE – Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis

shRNA – short haiprin RNA

SM – Sec1/Munc18

SNAP – Soluble NSF Attachment Protein

SNARE – Soluble N-ethylmAlemide sensitive Receptors

SNAREs – SNAP Receptor

SV – Synaptic Vesicles

TBE – Tris/Borate/EDTA

TEMED – TEtraMethylEthyleneDiamine

TER – Trans Epithelial Resistance

TEV – Tobacco Etch Virus Protease

TGN – Trans Golgi Network

TJ – Tight Junction

UDP – Uracil Diphosphate Glucose

UV – UltraViolet

VIPAR – VPS33B Interacting Protein involved in polarity and Apical Restriction

VPS – Vacuole Protein Sorting

WB – Western Blot
WT – Wild Type

YFP – Yellow Fluorescent Protein

ZO – Zonula Occludens

γGT – Gamma Glutamyl Transpeptidase

CHAPTER ONE:

INTRODUCTION

1.1 - ARC syndrome

Arthrogryposis, renal dysfunction and cholestasis (ARC) syndrome (OMIM 208085) is a rare autosomal recessive disease typically characterised by neurogenic arthrogryposis, renal tubular dysfunction and neonatal cholestatic jaundice. Other features of the disease may vary from one patient to another, these include: severe metabolic acidosis, worsening nephrogenic diabetes insipidus, and low gamma glutamyl transpeptidase (γGT) activity (Gissen et al., 2004).

In most cases, clinical features are accompanied by ichthyosis, platelet α-granule biogenesis defect (that often leads to bleeding tendency) and severe failure to thrive. Other clinical features include mild dysmorphic features and absent corpus callosum (Gissen et al., 2006).

Patients with ARC rarely exceed the age of one year due to recurrent infections that lead to dehydration and acidosis (Tanaka et al., 1992). Those surviving longer develope liver cirrhosis and developmental delay (Eastham et al., 2001). A typical phenotype of a patient with ARC is shown in figure 1.1.

Figure 1.1 - Typical features of a patient with ARC syndrome can be seen, including joint contractures, dry skin and jaundice.

The two organs most affected by ARC syndrome are the liver and kidney. Immunostaining of liver and kidney biopsy samples from ARC patients have been performed using antibodies against markers that localise exclusively to the apical membrane of the aforementioned organs (Bull et al., 2006; Gissen et al., 2004). These markers include the carcinoembryonic antigen (CEA) and γ-GT (Gamma-glutamyl transpeptidase) for both liver and kidney, and dipeptidyl peptidase (CD26) for immunostaining of kidney only. Gissen et al (2004) observed that in individuals with ARC apical markers were mislocalised in the cytoplasm and/or basolateral domain. This was clearly distinct from the expression pattern found in age-matched control kidney and liver samples. As an example, immunostaining of CD26 in kidney biopsies from ARC patients and age-matched controls is reported in figure 1.2.

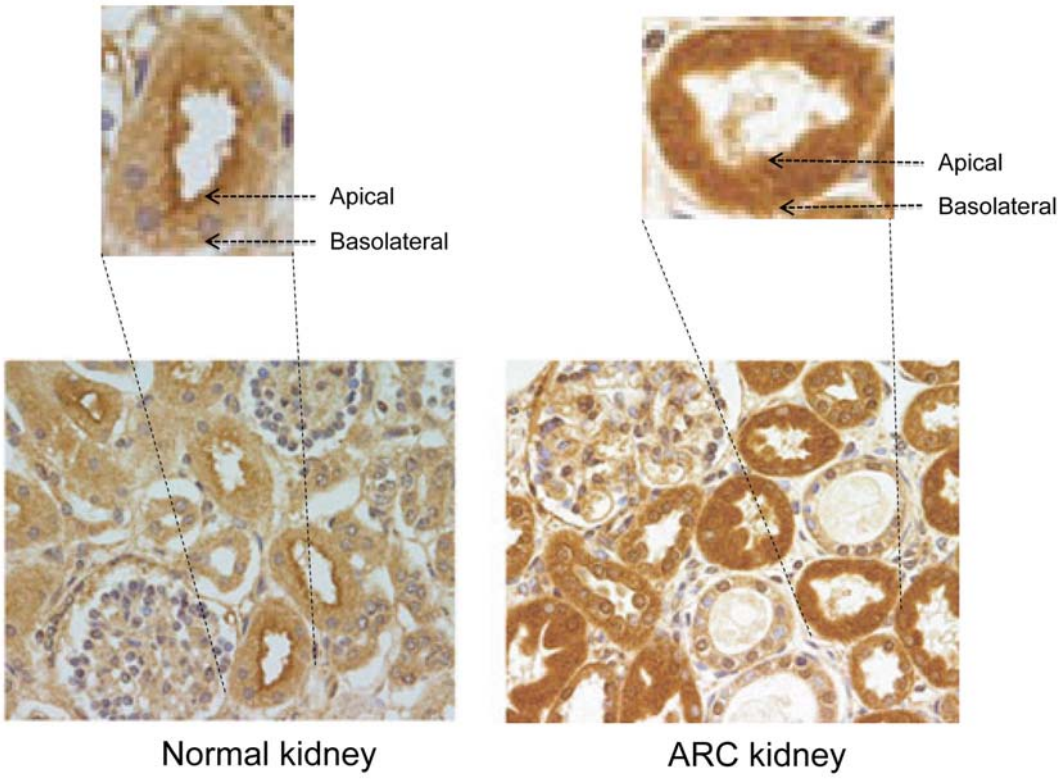


Figure 1.2 - Immunostaining of kidney biopsies from ARC patients and age matched controls. Staining was performed with an antibody against CD26. The zoomed-in micrograph shows immunostaining of CD26 at the apical membrane in the age-mathched controls but mislocalised to the cytoplasm in ARC kidney cells. Figure modified from (Gissen et al., 2004).

1.2 - VPS33B and ARC

Although the first case of ARC syndrome was reported in 1973 (Lutz-Richner and Landolt., 1973), the molecular basis of the disease was discovered only in 2004 when the genetic locus of ARC was mapped to chromosome 15q26.1. Different germline mutations were found in *VPS33B* (Vacuolar protein sorting 33B) gene (Gissen et al., 2004). Subsequent analysis of the ARC patients DNA suggested that germline mutations in *VPS33B* account for ~75 % of all reported cases (Cullinane et al., 2009; Gissen et al., 2006). This evidence suggested that other genes are involved in the

aetiology of ARC and their defects might cause the disease in the absence of mutations in *VPS33B*. The *VPS33B* genomic sequence (NC_000015.9) is 24,060 bp in length and is composed of 23 exons. Alternative splicing is predicted to produce nine transcript variants, four of which result in a protein product (figure 1.3). The longest processed mRNA (1,854 bp) encodes the VPS33B protein (NP_061138.3), which consists of 617 amino acids and has a molecular weight of 70.6 kDa.

The function of VPS33B has not yet been fully characterised in humans but the yeast homolog Vps33p is known to belong to the Sec1/Munc18 (SM) family of proteins (Peterson and Emr, 2001). SM proteins play an important role in conferring specificity to membrane fusion events by binding t-SNAREs of the syntaxin family (Chen et al., 2002; Gallwitz and Jahn, 2003; Toonen and Verhage, 2003).

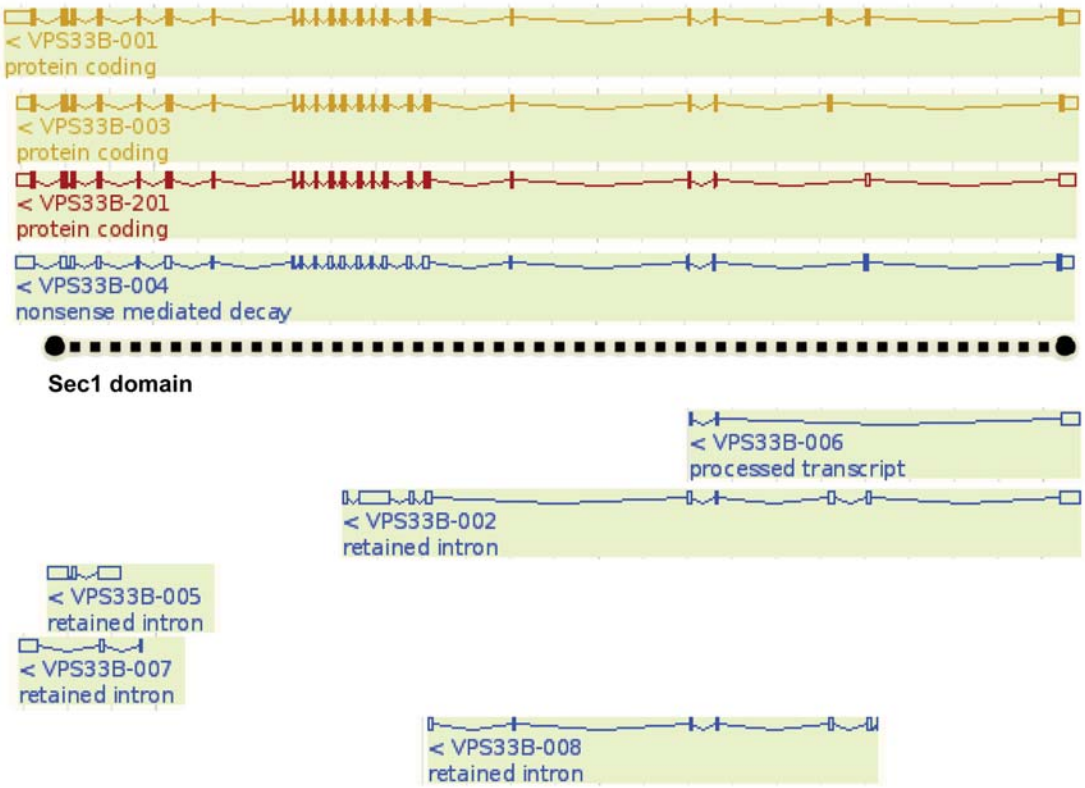


Figure 1.3 - VPS33B splice variants analysed with Ensemble http://www.ensembl.org/ Boxes are exons. Lines connecting boxes are introns. Filled boxes are coding sequences, and unfilled boxes are UTR (UnTranslated Region).
A gold coloured transcript is identical between Ensembl automated annotation and VEGA/Havana (http://vega.sanger.ac.uk/index.html) manual curation. A red transcript comes from either the Ensembl automatic annotation pipeline (http://www.ensembl.org/info/docs/genebuild/genome_annotation.html) or manual curation by the VEGA/Havana project. Blue transcripts are non-coding.
Nonsense-mediated decay transcripts contain nonsense mutations that prevent the expression of truncated or erroneous proteins. This type of transcript is annotated by VEGA/Havana. The dashed line indicates the Sec1 domain; it characterizes most of the coding sequence of the transcripts that result in a protein product.

Vps33p is a Class C vacuole protein sorting protein whose role is crucial for vacuolar biogenesis in *Saccharomyces cerevisiae* (Peterson and Emr, 2001). As a result Vps33p mutants show severe defects in endocytosis (Subramanian et al., 2004) associated with a block in intracellular trafficking of proteins. This is shown by a missorting of vacuolar hydrolases and abnormal secretion (Banta et al., 1990).

Two homologues of the yeast Vps33p have been identified in higher organisms: VPS33A and VPS33B (Gissen et al., 2005). Just like their yeast homolog, both VPS33A and VPS33B contain a Sec1 like domain (figure 1.3) that makes them part of the SM family of proteins. This suggests that VPS33B might play a crucial role in vesicular trafficking in mammalian organisms and that ARC phenotype may be secondary to abnormal organelle biogenesis (Gissen et al., 2005).

The function of VPS33A is however different from VPS33B. This suggestion comes from the hypo-pigmented phenotype of the *buff* (*bf*) mouse that carries a mutation in *VPS33A*. The *buff* mouse is considered a model for Hermansky - Pudlak syndrome (Chintala et al., 2009). Both *VPS33B* and *VPS33A* genes are ubiquitously expressed. In addition VPS33B is highly expressed in testis and has a low level of expression in the lung (Carim et al., 2000). *VPS33A* and *VPS33B* share only 31% identity and 51% similarity, which can explain the different functions performed by the two encoded proteins (Gissen et al., 2005). This is also consistent with the common origin of the genes and subsequent divergence.

The crystal structure of VPS33B is not yet known, therefore it is difficult to correlate mutations identified in patients with ARC patients with changes in the three-dimensional folding of the protein, its stability, or the binding affinity for possible interactors. However, prediction studies have been performed using the previously reported crystal structure of the interaction between neuronal-SEC1 and Syntaxin1 proteins in rat (Gissen et al., 2005; Misura et al., 2000). In this regard, it is important to underline that VPS33B and SEC1 share only 15% of identity (section 3.1). This prediction studies suggest that L30P, a missense mutation that causes a severe ARC

phenotype, might interfere with the N-terminal binding site for t-SNARE by disrupting protein-protein interactions (Gissen et al., 2005).

According to recent studies (Bach et al., 2008; Wong et al., 2011) VPS33B is autophosphorylated on a tyrosine residue. The tyrosine phosphatase PtpA of *M.tuberculosis* directly interacts and dephosphorylates VPS33B on its tyrosine residue and this blocks the phagosome maturation in *M.tuberculosis*-infected macrophages.

1.3 - VIPAR and ARC

Two different studies demonstrate that VPS33B interacts with a protein encoded by the gene initially termed *C14orf133* (Cullinane et al., 2010; Zhu et al., 2009). Thus the gene was named *VIPAR* (VPS33B-interacting protein involved in polarity and apical protein restriction) **(**NC_000014.8) (Cullinane et al., 2010) but confusingly was also carrying names *VPS16B* and *hSPE-39 (Spermatogenesis-defective protein homologue 39*) because its role in vesicular trafficking during spermatogenesis (L'Hernault and Faundez, 2011; Zhu et al., 2009). The final and official name of the gene is *VIPAS39*, which is a combination of *VIPAR* and *Spe39.* In this thesis *VIPAS39* is used for the name of the gene while VIPAR for the name of the protein. The genomic sequence of *VIPAS39* is 30,966 bp long. Alternative splicing results in twelve transcript variants among which seven are decoded to produce a protein (figure 1.4). The first four variants (named *VIPAS39-005, -001. -201, -003*) encode the same protein of 493 amino acids (isoform 1) and between them differ in the 5' UTR (figure 1.4). Variant *VIPAS39-002* and *-004* lack an alternate in-frame exon in the 5' coding region compared to variant

VIPAS39-005. The resulting protein contains 444 amino acids (isoform 2) and is shorter than isoform 1 (figure 1.4).

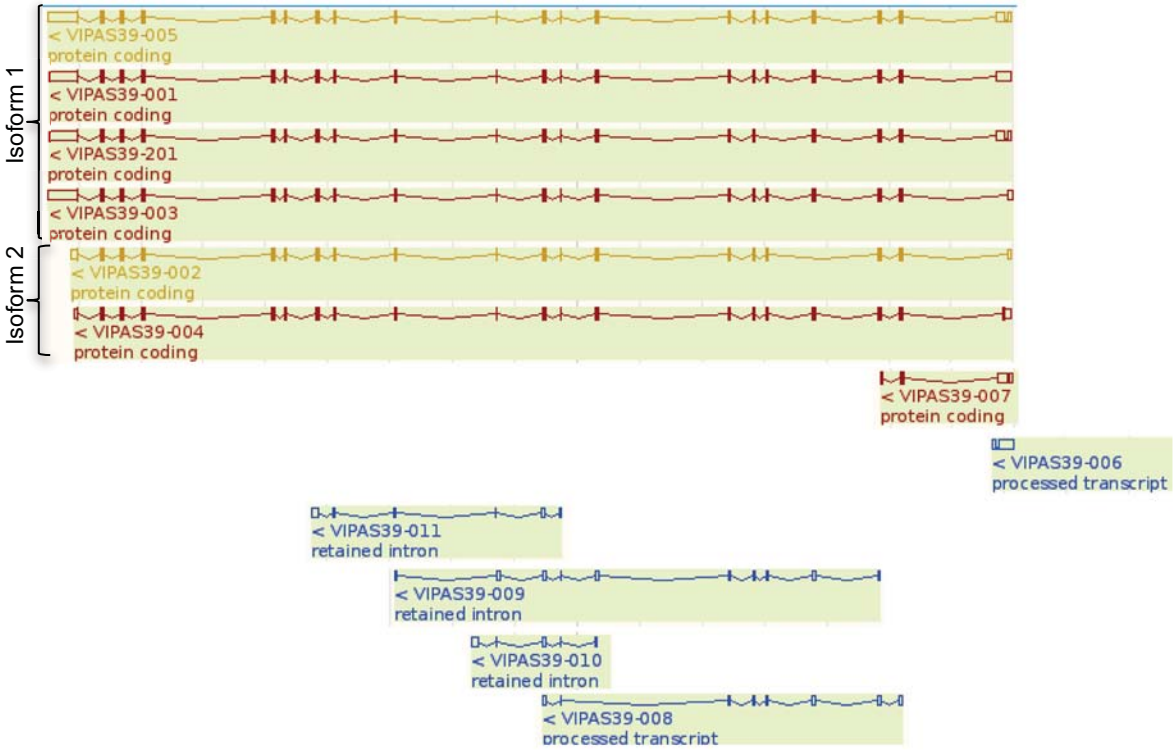
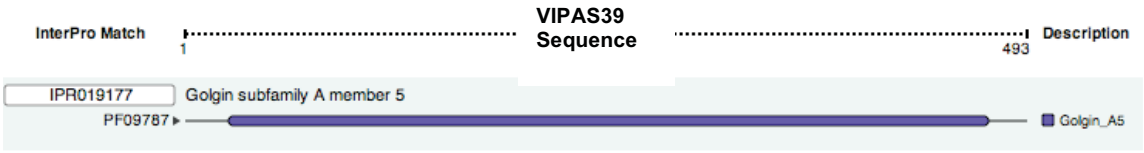


Figure 1.4 - VIPAS39 splice variants analysed with Ensembl (http://www.ensembl.org/)
Boxes are exons. Lines connecting boxes are introns. Filled boxes are coding sequence, and unfilled boxes are UTR (UnTranslated Region).
A gold coloured transcript is identical between Ensembl automated annotation and VEGA/Havana (http://vega.sanger.ac.uk/index.html) manual curation.
A red transcript comes from either the Ensembl automatic annotation pipeline (http://www.ensembl.org/info/docs/genebuild/genome_annotation.html) or the manual curation by the VEGA/Havana project. Blue transcripts are non-coding.
The protein coding variants generate two different isoforms of VIPAS39 indicated in figure.

The precise role of the VPS33B/VIPAR in vesicular trafficking is not completely understood. Studies of *in-vitro* interaction have shown that the *C. elegans* homologue of VIPAR interacts with both VPS33A and VPS33B (Zhu et al., 2009). However experiments performed in our laboratory did not reveal appreciable evidence for interaction between VPS33A and VIPAR (Smith et al., 2012). The same *in-vitro* study

(Zhu et al., 2009) has shown that VIPAR is part of the homotypic fusion and vacuole protein sorting (HOPS) complex (section 1.4.4.4.1) which regulates vesicle docking through its interaction with SNARE proteins. Published data from our laboratory suggest that VIPAR may interact with VPS18 (the core protein of the HOPS complex) only in the absence of VPS33B. Therefore the VPS33B–VIPAR complex is unlikely to be part of the HOPS complex (Smith et al., 2012). The homolog of VIPAR in *Drosophila* (FOB) is thought to play an important role in the fusion of phagosomes with late endosomes/lysosomes in fly macrophages (Akbar et al., 2011). Defects in *Drosophila*'s immune response to non-pathogenic bacteria support this hypothesis. Other studies have shown that VPS33B/VIPAR might be involved in the trafficking from the trans-Golgi network in Dami cells (human megakaryocytic cell line) (Urban et al., 2012). Furthermore electron micrographs of enterocytes derived from *vps33b* morpholino-injected Zebrafish larvae showed dilated stacks of Golgi cisternae in these cells (Matthews et al., 2005). Further evidence of the possible role of VIPAR in Golgi trafficking comes from bioinformatics analysis of VIPAR using the Pfam and SMART databases. This found a large Golgin subfamily A member 5 domain in the VIPAR protein sequence (figure 1.5). Golgin family members are coiled-coil motif proteins known to interact with Rab GTPases and mediate tethering of vesicles to Golgi membranes and cisternal membranes to each other (Burguete et al., 2008; Satoh et al., 2003)

A



B

Confidently predicted domains, repeats, motifs and features:

Name	Begin	End	E-value
Pfam:Golgin_A5	24	491	1.80e-154

*Figure 1.5 - Analysis of VIPAR protein sequence using Pfam **(A)** and SMART **(B)** databases predict a large Golgin subfamily A member 5 domain along its sequence.*

An abnormal expression and distribution of E-cadherin, one of the major components of adherens junctions, was highlighted in a patient with ARC (Cullinane et al., 2010). The same abnormality occurs in mouse inner medullary collecting duct (mIMCD-3) cell lines knockdown for *VPS33B* and *VIPAR*. It was found that in the knockdown cell lines E-cadherin was transcriptionally down-regulated. (Cullinane et al., 2010). Co-localisation and co-immunoprecipitation experiments performed in this study suggest that the VPS33B-VIPAR complex interacts with RAB11A and regulates indirectly the expression of E-cadherin at adherens junctions. (Cullinane et al., 2010)

The mislocalisation of proteins normally present in the apical portion of the renal tubular and bile canalicular membranes (showed by immunostaining of CEA, γGT, CEACAM5, BSEP) may be able to explain cholestasis, dehydration and urinary waste of amino acids and sugars in ARC patients. In addition E-cadherin down-regulation with consequent disruption of the adherens and tight junctions may be associated with malformation of luminal structures such as bile ducts and renal tubules (Cullinane et al., 2010; Zhao et al., 2010).

1.4 - Insights into vesicular transport

1.4.1 - Endocytic and biosynthetic pathway: the central role of TGN

Bioinformatics analysis in combination with published data discussed above and clinical features of individuals with ARC, suggests that one of the roles of VPS33B and VIPAR is to mediate vesicular trafficking and epithelial polarisation through the transport of proteins to the apical surface.

Proteins are transported in membrane-bound vesicular structures through and between different cellular compartments. Intracellular transport of proteins is divided into two routes according to the direction of the transport in relation to the cell membrane (Ang and Folsch, 2012; Erickson and Bocock, 2007; Gissen and Maher, 2007; Keller and Simons, 1997). They are known as the biosynthetic and endocytic pathways. In the biosynthetic route the *de novo* synthesised proteins are transferred from the cytosolic ribosomes to the Endoplasmic Reticulum (ER) where the correct folding of the proteins occurs (Sannerud et al., 2003). The cargo proteins are then packed into transport vesicles and sent towards Golgi cisternae (first *cis* and then *trans* Golgi) for post-translational modifications. In the Golgi apparatus the proteins are "labelled" with a sorting signal that determines their final destination within the cell (Gissen and Maher, 2007). In the endocytic pathway, molecules are internalised from the extracellular space and plasma membrane. Subsequently the cargo molecules, encapsulated in endocytotic vesicles, are transported to early sorting endosomes. In the early endosomes the molecule can either be recycled back to the plasma membrane or transferred to late endosomes and then delivered to lysosomes for degradation (figure 1.6) (De Matteis and Luini, 2008; Gissen and Maher, 2007; Gruenberg and Maxfield, 1995; Sachse et al.,

2002). The trans Golgi network (TGN) is a major sorting station that directs newly synthesised proteins from earlier Golgi compartments to different destinations (figure 1.6) (De Matteis and Luini, 2008; Gu et al., 2001). Apart from this central role in the secretory pathway, the TGN also receives extracellular materials from the endocytic pathway. The main cellular compartments that receive material from TGN include: the apical plasma membrane (1), the basolateral plasma membrane (2), recycling endosomes (3), early endosomes (4), late endosomes (5) and specialised compartments such as secretory granules (6) in secretory cells (figure 1.6) (Gu et al., 2001). Different components of trafficking machinery (such as coat proteins) are involved in ensuring that the vesicles are delivered to different destinations. Clathrin-coated vesicles mediate the transport between TGN and endosomes. The recruitment of clathrin on membranes, as well as the recognition of sorting signals in cargo molecules, is mediated by adaptor-coat proteins (AP). Three-dimensional tomographic studies have shown that TGN is composed of tubules emanating from the last two trans-Golgi cisternae and only the tubules deriving from the last cisterna are clathrin coated (Gu et al., 2001). Proteins are transported from TGN to endosomes by clathrin/AP-1-coated vesicles.

The AP-1 complex is a heterotetramer consisting of two large subunits termed γ and β, a medium-sized subunit (μ1), and a small subunit (ς1) (Orzech et al., 2001). The β-subunit is particularly important for clathrin binding (Shih et al., 1995), while the γ-adaptin is most likely involved in targeting (Page and Robinson, 1995). The μ1 subunit is formed of two chains: μ1A, the ubiquitous chain, probably involved in the recognition of tyrosine-based sorting signals encoded by cargo proteins (Kirchhausen, 1999), and the recently identified epithelium-specific μ1B. μ1B was found to be expressed in various polarised epithelial cell lines and is implicated in basolateral targeting of

specific receptors (Folsch et al., 1999) (section 1.4.2). The function of the ς subunit is unknown.

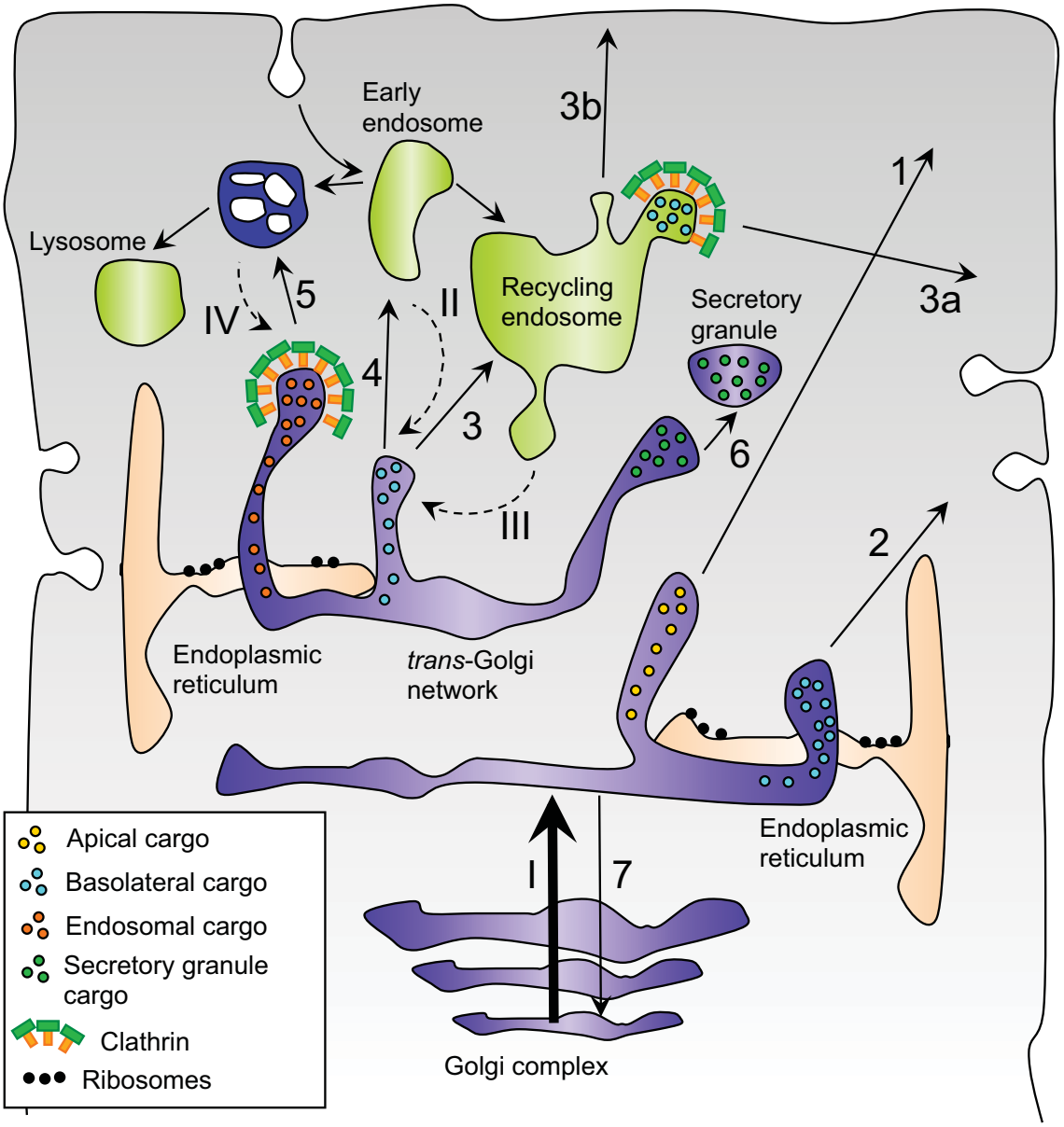


Figure 1.6 - In the biosynthetic route newly synthesised proteins derived from earlier Golgi compartments (I) are sorted from TGN towards different destinations (1–5). The exit routes from the TGN include those towards the apical plasma membrane (1), the basolateral plasma membrane (2), recycling endosomes (3), early endosomes (4), late endosomes (5) and specialized compartments such as secretory granules (6) in secretory cells. The TGN also receives also extracellular materials from the endocytic pathway (II–IV) and sends it back to earlier Golgi compartments (7). Figure modified from (De Matteis and Luini, 2008).

1.4.2 - Polarized cells and vesicular trafficking

The fundamental feature of cell polarity is the division of the plasma membrane into two discrete domains, one apical and one basolateral, each characterised by a particular distribution of proteins and lipids. The maintenance of two biochemically and functionally distinct domains in the membrane is tailored to the particular functions of each domain (Schuck and Simons, 2004). Examples of polarised cells include neurons, migrating cells and the growing bud in *Saccharomyces cerevisiae*. However, the best-studied type of cell polarity is found in epithelia, which are the most common type of tissue in animals. The epithelial apical membrane faces the luminal compartment and is specialised primarily in secretion or absorption (Rodriguez-Boulan and Nelson, 1989). The basolateral domain faces the extracellular fluid compartment and specialises in the establishment of cell-cell contacts, recognition and transduction of signals.

Polarised protein trafficking in epithelial cells is triggered by signals embedded within the sequence of the proteins themselves. The intracellular trafficking machinery, after reading and interpreting these sorting signals, segregates and packages membrane proteins into specialised transport vesicles destined for apical or basolateral delivery.

Once the polarity is established, highly specialised cell machinery is required for its maintenance; membrane proteins are brought to their appropriate sites of action and recycled from one domain to the other. The biosynthetic and endocytic pathways coexist in every type of cell. In polarised cells there is another particular route, called transcytosis (Bastaki et al., 2002). In this route, vesicles containing macromolecules are transferred from one domain of the plasma membrane to the opposite one. This mechanism is a strategy that polarised cells employ to interconnect different sides of the

cell, maintaining at the same time the different composition of these microenvironments (Schuck and Simons, 2004).

The cytoskeleton has a crucial role in generating and maintaining apical and basolateral polarity in epithelial cells. Microtubules play an essential role for the organisation of the apical membrane, whereas the actin cytoskeleton has a role in basolateral polarity (Rodriguez-Boulan E et al ., 2005).

A large number of diseases, including ARC syndrome, are known to be associated with failure of normal protein sorting to the correct cell surface or intracellular compartment (Carmosino et al., 2010).

1.4.3 - Steps in vesicular transport: from budding to fusion

Vesicular transport of a cargo protein from a donor to an acceptor compartment can be summarised into 3 major steps (Carmosino et al., 2010). In the first step the trafficking vesicle is assembled at the donor compartment (figure 1.7A). A specific v-SNARE is bound to the membrane of the donor compartment in proximity to a transmembrane receptor that has exclusive affinity for the cargo protein (Kirchhausen, 2000). This is the zone of the membrane where the vesicle budding is going to occur. The cargo protein is recognised by the transmembrane receptor and therefore internalised into the budding vesicle. The coat protein complexes assemble at the membrane that starts to increase its curvature and form a vesicle-like structure (Bonifacino and Lippincott-Schwartz, 2003). This process is mediated by Rab proteins, such as Rab1 and Rab9 (Bonifacino and Lippincott-Schwartz, 2003). Budding of transport intermediates from cellular membranes involves a variety of molecular machinery including protein coats,

membrane-sculpting domains such as ENTHs (epsin N-terminal homology) and BARs (Bin/Amphiphysin/Rvs), enzymes that translocate or modify lipids, actin nucleating promoting factors (NPFs), and molecular motors participate in membrane deformation (Farsad and De Camilli, 2003; Idrissi et al., 2012). Once the vesicle has budded from the donor compartment it moves along the actin cytoskeleton (Stamnes, 2002) (figure 1.7B) that guides the protein to the acceptor compartment. Motor proteins, GTP bound Rabs (such as Rab6 and Rab33) and Rab effector molecules are required in this process (Stenmark and Olkkonen, 2001). The vesicle loses its coat and becomes tethered to the acceptor compartment by the combination of tethering factors, such as the Homotypic fusion and vacuole protein sorting (HOPS) complex and Rab proteins (such as Rab1 and Rab5) (Yu and Hughson, 2010). Successively the t-SNARE at the acceptor compartment and the v-SNARE at the vesicle, assemble into a four-helix bundle called the "trans-SNARE complex" which promotes fusion (section 1.4.4.1) (Jahn and Scheller, 2006). This process is assisted by Rabs (such as Rab3 and Rab9) and Sec1/Munc18 (SM) proteins, as well as a number of other proteins (Sudhof and Rothman, 2009). The cargo protein is then released into the acceptor compartment (figure 1.7C).

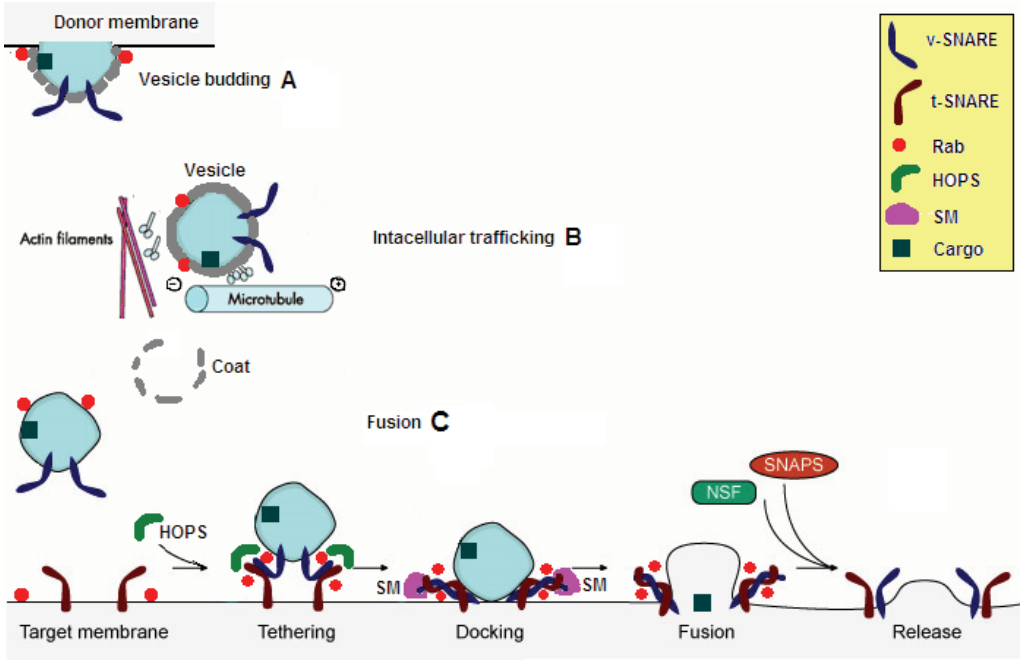


Figure 1.7 - Steps in vesicular transport of a cargo protein from a donor to an acceptor compartment. ***(A)*** *Assembly of the trafficking vesicle.* ***(B)*** *Intracellular trafficking.* ***(C)*** *Fusion of the vesicle to the target membrane.*

1.4.4 - Vesicular trafficking machinery

As mentioned above, VPS33B belongs to the SM family of proteins that confer specificity to fusion events by binding t-SNAREs of the syntaxin family (Dulubova et al., 2003; Toonen and Verhage, 2003). Altogether SNAREs, SM and other docking and fusion proteins generate very complex intracellular machinery.

1.4.4.1 - SNARE proteins

SNARE proteins (soluble N-ethylmaleimide-sensitive-factor attachment protein receptor) are central components of the membrane fusion machinery in cells. They directly mediate intracellular membrane fusion through the binding of SNAREs located

at the donor membrane to SNAREs placed at the acceptor membrane (Chen and Scheller, 2001; Jahn et al., 2003; Ungar and Hughson, 2003).

SNAREs were first biochemically identified in 1984 when Rothman and colleagues isolated an *N*-ethylmaleimide sensitive factor (NSF) and an adaptor protein called soluble NSF attachment protein (SNAP), both required for efficient transport of proteins *in-vitro* (Balch et al., 1984). The receptors of those molecules were first isolated from brain tissues and were termed SNAREs (SNAP receptors) (Sollner et al., 1993). Since then SNARE proteins have been found to be involved in every membrane trafficking event (Chen and Scheller, 2001; Jahn et al., 2003; Kavalali, 2002; Pelham, 2001). Rothman and colleagues (Sollner et al., 1993) named v-SNARE (v- vesicle) the SNARE protein located at the donor compartment, and t-SNARE (t- target) the SNARE protein located at the acceptor compartment. Another nomenclature categorises SNAREs into R-SNAREs (R- arginine) and Q-SNAREs (Q- glutamine) because of a single key amino acid residue located in the middle of the SNARE domain: most v-SNAREs have an arginine residue while t-SNAREs have a glutamine (Fasshauer et al., 1998).

The structure of SNARE proteins is characterised by a unique SNARE motif of about 60-70 amino acid residues (Fasshauer et al., 1998). The SNARE motif consists of a repeating heptad pattern of hydrophobic residues organised in a α-helical structure. Four α-helices assemble into a helix bundle that is stabilised by hydrophobic residues facing the core of the bundle (Antonin et al., 2002; Poirier et al., 1998; Sutton et al., 1998). The central arginines and glutamines form an ionic layer composed of a hydrogen-bonded network in which three Q-SNAREs bind to one R-SNARE. The R/Q nomenclature is a result of the polar layer identification (Fasshauer et al., 1998).

Q-SNAREs are further classified into Qa, Qb, or Qc subgroups according to the positions of the SNARE motif in the four-helix structure (Bock et al., 2001). As previously mentioned, the formation of the *trans*-SNARE complex is the driving force that brings together the two membranes, mediating their fusion. After fusion has occurred, the *trans*-SNARE complex is converted into a very stable *cis* complex where SNAREs are associated within the same membrane (Bock et al., 2001). The NSF/SNAP chaperone system utilises the energy of ATP hydrolysis to disassemble the *cis*-SNARE complex (figure 1.7C) (May et al., 2001). Thus SNAREs are recycled for another round of membrane fusion events.

Originally SNARE proteins were thought to be the component that confers specificity to the fusion reaction. This hypothesis, known as the SNARE hypothesis, was based on the identification of numerous SNARE proteins in eukaryotic cells that were thought to be required in distinct fusion processes (Sollner et al., 1993). Subsequent literature data showed that this is not absolute truth. SNARE assembly and function is regulated by numerous other proteins that provide specificity to the fusion reactions (Bock et al., 2001; Chen and Scheller, 2001). These SNARE-associated proteins can interact either directly or indirectly with SNAREs. The three main groups include Rab GTPases (1.4.4.3), SM proteins (section 1.4.4.2) and other tethering factors.

1.4.4.2 - SM proteins

SM proteins are soluble, 60–90 kDa proteins that interact directly with SNAREs and mediate every membrane fusion reaction (Malsam et al., 2008). There are several models that show different ways in which SNARE and SM can bind together, but in all

of these an SM protein binds specifically to a Qa-SNARE (Gupta and Brent Heath, 2002). The diverse binding models have been associated with the ambiguous role played by SM proteins in the SNARE complex assembly. *In-vivo* studies have demonstrated that this role might be either positive (enhance SNARE complex assembly) or negative (inhibit SNARE complex assembly) (Carr and Novick, 2000). The SM protein Vps33p (the yeast homolog of VPS33B) is an example of the positive role that SM proteins play in the SNARE complex formation (Sato et al., 2000; Seals et al., 2000). Recent published data show that Vps33p binds to the quaternary SNARE complex Vam3, Vam7, and Nyv1 on their SNARE domains (Lobingier and Merz, 2012). The identification of other interacting proteins that bind to the SM-SNARE complex can further help in elucidation of the role of specific SM proteins (Dulubova et al., 2003).

1.4.4.3 - Rab GTPases

Ras-Associated Binding (RAB) proteins are small GTPases that are considered a key regulatory factor in the intracellular traffic. RAB proteins belong to the RAS superfamily. Sec4, the first Rab family member to be discovered in yeast, encodes a small GTP-binding protein related to Ras (Salminen and Novick, 1987). Since then, more than 60 RAB proteins have been identified in mammals and more than 10 in yeast, highlighting the importance of RABs in cell homeostasis and intracellular trafficking (Pereira-Leal and Seabra, 2001). In agreement with this, mutations in genes encoding RABs have been associated with many human diseases such as immunodeficiency, cancer, and neurodegeneration, Carpenter and Griselli syndrome (Bem et al., 2011; Corbeel and Freson, 2008; Schwartz et al., 2007; Stein et al., 2003).

RAB proteins are specifically located to distinct organelles and associated vesicles (figure 1.8). They are involved in all the steps of intracellular trafficking described above including vesicle budding, traffic of the vesicle along the cytoskeleton and vesicular fusion (Tuvim et al., 2001).

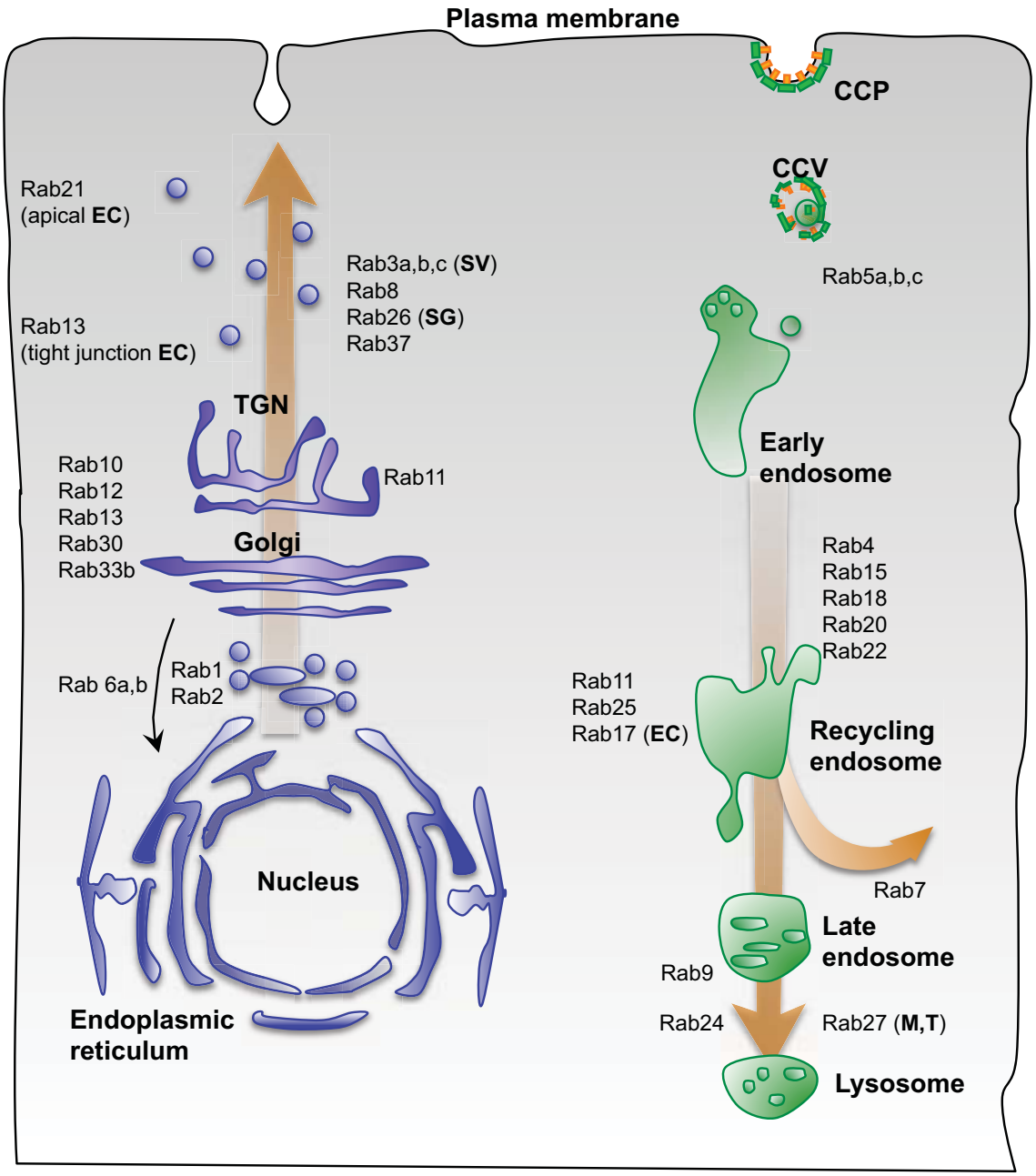


Figure 1.8 - The diversity of Rabs in intracellular vesicular pathways (CCV, clathrin-coated vesicle; CCP, clathrin-coated pit; EC, epithelial cells; M, melanosomes; SG, secretory granules; SV, synaptic vesicles; T, T-cell granules; TGN, trans-Golgi network.). Figure modified from (Zerial and McBride, 2001).

Experimental data show that direct interaction between RABs and SNARE proteins can enhance the formation of the trans-SNARE complex (Lian et al., 1994). However other published data suggest that this role may be indirect and requires RABs interaction with tethering factors (Grote and Novick, 1999; Hutagalung and Novick, 2011; Whyte and Munro, 2002). The role played by RABs in vesicular fusion is therefore not completely characterised. The diverse functions of RABs originate from their ability to interact with different RAB-associated/RAB-effector proteins. RABs regulate the activity of downstream effectors through the RAB GTPase cycle (figure 1.9) (Stenmark and Olkkonen, 2001). In the active form RABs bind GTP and activate downstream effectors. Conversely, the inactive form (GDP-bound) is unable to activate effector proteins. This on/off 'switch' mechanism is controlled by the RAB-associated proteins GEFs (Guanine nucleotide exchange factors) that catalyse the exchange of GTP for GDP, and GAPs (GTPase activating proteins), which stimulate the hydrolysis of GTP (Pereira-Leal and Seabra, 2000).

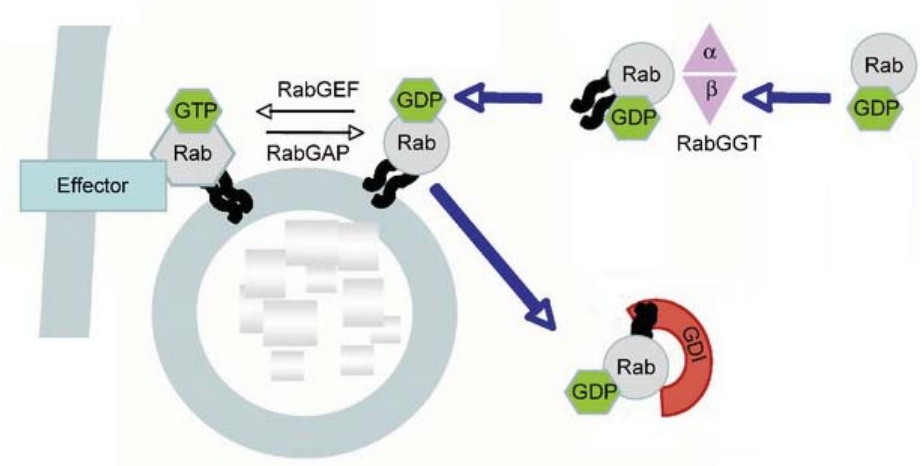


Figure 1.9 - RAB protein–GTPase cycles: GDP/GTP exchange and membrane association/dissociation (Adapted from (Corbeel and Freson, 2008)).

RAB proteins are synthesised as soluble proteins into the cytosol where they remain in an inactive state until required. In response to intracellular and extracellular signals, RABs associate with the lipid bilayer of membranes after being post-translationally modified. This modification is mediated by RAB geranylgeranyl transferase (RabGGT), which consists of two subunits (RabGGTa and RabGGTß) that catalyse the prenylation of two cysteines in the C-terminus of RABs (figure 1.9) *(Corbeel and Freson, 2008)*. The RAB protein can therefore bind to the donor membrane where it becomes subsequently activated by a specific GEF. In the active state, RAB interacts with its specific effectors, performing its function in intracellular trafficking. Once the required function is completed, the RAB protein is inactivated by a specific GAP and is removed from the membrane by a Rab-GDI (GDP dissociation inhibitor) (figure 1.9). The cycle is finally completed with the RAB protein returning to the cytosol and ready for another cycle once required.

1.4.4.3.1 - The RAB11 GTPase subfamily

Transport from the trans-Golgi network (TGN) to the cell surface, discussed in section 1.4.1, requires RAB8 or RAB11 (Ullrich et al., 1996), figure 1.6).

RAB11 has been detected on a variety of subcellular membranes. In non-polarised cells it was found associated with both Golgi and recycling endosomes. In addition, transport from sorting endosomes to the recycling compartment has been shown to require functional RAB11A (Ullrich et al., 1996). In polarised cells RAB11 has been localised to Golgi, recycling endosomes, as well as to apical vesicle populations (Goldenring et al., 1996). In regulated secretory cells RAB11 has been associated with TGN and TGN-

derived vesicles of both constitutive and the regulated secretory pathway (Urbe et al., 1993). RAB11 might play also a role in the exocytic membrane trafficing and is specifically required for TGN–to–plasma membrane transport (Chen et al., 1998).

The RAB11 GTPase subfamily comprises three members: RAB11A, RAB11B and RAB25/RAB11C (Kelly et al., 2012). Unlike universally expressed RAB11A and B, RAB25 expression seems to be limited to polarised epithelia such as gastrointestinal mucosa, kidney, and lung (Goldenring et al., 1993). RAB25 function is thought to mirror that of RAB11A; it associates with the apical recycling endosome (ARE) in polarised epithelial cells where it regulates the processes of transcytosis and recycling to the plasma membrane (Calhoun and Goldenring, 1997; Casanova et al., 1999). Similar to RAB11A, it has been shown that RAB11B associates with recycling endosomes (Schlierf et al., 2000) and is associated with the apical recycling compartment where it regulates targeting of proteins to the apical plasma membrane (Butterworth et al., 2012; Silvis et al., 2009).

Changes in RAB25 expression, including both up-regulation and down-regulation have been associated with breast, ovarian, colon and other types of cancer (Agarwal et al., 2009; Goldenring et al., 1993; Tong et al., 2012). This is promoted by the interaction between RAB25 and $\alpha 5\beta 1$ integrin that leads to loss of polarity and cell migration (Caswell et al., 2007).

1.4.4.4 - Tethering factors

Tethering factors are thought to be the initiators of the fusion reactions between two membranes (Pfeffer, 1999; Waters and Hughson, 2000; Whyte and Munro, 2002). In association with RAB proteins they stimulate the formation of the initial pairing between membranes that precedes the SNARE complex assembly. Indeed many tethering factors interact at the same time with SNAREs and RABs, thus facilitating the SNARE complex formation (Guo et al., 2000). Therefore tethering factors are other essential elements that confer specificity to the fusion reaction.

Despite their structural diversity, tethering factors are divided into two main groups: long coiled-coil proteins and large hetero-oligomeric complexes. The first group includes the Golgi protein p115/Uso1 (Waters et al., 1992) and the endosomal EEA1 (Simonsen et al., 1998). Hetero-oligomeric proteins of 4-10 subunits belong to the second group of tethering factors and include the COG (Ungar et al., 2002) and TRAPP (Sacher et al., 1998) complexes at the Golgi compartment, the Dsl1p complex at the ER (Reilly et al., 2001), the exocyst or Sec6/8 complex at the plasma membrane (Hsu et al., 1996) and the HOPS and CORVET complexes at the vacuole/lysosomes (Peplowska et al., 2007; Sato et al., 2000). Although it is known that many tethering factors interact with SNAREs (Ungar and Hughson, 2003), and some, such as the TRAPP complex and the Vam6 subunit of the HOPS complex, show RAB GEFs properties (Wang et al., 2000; Wurmser et al., 2000), the function of many domains and subunits of different tethering factors remains unknown.

This thesis is focusing on the proteins whose yeast homologues are involved in the HOPS complex, including the SM protein Vps33p.

1.4.4.4.1 - HOPS complex

The HOmotypic Fusion and vacuole Protein Sorting complex (HOPS) is a multi-subunit complex initially identified as a tethering factor for vacuolar fusion in yeast (Collins et al., 2005). There are 6 subunits that form the HOPS complex: Vps11, Vps16, Vps18, Vps33, Vps41 (Vam2) and Vps39 (Vam6) (Nakamura et al., 1997; Price et al., 2000) (figure 3.3). In yeast the HOPS complex has been shown to interact with the active form of the Rab protein Ypt7 (yeast Rab7) and mediate vesicle tethering by binding to SNAREs (Collins et al., 2005; Stroupe et al., 2006). This event precedes the trans-SNARE complex formation.

The individual subunits are characterised by completely different domains: as mentioned above Vps33p contains a Sec1 like domain while Vps11 and Vps18 have RING domains at the C-terminus (Rieder and Emr, 1997) (figure 1.10). Both Vps11 and Vps18 contain a clathrin domain, a region of about 140 amino acids long composed of multiple alpha helical repeats. Clathrin domains occur in the arm region of the clathrin heavy chain and are required for clathrin heavy chain self-assembly (figure 1.10).

A.

Source	Domain	Start	End
Pfam A	Clathrin	411	548
Pfam A	zf-C3HC4_2	818	861
Pfam A	VPS11_C	861	909

B.

Source	Domain	Start	End
Pfam A	Pep3_Vps18	291	436
Pfam A	Clathrin	619	771

Figure 1.10 - Analysis of VPS11 and VPS18 protein domains using Pfam database. ***A)*** *Domains identified in VPS11 include: -clathrin domain; -the Zinc finger C3HC4 type (RING finger), a member of clan RING (CL0229) with a total of 19 members; -the VPS11 C of about 50 amino acids in length.*
B) *Domains identified in VPS18 include: -clathrin domain; -the Pep3/Vps18/, a region found in a number of proteins identified as involved in golgi functions and vacuolar sorting of proteins. The members of Pep3/Vps18 contain a C-terminal ring finger domain.*

From the other side, Vps41 is involved in tethering (Price et al., 2000) as well as in the biogenesis of AP-3 vesicles (Rehling et al., 1999). Vps41 has recently been shown to be involved in the fusion of trans-Golgi network-derived lysosome-associated membrane protein carriers with late endosomes (Pols et al., 2013).

Data in literature have shown a direct interaction between Vps11 and activated Rab5 on early endosomes (Christoforidis et al., 1999). On the other hand, Vps39 has properties of a Rab7 GEF protein and has been shown to act at late endosomes (Wurmser et al., 2000). However, other studies have not confirmed Vps39 functioning as a Rab7 GEF and therefore the GEF activity of Vps39 remains controversial (Peralta et al., 2010).

These findings are consistent with the key role that the HOPS complex plays in the conversion of early into late endosomes.

The mammalian HOPS complex remains poorly understood. The human homologues of the yeast HOPS complex components include VPS11, VPS16, VPS18, VPS33A, VPS39 and VPS41 that localise to the endosomal compartment (Huizing et al., 2001). The above indicates a possible evolutionary conserved role of the HOPS complex in intracellular trafficking. Published data suggests that the HOPS complex is required for the Rab5 to Rab7 endosomal conversion and also for phagosome maturation in *C. elegans* (Kinchen et al., 2008). Mammalian cells contain two homologs of yeast Vam6: hVPS39 (Caplan et al., 2001) and TRAP-1 (Charng et al., 1998) which have both been implicated in Smad signalling at the endosome. As mentioned above, VPS16 and VPS33A might be part of the mammalian HOPS complex (Kinchen et al., 2008; Zhu et al., 2009) but at this point it is not known if this is the case for VPS33B and VIPAR.

1.4.4.4.2 - The CORVET complex

The CORVET (class C cORe Vacuole Endosome Tethering) complex has been recently identified as a novel complex involved in endosomal tethering (Peplowska et al., 2007). It shares high similarity with the HOPS complex as both complexes contain the class C VPS proteins. However the HOPS subunit Vam6/Vps39 is replaced by the homologue Vps3 and in the same way Vps8 substitutes for the HOPS subunit Vam2/Vps41 (Peplowska et al., 2007). Studies in yeast have shown that the CORVET complex interacts with Vps21 (Rab5 homologue). Activated Vps21 in association with Vps8 and Vps3 mediate endosomal clustering (Solinger and Spang, 2013).

It was suggested that HOPS and CORVET complexes could convert from one to the other and this process would be associated with the conversion of early (Rab5 positive) to late (Rab7 positive) endosomes. If this is correct then whilst the HOPS complex is crucial for conversion from early to late endosomes, the CORVET complex could mediate the retrograde conversion, i.e. from late to early endosomes (Solinger and Spang, 2013).

1.5 - Extracellular matrix

The extracellular matrix (ECM) is a complex three-dimensional network of precisely organised macromolecules that provide a substrate for cell anchorage, scaffolding and strength. Furthermore, it has a key role in regulating numerous cellular functions such as cell shape maintenance, cell migration and polarity as well as cell differentiation, proliferation and apoptosis (Aszodi et al., 2006; Aumailley and Gayraud, 1998) (figure 1.11). The cell-matrix interactions respond to extracellular stimuli and initiate intracellular signal transduction cascades, which ultimately regulate the expression of specific genes and affect cell behavior.

Depending on the type of cell it is associated with, the extracellular matrix can also perform other functions such as guidance of the cell migration during embryonic development, wound repair, fibrosis and tissue morphogenesis (de Vega et al., 2009).

Increased synthesis of ECM components can contribute to tumour growth and cancer progression (Cox and Erler, 2011). A better understanding of how the cells communicate with their environment is changing the way we view cancer.

Extracellular matrix includes the interstitial matrix, present in the spaces between cells, and the basement membrane, adjacent to epithelial cells, that anchors and supports the

tissue. The molecular composition of the ECM determines its physical and physiological properties and tailors the matrix for different functions.

1.5.1 - Extracellular Matrix Components

The ECM is a dynamic structure. Its constant remodeling implies continuous breakdown by proteases and active re-building (Lu et al., 2011). It has three major components and their different combinations give specificity to matrix function.

- Proteoglycans (heparan-sulphate, chondroitin-sulphate and keratan-sulphate) are highly glycosylated proteins that play a key role in keeping the ECM and cells hydrated.
- Insoluble collagen fibres, which provide strength and resilience.
- Soluble multi-adhesive extracellular matrix proteins, for example fibronectin and laminin, that connect cells with components of the ECM and therefore allow the cell to move.

1.5.1.1 - Adhesion molecules

Cell adhesion to ECM components occurs according to two main mechanisms: by focal adhesions, that connect ECM to actin filaments, and hemidesmosomes, that connect ECM to intermediate filaments (Ozawa et al., 2010).

Specific cell adhesion molecules (CAMs) also known as integrins, mediate this cell-to-ECM adhesion (Aplin et al., 1998). Indeed CAMs are transmembrane proteins with an extracellular domain that binds to a ligand, a transmembrane domain, and an intracellular domain that interacts with cytoplasmic proteins (figure 1.11).

The main classes of CAMs include integrins, cadherins, selectins and the immunoglobulin family. Cd36, cd44 and discoidin domain receptors (DDR) belong to the class of integrins that bind to collagens. CAMs, such as cadherins, mediate homotypic interactions by binding to CAMs on the surface of other cells.

In addition to their adhesive functions, cell adhesion molecules modulate signal transduction pathways by interacting with molecules such as receptor tyrosine kinases, Rho-GTPase and components of the Wnt signalling pathway (Aplin et al., 1998) (figure 1.11).

Adhesive interactions are finely regulated quantitatively and temporally, for example in some situations such as cell migration, a quick assembly and disassembly is required. In others, adhesive contacts (such as in epidermis) need to be strong to resist mechanical stress (Hotchin et al., 1995).

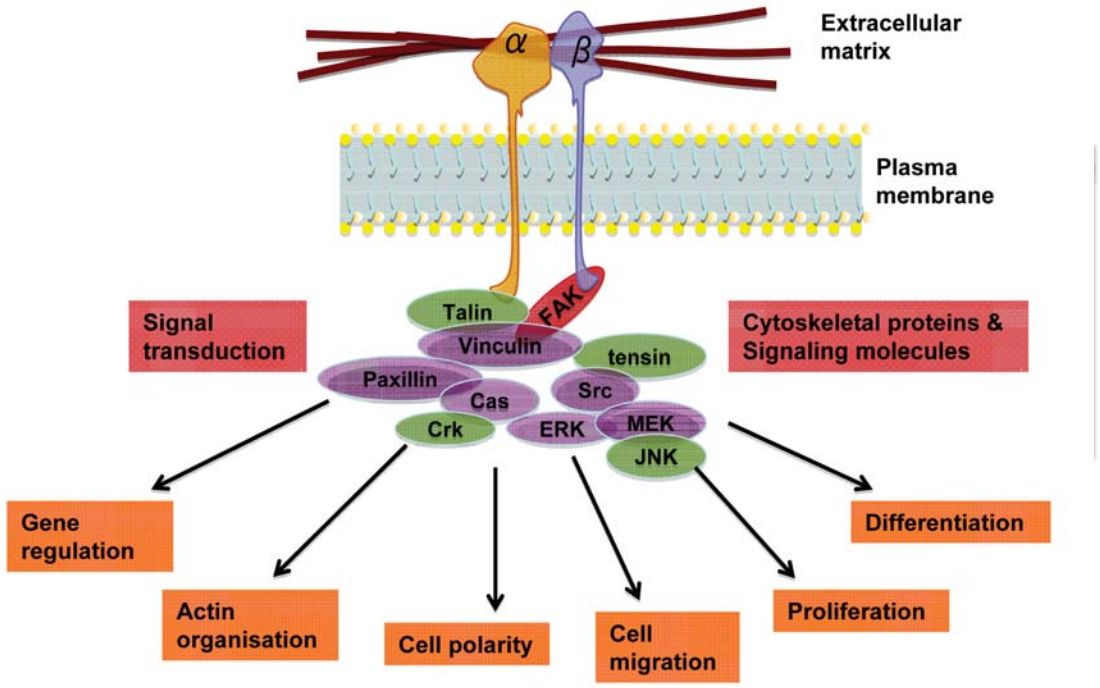


Figure 1.11 - Interactions between extracellular matrix proteins and integrins. Integrins mediate cell-to-ECM adhesion by interacting with multiple intracellular proteins that initiate signalling pathways regulating important cellular processes.

1.5.2 - Adherens Junctions

CAMs are essential not only for cell adhesion to ECM components but also for the physical cohesion between cells. This is a fundamental requirement for the formation and maintenance of tissues and organs. Various types of intercellular junctions exist between epithelial cells and each of those serve specific functions and are finely regulated during morphogenesis and tissue homeostasis. Adherens junctions and desmosomes are primarily involved in the initiation and stabilisation of cell-cell adhesion whereas tight junctions serve as physical and ion-selective diffusion barriers. Gap junctions form channels between cells and therefore allow intercellular communication (figure 1.12).

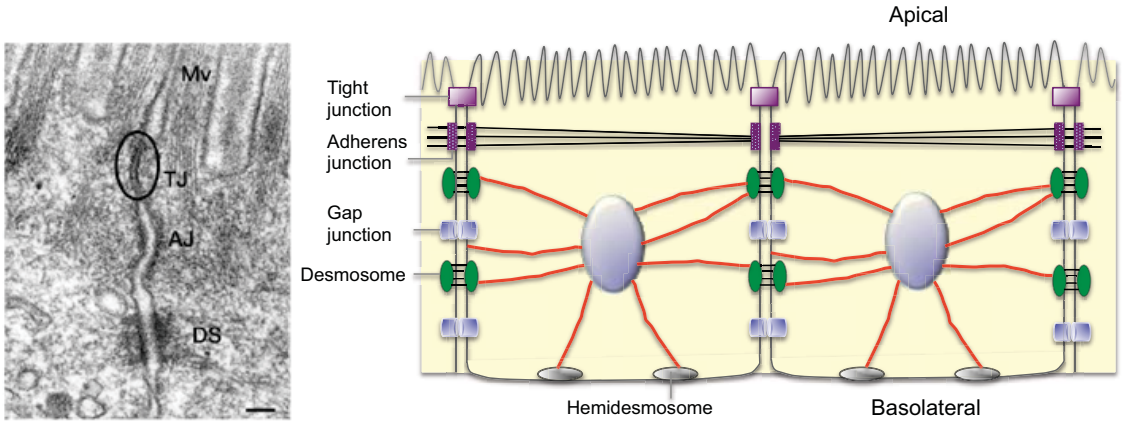


Figure 1.12 - Schematic drawing and electron micrograph of intestinal epithelial cells. The junctional complex, which is located at the most apical region of lateral membranes, is represented. The tight junction is circled in the electron micrograph of the junctional complex. (Mv, microvilli; TJ, tight junction; AJ, adherens junction; DS, desmosome.) Figure modified from (Tsukita et al., 2001).

Adherens junctions together with the apical localised tight junctions and the more basally found desmosomes form the Apical Junctional Complex (AJC) (figure 1.12). The AJC separates the apical from the basolateral domain of the cell.

Apart from providing adhesion between cells, adherens junctions have numerous functions including regulation of the actin cytoskeleton, intracellular signaling and transcriptional regulation (Hartsock and Nelson, 2008). Proteins that constitute adherens junctions consist of transmembrane receptors of the cadherin and nectin family that bind through their cytoplasmic domain to signaling molecules and actin binding proteins, such as vinculin, talin, tensin, and α-actinin, providing the transduction of the signal inside the cell and connection to the actin cytoskeleton (figure 1.11).

Classical cadherins, such as E-cadherin, are type I transmembrane proteins characterised by five Ca^{2+} binding extracellular motif, the so-called extracellular cadherin (EC) that interact with E-cadherin molecules from the neighboring cell (Pokutta et al., 1994).

E-cadherin is trafficked via the Rab11 positive recycling endosome (RE) in non-polarised and newly polarised cells. This process is important for E-cadherin trafficking in MDCK cells from a non-polarised to fully polarised state (Desclozeaux et al., 2008).

Based on phylogenetic evolution, cadherins have been classified in vertebrates into two major types (Oda and Takeichi, 2011). Type I cadherins are the major components of adherens junctions and include E-cadherin, expressed in most epithelial tissues and N-cadherin, expressed by neuroepithelial cells, neurons, and mesenchymal cells. The E- to N-cadherin switching is crucial during embryonic development (Wheelock et al., 2008), neuronogenesis, tissue remodeling and wound healing (Dady et al., 2012). This E- to N-cadherin switch occurs also in pathologic conditions such as in epithelial carcinomas with a strong impact on metastatic progression (Hao et al., 2012).

Type-II cadherins include cadherins-5, -6, -7, -8, -9, -10, -11, and -12 and are expressed by various cell types but little is known to date about the type II cadherins (Takeichi, 1991).

1.5.2.1 - Regulation of adherens junctions: EMT

Regulation of adherens junctions occurs on multiple levels. This includes alteration of their gene expression, post translational modifications, alterations in cytoskeletal dynamics, regulation of adherens junction trafficking as well as cleavage of their components by different proteases.

Epithelial to mesenchymal transition (EMT) is a biological process in which epithelial cells are trans-differentiated to a mesenchymal state. Mesenchymal properties include loss of polarity and loss of adherens junctions concomitant with cytoskeletal

rearrangements that enable cells to acquire the ability to invade and migrate (De Craene and Berx, 2013). Loss of E-cadherin, together with increased expression of vimentin, N-cadherin and metalloproteinases (MMPs) are key features of EMT. E-cadherin transcriptional repressors such as, for example Snail, are up-regulated during EMT resulting in reversible silencing of E-cadherin expression. TGF-β signalling has been shown to play an important role in EMT. In fact, the addition of TGF-β to epithelial cells in culture induces EMT (Xu et al., 2009). Published data show that the expression of Snail proteins is induced in response to TGF-β in cells that undergo TGF-β-induced EMT (Kokudo et al., 2008).

Recently published data demonstrate that native collagen IV induces down-regulation of E-cadherin expression and EMT (Espinosa Neira and Salazar, 2012).

1.5.3 - Collagens

Collagens are the most abundant proteins in the ECM and in the human body with 28 different types identified to date (Kadler et al., 2007). Thus, various types of collagen are the major structural elements of all connective tissues (Kadler et al., 1996). They provide tissue scaffold and framework and influence, for example, cell adhesion and migration, angiogenesis, tissue morphogenesis and repair.

Collagens can be classified according to their structure and molecular properties into many groups (table 1) including (Gelse et al., 2003; Kadler et al., 2007):

- fibril-forming collagens (such as types I, II, III and V) which represent the most abundant and widespread family of collagens
- fibril-associated collagens (FACIT) (for example types IX, XII and XIV)

- network-forming collagens (type IV)

- other groups - anchoring fibrils, transmembrane collagens, basement membrane collagens and others with unique functions

The tissue distribution of different collagens is summarised in Table 1. In particular type IV collagen is the major constituent of basement membranes and is crucial for its function.

Although the size, function and tissue distribution of collagens vary considerably, all members of this family are characterised by three α-chains each consisting of a regular arrangement of amino acids (Kuhn et al., 1985): Gly-X-Y. Position X may be any of various other amino acid residues, while either proline or hydroxyproline are found in position Y. 3-hydroxyproline is found at the X position in -X-4Hyp-Gly-sequences (O'Leary et al., 2011). The post-translational modifications that occur in proline and lysine residues are specific for collagen type and crucial for collagen stability and functions (see next sections).

Several proteins contain a collagenous domain, but are not classified as collagens. These include: macrophage scavenger receptors, adiponectin, acetylcholine esterase, the C1q subcomponent of complement, ectodysplasin A, emilins, surfactant proteins A and D, colmedins, and mannose binding protein (Franzke et al., 2005; Rautavuoma et al., 2004).

Type	Molecular composition	Genes (genomic localization)	Tissue distribution
Fibril-forming collagens			
I	$[\alpha1(I)]_2\alpha2(I)$	COL1A1 (17q21.31–q22) COL1A2 (7q22.1)	bone, dermis, tendon, ligaments, cornea
II	$[\alpha1(II)]_3$	COL2A1 (12q13.11–q13.2)	cartilage, vitreous body, nucleus pulposus
III	$[\alpha1(III)]_3$	COL3A1 (2q31)	skin, vessel wall, reticular fibres of most tissues (lungs, liver, spleen, etc.)
V	$\alpha1(V),\alpha2(V),\alpha3(V)$	COL5A1 (9q34.2–q34.3) COL5A2 (2q31) COL5A3 (19p13.2)	lung, cornea, bone, fetal membranes; together with type I collagen
XI	$\alpha1(XI)\alpha2(XI)\alpha3(XI)$	COL11A1 (1p21) COL11A2 (6p21.3) COL11A3 = COL2A1	cartilage, vitreous body
Basement membrane collagens			
IV	$[\alpha1(IV)]_2\alpha2(IV)$; $\alpha1-\alpha6$	COL4A1 (13q34) COL4A2 (13q34) COL4A3 (2q36–q37) COL4A4 (2q36–q37) COL4A5 (Xq22.3) COL4A6 (Xp22.3)	basement membranes
Microfibrillar collagen			
VI	$\alpha1(VI),\alpha2(VI),\alpha3(VI)$	COL6A1 (21q22.3) COL6A2 (21q22.3) COL6A3 (2q37)	widespread: dermis, cartilage, placenta, lungs, vessel wall, intervertebral disc
Anchoring fibrils			
VII	$[\alpha1(VII)]_3$	COL7A1 (3p21.3)	skin, dermal–epidermal junctions; oral mucosa, cervix,
Hexagonal network-forming collagens			
VIII	$[\alpha1(VIII)]_2\alpha2(VIII)$	COL8A1 (3q12–q13.1) COL8A2 (1p34.3–p32.3)	endothelial cells, Descemet's membrane
X	$[\alpha3(X)]_3$	COL10A1 (6q21–q22.3)	hypertrophic cartilage
FACIT collagens			
IX	$\alpha1(IX)\alpha2(IX)\alpha3(IX)$	COL9A1 (6q13) COL9A2 (1p33–p32.2)	cartilage, vitreous humor, cornea
XII	$[\alpha1(XII)]_3$	COL12A1 (6q12–q13)	perichondrium, ligaments, tendon
XIV	$[\alpha1(XIV)]_3$	COL9A1 (8q23)	dermis, tendon, vessel wall, placenta, lungs, liver
XIX	$[\alpha1(XIX)]_3$	COL19A1 (6q12–q14)	human rhabdomyosarcoma
XX	$[\alpha1(XX)]_3$		corneal epithelium, embryonic skin, sternal cartilage, tendon
XXI	$[\alpha1(XXI)]_3$	COL21A1 (6p12.3–11.2)	blood vessel wall
Transmembrane collagens			
XIII	$[\alpha1(XIII)]_3$	COL13A1 (10q22)	epidermis, hair follicle, endomysium, intestine, chondrocytes, lungs, liver
XVII	$[\alpha1(XVII)]_3$	COL17A1 (10q24.3)	dermal–epidermal junctions
Multiplexins			
XV	$[\alpha1(XV)]_3$	COL15A1 (9q21–q22)	fibroblasts, smooth muscle cells, kidney, pancreas,
XVI	$[\alpha1(XVI)]_3$	COL16A1 (1p34)	fibroblasts, amnion, keratinocytes
XVIII	$[\alpha1(XVIII)]_3$	COL18A1 (21q22.3)	lungs, liver

Table 1.1 - Collagen types with corresponding localisation in the genome and tissue distribution. Modified from (Gelse et al., 2003).

1.5.3.1 - Collagen type IV - the collagen of basement membranes

Type IV collagen is the major structural component of basement membrane (BM) that inter-connects the other components of the BM (laminins, proteoglycan and entactin/nidogens) into a stable reticulum aggregate (Khoshnoodi et al., 2008).

Type IV is now classified as a protein family of triple helical isoforms consisting of at least six genetically distinct chains: the classical al(IV) and a2(IV) chains and the later discovered a3(IV), a4(IV), a5(IV), and a6(IV) (Khoshnoodi et al., 2008).

Mutations of the major isoforms a1(IV) and a2(IV) are embryonically lethal in mouse (Content et al., 1980; Rautavuoma et al., 2004). The isoforms a3(IV)-a6(IV) show restricted, tissue-specific expression patterns. The recently identified chains have been shown to be directly involved in several collagen related diseases such as Goodpasture syndrome (Hudson et al., 1993), Leiomyomatosis (Hudson et al., 1993) and Alport syndrome (Hudson et al., 1993; Tryggvason, 1996). Alport syndrome, the most common form of hereditary glomerulonephropathy, is often caused by mutations in a3(IV), a4(IV), a5(IV), and a6(IV) genes that causes lamellation and splitting of the glomerulus basement membrane with consequent progressive loss of renal function (Hudson et al., 1993; Peissel et al., 1995). Apart from kidney disease, patients with Alport syndrome frequently develop sensorineural hearing loss, a common feature of ARC syndrome. Type IV collagen is an important component of inner ear structures and alterations in type IV collagen often result in abnormal inner ear function.

Like other type of collagens, the monomer constituent of the type IV collagen network is a triple helical molecule (figure 1.13). Three domains characterise each chain: the N-terminal 7S domain, a C-terminal non-collagenous domain (NC1), and the central triple helical part that consists of a long collagenous domain of 1400 residues of Gly-X-Y

repeats. This domain is interrupted by short non-collagenous sequences resulting in a flexible triple helix (Gelse et al., 2003; Hudson et al., 1993).

Non-collagenous domains of collagens IV, XV and XVIII have been shown to be involved in tumorigenesis and angiogenesis and they might also be important in many other biological processes (Ortega and Werb, 2002). Therefore, these domains are considered a target for potential therapeutic compounds.

Specific interactions between the domains are fundamental for the formation of a stable network characteristic of collagen IV. Each chain is highly hydroxylated and glycosylated. Hydroxylation is essential for the cross-linking of collagen molecules, thus ensuring the strength of collagen fibrils (Schegg et al., 2009). Although the role of glycosylation is not completely understood, it seems to be crucial for oligomerisation and proper secretion of collagen IV (Schegg et al., 2009).

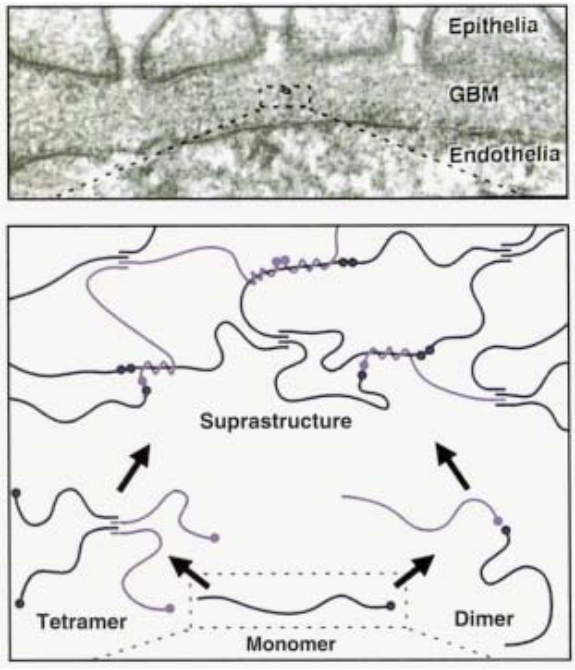


Figure 1.13 - Schematic representation of collagen IV structure.
The electron micrograph in the top panel shows that the basement membrane in the renal glomerulus (GBM) is located between the capillary endothelium and the epithelium of podocytes. The lower panel schematically represents the association of collagen helix in a reticulated suprastructure. The building block is a triple helical monomer that self-assemble in dimers by association of monomers at the C-termini, or tetramers when they associate at the N-termini. Figure modified from (Hudson et al., 1993).

1.5.3.2 - The Biosynthesis of Collagen

Transcription and mRNA processing

The biosynthesis of collagens begins with collagen gene transcription in the nucleus (figure 1. 14). There are 45 genes known encoding individual α-chains. Collagen molecules can be composed of the same or up to three different gene products forming homo- or hetero-trimeric molecules.

The transcription of collagens is finely regulated and depends on the cell type. Numerous growth factors can also control collagen transcription. For example in adult bones, collagen formation is stimulated by members of the TGF-β-superfamily as well

as insulin-like-growth factors (Chen et al., 2012). The potential to bind growth factors and cytokines makes collagen a suitable transport vehicle for therapeutic factor delivery. Most collagen genes undergo alternative splicing and possess multiple transcription initiation sites that generate different mRNA species.

Translation

Once the mRNA is mature it is transported to the cytoplasm (figure 1.14) and translated on ER membrane-bound ribosomes. The neo-translated protein is called pre-pro-collagen and it contains a signal sequence recognised by the corresponding receptors in the rough endoplasmic reticulum. This mediates the protrusion of pre-pro-collagen into the lumen of the ER (Gelse et al., 2003).

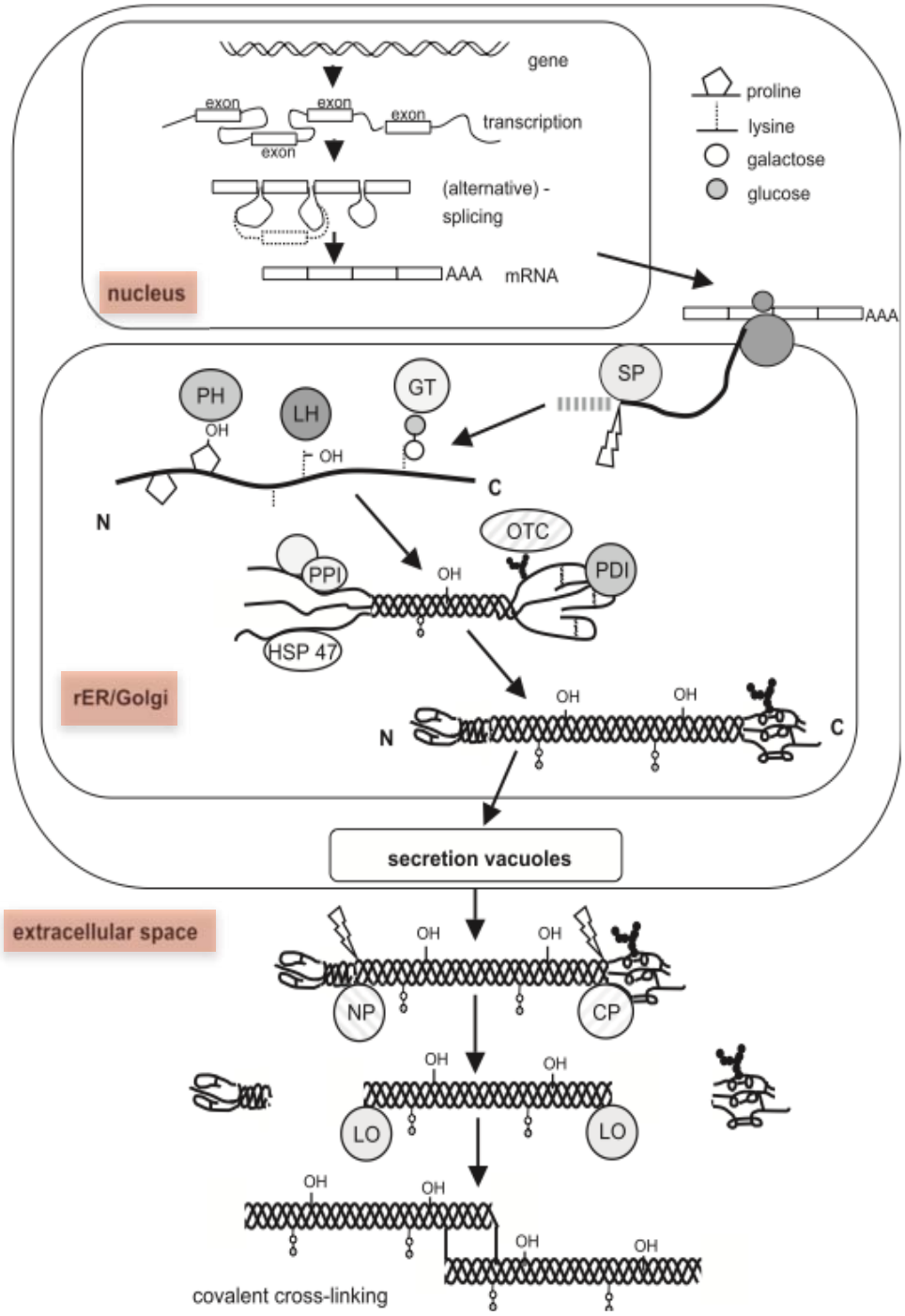


Figure 1.14 - Schematic representation of the biosynthesis of collagen: collagen gene transcription in the nucleus, mRNA processing, translation in the ribosomes and post-translational modifications in the ER/Golgi, secretion in the extracellular space and the final steps of fibril formation. (SP: signal peptidase; GT: hydroxylysyl galactosyltransferase; LH: lysyl hydroxylase; PH: prolyl hydroxylase; OTC: oligosaccharyl transferase complex; PDI: protein disulphide isomerase; PPI: peptidyl-prolyl cis-trans-isomerase; NP: procollagen N-proteinase; CP: procollagen C-proteinase; LO: lysyl oxidase; HSP47: heat shock protein 47, colligin1). Figure modified from (Gelse et al., 2003).

Posttranslational modifications and secretion into the extracellular space

Once in the ER the signal peptide is removed by a signal peptidase (figure 1.14) and the resultant molecule is called pro-peptide. Numerous and multi-step post-translational modifications occur at this stage. Hydroxylation of prolines is catalyzed by prolyl-3-hydroxylase and prolyl-4-hydroxylase. 4-hydroxyprolines are essential for intramolecular hydrogen bonds and therefore contribute to the stability and integrity of the triple helical domain (Tiainen et al., 2008). The function of 3-hydroxyproline is not known (Sipila et al., 2007).

Lysyl hydroxylases mediate the hydroxylation of lysine residues. The extent of lysine hydroxylation varies between tissues and collagen types (Kivirikko et al., 1973). Most lysines in type IV and VI collagens are hydroxylated and glycosylated (Mundy, 1996). Hydroxylysine residues are responsible for the formation of stable intermolecular crosslinking in collagen fibrils. Additionally they are sites for the attachment of glucosyl and galactosyl residues. This reaction is catalyzed by hydroxylysyl galactosyltransferase and hydroxylysyl glucosyltransferase enzymes respectively (figure 1.14) (see next section).

Before being transported to the Golgi, triple helices, now called procollagen, are formed in the ER. From the C-terminus of three-aligned α-chains, the triple helix is folded in a zipper-like manner progressing to the N-terminus. Molecular chaperones including PPI (peptidyl-prolyl cis-trans-isomerase) (Lang et al., 1987) and collagen-specific chaperones like HSP47 (Clarke et al., 1991) mediate the efficient folding of procollagen molecules (figure 1.14).

Procollagen is then transported in vesicles towards the Golgi apparatus. Other models have been proposed for collagen trafficking due to the small size of vesicles compared

to the large procollagen molecules (Stephens, 2012).

In the Golgi apparatus procollagen undergoes the last post-translational modifications before being secreted. Oligosaccharides are added (not monosaccharides as in the ER) before procollagen is packaged within the Golgi compartment into secretory vesicle and released into the extracellular space.

Fibril formation

Following their secretion, C- and N- propeptides are cleaved by specific proteases that belong to the family of Zn^{2+} dependent metalloproteinases (Prockop et al., 1998). The cleaved molecule is known as tropocollagen. Several feedback mechanisms have been proposed for the control of collagen expression mediated by the internalisation of released N- and C-propeptides but the potential mechanisms of regulation are mainly unresolved (Gelse et al., 2003). The solubility of procollagen decreases rapidly after the removal of the C-propeptide and it starts spontaneously aggregating into organised fibrillar structures (Silver et al., 2003). Extracellular enzymes such as lysil-oxidases catalyze the oxidative deamination of lysines and hydroxylysines producing aldehyde groups (Feres-Filho et al., 1995). This modification causes a spontaneous reaction of cross-linking between tropocollagen molecules. The resulting polymer is known as collagen fibril. The formation of cross-links is crucial for the physical and mechanical properties of collagen fibrils and its role in determining the strength of the collagen matrix. The knowledge of the structure, expression and function of the different types of collagen is important to better understand embryonic and fetal developmental processes as well as diseases based on defects of collagens.

1.5.4 - Lysyl hydroxylases: localisation and activities

Lysine residues in collagen-like peptides are enzymatically modified into hydroxylysine, galactosyl and glucosyl hydroxylysine by specific enzymes, namely lysyl hydroxylases (LHs), galactosyl and glucosyltransferase (GT and GGT respectively) (Myllyla et al., 2007) (figure 1.15).

LHs belong to the family of 2-oxoglutarate dioxygenases and therefore their genes are named *Plod* (procollagen lysine 2-oxoglutarate 5-dioxygenase). Like other collagen hydroxylases LHs use Fe^{2+}, 2-oxoglutarate, molecular oxygen and ascorbate to catalyse the hydroxylation of lysine residues. One atom from O_2 is incorporated into the hydroxyl group formed in the lysine residue, while the other atom reacts with 2-oxoglutarate to form succinate and carbon dioxide (Myllyla et al., 2007) (figure 1.15).

The resultant hydroxylysyl groups are attachment sites for collagen galactosyltransferases forming galactosylhydroxylysine residues in collagens that, as mentioned previously, are critical for its stability. Galactosylhydroxylysines can be further modified to glucosylgalactosylhydroxylysines by the addition of a glucosylic group and this reaction is catalyzed by collagen glucosyltransferases. The hydroxylysinegalactosylation and hydroxylysineglucosylation enzymes use UDP-galactose or UDP-glucose respectively in the reaction where galactose or glucose is transferred to a hydroxylysine/galactosyl-hydroxylysine substrate (Myllyla et al., 2007) (figure 1.15).

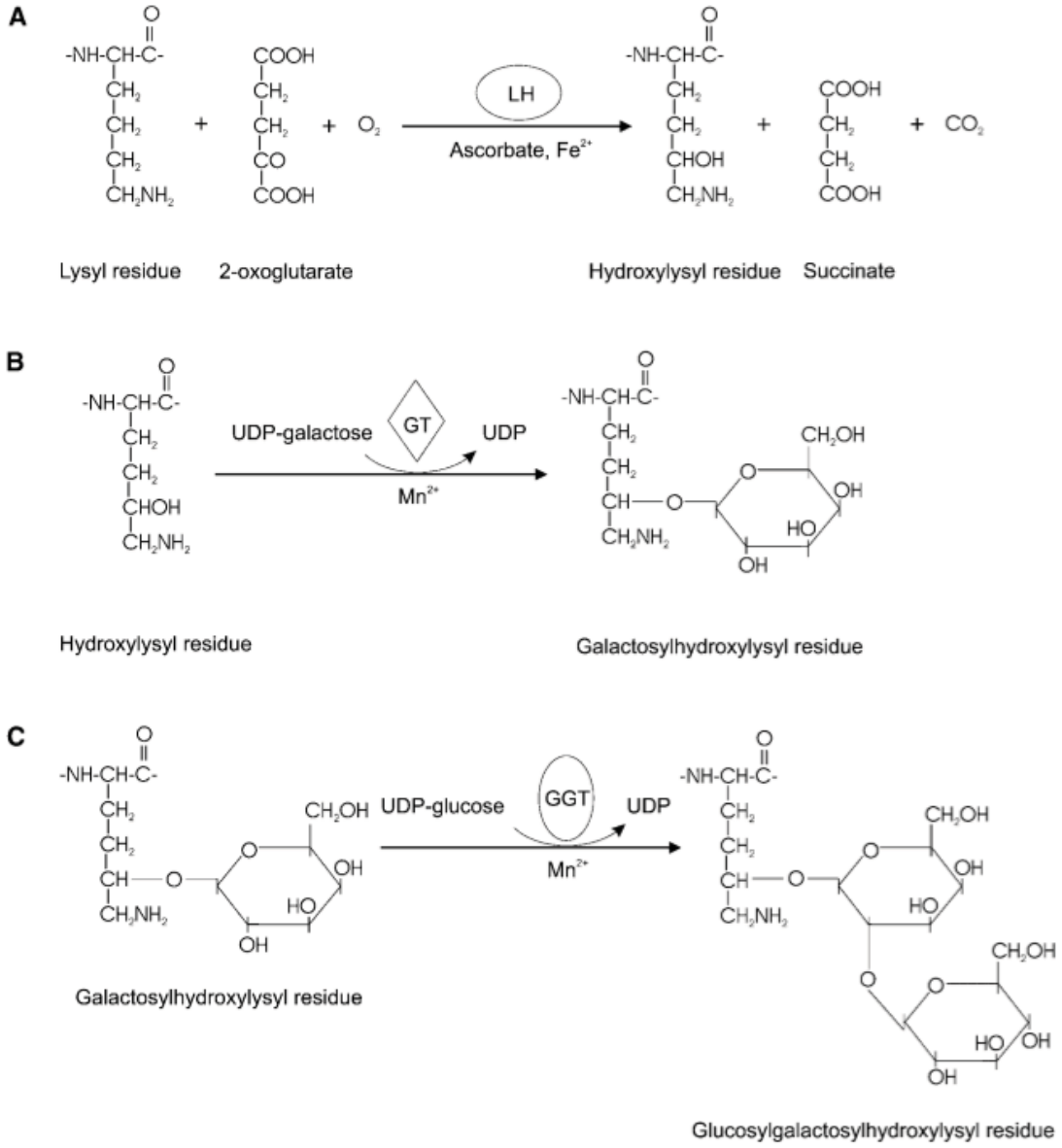


*Figure 1.15 - Enzymatic Reaction catalyzed by **(A)** lysyl hydroxylases (LHs) that use Fe^{2+}, oxoglutarate, molecular oxygen and ascorbate to catalyse the hydroxylation of lysine residues **(B)** collagen galactosyltransferases (GT) that use UDP-galactose and Mn^{2+} to transfer galactose to hydroxylysines and **(C)** collagen glucosyltransferases (GGT) that use UDP-galactose and Mn^{2+} to transfer glucose to galactosylhydroxylysine. Modified from (Myllyla et al., 2007).*

The first lysyl hydroxylase enzyme was initially purified from chicken embryos (Myllyla et al., 1991) and subsequently three lysyl hydroxylase isoforms LH1, LH2 and LH3, also named PLOD1 PLOD2 and PLOD3, have been characterised in vertebrates. The *D. melanogaster* and *C. elegans* have only one lysyl hydroxylase homologue each (Myllyharju and Kivirikko, 2004). The amino acid sequence homology between isoforms of the same species is around 60% in both mouse and human (Salo et al., 2006b). Surprisingly, the most recently identified, but evolutionarily the oldest of the PLOD isoforms, PLOD3, also possesses collagen glucosyltransferase and galactosyltransferase activity, in addition to the lysyltransferase, whereas PLOD1 and PLOD2 can only catalyze the formation of hydroxylysines (Salo et al., 2006b). The N-terminal part of PLOD3 has been shown to be responsible for the glycosyltransferase activity, whereas the lysyl hydroxylase activity is located at the C-terminus of the enzyme (Feres-Filho et al., 1995).

As mentioned previously, collagen glycosylation is essential for basement membrane formation. Data from the literature show that PLOD3 is crucial for embryonic development and basement membrane formation, possibly because of its unique property among PLOD enzymes of possessing collagen glucosyltransferase activity (Rautavuoma et al., 2004). In cancordance with the above statement Ruotsalainen et al showed that the amount of PLOD3 is higher in embryonic than in adult tissues (Salo et al., 2006a). The importance of PLOD3 activity for collagen type IV modification is supported by the finding that collagen IV isolated from Plod3 knockout mouse migrates faster on SDS-PAGE compared to controls, suggesting a lack of glycosylated hydroxylysines (Rautavuoma et al., 2004).

PLOD3 has also been shown to possess a critical role in the neuronal growth cone migration determined from its glucosyltransferase activity on myotomal type XVIII collagen, a ligand for neural-receptor protein tyrosine phosphatases that guide motor axons (Schneider and Granato, 2006).

The expression levels of lysyl hydroxylase isoenzymes differ between tissues and organisms (Myllyla et al., 2007). A summary of the expression pattern in selected cell lines in human and mouse tissues is shown in Table 2.

	mRNA levels		
Human Tissues	**PLOD1**	**PLOD2**	**PLOD3**
Heart	++	++	++
Brain	++	+	+
Placenta	++	++	++
Lung	++	-	+
Liver	++	++	+
Skeletal muscle	++	++	+
Kidney	++	+	+
Pancreas	++	++	++
Mouse Tissues			
Heart	++	++	++
Brain	-	-	+
Spleen	-	-	-
Lung	++	++	++
Liver	++	-	++
Skeletal muscle	++	+	+
Kidney	++	++	+
Testis	-	+	++

Table 1.2 - Expression of lysyl hydroxylase isoforms in adult human and mouse tissues. - no, + weak, ++ intermediate. The expression levels are comparable only within the same isoform. Modified from (Wang et al., 2000).

Subcellular localisation

PLOD isoforms are targeted to the rough endoplasmic reticulum via a signal peptide (of 18-25 amino acids) even though they lack the typical ER-retention signals such as KDEL or double lysine motif (Wang et al., 2012). PLOD1 was shown to be a peripheral protein in the ER, weakly bound to the luminal side of the membrane by weak electrostatic interactions with unknown proteins (Kellokumpu et al., 1994).

Interestingly, it has been shown that in addition to the ER, active PLOD3 is also found in the extracellular space, in serum and on the surface of cultured cells (Myllyla et al., 2007; Salo et al., 2006b). Furthermore the distribution of PLOD3 is tissue specific. Salo *et al* showed that PLOD3 was localised both intracellularly and extracellularly in mouse muscle, spleen, and kidney, whereas in mouse liver and pancreas was found only in the extracellular space (Salo et al., 2006b). The ability of PLOD3 to modify extracellular proteins is required for cell growth and viability and is thought to be important for matrix remodeling (Salo et al., 2006b).

The molecular weight of PLOD3 on the cell surface differs from the PLOD3 fraction found in the cell culture medium possibly because the two proteins are differently glycosylated. This difference suggests that PLOD3 may be secreted through two different pathways. PLOD3 secreted into the cell culture medium might pass through the Golgi complex for it to be further glycosylated, while PLOD3 that is trafficked and remains on the cell membrane might bypass the Golgi (Wang et al., 2012).

The mechanisms involved in PLOD3 trafficking remain unknown and more studies are required.

1.5.4.1 - Lysyl hydroxylase 3 deficiency

In general, lack of LH activity leads to changes in collagen structure with the consequential formation of highly fragile tissues. Mutations in *PLOD1* and *PLOD2* have been associated with the kyphoscoliotic subtype of Ehlers-Danlos (Giunta et al., 2009) and Bruck syndrome (Hyry et al., 2009) respectively.

Only one patient with mutations in *PLOD3* has been identified to date (Hyry et al., 2009). The two identified mutations were heterozygous with recessive transmission and were located in conserved regions of the PLOD3 protein sequence: one nucleotide transition c.668A→G (mutation 1) resulting in the amino acid substitution, p.Asn223Ser, and one nucleotide deletion c.2071 delT (mutation 2) causing a translational frameshift and generating a premature translational stop codon (Salo et al., 2008). The first mutation interfered with the glucosyltransferase activity of PLOD3 and the other with the lysyl hydrolase activity. Furthermore, the overall levels of PLOD3 protein were reduced (Salo et al., 2008). This resulted in a unique phenotype mostly characterised by several collagen related abnormalities that included: joint contracture, congenital scoliosis, scleral fragility, right-sided diaphragmatic eventration, knee joints appearing prominent, skin, hair, and nail abnormalities (Salo et al., 2008). The patient also showed profound bilateral sensorineural deafness, reduced visual acuity with myopia, spontaneous vascular rupture and global developmental delay. The common features between the PLOD3 deficient patient and individuals with ARC include: joint contractures, osteopenia, sensorineural deafness, growth retardation.

In mouse, PLOD3 null embryos were shown to be lethal at embryonic day 9.5 because of the severe deficiency in the synthesis of type IV collagen that lead to the fragmentation of BMs.

CHAPTER TWO:

MATERIALS AND METHODS

2.1 - Suppliers

All chemicals and tissue culture reagents were purchased from Sigma-Aldrich unless otherwise stated.

Tissue culture plastics were purchased from Corning. Lipofectamine 2000 transfection reagent was ordered from Invitrogen.

IMAGE clones were ordered from Geneservice (Cambridge, UK) and purified PCR primers were obtained from Sigma. The proofreading Platinum Taq Polymerase kit was ordered from Invitrogen. The T4 DNA ligase and buffer were purchased from Promega. Restriction endonucleases were purchased from New England Biolabs. Plasmid mini and maxi prep kits were purchased from QIAGEN.

Protein G coupled Dynabeads were ordered from Invitrogen.

Suppliers for antibodies are listed in table 2. Anti-VPS33B antibody was a gift from Guang-dan Zhu from Emory University, Atlanta**.**

2.2 - Cloning system

PCR inserts for all constructs used in this thesis were first amplified using a proofreading DNA polymerase with primers containing restriction sites on the 5' (forward primer) and 3' end (reverse primer) for ligation into the expression vector. Primer sequences used for cloning of plasmid constructs used in this thesis are shown in table 2.1.

Cloning primers	Direction	Sequence
mCherry-Plod3-F	Forward	5′-AAAGAATTCTGATGACCTCCTCGGGGCCTGGACC-3′
mCherry-Plod3-R	Reverse	5′-TTTGGATCCTTGGGGTCGACAAAGGACACCATGAT-3′
mCherry-Plod1-F	Forward	5′-AAAGAATTCTGATGCGGCCCCTGCTGCTAC-3′
mCherry-Plod1-R	Reverse	5′-TTTGGATCCTTGGGATCGACGAAGGAGACTGCGAT-3′
Vipar-(N-)-F	Forward	5′-AAGAATTCGTGGATCCATGAATCGGACAAAGGGTGATGAGGAGG-3′
Vipar-(N-)-R	Reverse	5′- AAGGTACCTTAAGCGGCCGCCACTGTATCACTGGGGCTCCAGTCATT-3′
VIPAR-(C+)-F	Forward	5′- AAGAATTCGTGGATCCGATACAGTGCGACGTCTCCGGAAG-3′
VIPAR-(C+)-R	Reverse	5′-AAGGTACCTTAAGCGGCCGCATTCTTCCATCGAATTTGCGAGCTGC-3′
VIPAR-(C-)-F	Forward	5′- AAGAATTCGTGGATCCCTAGAGAGATTCCGCTCCTTACAGGAC-3′
VIPAR-(C-)-R	Reverse	5′- AAGGTACCTTAAGCGGCCGCATTCTTCCATCGAATTTGCGAGCTGC-3′

Table 2.1 - List of primes used for cloning of different constructs

2.2.1 - Proofreading Polymerase Chain Reaction

PCR reactions were performed on a GeneAmp 9700 thermocycler using Platinum® Taq DNA Polymerase High Fidelity (Invitrogen) with proofreading activity. The reaction was set up as follows:

Component	Volume (µL)	Final Concentration
10X High Fidelity PCR Buffer	5	1X
10 mM dNTP mixture	1	0.2 mM each
50 mM MgSO	2	2 mM
Forward Primer (10 µM)	1	0.2 µM
Reverse Primer (10 µM)	1	0.2 µM
Template DNA (35 ng/ul)	1	0.7 ng/ul
Platinum ® *Taq* High Fidelity	0.2	1.0 unit
Nuclease Free Water	**to 50**	Not applicable

Table 2.2 - Standard PCR reaction used for cloning of plasmid constructs

The reaction was then subjected to the amplification conditions shown below:

Initial denaturation: 94 ºC for 2 minutes
25–35 cycles of:
Denature: 94 ºC for 30 seconds
Anneal: 60 °C for 30 seconds
Extend: 72 ºC for 1 minute per kb of PCR product
Final extension: 72 ºC for 10 minute

2.2.2 - Agarose Gel Electrophoresis

10 µL of the total PCR reaction volume (50 µL) was used to check the success and the specificity of each reaction. The remaining 40 µL was PCR-purified before restriction digestions were set up.

A 1 % agarose gel was prepared by dissolving 1 % w/v agarose in 1x TBE buffer (90 mM tris-base, 90 mM boric acid, and 2 mM EDTA, pH=8.0) and heating, which was then supplemented with and 10 ng/mL of ethidium bromide (EtBr).

6x DNA loading buffer (2 mM EDTA, 50 % glycerol and 0.1 % orange G) was added to the PCR product before loading on the agarose gel. A DNA ladder was loaded on the same gel as a size standard. The electrophoresis was run at 110 Volts and the progress was monitored by the migration of orange G dye.

2.2.3 - PCR purification

The PCR products were purified using the QIAquick PCR Purification Kit.

5 volumes of Buffer PB were added to 1 volume of PCR sample and the mix was placed in a QIAquick spin column before centrifuging for 1 min at 13,000 rpm. The flow-through was discarded and the column was washed with 750 µL of Buffer PE and the

centrifugation step repeated as above. After the flow-through was discarded, the column was centrifuged for an additional minute to remove residual ethanol. 50 µL of nuclease free water were added to the centre of the column, which was then centrifuged for 1 min at 13,000 rpm. The concentration of the eluted DNA was measured by the NanoDrop ND-1000 Spectrophotometer.

2.2.4 - Restriction Digestion

The purified products and appropriate vectors were separately digested with the appropriate restriction endonucleases reported in the table below.

Construct	5' Enzyme	3' Enzyme	Vector	Buffer
mCherry-PLOD1	EcoRI-HF	BamHI	pmCherry-N1	4 + BSA
mCherry-PLOD3	EcoRI-HF	BamHI	pmCherry-N1	4 + BSA
VIPAR N-	EcoRI-HF	KpnI-HF	pCMV-Myc	4 + BSA
VIPAR C+	EcoRI-HF	KpnI-HF	pCMV-Myc	4 + BSA
VIPAR C-	EcoRI-HF	KpnI-HF	pCMV-Myc	4 + BSA

Table 2.3 - List of restriction endonucleases and buffers used during the cloning of specific constructs in the respective vectors

The reaction was set up using 1 µg of both vector and insert, 10 U of each restriction endonuclease (New England Biolabs), 10x specific buffer and 100x BSA where applicable, in a total reaction volume of 50 µL. Restriction digestions were carried out at 37°C for 4 hours and subsequently run into a 1 % agarose gel as described in section 2.2.2.

2.2.5 - Agarose Gel Extraction

Once visualised in the agarose gel, the bands corresponding to the amplified gene size were gel extracted using the QIAGEN gel extraction kit. The correct band was excised from the gel with a clean scalpel. 3 volumes of Buffer QG were added to 1 volume of the gel slice (1 volume ~ 1 weight) and then placed in a heating block at 55 °C until the gel had completely melted. Subsequently, one volume of 100 % isopropanol was added and the resulting sample was applied to a QIAquick spin column and centrifuged for 1 min at 13,000 rpm. The following wash and elution steps are common to the PCR purification protocol (section 2.2.3).

2.2.6 – Ligation

The gel extracted DNA insert and vector have the same cohesive ends as a result of the restriction digestion. To ligate the insert into the expression vector the ligation reaction was set up as below, in a reaction volume of 5 µL and incubated either 3 hours at 16 °C or overnight at room temperature.

Component	**Volume (µL)**
10X Ligation Buffer	0.5
T4 DNA Ligase (3 U/ µL)	0.35
Insert DNA	1
Vector DNA	0.5
Nuclease Free Water	**to 5**

Table 2.4 - Ligation reaction

2.2.7 - Transformation

1 µL of the above ligation products was added to 20 µL of DH5α *Escherichia coli* competent cells and incubated on ice for 30 min. Cells were then subjected to heat shock at 42 °C for 45 seconds and immediately placed on ice for 2 minutes. Cells were then allowed to grow by adding 500 µL of SOC medium (Invitrogen) and incubating in an orbital shaker at 37 °C for an hour at 220 rpm. Subsequently cells were centrifuged at 5000 rpm for 10 minutes and 360 µL of the supernatant was removed, while the remaining 200 µL was resuspended and spread onto LB agar plates supplemented with the appropriate antibiotic reported in the table 2.5. Plates were incubated overnight at 37 °C.

Antibiotic	**Final Concentration**	**Vector**
Ampicillin	100 µg/mL	pCMV-Myc
Kanamycin	25 µg/mL	pmCherry-N1

Table 2.5 - List of cloning specific antibiotics used in this thesis

2.2.8 - Colony Screening

Bacterial colonies were screened for insert incorporation into the plasmid by Colony PCR. Single colonies (~5 per plate) were resuspended in 10 µL nuclease free water and 1 µL of the mixture was used as a template in the PCR reaction. 5 minutes of the initial denaturation was set up to allow bacterial lysis. The forward and reverse primers were complementary to the vector and insert sequence respectively to ensure the selection of plasmids containing the insert. One of the colonies corresponding to a PCR positive reaction was transferred to 5 mL of Luria Broth (LB) medium supplemented with the

same antibiotic concentration used for plates. The small culture was allowed to grow overnight in a shaking incubator (37 °C, 220 rpm).

2.2.9 - Plasmid DNA Purification, Miniprep

Plasmids were extracted from overnight cultures using the QIAGEN mini-prep kit. 5 mL of bacterial cultures were pelleted by centrifugation at 5,500 rpm for 15 minutes. The resulting pellet was resuspended in 250 µL of Buffer P1 and transferred to a 1.5 mL microcentrifuge tube before addition of 250 µL of Buffer P2 to allow cell lysis. The tube was carefully inverted several times until a homogeneous blue suspension was achieved. Subsequently 350 µL of the alkali Buffer N3 was added, and the solution was immediately mixed by inverting the tube until the suspension turned colourless. The mixture was then centrifuged at 13,000 rpm for 10 minutes to pellet cellular debris, including genomic DNA. The supernatant, which contained the plasmid DNA, was transferred to a QIAprep spin column and centrifuged for 1 minute at 13,000 rpm. The column, with bound plasmid DNA, was washed with 750 µL of Buffer PE and centrifuged as above. The flow through was discarded and the column was centrifuged for a further 1 minute at 13,000 rpm to wash away any residual buffer. The QIAprep column was then placed into a clean 1.5 mL tube and 50 µL of nuclease free water was added to the centre of the column and left to stand for 1 minute. The column was then centrifuged at 13,000 rpm for 1 minute to elute the plasmid DNA.

2.2.10 - Plasmid Sequencing

The plasmid DNA was subsequently sequenced to verify sequence accuracy using specific primers mapping in each vector. The dideoxy termination method was used to determine the exact sequence of the insert and to verify that it was in-frame. Therefore, it was essential to use primers (in the poly-linker region) that would demonstrate the sequence of the insert/plasmid interface.

The DNA was sequenced bi-directionally and the forward and reverse primers were kept in separate reactions as shown below.

Reagent	Volume per reaction (µL)
5x Sequencing Buffer	2.5
BigDye	0.5
Primer (2 µM)	1.6
Plasmid (0.5 µg/µL)	1
Nuclease Free Water	to 10

Table 2.6 - Sequencing reaction

The reaction was then subjected to the following amplification conditions:

Initial denaturation: 96 ºC for 30 seconds
30 cycles of:
Denature: 96 ºC for 30 seconds
Anneal: 50 °C for 15 seconds
Extend: 60 ºC for 4 minutes

2.2.10.1 - Sequencing Reaction: precipitation and analysis

The ethanol precipitation method was used to purify sequencing products from the rest of the reaction mix (enzymes, ddNTPS). 1 µL of 0.25 M EDTA and 32 µL of 100 % ethanol were added to the PCR reaction and incubated at room temperature in the dark for 10 minutes. The precipitated DNA was pelleted by centrifugation at 4000 rpm for 30 minutes at 4 °C. The supernatant was then carefully removed by centrifuging the plate

upside down at 500 rpm for 30 seconds. To wash the DNA pellet 125 µL of 75 % ethanol were added to each reaction and the centrifugation step was repeated as above. The supernatant was again removed by centrifuging the plate upside down and the DNA pellet was allowed to air dry for 10 minutes. 10 µL of HiDi formamide were then added to each sample before incubating at 95 °C for 5 minutes to denature the DNA. The plate was placed on ice for 2 minutes and the DNA was sequenced using the ABI Prism 3700. Sequences were analysed with BioEdit Sequence Alignment Editor.

2.3 - Protein harvesting and quantification

2.3.1 - Mammalian protein extraction

For knockdown assessment, mIMCD-3 clones were grown to confluence in 25 cm2 tissue culture flasks, washed with 10mL ice-cold PBS and scraped into 150 µL of RIPA lysis buffer (50mM Tris/HCl pH7.5, 1mM EDTA, 150mM NaCl, 0.1% sodium dodecyl sulfate (SDS), 0.5% deoxycholic acid, 1% Igepal CA-360 (Mammalian Cell Lysis Kit, Sigma), and protease inhibitor cocktail (Roche)) and placed on ice for 30 minutes. The suspension was then centrifuged at 13,000 rpm for 15 minutes at 4°C and the supernatant transferred to a clean tube.

For co-immunoprecipitation experiments HEK293 cells grown on 6-well plates were lysed as in section 3.3.2.

For cell-fractionation experiments HEK293, mIMCD-3 and HeLa cells were processed as in in sections 5.3.1 and 5.3.2.

2.3.2 - Protein detection and quantification

Protein concentration was determined either by using the NanoDrop 2000c UV-Vis for micro-volume protein quantitation or by the method of Bradford.

2.3.2.1 - Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins according to their electrophoretic mobility, which is dependent on the molecular weight of proteins.

The separation of the proteins was performed in polyacrylamide gels that consist of "stacking" and "resolving" gels. The reagent concentration of the resolving gel depends on the molecular weight of the proteins to be visualised. A 10 % resolving gel was used to visualise all proteins detected in this study. For detection of collagen IV by Western-blot a 4% resolving gel was used. Both resolving gels and stacking were prepared from the following components, volumes stated for one gel solution set up in a 1 mm thick glass plate (BioRad).

Reagent	**Volume (mL)**		
	Resolving gel (10 %)	**Resolving gel (4 %)**	**Stacking gel**
30 % acrylamide Protogel (Geneflow) (mL)	3.16	0.66	0.66
1.5/0.5 M Tris-HCl (pH=8.8/6.8)* (mL)	2.5	1.25	1.26
10 % Ammonium Persulfate (mL)	0.1	0.05	0.08
10 % SDS (mL)	0.1	0.05	0.06
TEMED (mL)	0.06	0.004	0.01
Nuclease-free water (mL)	**to 5**	**to 5**	**to 5**

*1.5 M Tris-HCl (pH=8.8) was used for the resolving gel and 0.5 M Tris-HCl (pH=6.8) for the stacking.

Table 2.7 - Tris-glycine SDS-Polyacrylamide gel preparation

The resolving gel was poured to within 2 cm of the top of the glass plates and overlaid with isopropanol. The resolving gel was left to polymerise before the isopropanol was removed and the stacking gel poured on the top. Combs were immediately added to the stacking gel to create sample wells. Once the stacking gel polymerised the combs were removed and the gel was placed into an electrophoresis chamber (BIORAD Mini). Before loaded onto the gel, protein samples were diluted 5-folds with 5x Lamelli loading buffer (65 mM Tris/HCl pH 8, 10 % (v/v) glycerol, 2.3 % (w/v) SDS, 0.01 bromophenol blue, 1 % dithiothreitol (DTT)) and denatured for 3 minutes at 100°C. The gel was run in Tris-glycine buffer (192 mM Glycine, 250 mM Tris-Base and 0.1 % SDS) for ~1 hour (till the loading dye had reached the bottom of the gel) at 120 V. A protein ladder (Page Ruler; Fermentas, York, UK) was loaded on the same gel to evaluate the correct size of the protein bands.

2.3.2.2 - Page Blue protein staining and Drying of SDS-PAGE Gels

SDS-PAGE gels were stained using PageBlue (Fermentas). The gel was first washed three times with 100 mL distilled water by microwaving each time on high power for 1 min. This step is required to remove the SDS from the gel as it interferes with the staining. Subsequently, the gel was stained for 60 minutes (or overnight) with 20 mL of PageBlue, under gentle agitation. Destaining was performed by washing the gel several times with distilled water until the protein bands were clearly visualised on the gel.
The gel was vacuum dried at 80 °C for 60 minutes on a Gel Master Gel Dryer Vacuum System (Welch).

2.3.3 - Western Blotting

2.3.3.1 - Electrophoretic Transfer to PVDF Membranes

To complement the gel staining, the resolved proteins were transferred from the SDS-PAGE gel (section 2.3.2.1) to a PVDF (polyvinylidene difluoride) (GE Healthcare) membrane for detection of proteins by Western blot. The PVDF membrane was first activated in methanol before the blotting sandwich was prepared in 1 x Transfer Buffer (192 mM Glycine, 250 mM Tris-Base and 20 % methanol) in the following order: fibre pad, filter paper, PVDF membrane, gel, filter paper, fibre pad.

The sandwich was then correctly placed into a transfer device in order to allow the electrotransfer of proteins from the gel to the PVDF membrane. The transfer was performed in 1x Transfer Buffer at 250 mA for 1.5 hours, using an ice pack to cool down the apparatus.

2.3.3.2 - Western Blot detection

During the detection process the membrane was first blocked in 10 % dried non-fat milk powder in PBST, on a roller, for 1 hour at room temperature (or alternatively overnight at 4 °C). The membrane was then incubated with the primary antibody diluted in 5 % milk solution for one hour at room temperature. Specific dilutions for all primary antibodies used for Western blotting in this thesis are reported in table 2.8. The membrane was then washed three times on a shaker (~200 rpm), for 10 minutes each time. After the wash, the membrane was blocked with properly diluted secondary

antibody (table 2.8) in 5 % milk solution and incubated at room temperature for one hour.

Antibody	Clone / Catalogue nr	Supplier	Primary Dilution	Secondary Dilution
m α myc	9E10	Sigma	1/5000	1/20000
m α HA	HA-7	Sigma	1/1000	1/2000
m α His	RL90	Sigma	1/3000	1/5000
Rb α VPS33B	-	Gift *	1/500	1/1000
Rb α VIPAR	HPA003589	Sigma	1/500	1/1000
Rb α PLOD3	11027-1-AP	proteintech	1/500	1/500
m α PDI	RL90	abcam	1/1000	1/2000
m α LAMP1	CD107a	Hybridoma Bank	1/5000	1/10000
Rb α Collagen I	NB600-408	Novus Biologicas	1/1000	1:10000
Rb α Collagen IV	ab6586	abcam	1/1000	1/10000
m α E-cadherin	4A2C7	Invitrogen	1/2000	1/2000
m α β-actin	AC15	Sigma	1/15000	1/20000

* gift from Guang-dan Zhu from Emory University, Atlanta

Table 2.8 - Antibodies and their dilutions used in this thesis for Western blotting. Other information provided includes the species the antibody was raised in (m- mouse, Rb- rabbit), the company along with the clone number for monoclonal antibodies or catalogue number for polyclonal antibodies.

Membrane washes were repeated as above. The secondary antibodies were all conjugated to Horse Radish Peroxidase (HRP) which allows detection once the membrane is incubated for 2 minutes with the enhanced chemo-luminescence plus (ECL+) western blotting detection system reagents (GE Healthcare) at a ratio of 1:40 for regent A:B. Exposure of the autoradiography film (Kodak) for the length of time that was a function of the signal strength and its subsequent development in an X-ray film developer (Konica Minolta) allowed the detection of protein bands.

2.4 - Immunoprecipitation

For immunoprecipitation experiments Protein G Dynabeads (Invitrogen) were first conjugated to the antibody according to the following protocol. 20 µL per reaction of well resuspended Dynabeads were washed twice with 500 µL of Citrate Phosphate Buffer (24.5 mM Citric Acid and 51.7 mM Dibasic Sodium Phosphate, pH=5.0). During all steps of the immunoprecipitation experiment, a magnet block was used to collect the beads. The solution was than pipetted out after 1 minute with the tube remaining on the magnet. 8 µg per reaction of the specific antibody was diluted in 20 µL of PBS, added to the dry beads and incubated for 1 hr at room temperature with rotation.

After removal of the supernatant, beads were washed three times with 500 µL of Citrate Phosphate Buffer followed by three washes with 1 mL of 0.2 M Triethanolamine.

The beads were then incubated with rotation for 30 min with 20 mM of fresh prepared Dimethyl Pimelimidate in order to covalently cross-link the antibody to the immobilised beads. The supernatant was removed and the reaction was quenched by incubating the beads with 1 mL of 50 mM Tris (pH=7.5) for 15 min with rotation at room temperature.

The antibody-conjugated beads were washed three times with 500 µL PBST followed by three washes with lysis buffer. Beads were then re-suspended in the 20 µL of lysis buffer per reaction and were ready to use for the immunoprecipitation experiment.

HEK293 cells were grown on 6 well plates and once they reached 80% confluence were transfected with a total of 4 µg of plasmid DNA per well. 6 hrs after transfection the medium was removed and replaced with normal medium where cells were allowed to recover for 42 hours prior to immunoprecipitation.

Cells were lysed as described in section 3.3.2 and after clarification of the lysates by centrifugation at 13,000 rpm for 15 min, the supernatant was added to 20 µL specific antibody conjugated to beads and incubated by rotation overnight at 4 °C.

Supernatants (flow-through) were collected and transferred to a clean tube. The beads were washed three times with lysis buffer and then resuspended in 20 µL of lysis buffer per reaction before the SDS-PAGE sample buffer was added and samples boiled at 100°C for 5 min. Beads were removed by centrifugation and samples were analysed by SDS-PAGE and Western blotting.

2.5 - Cell biology

2.5.1 - Cell lines culture conditions and maintenance

Human Embryonic Kidney (HEK) 293 and HeLa cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% complete fetal bovine serum (FBS), 2 mM L-glutamine, 1x MEM non-essential amino acid solution and 1x penicillin-streptomycin.

Murine inner medullary collecting duct, mIMCD-3 cells (CRL2123; ATCC, Manassas, VA) were grown in a 1:1 mix of DMEM and Ham's F12 medium, supplemented with 10% FBS and penicillin-streptomycin. All cells were grown in incubators at 37 °C in 5 % CO_2.

All cell cultures were passaged when they reached ~ 90% confluency. Cells were washed twice in sterile PBS and the detached from flasks using 2 mL of 1 x trypsin/EDTA per 75 cm^2 culture flask for 5 minutes at 37 °C. For mIMCD-3 cell lines

0.25% trypsin/EDTA was used. New flasks were prepared by diluting 1:10 the confluent culture.

Cell lines were frozen and stored by resuspending aliquots of 4 x 10^6 cells in 1 mL of freezing medium (DMEM supplemented with 10% DMSO, 20% FBS) in cryo-vials (1 1mL/vial). Vials were slowly frozen in isopropanol and left at –80°C overnight.

2.5.1.1 - Culture of fully polarised mIMCD-3 cells

In order to induce mIMCD-3 cells to fully polarise, cells were seeded onto 0.4 μm pore size transwell permeable supports at a density of 1 x 10^5 cells per cm^2. Media was changed every 2-3 days and cells were grown for 14 days.

mIMCD-3 cells were grown in medium containing 1.5 μg/mL of G418 sulphate (PAA Laboratories) when selection was required.

To monitor the polarisation process TER (Trans-Epithelial Resistance) was measured with an EVOM AC square wave current (+/- 20mA, 12.5Hz) resistance meter (World Precision Instruments, Stevenage, UK) after one week of growth. A 'blank' measurement was done for a filter with no cells. The confluence of the cellular monolayer was monitored by the increase in TER.

For mIMCD-3 culture growing on 3D cultures, Geltrex matrix was used as a support. Geltrex (Life Technologies) was prepared according to manufacturer protocol. Geltrex was thawed on ice and 50 μL was spread in each well of tissue culture treated 8-well μ-Slide to form a thick layer, and set at 37°C. 1x10^5 mIMCD-3 cells were mixed with 1mL of 2% Geltrex and 200 μL of the mixture was pipetted on the top of the Geltrex layer in each well. Slides and plates were incubated at 37°C in a humidified atmosphere

of CO_2 in air for 4 days. At day 4 the gel in slides was fixed in cold Methanol and stained (section 2.5.3).

2.5.2 - Transfection of HEK293, HeLa and mIMCD-3s using Lipofectamine™ 2000

The day before transfection HEK293 and mIMCD-3 cells were seeded in the culture vessel of choice, so they would reach 50-90% confluency the next day. Cells were then transfected with plasmid DNA constructs using Lipofectamine™ 2000 reagent (Invitrogen). Before transfection, growing medium was removed and replaced with transfection medium (serum free).

A total of 1 μg of DNA per 2 cm² area per well was added to 50 μL of transfection medium, and 10 μL of Lipofectamine 2000 was added to the same volume in separate tubes and incubated for 5 min at room temperature. The 2 solutions were then combined together and incubated at room temperature for 20 min.

The 100 μL of DNA-Lipofectamine 2000 complexes were added to the cells and, after gentle mixing, were incubated at 37 °C with 5% CO_2 for 6 hours.

Transfection medium was then replaced with normal growth medium and cells were allowed to recover for 18hrs before immunofluorescence microscopy experiments, or for 42 hrs when immunoprecipitation experiments were performed.

2.5.3 - Immunofluorescence microscopy

HEK293, HeLa and mIMCD-3 cells were seeded on glass coverslips, transfected 24 hrs prior to preparation for immunofluorescence, as described in the previous section. Cells were fixed in 4% paraformaldehyde-PBS (pH=7.0) for 20 min at 37°C. After washing twice in PBS, cells were permeabilised in 0.5 % Triton-X-100 (diluted in PBS) for 10 min, followed by blocking in Blocking Buffer (2% BSA in PBS).

Filters with mIMCD-3 cells were first excised from transwell supports before being processed. When Tight Juntions (TJ) and Adherens Junction (AJ) proteins were immunostained, mIMCD-3 cells were fixed using ice-cold methanol for 5 minutes and permeabilised with 0.1% saponin.

After blocking, filters and coverslips were further incubated in primary antibodies diluted in blocking buffer, for 1 hour at room temperature. Specific dilutions for all primary antibodies used for immunofluorescence in this thesis are reported in table 2.9.

Primary antibodies were fluorescently labeled using Zenon Rabbit IgG Labeling Kit (Invitrogen) when two antibodies derived from the same species were used in the same application. Cells were then washed 4x in PBS, incubated in the dark in secondary antibody solution at the required concentration (table 2.9). Goat anti-mouse IgG and anti-rabbit IgG conjugates, as well as donkey anti-sheep IgG conjugates with Alexafluor 488, 568 and 633 secondary antibodies were used. Nuclei were stained with DAPI or TO-PRO-3 (Invitrogen), both diluted 1/2000 in blocking buffer for 2-5 min at room temperature. Cells were than washed again 4x in PBS to remove the excess of secondary antibodies. Coverslips and filters were then mounted onto microscopy slides using ProLong Gold antifade solution (Invitrogen) (5 µL/coverslip).

Confocal analysis was performed and images were captured using an inverted Leica

TCS SP5 AOBS confocal microscope with a 63 x 1.4 oil immersion objective (Leica Microsystems) and the pinhole set to 1 Airy unit.

Series of Z-sections at 0.25 μm intervals were captured and merged into maximum projection images when required. Single optical sections were averages of 3 scans at 1024 x 1024 resolution. Single xz plane images were also obtained where necessary.

Images were colorized and contrast was enhanced using ImageJ software package.

Antibody	**Clone / Catalogue nr**	**Supplier**	**Primary Dilution**	**Secondary Dilution**
Sheep α TGN46	AHP500G	AbD serotec	1/1000	1/2000
m α p230	15/p230	BD Biosciences	1/5000	1/6000
Chicken α Calreticulin	ab14234	abcam	1/400	1/400
Rb α VIPAR	HPA003589	Sigma	1/50	1/100
Rb α PLOD3	11027-1-AP	proteintech	1/50	1/100
m α PDI	RL90	abcam	1/100	1/200
m α LAMP1	CD107a	Hybridoma Bank	1/100	1/100
m α AP-1	100/3	Sigma	1/100	1:100
Rb α Collagen IV	ab6586	abcam	1/200	1/400
m α E-cadherin	4A2C7	Invitrogen	1/100	1/200
m α Rab11a	47/Rab11	BD Biosciences	1/50	1/100
m α GM130	35/GM130	BD Biosciences	1/100	1/100
Rb α Claudin-1	51-9000	Invitrogen	1/400	1/500
Alexa Fluor® 488 Phalloidin	A12379	Invitrogen	-	1/2000

Table 2.9 - Antibodies and their dilutions used in this thesis for immunofluorescence experiments. Other information provided includes the species the antibody was raised in (m- mouse, Rb-rabbit), the company along with the clone number for monoclonal antibodies or catalogue number for polyclonal antibodies.

2.5.4 - mIMCD-3 stable knockdowns

mIMCD-3 stable knockdown cells lines were created by transfecting cells with plasmids containing shRNA hairpins for the specific gene. For Vps33b a pSM2c plasmid was used and a pRS vector for Plod3 and Vipar. The oligonucleotide sequences for all shRNAs are reported in table 2.10. Non-silencing hairpins were used as controls. Plasmid shRNA construct for Vps33b was purchased from Open Biosystems (V2MM_91195) while plasmids shRNA for Vipar (TR517186) and Plod3 (TR502609) were purchased from OriGene.

mIMCD-3 cells were transfected with 4 μg of plasmid and allowed to recover for 48 h before transfected cells were selected using 1.5 μg/ml of Puromycin. Individual clones were selected and knockdown levels of each clone were assessed by western blotting.

Target gene	**shRNA sequence**
Vps33b	CCCTGGCCAGTGCTGAAGTATT
Vipar	CGCGCAGTGAATAGCCTTCGAGACATCGT
Plod3	ATGACTCGTCCACCTTCACTCTCAATGTC

Table 2.10 - List of shRNA oligonucleotides used for shRNA transfection.

CHAPTER THREE:

VPS33B and VIPAR interact forming a stable binary complex

3.1 - INTRODUCTION AND OVERVIEW

Genetic investigations on patients with ARC syndrome identified mutations in genes encoding two proteins: VPS33B or VIPAR. Mutations in *VPS33B* account for approximately 75% of ARC patients while mutations in *VIPAR* were found in all ARC cases where mutations in *VPS33B* were excluded (Cullinane et al., 2010). The ARC–LOVD database https://grenada.lumc.nl/LOVD2/ARC reports all known mutations and variants for *VPS33B* and *VIPAR* (Smith et al., 2012). A summary of the ARC–LOVD database reporting pathogenic mutations in *VIPAR* and *VPS33B* along their genomic sequence is shown in figure 3.1 A and C. ARC patients carrying mutations in *VPS33B* or *VIPAR* show no difference between them with respect to clinical symptoms, signs or disease course (Cullinane et al., 2010). This observation might suggest that VPS33B and VIPAR are strongly associated in the cell by regulating the same intracellular process or being part of the same intracellular pathway. By analysing the map of known pathogenic ARC-mutations along VPS33B and VIPAR genomic sequence (figure 3.1 A and C) was noticed that mutations are distributed uniformly along VPS33B genomic sequence while they don't seem to affect the first five exons of VIPAR. Because no structural information exists for VPS33B and VIPAR and it is not therefore possible to correlate the three-dimensional folding and misfolding of the protein in healthy and pathological conditions, I performed structural predictions for both proteins (figure 3.1 B and D, section 3.6). Predictions of protein secondary structure, globularity and disorder were performed using GlobPlot (http://globplot.embl.de/), FoldIndex (http://bip.weizmann.ac.il/fldbin/findex), IUPred (http://iupred.enzim.hu/), RONN (http://www.oppf.ox.ac.uk/RONN/) and HHPRED (http://toolkit.tuebingen.mpg.de/hhpred) (section 3.6). Disordered regions (DRs) are

defined as regions of proteins that lack any recognisable secondary structure elements, (α-helices, β-strands). These analyses suggest that both proteins contain globular regions with well-defined secondary structure elements. In addition, a large disordered protein segment has been consistently identified by multiple structure prediction software at the N-terminus of VIPAR (figure 3.1 B). Many disordered proteins function via binding to a structured partner and undergo a disorder-to-order transition. The N-terminal of VIPAR might therefore require a binding partner that determines its fold in physiologic conditions (Forman and Mittag, 2013). The absence of known pathogenic ARC mutations that affect the N-terminal disordered part of VIPAR (figure 3.1 C and D) might suggest that mutations in this region are either lethal (and therefore do not exist in living patients) or dispensable for VIPAR interaction with other binding partners such as VPS33B.

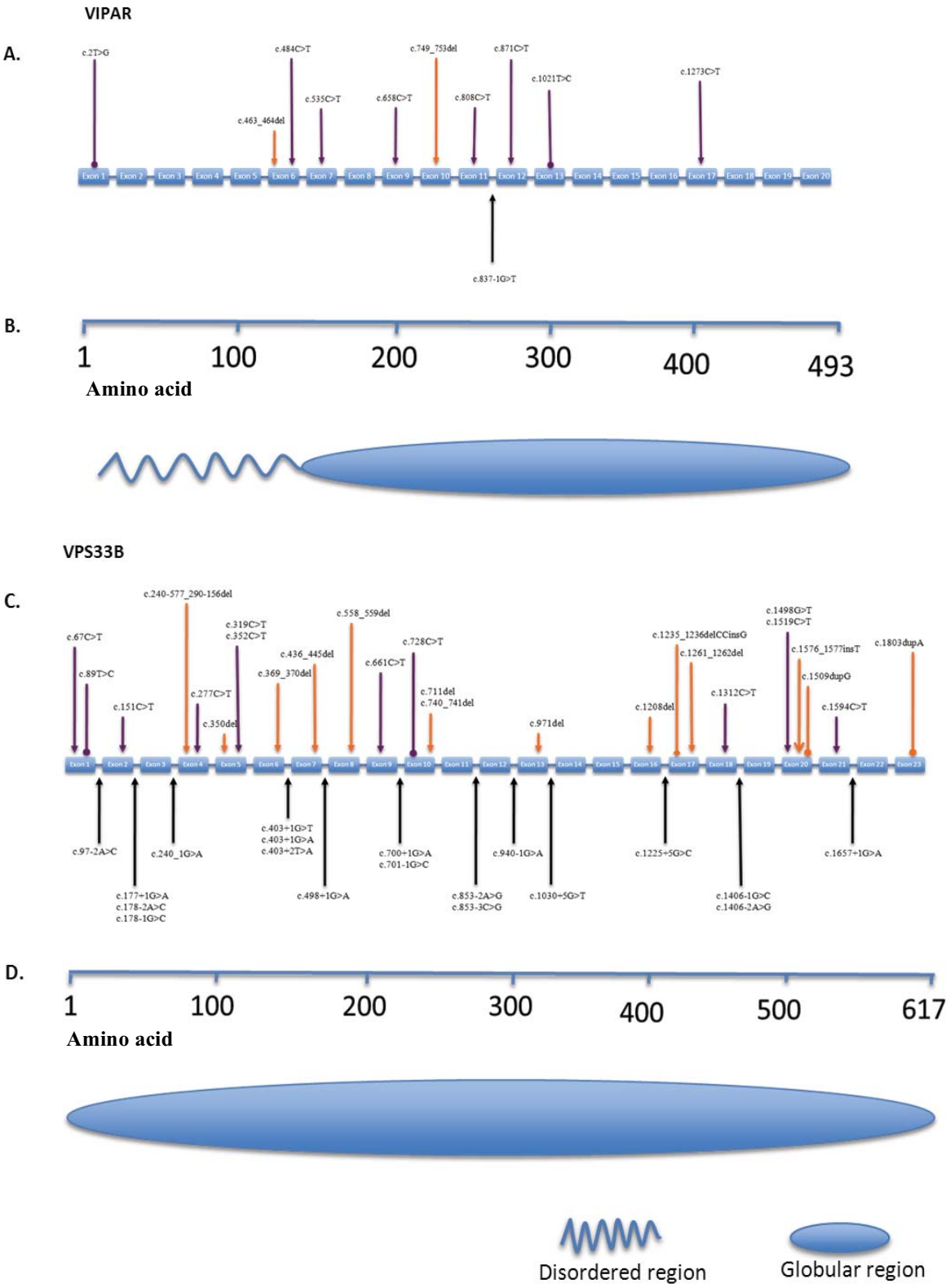
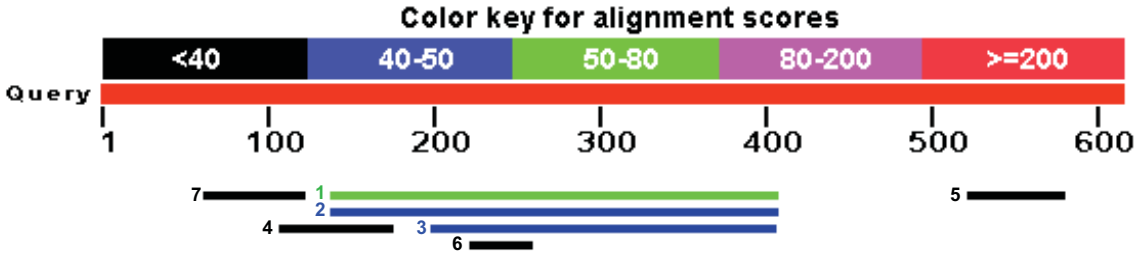

VIPAR
A.
c.2T>G
c.484C>T
c.749_753del
c.871C>T
c.1021T>C
c.1273C>T
c.535C>T
c.658C>T
c.808C>T
c.463_464del
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Exon 7
Exon 8
Exon 9
Exon 10
Exon 11
Exon 12
Exon 13
Exon 14
Exon 15
Exon 16
Exon 17
Exon 18
Exon 19
Exon 20
c.837-1G>T
B.
1
100
200
300
400
493
Amino acid
VPS33B
C.
c.240-577_290-156del
c.67C>T
c.319C>T
c.352C>T
c.558_559del
c.728C>T
c.436_445del
c.661C>T
c.89T>C
c.369_370del
c.277C>T
c.711del
c.740_741del
c.151C>T
c.350del
c.971del
c.1235_1236delCCinsG
c.1261_1262del
c.1498G>T
c.1519C>T
c.1803dupA
c.1576_1577insT
c.1509dupG
c.1208del
c.1312C>T
c.1594C>T
c.97-2A>C
c.240_1G>A
c.403+1G>T
c.403+1G>A
c.403+2T>A
c.700+1G>A
c.701-1G>C
c.940-1G>A
c.1225+5G>C
c.1657+1G>A
c.177+1G>A
c.178-2A>C
c.178-1G>C
c.498+1G>A
c.853-2A>G
c.853-3C>G
c.1030+5G>T
c.1406-1G>C
c.1406-2A>G
D.
1
100
200
300
400
500
617
Amino acid
Disordered region
Globular region

Figure 3.1 - Analysis of the structured and disordered regions along VPS33B and VIPAR protein sequence related to the map of ARC mutations along the respective genomic sequence. Distribution of pathogenic mutations along VIPAR ***(A)*** *and VPS33B* ***(C)*** *genomic sequence. Boxes represent the exons (not to scale). Variants below the genogram represent intronic mutations, affecting splice sites. Variants above the genogram represent mutations in exons. Purple pointers indicate substitutions; → indicates a nonsense mutation, —• indicates a missense mutation. Orange pointers indicate deletions/duplications/indels; → indicates a deletion, —• indicates a duplication, and —♦ indicates an indel. Analysis of protein sequence with prediction of structural disorder for full-length human VIPAR* ***(B)*** *and VPS33B* ***(D)****. Regions predicted as globular are shown as blue ovals. Figure modified from (Smith et al., 2012).*

The function of VPS33B has not yet been fully characterised in humans but the yeast homolog Vps33p is known to belong to the SM family of proteins (Peterson and Emr, 2001). SM proteins play an important role in conferring specificity to membrane fusion events by binding directly or indirectly to Soluble N-ethylmalemide sensitive factor Attachment Receptor (SNARE) of the syntaxin family (Chen et al., 2002; Gallwitz and Jahn, 2003; Toonen and Verhage, 2003). They are structurally highly conserved in evolution, and each SM protein is specialised in a single or a small group of trafficking steps. Sec 1 was one of the first SM proteins identified (SM for Sec1/Munc18). Aalto et al. (1992) noticed that in *Saccharomyces cerevisiae* there are two other proteins functionally related to Sec1, known as Slyl, and Slpl (Vps33) that altogether form the Sec1 family of proteins. Secl, Slyl, and Slpj (Vps33), which share a low 22% of sequence homology among them.

In yeast Sec1 (Sec for secretion mutant) is required for protein secretion to the cell exterior and precisely for the fusion of secretory vesicles with the plasma membrane and consequent release of the cargo protein (Aalto et al., 1991; Novick et al., 1980). Sec1 protein is not soluble and is found in a detergent-resistant subcellular fraction in yeast (Egerton et al., 1993). The Sly1 protein instead may act in endoplasmic reticulum

to Golgi transport (Ossig et al., 1991) and therefore is also involved in protein secretion. The third member of the Sec1 family, Vps33/Slpj, might be involved in transport from the Golgi complex to the vacuole (Wada et al., 1990). VPS33B is homologue of the yeast Vps33 protein and it is therefore not surprising that protein similarity predictions for VPS33B using pBLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) against the PDB indicate 22% of identity between VPS33B and the neuronal Sec1 (nSec1) from the Squid L. Pealei (figure 3.2). Therefore, we can conclude that nSec1 is the protein with the highest identity to VPS33B. The crystal structure of nSec1 from the Squid L. Pealei (PDB accession number 1FVF) indicates two main functional regions in nSec1 structure: one contains the Syntaxin 1a binding site, and opposite to this, a potential and yet unspecified effector-molecule binding-pocket which is present in both rat nSec1 and homologous squid s-Sec11a. This fold is conserved between VPS33B and nSec1 because the two proteins belong to the same family. Consequently, like Sec1, it is likely that VPS33B binds to yet unknown SNARE and effector molecules to carry out its function.



Description	Max score	Total score	Query cover	E value	Max ident	Accession
1. Chain A, Crystal Structure Analysis Of Neuronal Sec1 From The Squid L. Pealei	50.4	50.4	43%	3e-06	22%	1FVF_A
2. Chain A, X-Ray Crystal Structure Of Neuronal Sec1 From Squid	46.2	46.2	43%	6e-05	22%	1EPU_A
3. Chain A, Crystal Structure Of The Unc18-Syntaxin 1 Complex From Monosiga Brevicollis	42.4	42.4	33%	0.001	25%	2XHE_A
4. Chain A, The Crystal Structure Of The Putative Regulator From Escherichia Coli Cft073	31.2	31.2	11%	1.5	32%	3HFI_A
5. Chain A, X-Ray Structure Of The Death Domain Of The Human Mucosa Associated Lymphoid Tissue Lymphoma Translocation Protein 1	28.1	28.1	9%	8.4	31%	2G7R_A
6. Chain A, Crystal Structure Of Sly1p In Complex With An N-Terminal Peptide Of Sed5p	29.3	29.3	6%	9.3	31%	1MQS_A
7. Chain A, Crystal Structure Of Human Glcat-P Apo Form	28.9	28.9	10%	9.7	26%	1V82_A

Figure 3.2 - Database sequences identified by pBLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) to be similar to the human VPS33B protein sequence. In the diagram, the bar color for a hit refers to the alignment score, a mathematically derived value that reflects the degree of similarity between hit and query sequences. The higher the score, the more similar the two. The Color Key at the top of the graphical display gives the range of alignment scores assigned to each color. Numbers flanking each bar color correspond to the sequence identified by BLAST and listed in the table below.
The table reports: the description/title of matched database sequence, the highest alignment score (Max score) from that database sequence, the total alignment scores (Total score) from all alignment segments, the percentage of query covered by alignment to the database sequence, the best (lowest) Expect value (E value) of all alignments from that database sequence, the highest percent identity (Max ident) of all query-subject alignments, and the Accession of the matched database sequence.

No proteins with known crystal structure that are homologous to VIPAR are known. Consequently we cannot deduce structural information on VIPAR based on the homology with other proteins. To deduce functional information for VIPAR, pBLAST was executed against the UniProt/Swiss-Prot database and found 27% of identity between VIPAR and Golgin subfamily A member 5 (GOLGA5) protein. GOLGA5 also

known as ‘Cell proliferation-inducing gene 31 protein’ or ‘Golgin-84’ encodes a member of the golgin family of proteins, whose members localise to the Golgi. This protein is a coiled-coil membrane protein that has been postulated to play a role in vesicle tethering and docking (Sohda et al., 2010). Previous data indicate that Golgin-84 interacts with another tether, the conserved oligomeric Golgi (COG) complex that is important in vesicular trafficking and for maintaining Golgi structure and function (Sohda et al., 2010; Zeevaert et al., 2008). Translocations involving this gene have been found in tumour tissues (Ishizaka et al., 1989; Rabes, 2001).

Like VIPAR, GOLGA5, belongs to the Golgin_A5 family of proteins whose members share a conserved Golgin_A5 domain. Indeed bioinformatics analysis on VIPAR using the Pfam, SMART and pBLAST databases finds a large Golgin subfamily A member 5 domain in the VIPAR protein sequence (section 1.3). As mentioned previously for Sec1 and VPS33B, members belonging to the same family of proteins share the same fold between them. We can therefore conclude that VIPAR shares its protein fold with GOLGA5 and like GOLGA5 requires other binding proteins to carry out its function.

Co-localisation and co-immunoprecipitation experiments have shown that VPS33B and VIPAR are physically associated (Cullinane et al., 2010). However these types of experiments don’t give any information regarding the interaction between the two proteins. This can be either direct or indirect. Indeed positive results in co-immunoprecipitation experiments may indicate that two proteins interact directly or they may interact indirectly via one or more bridging molecules. The aim of this part of the work was to establish if VPS33B and VIPAR interact biochemically and directly between them. This was achieved by pull-down experiments using over-expressed VPS33B and VIPAR. Pull-down assays are *in-vitro* methods used for both confirmation

of previously suspected interactions (such as from co-immunoprecipitation experiments) as well as for the identification of unknown protein interactors after coupling this technique to mass-spectrometry. The technique requires large amounts of the proteins whose direct interaction is being tested and this is achieved by their over-expression in a protein expression system. This procedure eliminates confusing results that could arise from interaction of the bait with other endogenous interacting proteins that are not being tested in the study. Another purpose for the large-scale production of VPS33B was the generation of an anti-VPS33B antibody. Because the commercially available ani-VPS33B antibodies were not suitable for detection of the endogenous protein by WB and immunofluorescence and this was therefore a limiting factor in the progress of this work.

To test the direct interaction between VPS33B and VIPAR the two proteins were over-expressed in several expression systems (bacterial, insect and mammalian). Such efforts had to be undertaken because of the initial difficulties with obtaining sufficient amount of purified protein present in their native conformation, first in insect and then bacterial cells. (Appendix 1). Recombinant protein production in large scale using the mammalian systems was the best strategy used to express and successively purify VPS33B and VIPAR.

For a successful confirmation of their interaction, the over-expressed proteins must preserve their native conformation. Size exclusion chromatography (section 3.3.5) was used as a biochemical tool to test the stability of the proteins alone and in complex and also to further confirm the direct interaction between VPS33B and VIPAR.

The '*in cell*' confirmation of protein-protein interaction was performed using confocal microscopy with Fluorescence Lifetime Imaging in Förster Resonance Energy Transfer

(FRET-FLIM) (section 3.4.3). This technique also allows estimation of molecular distances separating the proteins.

Once the direct interaction between VPS33B and VIPAR was confirmed, the work was focused on the protein regions involved in this interaction. As mentioned above, no mutations have been identified in the N-terminus of VIPAR that corresponds to the predicted disordered region at the N-terminus of the protein. I therefore hypothesized that this region is not involved in the interaction with VPS33B and tested this hypothesis by using short constructs containing different regions of VIPAR (section 3.4.4).

Several protein prediction softwares such as PredictProtein (https://www.predictprotein.org/) predict a nuclear localisation sequence (NLS) at the interface between the disordered and the globular region of VIPAR (section 3.6, figure 3.13). This result was quite surprising because previous data have localised VIPAR to the cytosol of eukaryotic cells (Cullinane et al., 2010; Smith et al., 2012). I questioned the role of this sequence in the interaction with VPS33B and/or other interacting partners and performed protein-protein interaction experiments using short constructs of VIPAR that contain or lack the NLS.

If VPS33B and VIPAR interact directly, we can interrogate the nature of this complex. Can VPS33B and VIPAR co-exist in a stable binary complex or do they require other interacting partners for the formation of a stable and functional unit? To answer these questions was the second goal of this part of the study. As mentioned above, in order to address the question of whether VPS33B and VIPAR co-exist in a stable binary complex, pull-down experiments were performed followed by size exclusion chromatography. The addition of a MALS (Multi Angle Dynamic Light Scattering)

detector coupled downstream to a chromatographic system is a further detection method that was used to obtain accurate information regarding the molecular weight and aggregation state of the proteins and protein complexes (section 3.3.6).

Very recently, the EM structure of the HOPS (vacuolar/lysosomal homotypic fusion and vacuole protein sorting) complex in yeast (section 1.4.4.4.1) was published (figure 3.3) (Brocker et al., 2012). HOPS forms an elongated structure with flexible head formed by Vps16-Vps33 (the yeast homologue of VIPAR/VPS33B). Vps41 is localised to the remainder of the head and Vps39-Vps11 form the bulky tail while Vps18 constitutes the flexible linker that connects the head to the tail regions. This study shows that HOPS components can be isolated in sub-complexes formed by 1) Vps16-Vps33 that form a very tight sub-complex, 2) Vps39-Vps11-Vps18 and 3) Vps39-Vps11 (figure 3.3).

As mentioned previously (section 1.3), it is unlikely that VPS33B and VIPAR are part of the HOPS complex (Smith et al., 2012), however there might be some conditions in the cell under which VPS33B and VIPAR are involved in HOPS (Cullinane et al., 2010; Smith et al., 2012; Zlatic et al., 2011). The EM structure shows that the yeast homologues of VPS33B and VIPAR (Vps33 and Vps16) can exist as a tight sub-complex of HOPS. We can therefore hypothesise that in higher organisms this potential sub-complex has acquired new functions that go beyond its role of a subunit of the HOPS complex. It is therefore of primary importance to establish if in higher organisms VPS33B and VIPAR exist as a stable sub-complex, such as their yeast counterpart Vps33-Vps16.

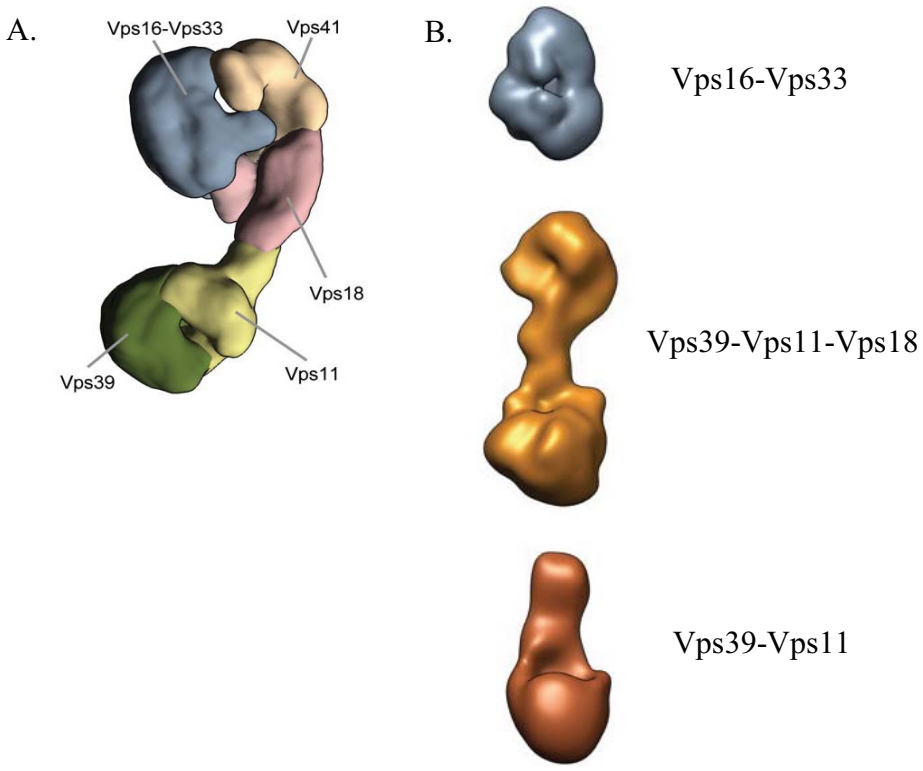


Figure 3.3 - Organisation of the HOPS complex. ***(A)*** *Segmented model of the HOPS tethering complex. The head domain consists of Vps16-Vps33 (grey) and Vps41 (sandy brown), and the bulky tail is composed of Vps11 (light olive green) and Vps39 (dark olive green). The head is connected to the tail by the Vps18 linker (pink).* ***(B)*** *Sub-complexes of HOPS: Vps16-Vps33 (gray), Vps39-Vps11-Vps18 (gold) and Vps39-Vps11 (brown). Figure modified from (Brocker et al., 2012).*

Furthermore, if VPS33B and VIPAR do not form part of the HOPS complex in physiological conditions, then it is important to identify the intracellular pathway that VPS33B and VIPAR are involved in. This can be achieved by identifying VPS33B and VIPAR interacting proteins and understanding how these interactions impact the phenotype of the cell model for ARC syndrome. For the identification of new interactors, the experimental approach undertaken consisted in performing pull-down experiments using purified VPS33B/VIPAR as a "bait" followed by mass spectrometry for the protein identification. Pull-down experiments are indeed more amenable to an

initial screening than other biochemical methods such as co-immunoprecipitation as they give higher sensitivity and specificity to the interactions detected by the screen.

3.2 - AIMS

Although previous studies suggested that VPS33B and VIPAR interact, the direct interaction was not unequivocally confirmed. The aim of the work described in this chapter was to biochemically demonstrate direct protein-protein interaction and successively determine the regions in VIPAR involved in the interaction.

Another aim of the work was to reconstitute the minimal VPS33B/VIPAR complex *in-vitro* and test its stability.

3.3 - METHODS

3.3.1 - Mammalian expression system

VPS33B and VIPAR were expressed using human embryonic kidney 293 cells that stably express Epstein-Bar virus nuclear antigen EBNA1 (HEK293-E cells, supplied by U-Protein Express BV (U-PE), Utrecht, The Netherlands). An overview of the whole process of expression and purification system used in this study is reported in figure 3.4. *VPS33B* and *VIPAR* genes were cloned using polymerase chain reaction adding 5' BamHI and 3' NotI restriction sites. The genes were transferred into a pCR8 TOPO cloning vector (Invitrogen) using the TOPO cloning protocol, that firstly adds adenine overhangs to the blunt PCR product, then ligates the product into a cloning vector that bears flanking thymines at 5' and 3'. After restriction digestion and agarose

electrophoresis, the bands corresponding to the VPS33B and VIPAR inserts were excised and purified using the Wizard SV Gel and PCR Clean-Up System (Promega), and successively ligated into a library of expression vectors bearing various tags for solubilisation and optimization of expression conditions. 2 μg of plasmid DNA (either using a single VPS33B or VIPAR expression plasmid or a mixture of 1 μg of a VPS33B and 1 μg of a VIPAR expression plasmids) were used to perform small-scale transient transfections of HEK293 (4 mL) grown in suspension. This procedure was performed by the U-Protein Express BV (Utrecht). Six days after transfection cells were harvested by centrifugation and the production of recombinant proteins was assayed for each expression condition by performing Western blots using antibodies against the expression tags selected for the expression tests. This procedure allowed us to identify the optimal combination of expression conditions and purification tags for the production of the proteins. After experimenting with more than 50 different combinations of expression vectors, the best expression condition was determined in a co-expression mixture of pUPE.02.09-VIPAR (N-His-$Strep_3$ tag, followed by a cleavage site for the Tobacco Etch Virus Protease (TEV) and pUPE.02.13-VPS33b (N-His-$cmyc_3$ tag, followed by TEV cleavage site). This combination was then used for a large scale (1 L) transient transfection of HEK293-E cells in suspension.

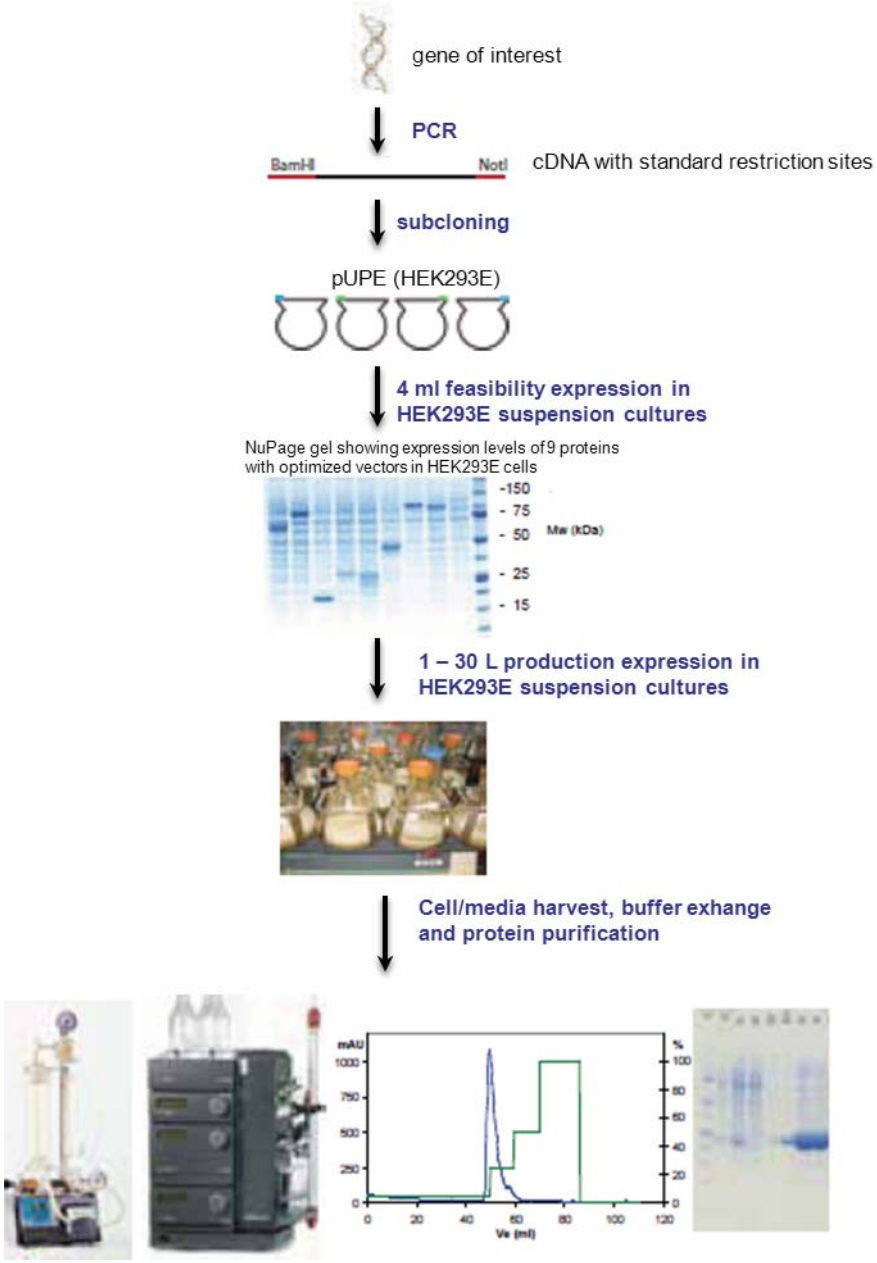


Figure 3.4 - Expression and purification platform illustration for VPS33B and VIPAR comprises the following steps: 1. Generation of coding cDNA via PCR using specific cDNA template and gene-specific primers containing standard restriction sites. 2. Ligation of this coding sequence into different expression vectors. 3. Small scale transient transfection of *HEK293E cells cultured in suspension. Six days post-transfection recombinant protein production was assayed and the most optimal expression vector was selected. 4. Large scale protein production in HEK293E cells cultured in suspension. 5. Purification and characterization of recombinant proteins. Figure modified from http://www.u-proteinexpress.com/media/brochure-r-PEx-technology.pdf*

3.3.2 - Mammalian protein extraction

The HEK293-E cells were grown in suspension for 6 days after co-transfection with VPS33B and VIPAR expression plasmids, and then centrifuged for 10 minutes at 1000 rpm. Medium was discarded, and a low salt method was used for cell lysis. The cell pellet was resuspended in 100 ml of cold lysis buffer (10 mM Hepes pH 8.5, 10 mM KCl, 5 mM MgCl2, 1 μg/mL DNase, 5mM DTT, Protease inhibitors). The mixture was incubated with gentle rotation at 4 °C for 5-10 minutes, then the Nonidet P-40 detergent was added to a final concentration of 0.3% to lyse the cells but preventing the breakage of the nuclear membrane. The suspension was incubated for another 15 minutes at 4 °C and then the lysate was clarified by centrifugation at 9.500 xg for 1 hour at 4 °C. The pellet, containing membrane fragments and insoluble material, was discarded.

3.3.3 - Mammalian protein purification

After centrifugation of the lysate, the soluble fraction was processed using liquid chromatography to isolate the VPS33B-VIPAR protein complex. The purification was performed using affinity chromatography, by taking advantage of the Strep tag used for the recombinant expression of VIPAR in the pUPE.02.09 expression vector. After adding NaCl to a final concentration of 250 mM, the clarified lysate was passed through a pre-packed StrepTrap FF 1 ml column (GE Healthcare) with a flow-rate of 0.25 ml/min. The column was then thoroughly washed with 50 ml of wash buffer (25 mM HEPES/NaOH pH 8, 250 mM NaCl, 5 mM DTT) to remove non-specifically bound contaminants, and then the strep-tagged proteins were eluted with wash buffer + 2.5

mM desthiobiotin. The elute fractions were analysed by SDS-PAGE and western blot (section 3.4.1, 3.4.2).

3.3.4 - Dialysis of Recombinant Proteins

Dialysis Membrane Tubing (Spectrapor) with 8 kDa molecular-weight cutoff (MWCO) was used to exchange the buffer solution in which the proteins were dissolved. Dialysis was performed to remove desthiobiotin from the protein elution pool. Proteins were dialysed against dialysis buffer (with the same composition of the StrepTrapcolumn wash buffer) overnight at 4 °C under agitation.

3.3.5 - Size exclusion chromatography

Size exclusion (also called gel filtration) chromatography separates molecules according to their differences in size as they pass through a packed medium (in general sepharose or acrylamide-based beads) in a glass column. The elution progress is monitored with continuous UV adsorbance detection using an Akta System (GE Healthcare), and the protein-containing fractions are collected separately.

The VPS33B-VIPAR complex eluted from the Strep column was concentrated using a Vivaspin 15 concentrator MWCO 30 kDa (Sartorius Mechatronics) to a maximum concentration of 5 mg/mL before gel filtration. Protein concentration was evaluated by measuring the absorbance at 280 nm using a Nanodrop ND-1000 spectrophotometer (Thermo Scientific) and applying a correction of 1.02% as determined from sequence-based calculations using the Expasy ProtParam tool (Expasy website reference). The

sample was injected into a *Superdex 200 5/150* gel-filtration column. Elution was performed with a flow rate of 0.25 mL/min using a buffer containing 25 mM Hepes pH 8, 300 mM NaCl, 5 mM DTT.

3.3.6 - Multi-Angle Laser Light Scattering (MALLS)

Multi-angle laser light scattering (MALLS) is an extremely sensitive technique for determining the absolute molar mass and the aggregation state of biomolecules in solution by measuring the intensity of a laser light scattered at different angles by molecular objects passing through the flow cell. This method is highly reproducible and does not require calibration with molecular weight standards. Our gel-filtration analysis was coupled to MALLS to determine the molecular weight and the oligomeric state of the VPS33B/VIPAR complex. 5-10 µL of purified protein complex was injected at 1 mg/mL. A single gel filtration peak indicated that VPS33B and VIPAR form a stable complex that does not dissociate in the column at a salt concentration of 300 mM. MALLS analysis was performed using a three-angle light scattering detector (mini-DAWN TREOS, Wyatt technology) and a refractive index detector (RID-10A, Shimadzu). Data points were collected every 0.5 seconds at a flow rate of 0.1 mL/min. Data analysis was carried out using the ASTRA software (Wyatt technology).

3.4 - RESULTS

3.4.1 - VPS33B interacts directly with VIPAR forming a binary complex

A biochemical approach was used to establish whether there is direct interaction between VPS33B and VIPAR. It consisted of large-scale VPS33B and VIPAR protein production followed by pull-down experiments using *in-vitro* purification systems. Pull-down experiments are strong indication of direct protein-protein interaction because of being performed with purified samples present in not-endogenous amounts. The mammalian cell system was used for the large-scale expression of VPS33B and VIPAR (section 3.3.1). This was done in collaboration with Professor Gros and Dr Federico Forneris from the Crystallography laboratory of the University of Utrecht, who commonly use the mammalian expression system to over-express human proteins. Previous attempts to perform pull-down experiments and reconstitute *in-vitro* the VPS33B-VIPAR complex in insect and bacterial cell system were not successful. Even though purification of VPS33B alone was successful in insect cells (Appendix 2) the protein formed aggregates and therefore was not suitable for further analysis.

VPS33B and VIPAR were expressed on large scale in suspension growing human embryonic kidney 293 cells that stably express Epstein-Bar virus nuclear antigen EBNA1 (HEK293-E cells, supplied by U-Protein Express BV (U-PE), Utrecht, The Netherlands) (section 3.3.1). The expression and purification of VPS33B and VIPAR alone and in complex was performed. Initially, the reconstitution of VPS33B-VIPAR protein complex was attempted from individually produced proteins. However, only very low expression levels of VPS33B and VIPAR were detected and therefore it was

found impossible to purify these proteins alone using the mammalian expression system.

The best results were achieved from the co-expression and successive co-purification of recombinant VIPAR (in fusion with His6-Strep3-TEV tag) with recombinant VPS33B (in fusion with His6-cMyc3-TEV tag). This led to an enormous increase in the expression of both proteins that reached 2-3 mg/L. Several publications showed that an increase in protein expression and solubility can be achieved when proteins that normally exist in stable protein complexes are co-expressed together (Hanzlowsky et al., 2006; Kerrigan et al., 2011) indicating that this might be the case for VPS33B and VIPAR.

After co-expression in HEK293-E cells, VPS33B was co-purified with VIPAR by affinity purification using the StrepTactin™ Sepharose™ resin that specifically binds the Strep3 tag of VIPAR. The co-elution of VPS33B in stoichiometric amount compared to VIPAR (figure 3.5) is an experimental proof of the direct interaction between VPS33B and VIPAR. Indeed only VIPAR, provided with the Strep-tag, is capable of affinity binding StrepTactin resin. The presence of VPS33B in the elution fractions might be explained either by its binding to VIPAR or by a non-specific binding to StrepTactin resin, although this second hypothesis is clearly excluded considering that the proteins co-elute stoichiometrically. This second hypothesis was further excluded by loading His6-cMyc3-VPS33B on the StrepTactin resin and analysing the bound fractions by SDS-PAGE and subsequent Western blot with an anti-His6 antibody. Western analysis was negative demonstrating that His6-cMyc3-VPS33B does not bind to StrepTactin resin. It was therefore possible to conclude that His6-cMyc3-VPS33B co-elutes with His6-Strep-VIPAR because the two proteins directly

interact. The identity of the proteins after purification was confirmed by Western-blot using an anti-His6 antibody that specifically recognises the His6-tag of recombinant VPS33B and VIPAR (figure 3.5).

The purified complex was concentred with Vivaspin 15 concentrator (Sartorius Mechatronics) before being subjected to size exclusion chromatography (gel-filtration) on the *Superdex 200 5/150* column for VPS33B-VIPAR complex analysis (figure 3.5; section 3.4.2).

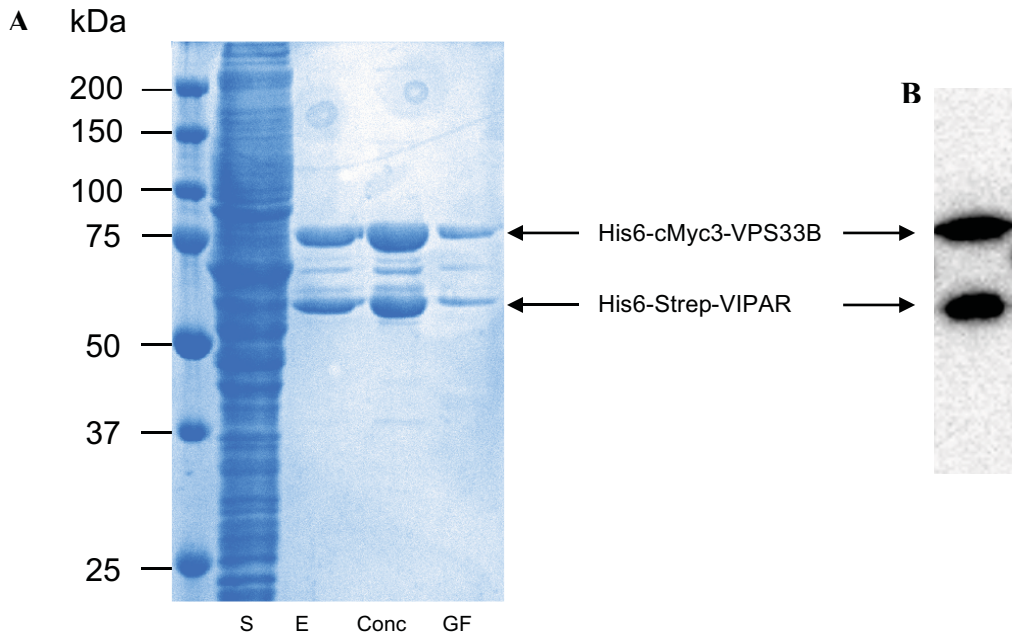


Figure 3.5 - Recombinant VPS33B-VIPAR complex analysis. ***(A)*** *SDS-PAGE for His6-cMyc-VPS33B / His6-Strep-VIPAR complex purification: S (soluble fraction), E (StrepTactin Sepharose elution), Conc (concentrated elution pool), GF (elution from the peak in gel filtration).* ***(B)*** *Western blot for the soluble fraction (S) with anti-His6 antibody*

3.4.2 - VPS33B and VIPAR form a stable complex

After establishing that VPS33B interacts directly with VIPAR, the next goal was to determine the stability of the complex *in-vitro*. Size exclusion chromatography (or gel filtration) was used on *Superdex 200 5/150* (figure 3.6) to further confirm the formation of a binary complex and to test its stability. In gel filtration, proteins or protein complexes are separated from other species by size as they pass through the chromatographic column (section 3.3.5). Therefore a single peak in gel filtration is equivalent to one protein or one protein complex. The VPS33B-VIPAR purified and concentrated complex (figure 3.5 A) elutes as a single peak in gel filtration that indicates that proteins are in complex. This complex is quite stable and aggregation phenomena were excluded because protein aggregates, too large to enter the exclusion range of the gel, normally elute in the void/dead volume of the chromatography column. The eluted fractions were analysed on SDS-PAGE confirming the presence of both proteins in the elution peak (figure 3.6).

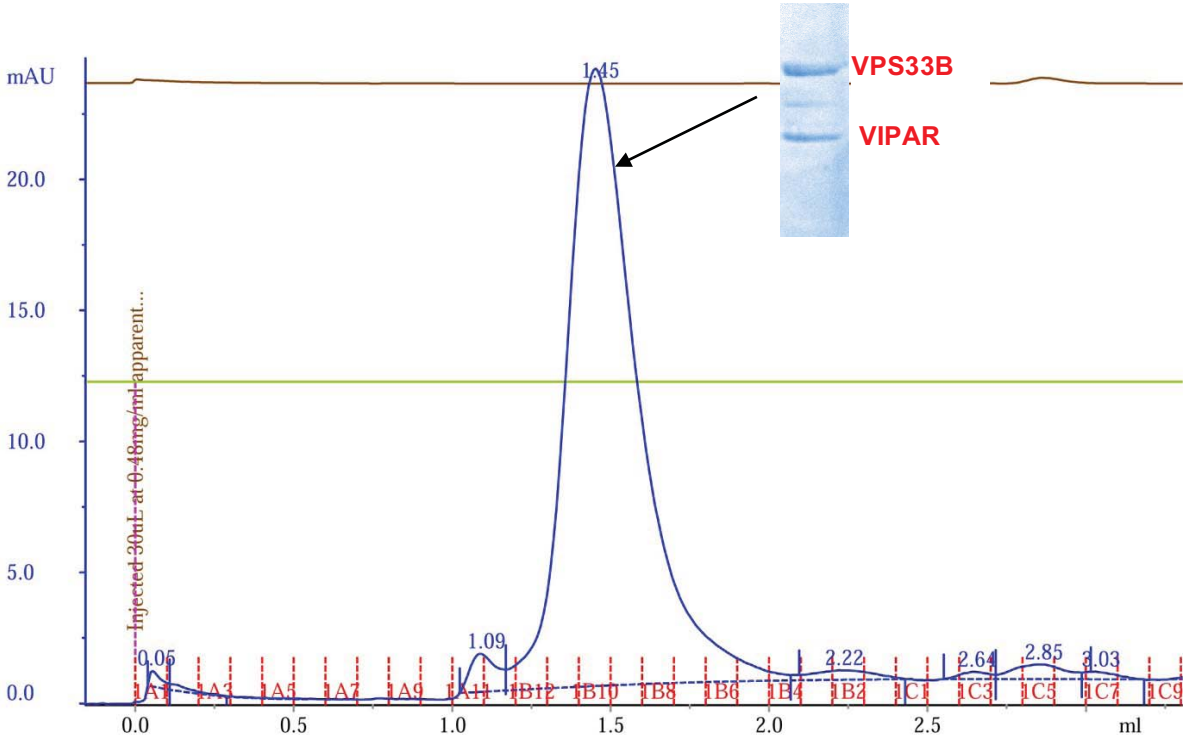


Figure 3.6 - Gel filtration chromatogram relative to VPS33B-VIPAR complex on the Superdex 200 5/150 column

For an exact and unambiguous determination of the molecular weight of the VPS33B-VIPAR protein complex, the gel-filtration was coupled to MALS (Multi Angle Dynamic Light Scattering molecular weights) (figure 3.7). The system detects a molecular weight of 280 kDa in correspondence of the VPS33B-VIPAR elution peak (red line) that might correspond to the dimer of VPS33B-VIPAR. The almost horizontal orientation of the red line indicates the complex is quite homogeneous.

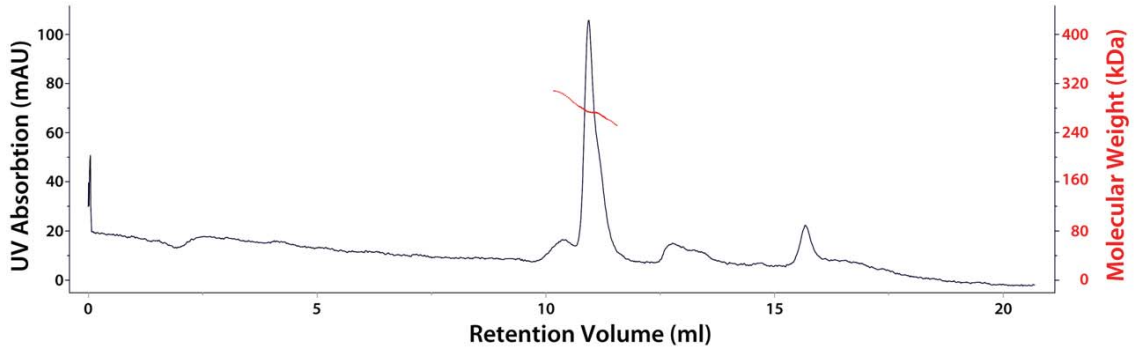


Figure 3.7 - Evaluation of the oligomeric state of the VPS33B-VIPAR complex. SEC-MALLS elution profile of the VPS33B-VIPAR complex using a Superdex 200 10/300 analytical column. The complex elutes as a single peak with a molecular weight of 280 kDa.

3.4.3 - VPS33B and VIPAR interact in the cell with Fluorescence Resonance Energy Transfer (FRET)

In-cell analysis of VPS33B-VIPAR interaction was performed using confocal microscopy with Fluorescence Lifetime Imaging in Förster Resonance Energy Transfer (FRET-FLIM). In FRET-FLIM, the occurrence of FRET can be imaged using fluorescence lifetime imaging microscopy (FLIM). For this purpose, HEK293 cells were transfected with the fluorescent FRET couple GFP-mCherry. FRET occurs when the lifetime of the GFP (donor) is reduced in the presence of an acceptor (mCherry) sample. The fluorescence lifetime of the donor was measured in a sample containing the donor only (first panel in figure 3.8) and compared it to samples where both donor and acceptor are present (figure 3.8). mCherry-tagged VPS33B and GFP-tagged VIPAR were used (fourth panel in figure 3.8). In the negative controls the donor GFP-VIPAR was coupled first to untagged-VPS33B (second panel in figure 3.8) and secondly to the empty mCherry vector only (third panel in figure 3.8). In panels in figure 3.8 the lifetime is shown as different colours. The more purple the image, the shorter the

lifetime, indicating the presence of FRET. The average fluorescence lifetime is calculated for each sample and is represented in the histograms in figure. 3.8. The raw data from the FRET analysis are reported in Appendix 3.

As mentioned previously, a reduction in the lifetime of GFP is associated to FRET and figure 3.8 indicates that the lifetime of GFP-VIPAR is reduced only when mCherry-VPS33B is present, and not in any of the negative controls (figure 3.8). For a positive FRET result the 2 fluorophores must be between 1 and 10 nm apart, matching the scale at which protein-protein interactions take place. This result unequivocally demonstrates that VPS33B and VIPAR interact in the cell.

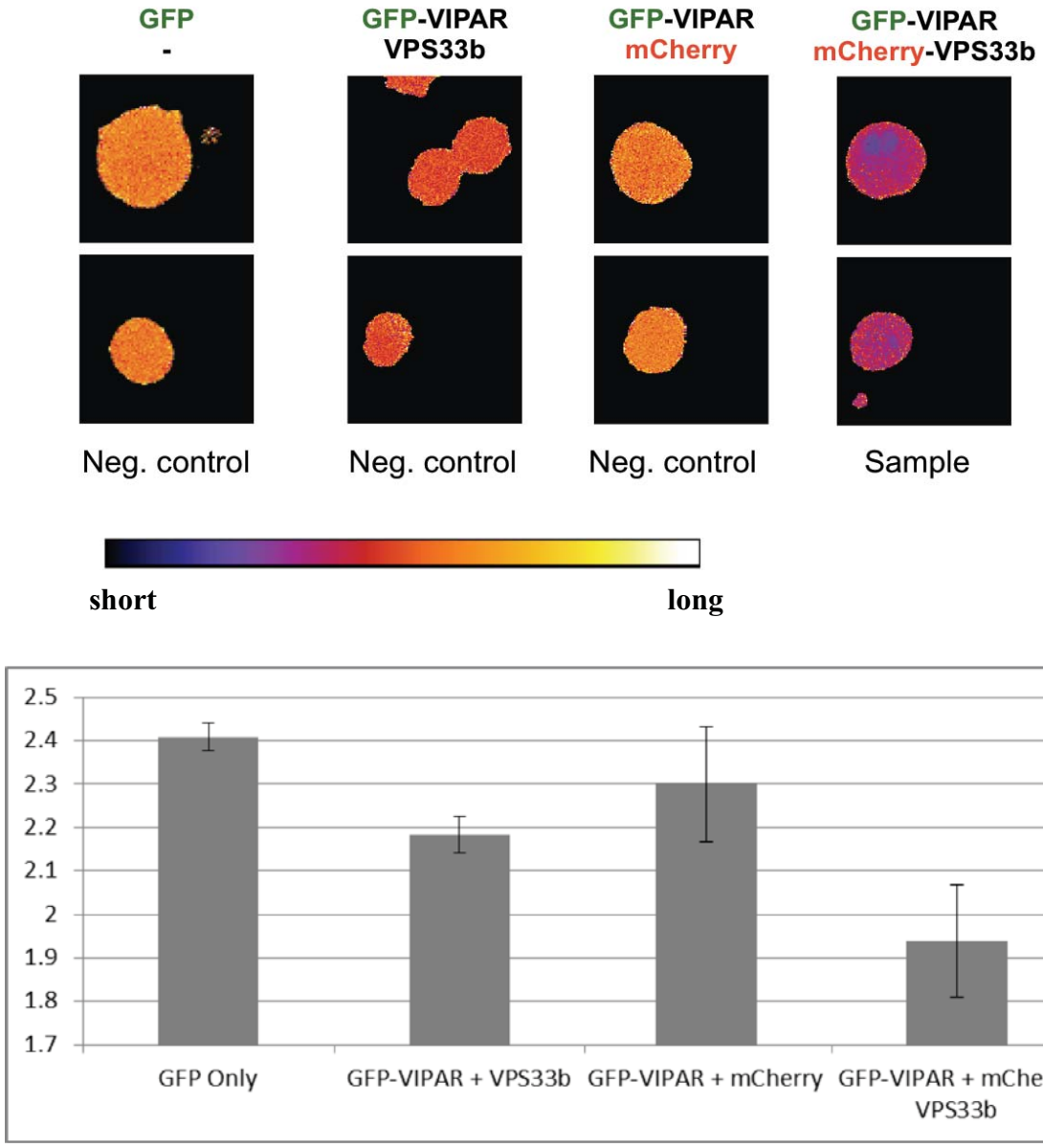


Figure 3.8 - FRET-FLIM of VPS33B-VIPAR interaction. Images are false colour lifetime images; the graphs are histograms of the lifetime in the images. Imaging experiment performed using a confocal microscope coupled to a FLIM (fluorescence life time imaging) detector. The distribution of the mean fluorescence lifetime of the GFP donor is displayed using a continuous pseudocolour scale ranging from 'short' (towards purple) to 'long' (towards orange). For each dataset n>5. Significant FRET values (P<0.01) were observed when comparing GFP-VIPAR / mCherry couple with the GFP-VIPAR / mCherry-VPS33B (P-value = 0.003254, N=3). Experiment performed in collaboration with Federico Forneris.

3.4.4 - VPS33B interacts with the globular C-terminal domain of VIPAR

Two main regions were identified in VIPAR from exploring its protein sequence for globularity and disorder (section 3.1), one disordered region at the N-terminus (N) and a globular structured region at the C-terminus (C) (figure 3.1). Unlike in the C-region, no mutations were identified in the genomic sequence of *VIPAR* that are located in the disordered N-region. Therefore I investigated which of these regions of VIPAR are involved in the interaction with VPS33B. For this purpose two constructs of VIPAR were used (N and C) including respectively the N- and C-terminal part of the protein, and co-immunoprecipiation experiments were performed with VPS33B. Furthermore, as mentioned in section 3.1, a nuclear localisation sequence (NLS) is predicted at the interface between the disordered and the globular region of VIPAR (section 3.6, figure 3.13). I investigated if this sequence is crucial or not for the interaction with VPS33B by using two constructs for the C-region of VIPAR, C- and C+, that respectively lack or include the NLS. The N-construct used in this study does not contain the NLS (N-). The co-immunoprecipitation experiment revealed that VPS33B interacts with the C-region of VIPAR only. No interaction was detected with the N-region (figure 3.9). This finding might explain the lack of known ARC mutations involving the N-terminal of VIPAR. Mutations in this region could be dispensable for VIPAR interaction with VPS33B. Another possibility is that mutations in the N-region of VIPAR are embryonically lethal because they abolish the interaction with other potential binding partners that are crucial for VIPAR's protein function. The hypothesis that the N-region of VIPAR might be involved in the binding of other proteins is supported by the evidence that disordered regions require the binding of structured partners *in-vivo* to undergo a disorder-to-order

transition. Consequently I hypothesised that also the N-region of VIPAR, not involved in the interaction with VPS33B, binds to other interacting proteins *in-vivo* for its stability.

From figure 3.9 I concluded that C+ and C- constructs bind with similar affinity to VPS33B indicating that NLS is not crucial for this interaction (figure 3.9).

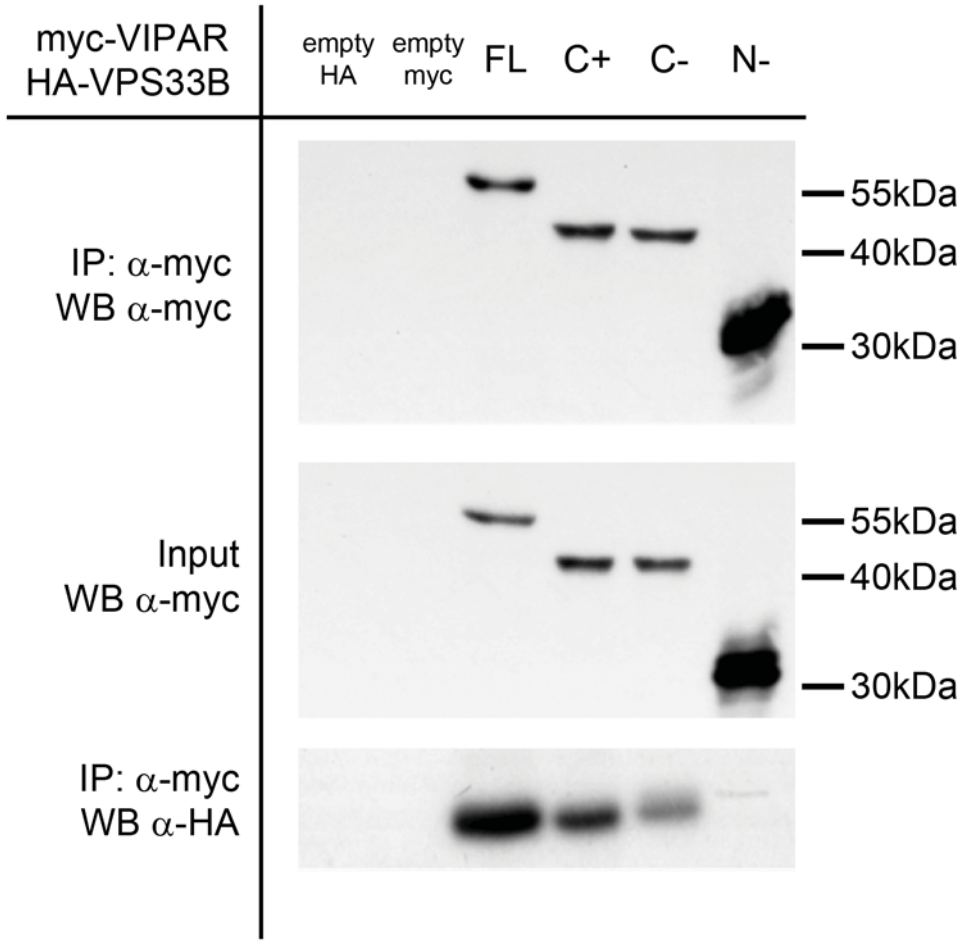


Figure 3.9 - Co-immunoprecipitation of myc-VIPAR constructs with HA-VPS33B. HEK293 cells were co-transfected with HA-empty or HA-VPS33B and myc tagged VIPAR full-length (FL), N-terminal construct of VIPAR without NLS (N-), C-terminal construct of VIPAR without NLS (C-), C-terminal construct of VIPAR with NLS (C+) and myc-empty vector. Immunprecipitation was performed with anti-myc antibody and it reveals a strong interaction between VPS33B and FL, C- and C+ constructs of VIPAR. No interaction with VPS33B was detected for the N- construct. None of these proteins were pulled down with HA-empty or myc-empty.

3.5 - DISCUSSION

Germline mutations in VPS33B or VIPAR have been identified in most patients with ARC syndrome. They share the same clinical features and this suggests that VPS33B and VIPAR are involved in the same intracellular pathway in higher organisms (Cullinane et al., 2010). Published data have confirmed VPS33B and VIPAR association by co-localising and co-immunoprecipitating the proteins (Cullinane et al., 2010). So far there has been no incontrovertible evidence regarding either the direct or indirect interaction between the proteins. I performed pull-down experiments and demonstrated that VPS33B directly interacts with VIPAR *in-vitro*. This method however, despite giving clear and unambiguous results, lacks spatial and physiological information since the proteins expressed are not in their natural environment. Positive results in FRET-FLIM experiments excluded the possibility of *in-vitro* artefacts associated with the pull-down system.

Through biochemical approaches including gel-filtration and SEC-MALLS I unequivocally demonstrated that VPS33B and VIPAR form a stable binary complex *in-vitro* with a molecular weight of 280 kDa that might correspond to the dimer of VPS33B-VIPAR.

These data are consistent with the suggestion by Bröcker et al, 2012, who published the EM structure of the HOPS complex in yeast where VPS33B and VIPAR yeast homologues Vps33 and Vps16 were shown to form a tight sub-complex of HOPS. As mentioned previously, unlike their yeast counterparts, VPS33B and VIPAR are unlikely to form a part of the HOPS complex in physiological conditions (Smith et al., 2012). It was therefore hypothesised that in higher organisms the VPS33B-VIPAR binary sub-complex has acquired new functions separate from HOPS, by interacting with binding

partners involved in different intracellular pathways. Consequently our next goal was to identify specific intracellular pathways where VPS33B/VIPAR are involved by identifying new interacting proteins of the complex. This will allow elucidation of the role of VPS33B/VIPAR in the physiological and pathological conditions.

Another evidence of the possible existence of other VPS33B/VIPAR interactors comes from protein predictions of globularity and disorder. This predicts a presence of a large disordered protein segment at the N-terminus of VIPAR and it is known that disordered proteins require binding to a structured partner *in-vivo* for their stability. Previously I showed that, unlike the C-terminal, the N-terminal region of VIPAR does not co-immunoprecipiate with VPS33B and therefore is dispensable for this interaction. Therefore is possible to conclude that other interactor proteins are likely to be involved in the binding to the N-region of VIPAR *in-vivo* and this hypothesis will be further discussed in chapter 4.

3.6 - Supplementary Bioinformatics analysis of VPS33B and VIPAR proteins

The crystal structure of VPS33B and VIPAR is not yet known. Therefore a preliminary bioinformatics analysis was performed to better characterise the proteins subject of this study.

Disordered regions in proteins often lead to difficulties in protein expression, purification and crystallisation. Several types of software for prediction of intrinsically unstructured/disordered proteins were used including "GlobPlot" (http://globplot.embl.de/) "FoldIndex©" (http://bip.weizmann.ac.il/fldbin/findex) and "IUPred" (http://iupred.enzim.hu/). All analyses confirmed that the largest predicted disordered region is located within the N-terminal portion of VIPAR (figure 3.10 A and B). This region is likely to lack any recognisable secondary structure elements, (α-helices, β-strands). This analysis suggests that the large N-terminal portion of VIPAR, as well as some small regions located along the whole sequence of VPS33B (figure 3.11 A and B), are potentially aggregation-susceptible in the absence of interacting partners.

Furthermore, a bioinformatics analysis specifically searching for "coiled-coil" regions in both proteins was performed using the available on-line software "Coils" (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_lupas.html). Proteins with coiled-coil regions are often insoluble because such regions might be responsible for aggregation in absence of interacting partners. VPS33B and in particular VIPAR are predicted to contain coiled-coil regions (figure 3.10 C and 3.11 C) that renders the expression and purification of those proteins more challenging.

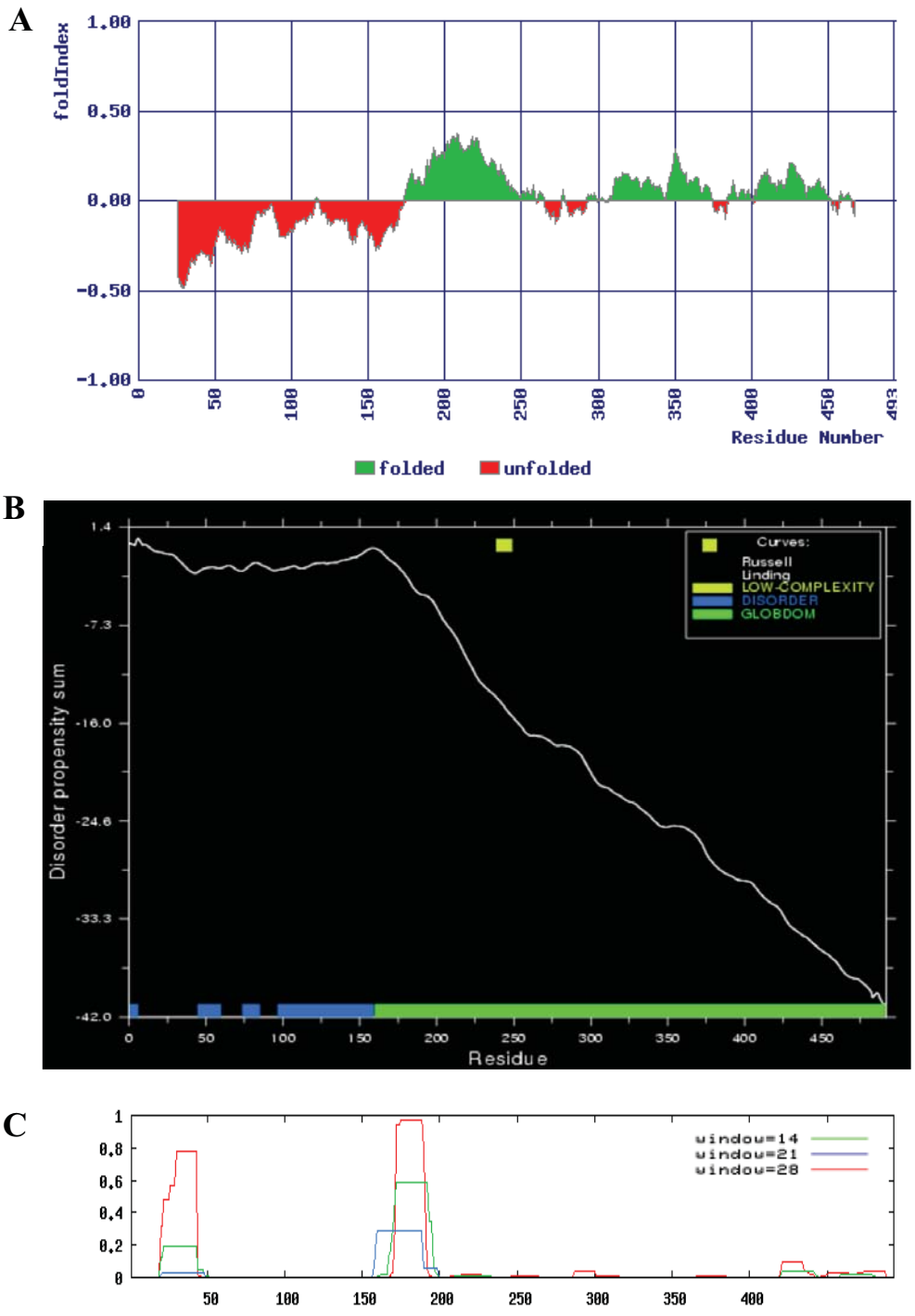


Figure 3.10 - Prediction of secondary structures and disordered regions for VIPAR. ***(A)*** *FoldIndex© for VIPAR plotted with window size 51 (scanning windows of 51 aminoacids). The software gives an estimation of the probability of protein folding along the sequence* ***(B)*** *Prediction of the intrinsic protein disorder, domain and globularity for VIPAR using GlobPlot. "Up-hills" on the graph are equivalent to disorder regions (blue segments) and "down-hills" regions correspond to putative globular domains (green segments).* ***C)*** *Prediction of coiled coil regions along VIPAR sequence obtained in scanning windows of 14 (green line), 21 (blue line) and 28 (red line) residues with Coils software.*

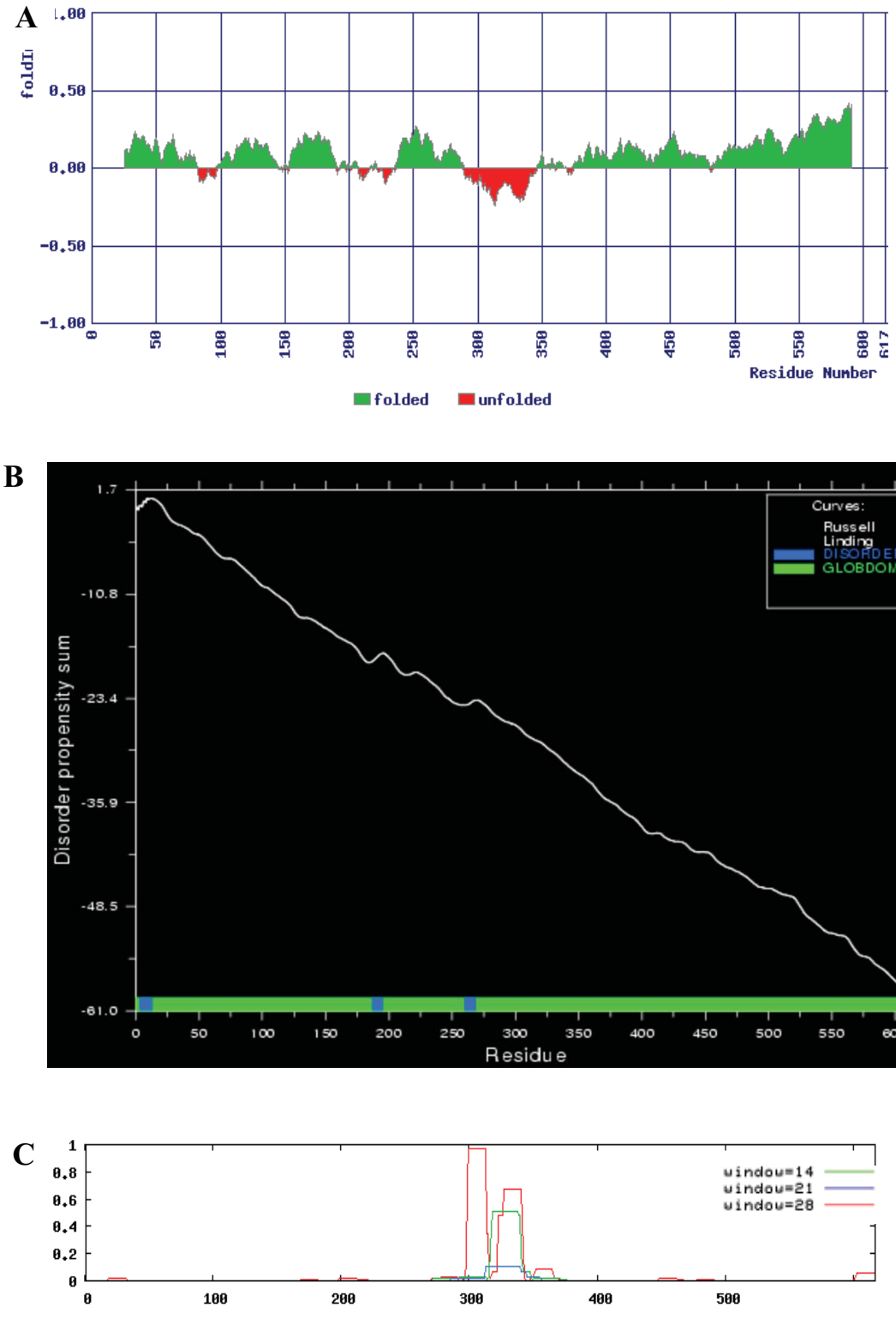


Figure 3.11 - Prediction of secondary structures and disorder regions for VPS33B. ***(A)*** *FoldIndex© for VPS33B plotted with window size 51.* ***(B)*** *Prediction of the intrinsic protein disorder, domain and globularity for VPS33B using GlobPlot.* ***(C)*** *Prediction of coiled coil regions along VPS33B sequence obtained in scanning windows of 14 (green line), 21 (blue line) and 28 (red line) residues with Coils software.*

Another important parameter to take into consideration during protein expression in a different system from the one where the protein is normally expressed, is the codon usage (it refers to differences in the frequency of occurrence of synonymous codons in coding DNA). Low protein expression or formation of insoluble aggregates may be attributed to differences in the codon usage between the target protein and the expression host. The available on-line software "the rare codon calculator" (http://www.codons.org/) was used to calculate the %Max (percentage of common codons) and the % Min (percentage of rare codons) for each amino acid. In figure 3.12 blue bars correspond to clusters of common codons while red bars indicate clusters of rare codons. As evident from figure 3.12, when the "*E. coli* codon bias" was used, it was found that both VPS33B and VIPAR contain large rare codon clusters for *E.coli* along their sequence (figure 3.12 A and B). However the predicted number of clusters of rare codons was significantly reduced for VPS33B when the "*Homo sapiens* codon bias" was used (figure 3.12 C), while some rare codon clusters were still predicted along VIPAR transcript.

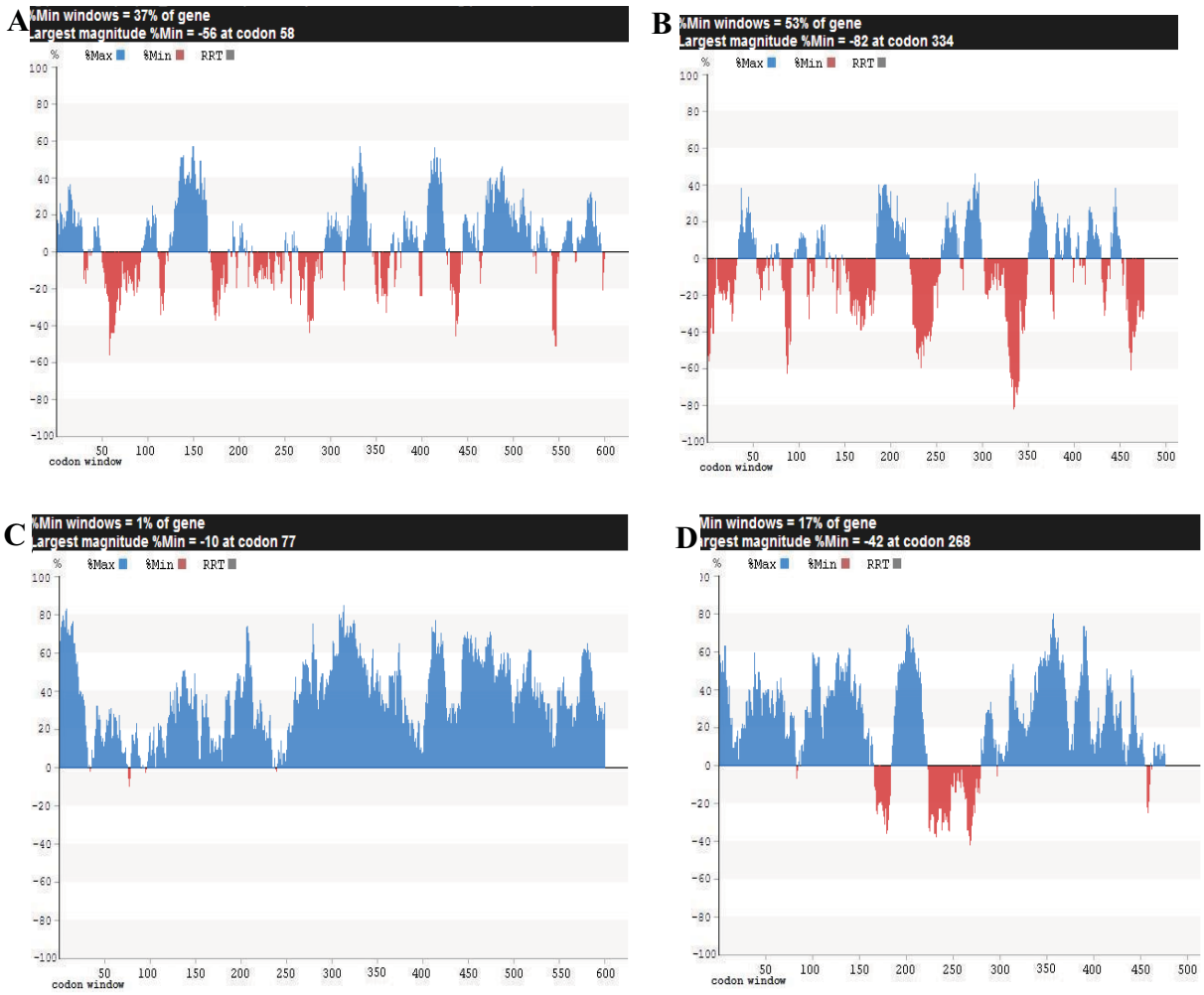
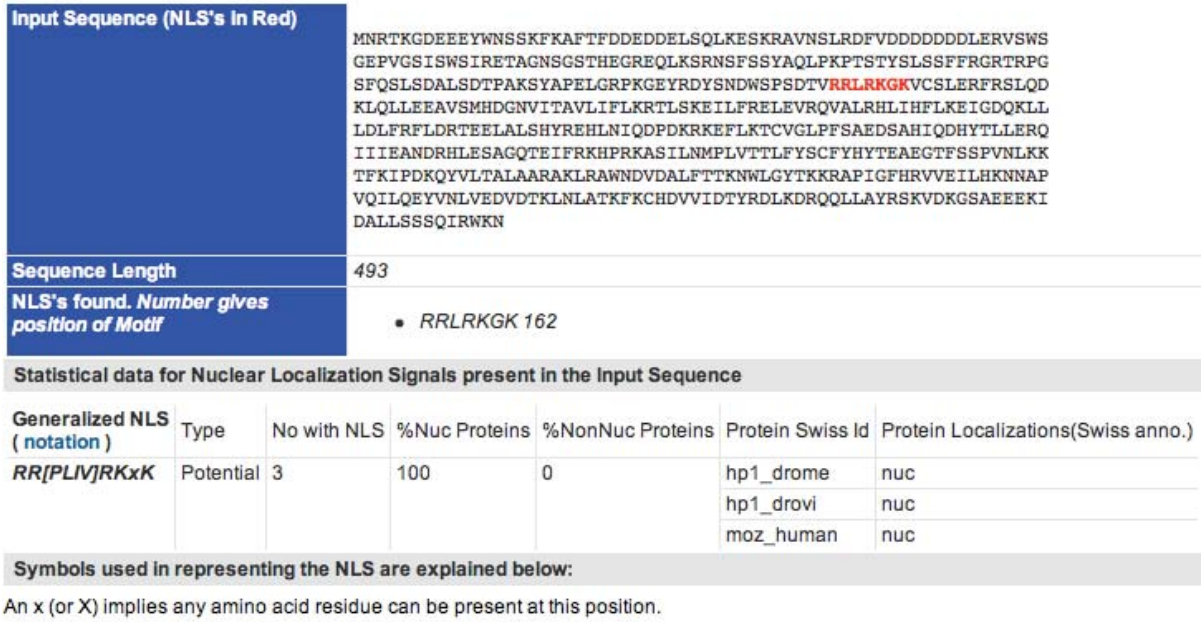


Figure 3.12 - Codon clustering calculation using the rare codon calculator software for VPS33B and VIPAR genes expressed in E. coli (A and B panel respectively) and Homo sapiens (C and D). Blue %Max bars correspond to clusters of common codons; red %Min (negative) bars correspond to clusters of rare codons.



Input Sequence (NLS's in Red)	MNRTKGDEEEYWNSSKFKAFTFDDEDDELSQLKESKRAVNSLRDFVDDDDDDDLERVSWS GEPVGSISWSIRETAGNSGSTHEGREQLKSRNSFSSYAQLPKPTSTYSLSSFFRGRTRPG SFQSLSDALSDTPAKSYAPELGRPKGEYRDYSNDWSPSDTV**RRLRKGK**VCSLERFRSLQD KLQLLEEAVSMHDGNVITAVLIFLKRTLSKEILFRELEVRQVALRHLIHFLKEIGDQKLL LDLFRFLDRTEELALSHYREHLNIQDPDKRKEFLKTCVGLPFSAEDSAHIQDHYTLLERQ IIIEANDRHLESAGQTEIFRKHPRKASILNMPLVTTLFYSCFYHYTEAEGTFSSPVNLKK TFKIPDKQYVLTALAARAKLRAWNDVDALFTTKNWLGYTKKRAPIGFHRVVEILHKNNAP VQILQEYVNLVEDVDTKLNLATKFKCHDVVIDTYRDLKDRQQLLAYRSKVDKGSAEEEKI DALLSSSQIRWKN
Sequence Length	*493*
NLS's found. *Number gives position of Motif*	• *RRLRKGK 162*

Statistical data for Nuclear Localization Signals present in the Input Sequence

Generalized NLS (notation)	Type	No with NLS	%Nuc Proteins	%NonNuc Proteins	Protein Swiss Id	Protein Localizations(Swiss anno.)
RR[PLIV]RKxK	Potential	3	100	0	hp1_drome	nuc
					hp1_drovi	nuc
					moz_human	nuc

Symbols used in representing the NLS are explained below:

An x (or X) implies any amino acid residue can be present at this position.

Figure 3.13 - Results of Nuclear Localisation Signal Prediction (NLS) in VIPAR using PredictProtein software (http://www.predictprotein.org/). NLS's found in the query sequence are highlighted in red in the output report. The identified NLS motif is found in 3 families (protein Swiss Id) with known nuclear localisation (nuc).

CHAPTER FOUR:

PLOD3 is a novel interactor for VPS33B/VIPAR

4.1 - INTRODUCTION AND OVERVIEW

Most biological processes are regulated by multiprotein complexes whose characterisation, through the identification of specific protein-protein interactions, is a key goal in cell biology.

As mentioned in the previous chapter, the main goal of this work is to understand the role of VPS33B and VIPAR in the cell as well as the underlying intracellular mechanisms involved. This can be achieved by identifying VPS33B and VIPAR interacting proteins which are part of possible known intracellular pathways and understanding how these interactions impact the cell model phenotype for ARC. To screen for new protein-protein interactions involving VPS33B and VIPAR, a pull-down experiment was performed by using the previously purified VPS33B-VIPAR complex as a 'bait' to 'fish' for novel interacting proteins. This technique was coupled to mass spectrometry (MS) for the identification of endogenous proteins bound to the VPS33B/VIPAR protein complex. MS was performed at the UCL Mass Spectrometry core facility. The specific type of MS employed Liquid chromatography electrospray ionisation tandem mass spectrometry (LC-ESI-MS-MS) where the nanoflow liquid chromatography is coupled with electrospray ionization to analyse peptides in complex mixtures obtained from an enzymatic digest, such as trypsin digestion (section 4.3.2). Positive results in MS are strictly dependent on the amount of protein in the original sample, therefore coupling this technique to the pull-down, performed with large-scale over-expressed proteins was crucial for the success of the whole experimental procedure. This work led to the identification of Lysyl Hydroxylase 3 (LH3), also named Procollagen-Lysine 2-Oxoglutarate 5-Dioxygenase 3 (PLOD3) as a potential binding partner for the VPS33B/VIPAR protein complex. In the next chapters the work

to establish the validity of this interaction and its physiological and pathological relevance is described.

PLOD3 is one of the lysyl hydroxylase (LH) enzymes that catalyse the hydroxylation of lysyl residues in collagen-like peptides. It belongs to the 2OG-Fe(II) oxygenase superfamily that catalyse a wide range of reactions including hydroxylations, desaturations and oxidative ring closures. They all require ferrous ion, 2-oxoglutarate, molecular oxygen and ascorbate for their activity. In the lysyl hydroxylase reaction one atom of oxygen is incorporated into the Lys residue and another one into 2-oxoglutarate, which is decarboxylated to form succinate and CO_2 (Turpeenniemi-Hujanen et al., 1980) (figure 4.1). Ascorbate is needed to regenerate the enzyme after several reaction cycles.

In vertebrates there are three lysyl hydroxylase isoforms PLOD1, PLOD2 and PLOD3 while only one (PLOD) is present in invertebrates. *D. melanogaster* dPLOD is homologue to PLOD3, evolutionarily the oldest of the lysyl hydroxylases isoforms (Ruotsalainen et al., 1999; Salo et al., 2008). All isoforms are capable of LH activity but PLOD3 (and its *D. melanogaster* homologue dPLOD) has additional galactosyltransferase (GT) activity on hydroxylated lysines and glucosylgalactosyltransferase (GGT) activity on galactosylhydorxylated lysines (figure 4.2) (Hudson et al., 1993; Rautavuoma et al., 2002; Ruotsalainen et al., 2006; Wang et al., 2002b). The aforementioned post-translational modifications occur during collagen biosynthesis (section 1.5.3.2) and are critical for the stability of intermolecular crosslinking of collagens and proteins with collagenous domains (Kivirikko and Pihlajaniemi, 1998; Sricholpech et al., 2012). The LH and GGT activity have been mapped to specific regions of PLOD3 protein sequence by using truncated forms that

progressively lose LH or GGT activity (Heikkinen et al., 2000; Wang et al., 2002b). These studies located the LH activity in the carboxy-terminal portion of PLOD3 (Myllyla et al., 2007), the most conserved region among the lysyl hydroxylases isoforms, and the GT and GGT activity in the amino-terminal part (Myllyla et al., 2007; Wang et al., 2002a; Wang et al., 2002b). Thus, defects in the activity of one catalytic domain does not necessarily affect the activity of the other. For example, missense mutations affecting the LH activity of PLOD3 do not affect the GGT activity (Wang et al., 2002b).

PLOD3 activity is crucial for the cell secretion of type-IV collagen, the main collagen of the basement membranes in humans, mice, zebrafish and *Caenorhabditis elegans*.

In this chapter, work to confirm the MS results is described. Co-immunoprecipitation was used to validate *in-vitro* interactions while co-localisation with confocal microscopy, as well as Fluorescence Lifetime Imaging in Förster Resonance Energy Transfer (FRET-FLIM) were used for *in-cell* confirmation of protein-protein interactions. Additionally, experiments to show whether endogenous PLOD3 co-localises with VIPAR and VPS33B were performed. These eliminate potential false positive results derived from over-expression artefacts. Another purpose for using FRET-FLIM technique was to estimate intermolecular distances on an Angström scale separating the proteins between them. From this data it is also possible to glean information regarding the type of interaction between the proteins in this study. Indeed the shorter the distance the more likely it is that a physical direct interaction is occurring between the proteins.

Section 3.4.4 provides evidence that the N-terminal domain of VIPAR, which was previously shown to be disordered (section 3.1), is not involved in VPS33B interaction.

Thus this domain could to participate in the interaction with other proteins, which would facilitate its transition from "disordered" to "ordered". In this chapter the experiments to determine the possibility that the N-terminus of VIPAR is involved in the binding of PLOD3 are described.

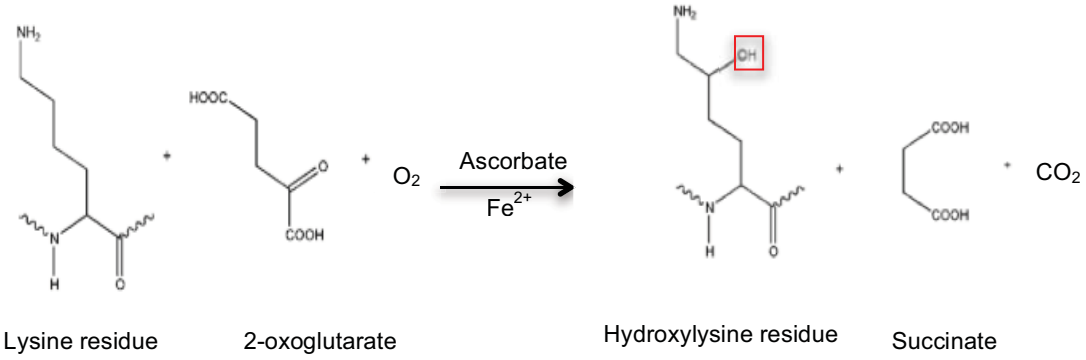


Figure 4.1 - Hydroxylation reaction catalyzed by lysyl hydroxylases. Fe^{2+}, 2-oxoglutarate, O_2 and ascorbate is required in the lysyl hydroxylase reaction where one atom of oxygen is incorporated into the hydroxyl group of the lysyl residue in the peptide substrate and another one into 2-oxoglutarate, which is decarboxylated to form succinate and liberates CO_2.

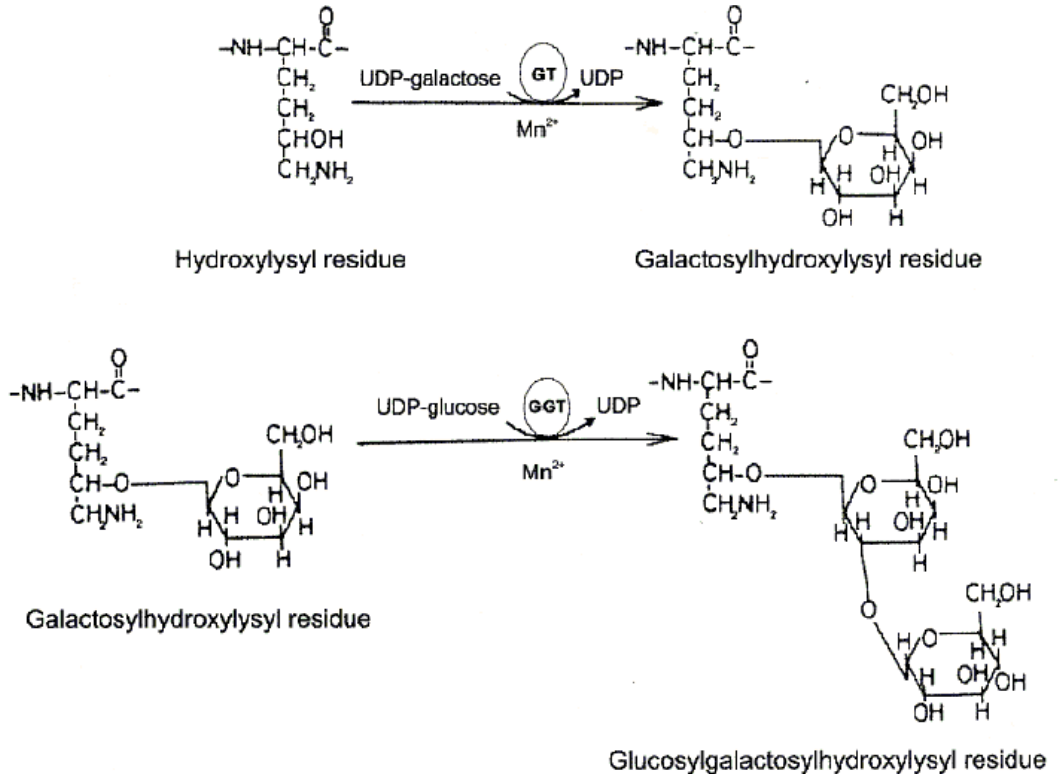


Figure 4.2 - Galactosylation and glycosylation reactions catalysed by peptidyl GT and GGT (Modified from (Kivirikko and Myllyla, 1983))

4.2 - AIMS

The first aim of the work described in this chapter was to identify new interacting proteins for VPS33B and VIPAR in order to further investigate the role of VPS33B-VIPAR complex in the cell.

The second aim was to validate the novel interaction experimentally and provide a hypothesis for the biological importance of this interaction.

4.3 - METHODS

4.3.1 - Strep-tactin pull-down assay

The pull-down assay was performed using the VPS33B/VIPAR purified complex as 'bait' for the identification of new interactors (figure 4.3). VPS33B/VIPAR was purified as in section 3.3.3 where the complex was captured on a StrepTrap FF 1 ml column taking advantage of the Strep tag on recombinant VIPAR that specifically binds the StrepTrap FF. This system generates a secondary affinity support for the binding of other proteins that specifically interact with the bait protein once this is incubated with a cell lysate that contains putative interactor proteins (prey) figure 4.3.

After the VPS33B/VIPAR complex was immobilised on the column (section 3.3.3), HEK293-E lysate was passed through the column over night with a flow-rate of 0.1 mL/min. The column was then washed again with 20 mL of wash buffer (25 mM HEPES/NaOH pH 8, 250 mM NaCl, 5 mM DTT) to remove the non-specifically bound contaminants, before the strep-tagged proteins were eluted with a wash buffer + 2.5 mM desthiobiotin (figure 4.3). The collected elution fractions containing both the 'bait' and putative 'prey' proteins were pooled and concentrated using a Vivaspin 15 concentrator MWCO 30 kDa (Sartorius Mechatronics) before being analysed by mass spectrometry for the determination of the 'prey' identity (figure 4.3). This step allowed a further enrichment in the 'prey' proteins that facilitated its MS identification.

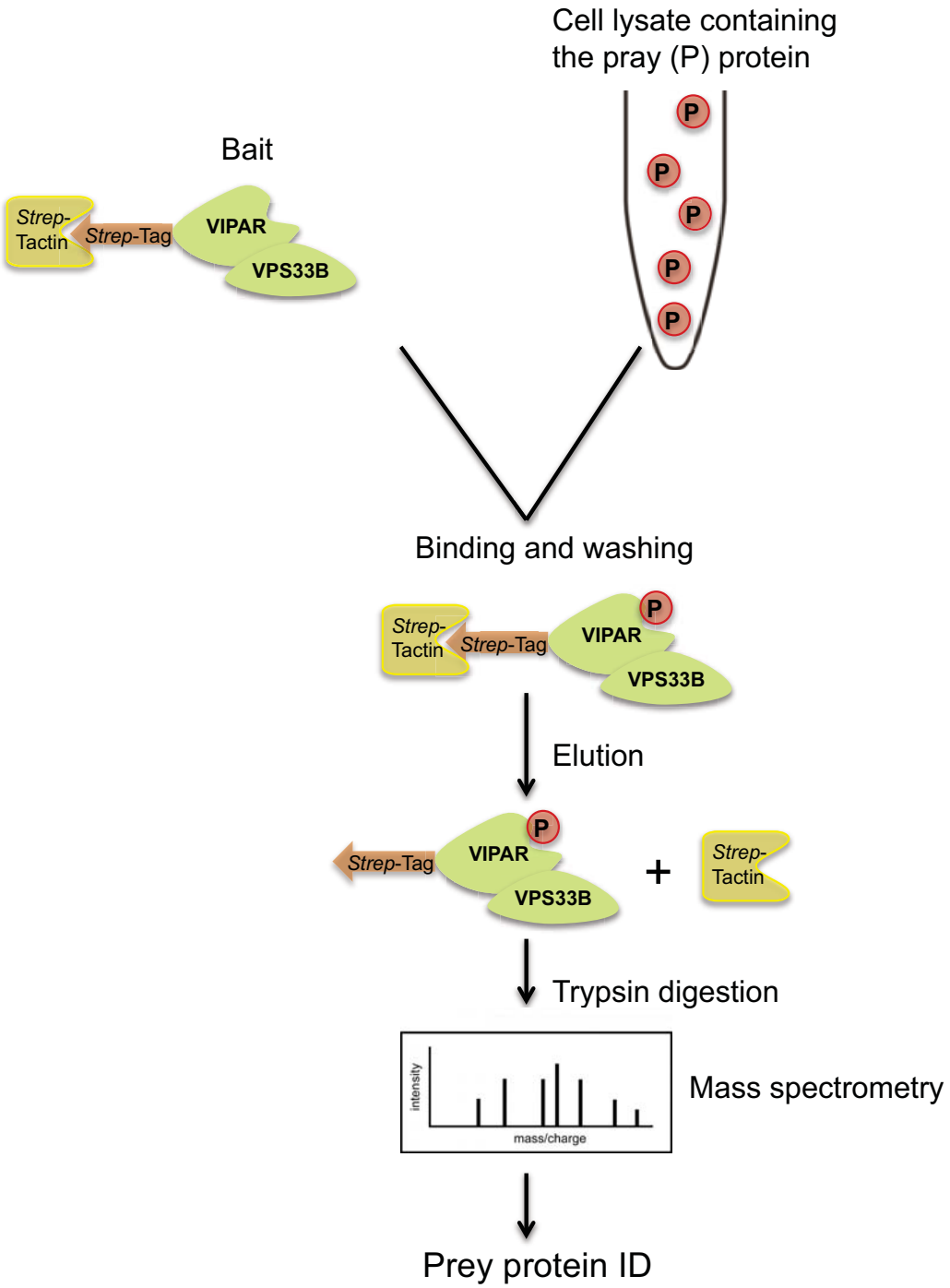


Figure 4.3 - Schematic representation of the strep-tactin pull-down assay used as a screening tool for the identification of new VPS33B/VIPAR interactors. The affinity system consists of strep-tagged VIPAR (in complex with VPS33B) immobilised on a strep-tactin affinity column. This protein complex was used as a "bait" to capture putative binding partners (prey) once the immobilised bait protein was incubated with HEK293-E cell lysate. After washing, the bound protein complexes were selectively eluted and the elution fraction containing both the bait and putative 'prey' proteins were analysed by mass spectrometry

4.3.2 - Liquid chromatography electrospray ionisation tandem mass spectrometry ESI-LC-MS/MS

Mass spectrometry on the pull-down elution was performed in collaboration with Wendy Heywood (Bennett et al., 2010; Bennett et al., 2012).

Proteins in the concentrated elution pool were firstly precipitated and then denatured by re-suspending in 40 μL of 6 M urea containing buffer (100 mM Tris, pH 7.8). To reduce disulphide bridges on proteins, 3 μL of 100 mM Tris–HCl buffer, pH 7.8, containing 5 M DTE (Dithiothreitol) were added and incubated at room temperature for 60 min. Free thiol groups were carboamidomethylated by incubating with 6 μL of 100 mM Tris–HCl, pH 7.8, containing 5 M iodoacetamide for 30 min at room temperature. This mixture was then diluted with H_2O to a final volume of 400 μL and vortexed, before 1 μg of sequence grade trypsin was added to the solution. Trypsin digestion was carried overnight at 37 °C in water bath. This trypsin digest was analysed by liquid chromatography electrospray ionisation tandem mass spectrometry (ESI-LC-MS/MS).

ESI-LC-MS/MS couple liquid chromatography to electrospray ionization for the analysis of single fragments resulted from the trypsin digestion. This allows the measurement of two masses, hence the term tandem mass spectrometry or MS/MS: firstly the Nanoflow-HPLC chromatography separates peptides that are directly eluted into an electrospray ion source in the mass spectrometer (figure 4.4). This allows the protonation of single peptides whose charge and first mass (MS-1) is measured by the instrument (figure 4.4). The second mass is measured when single protonated fragments pass through a second cell where Argon is used to further fragment single peptides by collision-induced dissociation (CID). After their separation, the mass of single fragmented ions is recorded (MS-2) (figure 4.4).

The fragmentation pattern of a single peptide is based on its sequence and can therefore be predicted. Predicted fragment ions for all peptides are available on databases such as the Swiss Prot database and experimental data for unknown peptides can therefore be searched against those databases using ProteinLynx Global Server (PLGS version 2.2.5, Waters, Milford, MA, USA) in order to identify the protein from which the peptide was originated. The Protein Lynx Global Server (PLGS 2.2.5) software groups together several peptides derived from the digestion of the same protein and compares it to theoretical peptides derived from an 'electronic' tryptic digestion of the protein database and assigns a PLGS Score (table 3.1), calculated by using a Monte Carlo algorithm. This algorithm measures the accuracy for assignment. A higher score implies greater confidence of protein identity (Wright et al., 2009).

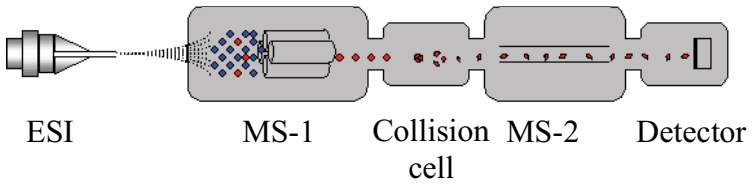


Figure 4.4 - Liquid chromatography electrospray ionisation tandem mass spectrometry ESI-LC-MS/MS illustration. Peptides separated by HPLC (blue and red dots) are eluted into an electrospray ion source (ESI) in the mass spectrometer where their first mass (MS-1) is measured. These protonated fragments pass through a second cell (collision cell) where Argon is used to further fragment peptides and their second mass is measured (MS-2). Modified from http://qbab.aber.ac.uk/roy/mss/qtof.htm

In order to determine the identity of the proteins contained in our pull-down elution concentrated fraction, a nanoAcquity HPLC and QTOF Premier mass spectrometer (Waters Corporation, Manchester, U.K.) was used (Bennett et al., 2010; Bennett et al., 2012). Peptides were trapped and desalted before a reverse phase separation was applied using a Symmetry C18 5 μm, 5 mm × 300 μm precolumn. Temperature in all columns

was constantly kept at 35 °C. The separation of peptides prior to mass spectrometry was performed using a 15 cm × 75 μm C18 reverse phase analytical column. Peptides were loaded onto the precolumn at a flow rate of 4 μL/min in 0.1% formic acid for 4 minutes and eluted before being separated on the analytical column using a gradient of 3–40% acetonitrile (0.1% formic acid) for 90 min and at a flow rate of 300 nL/min. The column was washed and regenerated at 300 nL/min for 10 min using a 99% acetonitrile (0.1%) rinse in order to remove all nonpolar and non-peptide material. The column was then re-equilibrated for 20 min at the initial starting conditions. Peptides were analysed in positive ion mode using a Q-Tof Premier mass spectrometer (Waters Corp., Manchester, U.K.) and in v-mode with a resolving power of 10 000 fwhm. The TOF analyser was calibrated with [glu1]-fibrinopeptide B fragments obtained using a collision energy of 25 V and over the mass range 50–2000 *m*/*z*. Post calibration of data files were corrected using the doubly charged precursor ion of [glu1]-fibrinopeptide B (785.8426 *m*/*z*) with a sampling frequency of 30 s. Accurate mass LC-MS data were collected in a data independent and alternating, low and high collision energy mode. Each low/high acquisition was 1.5 s with 0.1 s interscan delay. Low energy data collections were performed at constant collision energy of 4 V, high collision energy acquisitions were performed using a 15–40 V ramp over a 1.5 s time period and a complete low/high energy acquisition achieved every 3.2 s (Bennett et al., 2010; Bennett et al., 2012).

4.4 - RESULTS

4.4.1 - Mass spectrometry analysis of VPS33B-VIPAR binding proteins

To identify novel interactors for the VPS33B-VIPAR protein complex, a pull-down assay was set-up combined with a mass spectrometry analysis of the outcome elution fraction (section 4.3.1, 4.3.2). Mass-spectrometry was performed in collaboration with Wendy Heywood. Trypsin was used to digest the protein mixture into peptides that were then subjected to Liquid Chromatography Electrospray Ionisation Tandem Mass Spectrometry. ESI-LC-MS/MS outcome results were analysed by The Protein Lynx Global Server (section 4.3.2). The proteins identified by this analysis are reported in Table 3.1. ESI-LC-MS results confirmed the presence of the 'bait', VPS33B and VIPAR, in the pull-down elution mixture. H2B-1A and all keratin proteins were excluded as potential positive results since these represent the most common nonspecific contaminant in the field of protein mass spectrometry and were therefore removed from further analysis. Interestingly PLGS 2.2.5 analysis identified PLOD1 and 3 (Procollagen lysine 2 oxoglutarate 5 dioxygenase 1 and 3) as potential interactors for the VPS33B-VIPAR complex. As shown by this analysis, 58 theoretical peptide sequences (theoretical peptides) were identified from the Protein Lynx Global Server for PLOD3, amongst which 12 sequences showed homology to predicted protein fragments. The PLGS score for these proteins suggests reliable identification.

The experimental work described in this chapter is mainly aimed at confirming or excluding the possibility that PLOD3 or PLOD1 interact with VPS33B or VIPAR.

Accession	Entry	Description	mW (Da)	pI (pH)	PLGS Score	Peptides	Theoretical Peptides	Coverage (%)
H2B1A_HUMAN	Q96A08	Histone H2B type 1 A OS Homo sapiens GN HIST1H2BA PE 1 SV 3	14158	10.739	1936.133	6	12	11.811
VP33B_HUMAN	Q9H267	Vacuolar protein sorting associated protein 33B OS Homo sapiens GN VPS33B PE 1 SV 2	70540	6.2706	1872.545	68	54	33.7115
VIPAR_HUMAN	Q9H9C1	VPS33B interacting protein OS Homo sapiens GN VIPAR PE 1 SV 1	56970	7.0488	755.4543	46	47	31.4402
K1C9_HUMAN	P35527	Keratin type I cytoskeletal 9 OS Homo sapiens GN KRT9 PE 1 SV 3	62026	4.9574	441.1904	6	40	16.3724
K2C1_HUMAN	P04264	Keratin type II cytoskeletal 1 OS Homo sapiens GN KRT1 PE 1 SV 6	65998	8.2661	373.0901	25	4	17.0807
PLOD1_HUMAN	Q02809	Procollagen lysine 2 oxoglutarate 5 dioxygenase 1 OS Homo sapiens GN PLOD1 PE 1 SV 2	83497	6.4755	194.2717	8	59	15.4058
PLOD3_HUMAN	O60568	Procollagen lysine 2 oxoglutarate 5 dioxygenase 3 OS Homo sapiens GN PLOD3 PE 1 SV 1	84731	5.6232	162.6589	12	58	9.3496
K1C10_HUMAN	P13645	Keratin type I cytoskeletal 10 OS Homo sapiens GN KRT10 PE 1 SV 6	58791	4.9552	140.7099	15	35	16.0959

Table 4.1. Proteins identified by ESI-LC-MS in the VPS33B/VIPAR pull-down assay. Data analysed using Protein Lynx Global Server (PLGS 2.2.5). Indicated are the SwissProt accession number, entry and description, expected molecular weight (mW), expected isoelectric point (pI), PLGS score, number of peptides identified, number of calculated theoretical peptides, percentage of sequence coverage

4.4.2 - PLOD1 does not co-localise with VPS33B-VIPAR complex

Preliminary confocal microscopy experiments were performed to test the possibility that PLOD1 or PLOD3 co-localise with the VPS33B/VIPAR complex. HEK293 cells were used to over-express the proteins for this preliminary test. This cell line was used to remain consistent with the pull-down and mass spectrometry experiment. Over-expressed mCherry-PLOD1 shows a diffuse cytoplasmic distribution, most likely corresponding to ER (Salo et al., 2006a; Schneider and Granato, 2007) that does not co-localise with over-expressed YFP-VPS33B/myc-VIPAR clusters. In contrast, over-expressed mCherry-PLOD3 forms distributed cytoplasmic clusters, which co-localise

with over-expressed YFP-VPS33B/myc-VIPAR (figure 4.5). These findings suggest that PLOD1 might not constitute a true interactor for VPS33B/VIPAR but PLOD3 might be a novel interactor for this protein complex. This second hypothesis has been further investigated in this chapter.

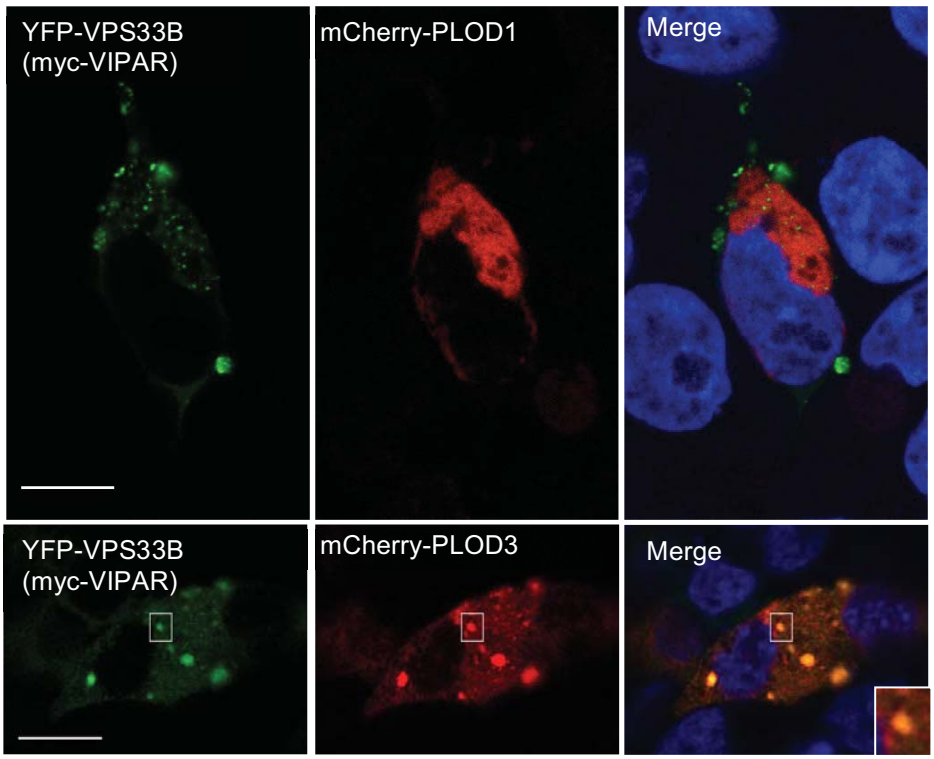


Figure 4.5 Co-localisation of VPS33B-VIPAR with PLOD1 and PLOD3. Confocal immunofluorescence of HEK293 cells co-transfected with myc-VIPAR (unstained), YFP-VPS33B (green) and mCherry-PLOD1 (red) show no co-localisation between the proteins (upper panel). HEK293 cells co-transfected with myc-VIPAR (unstained), YFP-VPS33B (green) and mCherry-PLOD3 (red) show co-localisation between the proteins in distinct clusters (lower panel). All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.

4.4.3 - Confirmation of the PLOD3-VIPAR interaction

4.4.3.1 - Reciprocal co-immunoprecipitation of PLOD3 and VIPAR

As PLOD1 is unlikely to be a real interactor of the VPS33B/VIPAR complex, this study was focussed on PLOD3. In order to further substantiate the evidence of an interaction between PLOD3 and VPS33B-VIPAR protein complex, the possibility that PLOD3 co-immunoprecipitates with VPS33B and/or VIPAR was examined. Endogenous PLOD3 co-immunopreicipitated with myc-VIPAR and VPS33B/VIPAR complex (figure 4.6). A very weak co-immunoprecipitation was detected with HA-VPS33B, while endogenous PLOD3 was not pulled-down with empty-HA or empty-myc. This result further confirms the pull-down and mass spectrometry experiment, as well as the preliminary co-localisation of the proteins by confocal microscopy, and suggests that PLOD3 interacts *in-vitro* with VIPAR. The weak binding to HA-VPS33B might be explained by the indirect binding of endogenous PLOD3 to endogenous VIPAR that in turn interacts with over-expressed HA-VPS33B.

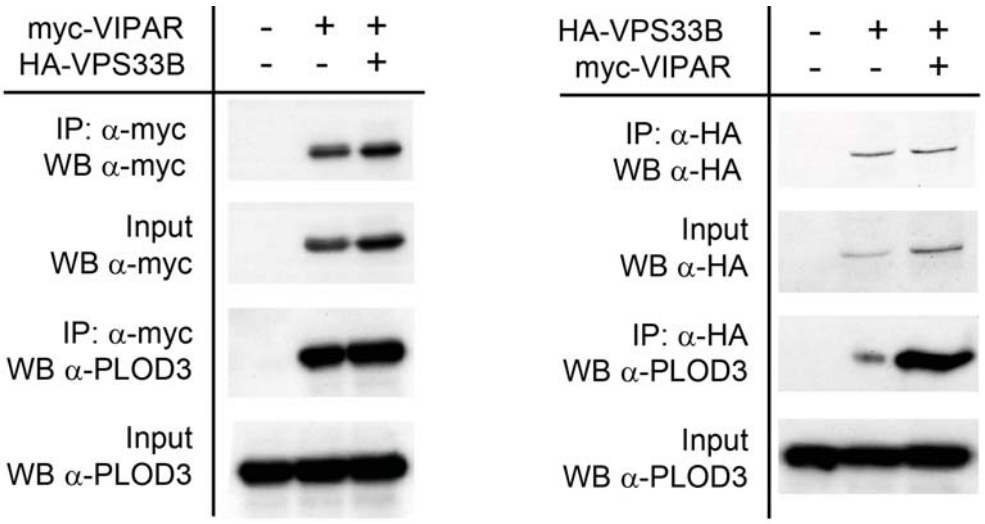


Figure 4.6 - Co-immunoprecipitation of myc-VIPAR with endogenous PLOD3. HA-VPS33B and myc-VIPAR were expressed in HEK293 cells separately and in complex before immunoprecipitation was performed with anti-HA or anti-myc antibody. One control was performed for each of the sets by transfecting with empty-myc and empty-HA vectors. The immunoprecipitates (bound fraction) and the flow-through (wash fraction) were separated by SDS-PAGE, followed by Western blotting using anti-PLOD3 antibody to detect the presence of endogenous PLOD3 in the fractions. The presence of HA-VPS33B and myc-VIPAR in the respective bound fractions was confirmed by western blot with anti-HA and anti-myc antibodies. Available antibodies to PLOD3 are poor at immunoprecipitating endogenous levels of the protein and therefore the reverse co-immunoprecipitation experiment was not successful.

4.4.3.2 - PLOD interacts with VPS16 in *D. melanogaster*

As mentioned previously (section 4.1) there are three lysyl hydroxylase isoforms in vertebrates and PLOD3 is evolutionarily the oldest amongst them, besides being the only isoform capable of both LH and GGT activity. In *D. melanogaster* and other invertebrates only one isoform is present (Plod) which is a homologue of PLOD3.

The novel interaction PLOD3-VIPAR has not been described previously. The only evidence in literature for this protein-protein interaction comes from the fly database: http://flybase.org/. From anti-tag co-immunoprecipitation and mass-fingerprinting data,

Plod was found to be physically associated with Vps16b (the fly homologue of VIPAR) (Guruharsha et al., 2011) (figure 4.7). This is a further indication that in higher organisms PLOD3 might be involved in the interaction with VIPAR.

protein-protein		
Interacting group	Assay	References
Vps16B - Plod	anti tag coimmunoprecipitation, peptide massfingerprinting	(Mintseris et al., 2009.11.23)

Figure 4.7 - Plod protein-protein interactions in D. melanogaster. http://flybase.org/ Anti-tag co-immunoprecipitation and massfingerprinting data show that Plod is physically associated with Vps16b

4.4.3.3 - Significant co-localisation between VPS33B/VIPAR complex and PLOD3

For evidence of PLOD3-VIPAR interaction within the cell, immunofluorescence experiments were performed with both over-expressed and endogenous proteins.

Co-localisation between proteins was quantified using Coloc_2 and JACoP (Just Another Colocalization Plugin). This are two different ImageJ / Fiji plugins that analyse the pixel intensity correlation over space methods of Pearson, Manders, Costes, Li and more. They perform scatterplots, analysis, automatic thresholding and statistical significance testing. Co-localisation is quantified by the Manders' M1 and M2 coefficients. M1 signifies the correlation of pixels from the first channel (green) overlapping the second (red) while M2 indicates the overlap coefficient of the second channel to the first. M1 and M2 values range between zero and one and values above 0.6, meaning 60% of pixel overlapping between two channels, are considered proof of

co-localisation. A threshold is set for both M1 and M2, which assumes that pixels below the threshold are not correlated.

Another value to consider when performing this type of co-localisation analysis is the P-value, which represents the percentage of random chance images that had better correlation than the real image. Indeed it is common that in images in which the signal is localised all over the cell in both channels there will be some random overlap that shouldn't be confused with true co-localisation with biological meaning. The P-value should be 1.00 meaning that none of the randomised images had better correlation than the real image and therefore the measured co-localisation in the real image is not likely to be better than random chance.

To test for PLOD3 co-localisation with VPS33B/VIPAR, HeLa cells were transfected with mCherry-PLOD3, YFP-VPS33B and myc-VIPAR (unstained). Over-expressed proteins co-localise (figure 4.8) as demonstrated by Mander's co-localisation coefficients M1 and M2 and P1 value (figure 4.8). To exclude the possibility that this finding was an over-expression artefact, the co-localisation experiments were repeated using a recently available antibody against endogenous VIPAR (Sigma) suitable for immunofluorescence experiments together with an anti- rabbit PLOD3 antibody against the endogenous protein. Zenon® Labeling kit was used to produce a fluorescently labelled anti-PLOD3 antibody that can therefore be used in combination with anti-VIPAR also produced in rabbit. Endogenous PLOD3 co-localises with endogenous VIPAR in the perinuclear region similar to the over-expressed proteins (figure 4.8). Because no anti-VPS33B antibody suitable for immunofluerescent experiments was commercially available, YFP-VPS33B was used to test co-localisation with endogenous

PLOD3. Also YFP-VPS33B, similar to endogenous VIPAR, co-localises with PLOD3 in the perinuclear area.

Investigations to determine the intracellular localisation of the interaction between VIPAR/VPS33B and PLOD3 will be described in chapter 5. The co-immunopreciptation and co-localisation experiments further validate the pull-down and mass spectrometry results that lead to the identification of PLOD3 as an interactor of VPS33B/VIPAR complex.

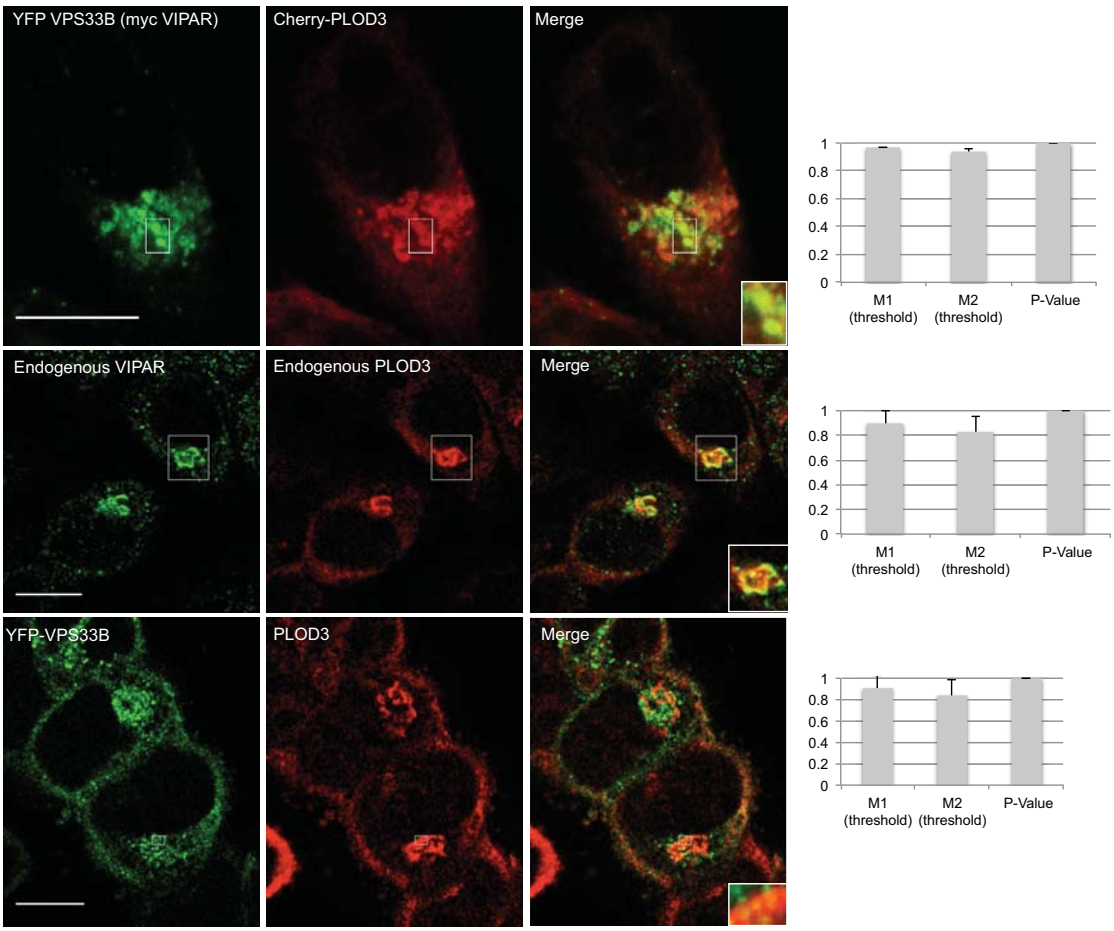


Figure 4.8 - Co-localisation of VPS33B-VIPAR with PLOD3. Confocal immunofluorescence of HeLa cells co-transfected with myc-VIPAR (unstained) YFP-VPS33B (green), and mCherry-PLOD3 (red) show co-localisation of the proteins. HeLa cells stained with anti-PLOD3 and anti-VIPAR antibodies show co-localisation of the endogenous proteins. Endogenous PLOD3 co-localises also with over-expressed YFP-VPS33B. For each co-localisation analysis of the Manders' overlap coefficients M1 and M2, as well as the P-value, are represented as the average of n>6 cells. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.

4.4.3.4 - Further evidence for VIPAR-PLOD3 intracellular interaction from FRET experiments

The same approach used for *in-cell* analysis of VPS33B-VIPAR interaction (section 3.4.3) was used also to test the newly identified PLOD3-VIPAR interaction. It consists of the Fluorescence Lifetime Imaging in Förster Resonance Energy Transfer (FRET-FLIM) technique that in addition to being a useful tool to test protein-protein interactions is also used to measure molecular distances separating proteins. This experiment provides the information necessary to deduce the type of interaction between the proteins i.e. direct or indirect. As mentioned previously (section 3.4.3) FRET occurs when the lifetime of the donor (GFP) is reduced by the presence of an acceptor (mCherry) in the sample, and this is imaged using fluorescence lifetime imaging microscopy (FLIM). The lifetime of GFP-VIPAR is significantly reduced only when mCherry-PLOD3 is present (fifth panel figure 4.9) as compared to the lifetime of GFP only (negative control, first panel figure 4.9) or GFP-VIPAR when coupled to untagged-VPS33B (negative control, second panel in figure 4.9) or empty mCherry vector (negative control, fourth panel in figure 4.9). The occurrence of FRET between GFP-VIPAR and mCherry-VPS33B described in section 3.4.3 was used as a positive control in this analysis (third panel figure 4.9). The average fluorescence lifetime is calculated for each sample and is represented in the histograms in figure. 4.9. The raw data from the FRET analysis are reported in Appendix 3.

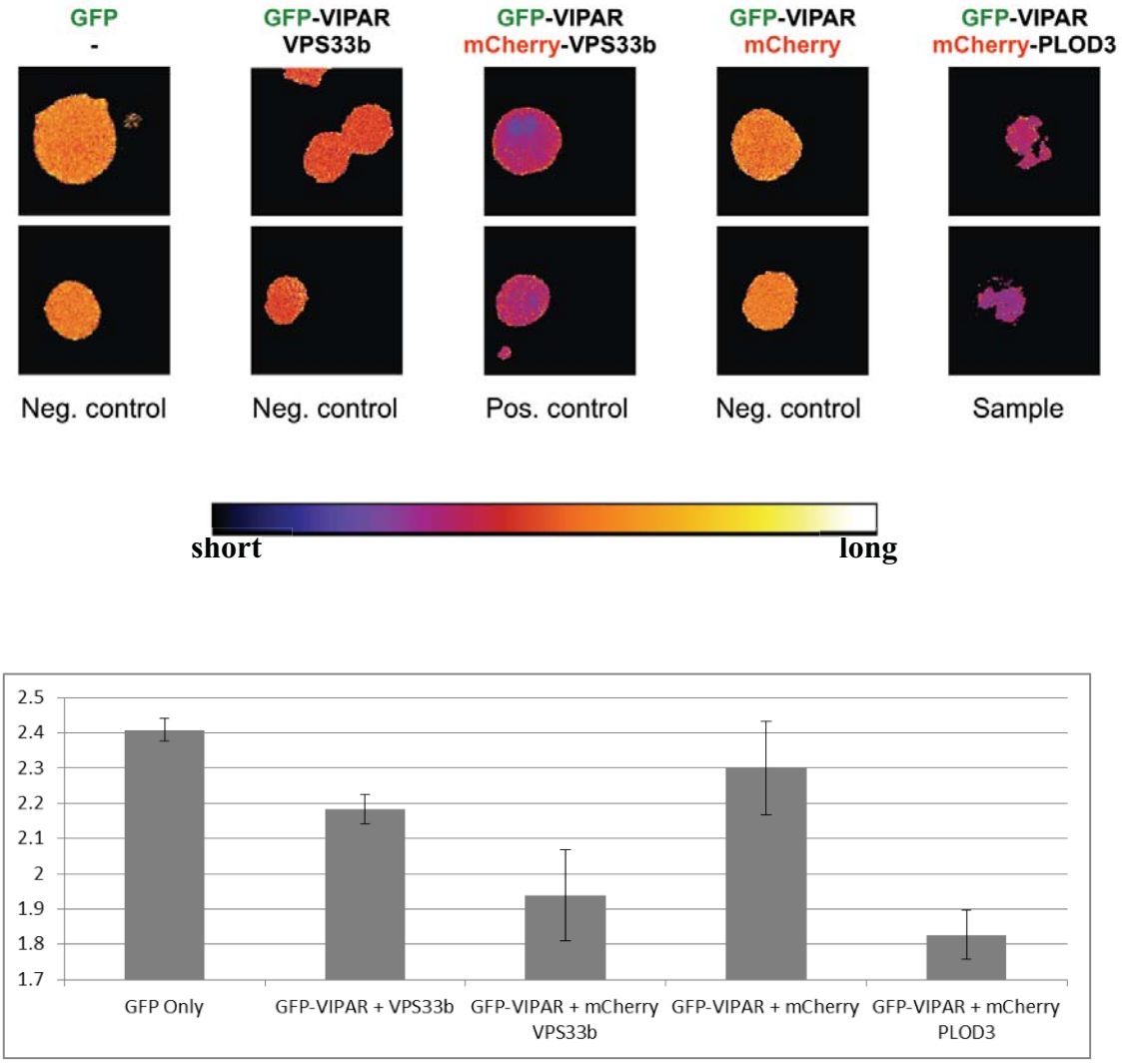


Figure 4.9 - FRET-FLIM of PLOD3-VIPAR interaction. To measure the occurrence of FRET between PLOD3 and VIPAR, HEK293 cells were transfected with GFP-VIPAR and mCherry-PLOD3 and the fluorescent lifetime of the donor (GFP) was measured by using a confocal microscope coupled to a FLIM (fluorescence life time imaging) detector. This was compared to the fluorescence lifetime of the donor in a sample where only the donor was present. In the panels above the lifetime is shown as different colours that are false colour lifetime images. Colours with a shorter wavelength (towards purple) correspond to a shorter lifetime, indicating the presence of FRET. Negative controls were performed by transfecting cells with GFP-VIPAR and untagged-VPS33B (second panel) or with GFP-VIPAR and empty-mCherry vector (fourth panel). The average fluorescence lifetime is quantified for each sample and reported in the histogram above. For each dataset n>5. Significant FRET values (P<0.01) were observed when comparing the GFP-VIPAR / mCherry couple with the GFP-VIPAR / mCherry-PLOD3 (P=0.009003, N=3). Experiment performed in collaboration with Federico Forneris.

As mentioned in section 3.4.3, for a positive FRET the distance between 2 fluorophores must be between 1 and 10 nm, the scale at which protein-protein interactions take place. This technique can also be used to measure the nanometer proximity of fluorescent labels (mCherry and GFP) attached to proteins, in this case GFP-VIPAR / mCherry-PLOD3 and mCherry-VPS33B / GFP-VIPAR. The excited state energy of the "donor" label (GFP) is transferred to the "acceptor" label (mCherry) with an efficiency (E) that is a function of the distance between them (R) (figure 4.10): $E = 1 / (1 + (R/R0)6)$ where R0 is the Förster distance between donor and acceptor at which the FRET efficiency is 50%. For the GFP/mCherry couple, the Förster distance R0 is 5.1nm.

The simplest way to measure FRET is by the decrease in the intensity of the donor (i.e. 57% reduction) and the corresponding increase in acceptor intensity. Still, for intermolecular FRET in an imaging experiment, the concentration of donor and acceptor for each pixel is not known. It is therefore difficult to attribute a certain donor intensity, or donor/acceptor intensity ratio to FRET. Therefore, FRET measurement is frequently done by monitoring changes in the fluorescence lifetime of the donor. After excitation, the fluorescence intensity I(t) decays exponentially (typically with a lifetime τ of a few nanoseconds). In case of FRET, the efficiency of the transfer can be calculated from the lifetime of the donor in presence (τDA) and absence (τD) of the acceptor: $I(t)=I0e-t/\tau$ and $E= 1 – \tau DA/\tau D$ (i.e. 57% FRET efficiency = 57% reduction in lifetime). The concentration of the donor will not change the lifetime, only the accuracy of the measurement.

In FRET/FLIM only the "donor" channel is imaged. For GFP-VIPAR/mCherry-VPS33B an E=16% decrease in the lifetime of GFP was measured. This reaches E=21% in the case of GFP-VIPAR/mCherry-PLOD3. By applying this data to the above

mentioned formula E = 1 / (1 + (R/R0)6) and resolving on R, the average distance between GFP and mCherry was calculated and it corresponds to 72.3 +/- 4 Å in the case of GFP-VIPAR/mCherry-VPS33B and 67.2 +/- 2.5 Å for GFP-VIPAR/mCherry-PLOD3. This data are consistent after excluding the possibility of FRET signal between GFP-VIPAR in presence of VPS33B lacking the mCherry tag (negative control for GFP lifetime). It is therefore possible to assume that intra-molecular distances separating VIPAR from PLOD3 are shorter than those separating VPS33B from VIPAR. From these data we can conclude that PLOD3 and VIPAR are likely to interact directly between them. Indeed the distance separating GFP from mCherry in GFP-VIPAR/mCherry-PLOD3 (67.2 +/- 2.5 Å) is too short to allow the presence of other intermediate interactors.

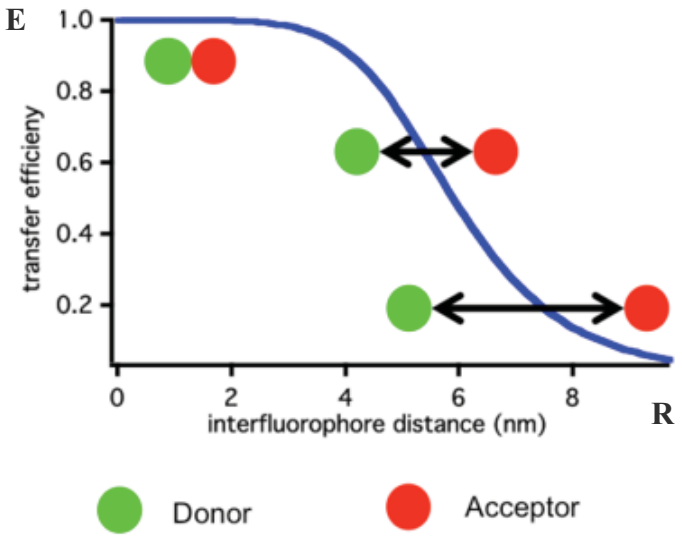


Figure 4.10 - A plot of the FRET efficiency (E) as a function of the distance (R) between a donor fluorophore (green sphere) and an acceptor fluorophore (red sphere) according to the E = 1 / (1 + (R/R0)6) where R0 is the Förster distance between donor and acceptor at which the FRET efficiency is 50%.

4.4.3.5 - PLOD3 is likely to interact with the N-terminal domain of VIPAR

As mentioned in section 3.1, VIPAR consists of two main regions, one N-terminus (N) disordered region and a globular structured region at the C-terminus (C). It was described in section 3.4.4 that the C region only is involved in the interaction with VPS33B. The findings above led to the hypothesis that the N-region of VIPAR might be involved in the binding of other interactors, which stabilise its disordered structure *in-vivo.* Indeed, disordered regions in proteins require binding of structured partners to undergo a disorder-to-order transition *in-vivo*. Thus, based on the identification of PLOD3 as a novel interactor of VIPAR it was suggested that the N-terminal region of VIPAR, which is not involved in the interaction with VPS33B, might directly interact with PLOD3. This hypothesis was tested by the co-immunoprecipition of PLOD3 with the same short constructs of VIPAR that were used to test the interaction with VPS33B: the N- and C-terminus containing constructs. Furthermore, two constructs for the C region of VIPAR were used, C- and C+, that respectively lack or include a nuclear localisation sequence (NLS) (section 3.4.4) that, as mentioned in section 3.1, is predicted to be localised at the interface between the disordered and the globular region of VIPAR. The N-terminal construct used in this study does not contain the NLS (N-). NLS was shown to be dispensable for the interaction of VIPAR with VPS33B (section 3.4.4). To investigate whether NLS is crucial for the VIPAR-PLOD3 interaction the following experiments were performed. The co-immunoprecipitation experiment (figure 4.11) revealed that PLOD3 interacts mainly with the N-region of VIPAR while only weak interaction was detected with its C-region with no major preference for the C- or C+ (figure 4.11) indicating that NLS is not crucial for this interaction. Furthermore

endogenous PLOD3 binds to full-length myc-VIPAR and also myc-VIPAR in complex with HA-VPS33B (figure 4.11), which confirmed previous results (section 4.4.3.1).

This experimental result confirmed the initial hypothesis that predicted the N-terminal disordered region of VIPAR to be involved in the binding of other proteins that stabilise its structure *in-vivo.*

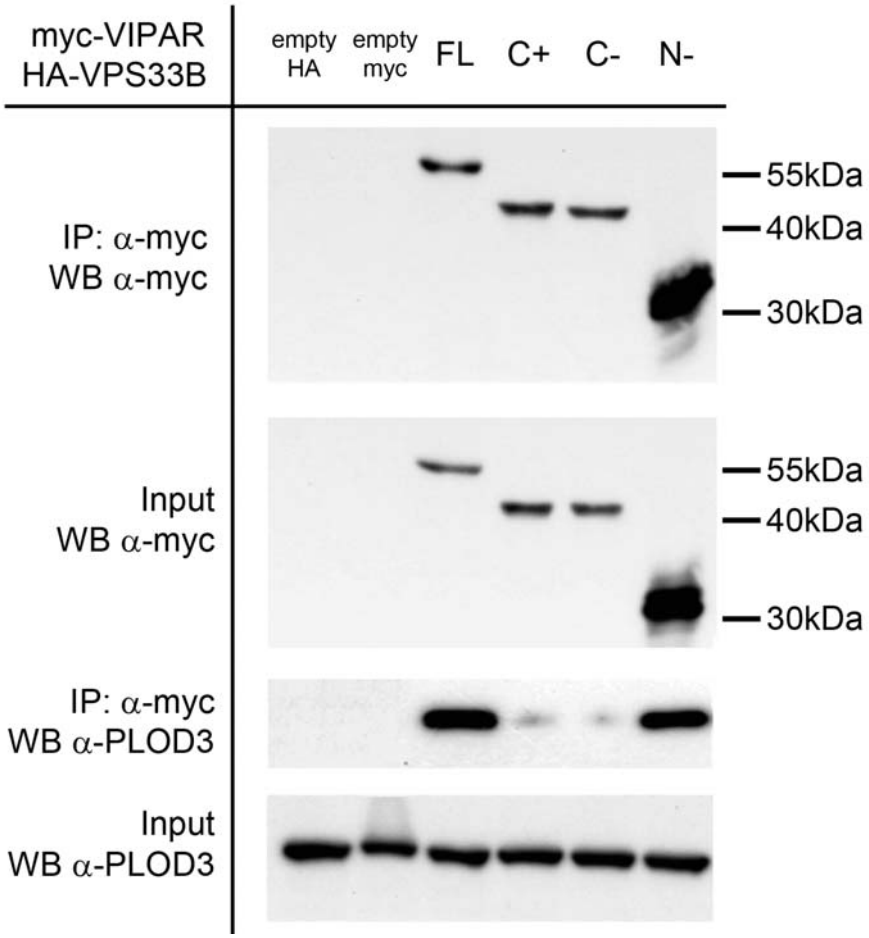


Figure 4.11 - Co-Immunoprecipitation of myc-VIPAR constructs with endogenous PLOD3. HEK293 cells were co-transfected with myc-empty vector (-) and myc tagged VIPAR full-length (FL), N-terminal construct of VIPAR without NLS (N-), C-terminal construct of VIPAR without NLS (C-), C-terminal construct of VIPAR with NLS (C+) and HA-VPS33B/myc-VIPAR complex (CPLX). Immunprecipitation was performed with anti-myc antibody and it reveals a strong interaction between endogenous PLOD3 and N- construct of VIPAR. Endogenous PLOD3 binds also to full-length myc-VIPAR and myc-VIPAR when in complex with HA-VPS33B. A weaker interaction is detected between PLOD3 and the C-terminal constructs of VIPAR (C+ and C-). None of these proteins were pulled down with myc-empty vector. Available antibodies to PLOD3 are poor at immunoprecipitating endogenous levels of the protein and therefore the reverse co-immunoprecipitation experiment was not successful.

4.5 - DISCUSSION

A pull-down experiment followed by mass spectrometry analysis led to the identification of two proteins as potential binding partners for the VPS33B/VIPAR protein complex. After exclusion of the likely false positives, the interaction with two proteins was further tested and included: Procollagen-Lysine 2-Oxoglutarate 5-Dioxygenase 1 (PLOD1) and Procollagen-Lysine 2-Oxoglutarate 5-Dioxygenase (PLOD3). Initial co-localisation experiments were performed to validate these results and no co-localisation was seen between over-expressed mCherry-PLOD1 and over-expressed YFP-VPS33B/myc-VIPAR. Furthermore PLOD1 localisation in the cell has been reported to be exclusively at the Endoplasmic Reticulum (Yeowell and Walker, 2000; Yeowell et al., 2000) and that doesn't match the previously published data for VPS33B/VIPAR intracellular localisation, which identified this protein complex at recycling enodosomes, late endosomes, lysosomes and Golgi complex (Akbar et al., 2011; Cullinane et al., 2010; Flaumenhaft, 2012; Zhu et al., 2009). Thus it was assumed that it was unlikely that VPS33B/VIPAR can interact with an exclusively ER localised protein such as PLOD1. Altogether these observations led to excluding an interaction between PLOD1 and VPS33B-VIPAR. The presence of this protein in the mass spectrometry results might be explained by its high homology to PLOD3 and consequent fail in the distinction between the isoforms by mass-spec. However more experiments, such as co-immunoprecipitation, are required to exclude the possibility that PLOD1 is an interactor of VPS33B/VIPAR.

PLOD3 is the only LH isoform that is found in several locations including intracellularly, the cell surface and extracellularly (Myllyla et al., 2007; Wang et al., 2009; Wang et al., 2012). These locations suggest that different pathways may be

involved in PLOD3 trafficking and therefore an interaction with VPS33B/VIPAR complex is more plausible. In this chapter several biochemical and cell-biology techniques are described that were used to validate the second putative interactor identified by mass-spectrometry: PLOD3. All different experimental approaches used in this section (co-localisation, co-immunoprecipitation, FRET-FLIM) confirmed that PLOD3 is a novel interactor of VIPAR *in-vitro* and these proteins are likely to interact *in-vivo*. PLOD3 does not co-immunoprecipitate with VPS33B on its own but only when VPS33B is in complex with VIPAR. This suggests a high specificity of PLOD3 binding to the complex. To further direct this analysis and investigate the regions of VIPAR that are specifically involved in the interaction with PLOD3, several short constructs of VIPAR were used and PLOD3 was shown to interact with the N-terminal region of VIPAR. The N-terminal is a predicted disordered region of VIPAR not involved in the interaction with VPS33B. PLOD3 might therefore be crucial to stabilise this region *in-vivo*. In addition to the FRET-FLIM analysis, a shorter average distance is calculated between GFP and mCherry in GFP-VIPAR/mCherry-PLOD3 than GFP-VIPAR/mCherry-VPS33B. The proximity of the GFP-mCherry fusion tags might be related to the regions of VIPAR involved in the interaction with PLOD3 and VPS33B (figure 4.12). Indeed the C-terminal of VIPAR only is involved in the interaction with VPS33B. The N-terminal region was found to interact with PLOD3 and therefore the fusion tags are expected to be closer in the case of PLOD3-VIPAR. This observation indicates that the results obtained are consistent and are confirmed by different technical approaches.

Co-localisation experiments performed with antibodies against endogenous PLOD3 and VIPAR found these proteins to co-localise in the perinuclear region of HeLa cells. In

the next chapter the intracellular localisation of these proteins will be investigated as well as the role of this novel interaction in physiological and pathological conditions.

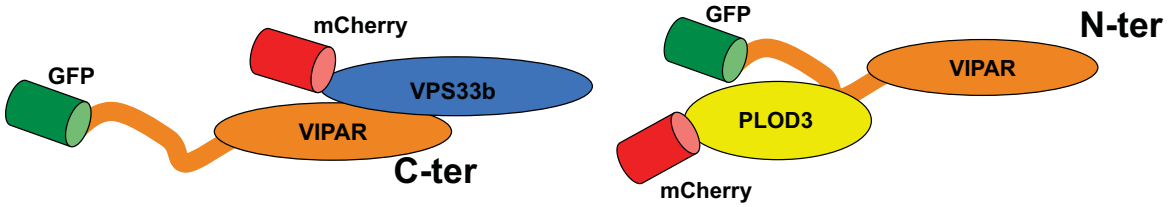


Figure 4.12 - Diagram for mCherry-VPS33B/ GFP-VIPAR and mCherry-PLOD3/GFP-VIPAR interaction showing proximity of the mCherry-GFP couple related to FRET signal. VPS33B interacts with the C-terminal of VIPAR only and therefore the mCherry-GFP distance in this couple is higher when compared to the mCherry-GFP located respectively on PLOD3 and VIPAR where PLOD3 mainly interacts with the N-terminal of VIPAR.

CHAPTER FIVE:

VPS33B/VIPAR is involved in PLOD3 intracellular delivery: a novel mechanism for Collagen trafficking

5.1 - INTRODUCTION AND OVERVIEW

In the previous chapter PLOD3 was identified as a novel interactor for VIPAR and this interaction was confirmed through several different experimental approaches. In this chapter the relevance of this interaction in ARC syndrome will be examined.

Endogenous PLOD3 and VIPAR were found to co-localise in the perinuclear region in HeLa cells (section 4.4.3.3). In this chapter the intracellular localisation of these proteins was investigated by performing co-localisation studies with different cellular markers.

During their synthesis, all lysyl hydroxylase homologues are targeted to the rough endoplasmic reticulum via a signal peptide, even if they lack the typical ER-retention signals such as KDEL or double lysine motif (Kellokumpu et al., 1994). PLOD1, whose topology and localisation has been better characterised among the homologues, was shown to be a luminal protein of the ER. PLOD1 completely co-localises with the protein disulfide isomerase (PDI), an ER resident KDEL-containing protein (figure 5.1) and also co-sediments with it after sucrose density gradient fractionation (Kellokumpu et al., 1994). Despite lacking ER-specific retention motifs, PLOD1 is a peripheral membrane protein of the ER that is retained in its lumen by electrostatic interactions with unknown membrane proteins (Kellokumpu et al., 1994). This represents the first identified protein with this type of association in the ER (Kellokumpu et al., 1994).

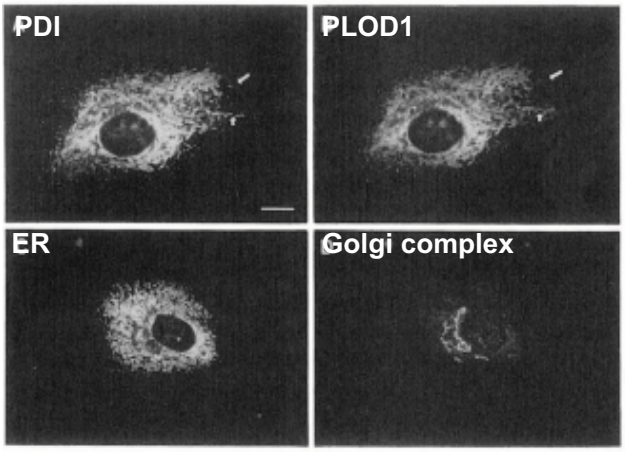


Figure 5.1 - PLOD1 and PDI endogenous staining in human skin fibroblasts. The two proteins show almost identical localisation except in a few places marked by arrows. ER and Golgi complex are stained in the lower panels. The Golgi complex is seen as an unstained area in PLOD1, PDI and ER staining panels. Bar, 10 pm. (Modified from (Kellokumpu et al., 1994))

PLOD2 and PLOD3 have also been thought to localise to the ER because of their homology to PLOD1 and because the hydroxylation reaction of lysines is known to occur in the ER prior to collagen triple-helix formation (Salo et al., 2006b). Overexpressed PLOD3 in COS-7 cells has been localised to the ER, however Golgi staining was observed in some cells (figure 5.2 second panel) (Heikkinen et al., 2000).

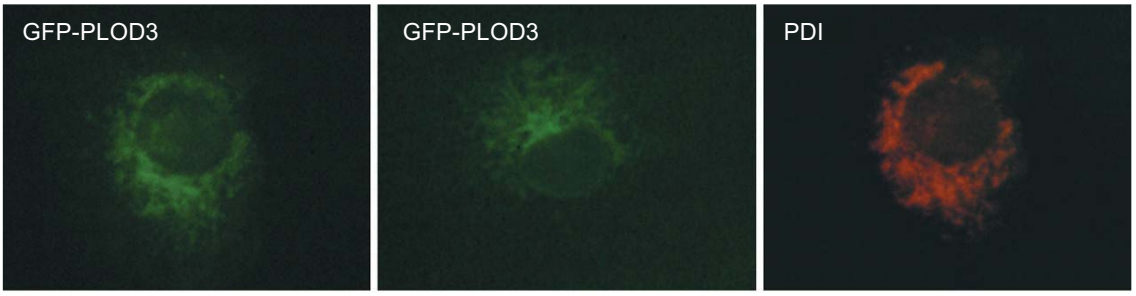


Figure 5.2 - Intracellular distribution of GFP-PLOD3 in COS-7 cells 48h after transfection (green). Rough endoplasmic reticulum is visualised by staining the cells with PDI (protein disulfide isomerase) antibody (red) (modified from (Heikkinen et al., 2000)).

A part of this chapter describes the co-localisation experiments that were performed to identify the intracellular localisation of endogenous PLOD3. Furthermore, the site of interaction between VIPAR and PLOD3 was sought using confocal microscopy and numerous intracellular markers. Additionally, different cellular fractionation experiments were performed to elucidate the topology of this interaction. PLOD3 is thought to be located in the lumen of ER, where the modifications of lysine residues in collagens take place (Myllyla et al., 2007). To date, the topology of VPS33B/VIPAR is not known but the yeast homologue Vps33p is cytosolic (Gerhardt et al., 1998) and it interacts directly with Vps16 (VIPAR homologue) constituting the side of the head of the HOPS complex (section 3.1). If both VPS33B and VIPAR were located in the cytosol, the direct interaction between VIPAR (cytosolic) and PLOD3 (luminal), that was established from FRET-FLIM measurements (section 4.4.3.4), would be topologically impossible. It was therefore crucial to investigate the topology of PLOD3/VIPAR complex to further validate this interaction and FRET-FLIM data.

It is known that PLOD3 localisation is different from the other PLOD isoforms. Apart from being found intracellularly, PLOD3 is secreted into the extracellular space surrounding some tissues such as kidney, spleen, muscle and liver (Salo et al., 2006a; Salo et al., 2006b). Interestingly PLOD3 is localised exclusively extracellularly in adult mouse liver but only intracellularly in the livers of mouse embryos at E13.5 and E15.5 (Salo et al., 2006a). The ability of PLOD3 to modify proteins in the extracellular space has been linked to a potential crucial role of the enzyme in matrix remodelling (Salo et al., 2006b).

Despite being ubiquitous, the expression of PLOD3 differs in tissues and development. During mouse embryogenesis PLOD3 is widely expressed but becomes more restricted

to specific cells within adult tissues such as brain, lungs, spleen, muscle, pancreas, kidney and liver (Salo et al., 2006a). Interestingly these are some of the most affected organs in ARC syndrome pathogenesis (section 1.1). Therefore mouse inner medullary collecting duct (mIMCD-3) cells lines were chosen to study the impact of VPS33B/VIPAR deficiency upon PLOD3 localisation. Stable knockdown of Vps33b/Vipar in mIMCD-3 cell lines represent the polarized epithelial cell model for ARC syndrome (Cullinane et al., 2010)

Apart from intra- and extracellular localisation, PLOD3 is also located on the cells surface in association with collagenous proteins (Salo et al., 2006b). PLOD3 found on the cell surface and intracellularly has different molecular weight from PLOD3 found in the cell culture medium, which is though to be due the two isoforms being differently glycosylated. This led to the hypothesis that PLOD3 may be secreted from cells through two different pathways. The first one passes through the Golgi complex, where the protein is further glycosylated and then secreted in the medium, while the second route bypasses the Golgi and directs PLOD3 to the cells surface where the protein is consequently found to be under-glycosylated (Wang et al., 2012). The mechanisms involved in PLOD3 trafficking remain nevertheless unknown. I hypothesised that VPS33B/VIPAR are directly involved in one of the routes of PLOD3 trafficking throught the direct interaction between VIPAR and PLOD3.

In order to understand the relevance of PLOD3-VIPAR interaction in the pathogenesis of ARC syndrome I analysed the phenotype of mIMCD-3 knockdown cells lines for Vps33b/Vipar and its relation to a possible mis-trafficking of PLOD3 in these cells. Knockdown of Vps33b-Vipar in mIMCD-3 causes epithelial polarization defects with cells undergoing a typical epithelial to mesenchymal transition (EMT) process

(Cullinane et al., 2010). Vps33b-Vipar knockdown cells lose their polarity and cell-cell contacts, they do not grow in a monolayer but instead grow over the neighbouring cells, become migratory, and divide at a faster rate. Loss of polarity in epithelia results in mislocalisation of membrane proteins (Fialka et al., 1999; Tanos and Rodriguez-Boulan, 2008) which has been shown in ARC patient liver and kidney biopsies as well as in mIMCD-3 knockdown cell lines for Vps33b and Vipar (section 1.1 and (Cullinane et al., 2010)). Loss of polarity in Vps33b-Vipar knockdown cell lines is associated with a loss of E-cadherin and, to a lesser extent, claudin-1 expression at adherens and tight junctions (Cullinane et al., 2010). In these cell lines E-cadherin is downregulated at a protein and mRNA level and this is associated to a reduction in E-cadherin promoter activation (Cullinane et al., 2010). The down-regulation of E-cadherin has been demonstrated also in Morpholino vipar injected *Zebrafish* larvae and in ARC patients liver biopsies. Loss of E-cadherin is one of the known causes of loss in cell polarity in epithelia (Desai et al., 2009; Theard et al., 2007). Therefore investigating the mechanism through which the VPS33B/VIPAR complex indirectly regulates the expression levels of E-cadherin, is crucial in terms of understanding the pathogenesis of ARC syndrome. In order to understand the impact of the novel identified PLOD3-VIPAR interaction in the phenotype of mIMCD-3 knock down cell lines, I analysed the downstream targets of PLOD3 enzyme. As described previously, PLOD3 is a lysyl hydroxylase (LH) enzyme that catalyses the hydroxylation of lysyl residues in collagen-like peptides as well as the galactosylation of hydroxylysines and glucolsylation of galacotosylhydroxylysines (sections 1.5.4, 4.1). A considerable number of proteins not classified as collagens contain a collagenous domain and therefore constitute a potential substrate for PLOD3 (section 1.5.3). Among them adiponectin has been shown to be

directly regulated by PLOD3 with regards to its amount and oligomerisation status in the cell (Ruotsalainen et al., 2012). However collagen represents the most abundant protein in mammals and it constitutes therefore the major substrate of PLOD3 and consequently the better characterised target in PLOD3 deficiency (Rautavuoma et al., 2004; Sipila et al., 2007; Sricholpech et al., 2011). Furthermore, the extent of hydroxylated and glucosyl-galacotosylated lysines varies between collagen types. Around 90% of all lysines are hydroxylated and hydroxylysines glycosylated in types IV and VI collagen but less than 20% of the lysines are hydroxylated in type III collagen. In this respect it has been shown that PLOD3 binds preferentially to type IV, VI, XI, and XII collagen peptides, despite there being no strict sequence specificity for binding of lysyl hydroxylase isoforms to specific collagenous peptides (Risteli et al., 2004). Taken together, these data from literature suggested that the best candidate for the study of possible effects of PLOD3 mistrafficking in Vps33b/Vipar knockdown mIMCD-3 cells is collagen type IV. This represents the major structural component of basement membrane (BM) that defines the other components of the BM itself such as, laminins, nidogens or perlecan, into highly organised architecture (section 1.5.3.1). The basement membrane underlies epithelium and endothelium and is therefore crucial for the renal function where it covers the entire outer surface of each individual nephron and collecting duct and is involved in filtration, cell adhesion, migration, and differentiation (Miner, 1999). Consequently defects in PLOD3 trafficking could be reflected in collagen IV structure/expression of inner medullary collecting duct cells (mIMCD-3) knockdown for Vps33b/Vipar and this possibility was investigated in this chapter. Additionally, a stable knockdown line for Plod3 was created in mIMCD-3 cells and compared to Vps33b/Vipar knockdowns in terms of its phenotype, collagen IV and

E-cadherin expression. In order to understand the correlation between the phenotype, collagen IV and E-cadherin expression in knockdown cells for Vps33b/Vipar or Plod3, it is crucial to underline the importance of collagen hydroxylation and their further galactosylation/glycosylation in collagen structure and secretion. The formation of hydroxylysines is crucial for collagen crosslinking, which stabilises the collagen fibril into supramolecular structures (Knott and Bailey, 1998; Reiser et al., 1992). The role of the glycosylation of hydroxylysine residues is not completely understood and sometimes controversial. Data from literature indicate that glycosylation has a role in collagen fibril formation (Notbohm et al., 1999) but this has been contradicted by other data (Batge et al., 1997). Glycosylation of type II collagen has been implicated in autoimmunity (Myers et al., 2004; Van den Steen et al., 2004) and the glycosylation of the collagenous domain in adiponectin is involved in the insulin-sensitising activity of the hormone (Wang et al., 2002c). Abnormal glycosylation in collagen IV has been shown to be involved in metastasis by preventing the adhesion of melanoma cells and consequent invasion (Lauer-Fields et al., 2003). The invasion of the basement membrane is indeed a critical step in the metastatic process (Friedl and Wolf, 2003; Wolf and Friedl, 2006). Collagen IV of the basement membrane contains glycosylated hydroxylysine residues that bind the tumour cell through the CD44/chondroitin sulphate proteoglycan receptors located on its surface. This binding is dramatically decreased when collagen IV residues were glycosylated, compared to non-glycosylated. These findings suggest a role for glycosylated hydroxylysines of collagens in cell-extracellular matrix interactions (Lauer-Fields et al., 2003).

Glycosylation of hydroxylysine residues is crucial for the formation of the embryonic basement membranes as shown by the embryonic lethal phenotype of the PLOD3

knockout mice (Rautavuoma et al., 2004). This study has shown that the embryonic lethality is caused by the block of only the GGT activity of PLOD3. Mice that carried a mutation that suppressed the LH activity of the enzyme were viable but with defects in the ultrastructure of the basement membrane and in collagen fibril organization whereas mice lacking the GGT activity of PLOD3 were embryonically lethal at E9.5 (Ruotsalainen et al., 2006). This study suggested that PLOD3 is the main isoform involved in the biosynthesis of type IV collagen during early vertebrate development, because the two other isoforms, PLOD1 and PLOD2 were not able to compensate for the lack of PLOD3. Electron microscopy (EM) of the E9.5 embryos revealed fragmentation of basement membranes with detachment in some areas. Newly synthesised type IV collagen was aggregated in secretion vesicles on the way to the cells surface and amorphous type IV collagen particles were evident both intracellularly and extracellularly. Part of the intracellular amorphous collagen IV was probably degraded in the ER or sent for degradation to lysosomes. Enlarged Golgi complexes and lysosomes as well as dilation of the endoplasmic reticulum were shown by immuno EM in PLOD3 null embryos (Rautavuoma et al., 2004).

Successive work demonstrated that the mechanism behind the fragmentation of BM in PLOD3 knockout embryos is related to both secretion and intracellular degradation of type IV collagen (Sipila et al., 2007). Intracellular accumulation, delayed secretion and intracellular degradation of type IV collagen was also seen in PLOD3 knockout fibroblasts. It is therefore possible to conclude that glycosylation of hydroxylysine residues is critical for assembly and secretion of highly glycosylated collagen types (Sipila et al., 2007). Altogether these data demonstrated for the first time a novel crucial role of glycosylation of hydroxylysines for collagens (Lauer-Fields et al., 2003;

Rautavuoma et al., 2004; Ruotsalainen et al., 2006; Sipila et al., 2007). Complete loss of function in PLOD3 has deleterious effects on cells and tissues. There are no known patients with homozygous mutations in *PLOD3* and the only report of a known case of PLOD3 deficiency describes a patient compound heterozygous for two mutations, one of which led to a reduction in GGT and the other to a reduction in LH activities (Salo et al., 2008). Severe connective tissue abnormalities were seen in this patient with features that overlap with a number of collagen defects including joint contractures, growth retardation, thin skin with poor healing, osteopenia and talipes equinovarus (Salo et al., 2008).

Partial loss of PLOD3 enzymatic activity in human skin fibroblast causes changes in the deposition and organisation of the Extracellular Matrix as well as the cytoskeletal arrangement of the cell (Risteli et al., 2009). Depletion of GGT activity of PLOD3 in HT-1080 cells by siRNA lead to arrest of growth followed by lethality in these cells. The glycosyltransferase activity of PLOD3 is therefore crucial for cell growth and viability and a complete loss of function of the enzyme is not compatible with life (Wang et al., 2009). Additionally, mutations in *Caenorhabditis elegans* Plod, the only isoform present in lower organism, leads to embryonic lethality due to disruption in collagen IV secretion and subsequent basement membrane malformation (Norman and Moerman, 2000).

As mentioned in the previous chapter (section 4.1) the LH activity of PLOD3 is located in its carboxy-terminal (Myllyla et al., 2007) while the GT and GGT activity in the amino-terminal part of the enzyme (Myllyla et al., 2007; Wang et al., 2002a; Wang et al., 2002b). Furthermore the LH active site mediates retention of PLOD3 inside the cell while the GGT active site is required for its secretion into the extracellular space (Wang

et al., 2012). I hypothesised that the VPS33B/VIPAR complex, directly involved in PLOD3 trafficking, might therefore indirectly participate in collagen trafficking inside the cell. To test this hypothesis, the co-localisation between collagen IV and PLOD3 was compared between wild-type mIMCD-3 cells and those with Vps33b/Vipar stable knockdown.

As explained above, the phenotype of Vps33b/Vipar knockdown mIMCD-3 is characterised by loss in E-cadherin expression and Epithelial to Mesenchymal Transition (EMT). Data in literature indicate that native collagen type IV induces down-regulation of E-cadherin and EMT in mammary epithelial cells MCF10A (Espinosa Neira and Salazar, 2012). Furthermore, collagen type IV induces cell migration through a DDR1 and CD9-dependent pathway (Castro-Sanchez et al., 2011; Espinosa Neira and Salazar, 2012) and is involved in tumour onset and progression (Agarwal and Ballabh, 2013; Hamano et al., 2003; Ohlund et al., 2013; Rider et al., 2013).

I hypothesized that loss of E-cadherin and EMT phenotype observed in Vps33b/Vipar knockout mIMCD-3 cell lines was caused by an abnormal expression/structure of collagen type IV caused by an abnormal PLOD3 trafficking in these cells.

5.2 - AIMS

In the previous chapter immunofluorescence experiments showed intracellular co-localisation between PLOD3 and VIPAR in HeLa cell lines. One of the goals of this chapter is to identify the intracellular compartment where the interaction of these proteins takes place in addition to determining the topology of VPS33B/VIPAR/PLOD3 interaction.

Identifying the biological relevance of PLOD3-VIPAR interaction and its role in the pathogenesis of ARC syndrome is the main goal of this work. Therefore the main aim of this chapter is to identify possible abnormalities related to PLOD3 mistrafficking in Vps33b/Vipar knockdown mIMCD-3 cell lines that can elucidate the mechanisms behind the EMT phenotype in these cells.

5.3 - METHODS

5.3.1 - Separating cytosol from crude membrane fractions in mIMCD-3 and HeLa cells

Membrane homogenisation of mIMCD-3 and HeLa cells was performed in the cold room with samples and homogeniser kept on ice. A 75cm^2 flask of HeLa or mIMCD-3 cells were transferred on ice and rinsed once with 10 mL of ice-cold PBS followed by a second wash with 4 mL of Homogenisation Buffer (HB) (20 mM HEPES-KOH, pH 7.5, 5 mM EDTA, 250 mM sucrose supplemented with protease inhibitors).

Cells were scraped on ice with 1mL of HB per T-75 flask and pelleted by centrifugation at 500 g for 5 min at 4 °C. The supernatant was discarded and the cell pellet was re-suspended in HB (~ 4 times the volume of cell pellet).

For the homogenisation protocol each cell suspension was passed 12 times through a ball-bearing homogeniser (10 μm clearance). The homogenates were recovered from the chamber and centrifuged at 100,000 xg for 5 min at 4°C for coarse fractionation, that is membrane fraction *versus* soluble fraction (or cytosol). The supernatant (cytosol) was recovered and kept on ice. The pellet (membrane fraction) was re-suspended in an equal volume of HB until no particulate was visible. To help this process cells were sonicated for 10 seconds at 40% amplitude. Both fractions were analysed by WB.

5.3.2 - Cell fractionation

ProteoExtract® Subcellular Proteome Extraction Kit (Merck4Biosciences) was used for the extraction of subcellular proteomes from adherent growing HEK293 cells. The kit takes advantage of the different solubilities of subcellular compartments in select buffers used for cell fractionation into four fractions 1. Cytosolic fraction, 2. Membrane/organelle protein fraction, 3. Nucleic protein fraction 4. Cytoskeletal fraction.

A T25 flask of adherent growing HEK293 cells were washed once with 2mL of Wash buffer before cells were scraped with 2mL of Wash buffer and pelleted by centrifugation for 10 min at 100-330 xg at 4°C. The Wash buffer was discarded and the pellet re-suspended in 1mL of Extraction buffer supplemented with 5 μL of Protease Inhibitor Cocktail. Cells were gently re-suspended and incubated for 10 minutes at 4°C

with gentle agitation before centrifugation was carried out for 10 minutes at 700 xg. The collected supernatant constitutes the cytosolic fraction. The pellet was re-suspended in 1 mL of Extraction buffer II supplemented with 5 μL of Protease Inhibitor Cocktail and incubated for 30 minutes at 4°C with gentle agitation then centrifuged for 10 minutes at 5500 xg. The supernatant, constituting the membrane/organelle fraction, was transferred to a clean tube. The pellet was further re-suspended in 500 μL of extraction buffer III supplemented with 5 μL of Protease Inhibitor Cocktail and 1.5 μL of Benzonase Nuclease and incubated for 10 minutes at 4°C with gentle agitation before centrifugation was carried out for 10 minutes at 6,800 xg at 4°C. The supernatant contains the nuclear fraction and was therefore collected while the pellet was re-suspended in 500 μL of Extraction buffer IV supplemented with 5 μL of Protease Inhibitor Cocktail. This mixture constitutes the fourth fraction that contains the cytoskeletal matrix proteins. All fractions were analysed by Western Blot and stored at -80°C.

5.4 - RESULTS

5.4.1 - Endogenous PLOD3 does not co-localise with PDI: PLOD3 intracellular localisation is different from PLOD1

As mentioned above (section 5.1) the intracellular localisation of PLOD1 is the best characterised among lysyl hydroxylase isoenzymes. PLOD1 shows an identical co-localisation with the ER resident protein disulfide isomerase (PDI) in human skin fibroblasts (figure 5.1). The same co-localisation has been reported for a GFP tagged construct of PLOD3 which co-localises with PDI and in some cells with the Golgi

complex (figure 5.2). Here an antibody against PLOD3 was used in order to establish the intracellular localisation of the endogenous protein in human skin fibroblasts, the same cell type used to demonstrate the complete co-localisation between endogenous PLOD1 and PDI proteins. No co-localisation with PDI was seen for endogenous PLOD3 (figure 5.3 A). This represents a novel finding regarding PLOD3 intracellular localisation that differentiates it from other PLOD isoforms. PLOD3 staining matched with the unstained perinuclear area in the PDI staining (figure 5.3 A). This area resembles in shape the unstained area in the PDI and PLOD1 staining panels that correspond to the Golgi complex (figure 5.1). Therefore TGN-46, a Trans-Golgi-Network (TGN) protein marker was used to co-localise endogenous PLOD3 in human skin fibroblasts. PLOD3 expression pattern completely overlaps with TGN-46 with the proteins co-localising in several areas (figure 5.3 B). Co-localisation of PLOD3 was observed also with RAB11A (figure 5.3 C). As mentioned previously (section 1.3) VPS33B interacts with RAB11A involved in apical membrane protein sorting (Cullinane et al., 2010). RAB11A is found associated to the recycling endosomes (Ullrich et al., 1996), TGN and post-Golgi vesicles (Chen et al., 1998; Urbe et al., 1993). Other data suggest that Rab11 influences endosome to TGN trafficking by regulating membrane distribution in the early endosomal pathway (Wilcke et al., 2000). Therefore I postulated that VPS33B/VIPAR are involved, through RAB11A, in the trafficking of PLOD3 from/to TGN. In order to support this hypothesis it is crucial to determine the intracellular localisation of endogenous VPS33B/VIPAR that is to date not known.

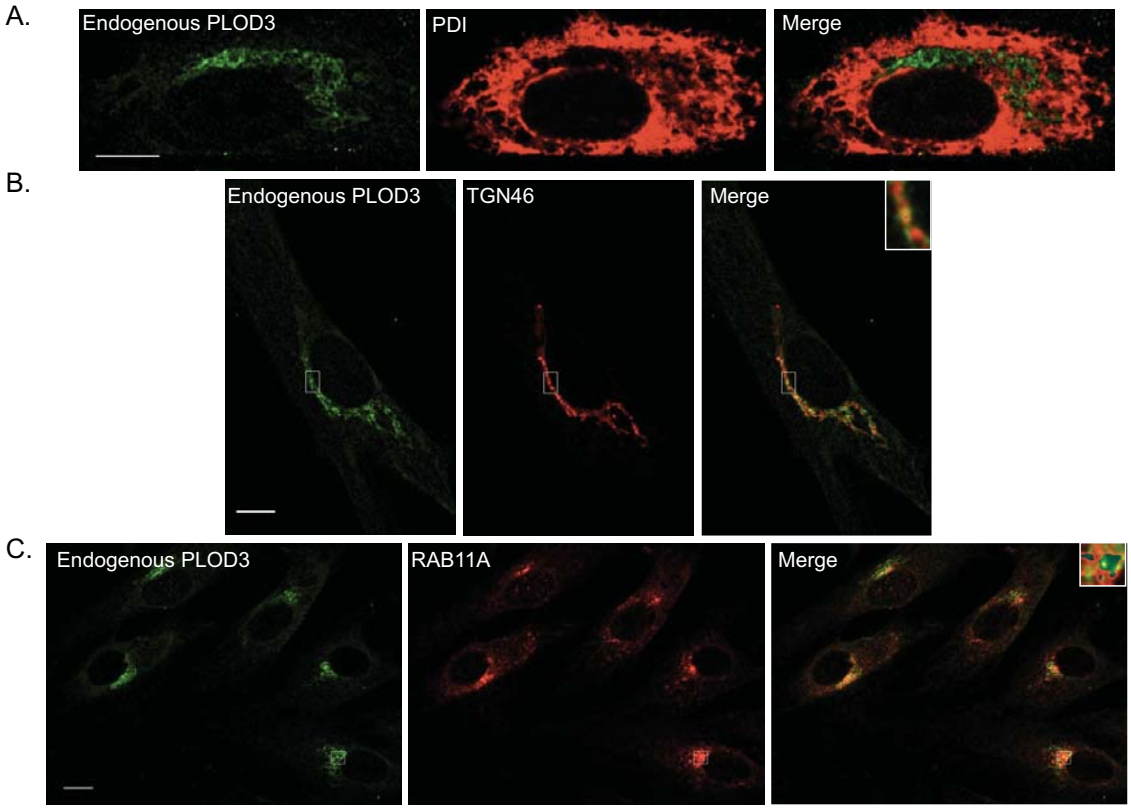
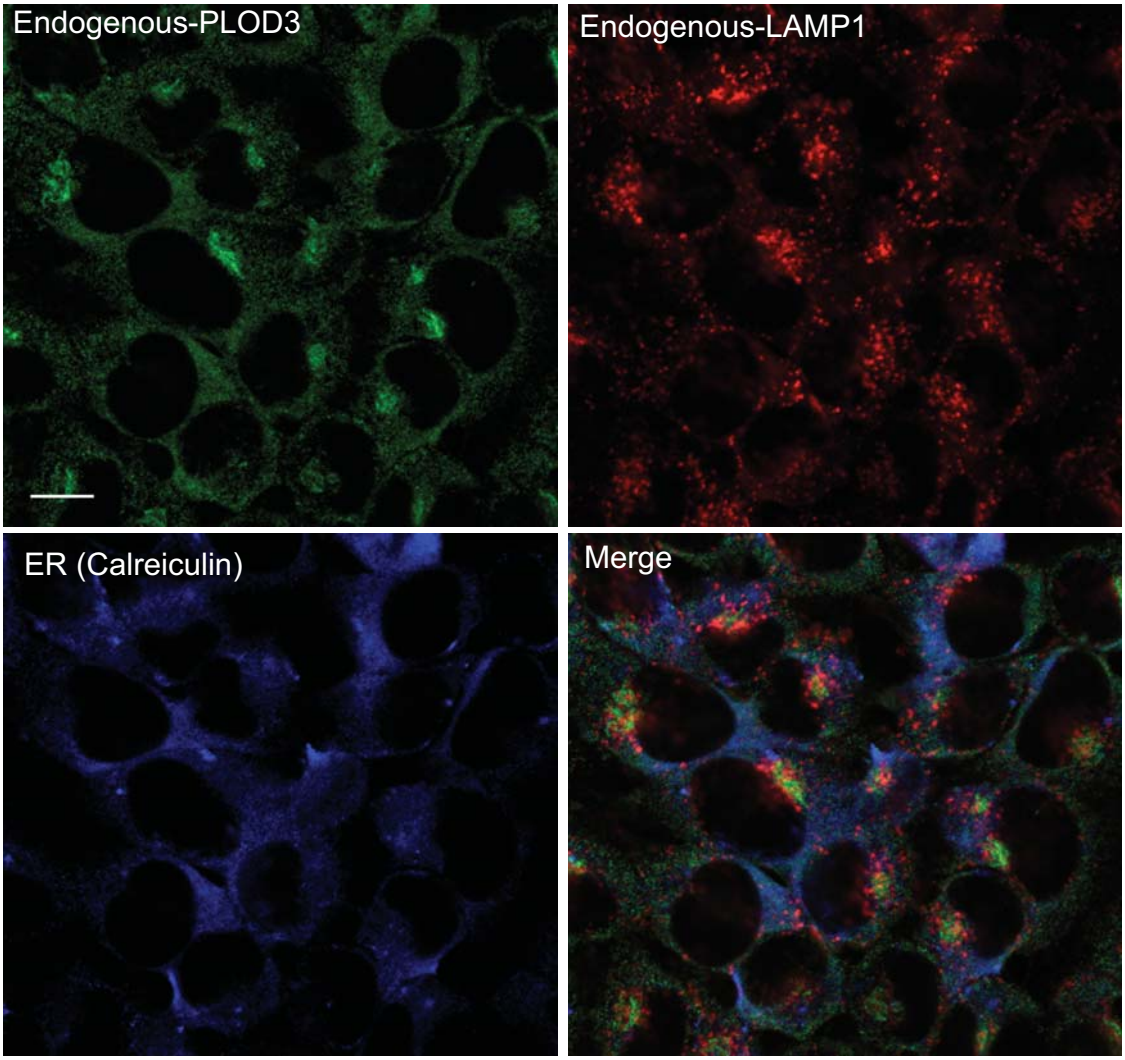


Figure 5.3 - Analysis of endogenous PLOD3 localisation in human skin fibroblasts. ***(A)****. Co-localisation analysis of endogenously expressed PLOD3 with endogenous protein disulphide isomerase (PDI) show no co-localisation between the two proteins. PLOD3 staining overlaps with the unstained area of PDI. PLOD3 co-localises with endogenously stained TGN46* ***(B)*** *and RAB11A* ***(C).*** *All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.*

5.4.2 - Endogenous PLOD3 and endogenous VIPAR co-localise with TGN markers and RAB11A in HeLa cells

To validate the novel finding that PLOD3 is found in the TGN of human skin fibroblasts, a different cell type was used to further investigate PLOD3 intracellular localisation. It consists of HeLa cells that were co-stained with an antibody against endogenous PLOD3 and antibodies against molecules of known cellular compartments (figure 5.4). As shown above, PLOD3 does not co-localise with the ER protein PDI. In order to exclude the possibility that PLOD3 is located in a different compartment of the ER, an antibody against Calreticulin, a more generic ER marker, was used. PLOD3 did not co-localise with Calreticulin in HeLa cells (figure 5.4 A) and it is therefore possible to conclude that PLOD3, unlike PLOD1, is not an ER resident protein. PLOD3 also did not co-localise with LAMP-1, a marker of the lysosomal compartment (figure 5.4 A) therefore it is unlikely that this compartment is involved in PLOD3 trafficking in HeLa cells. The previously observed PLOD3 localisation in the TGN of human skin fibroblasts was confirmed also in HeLa cells where PLOD3 co-localises with TGN-46 and p230, a different TGN marker used to further validate this data (figure 5.4 B). Similarly to what observed in fibroblasts, PLOD3 also co-localises with RAB11A in HeLa cells (figure 5.4 B).

A.



B.

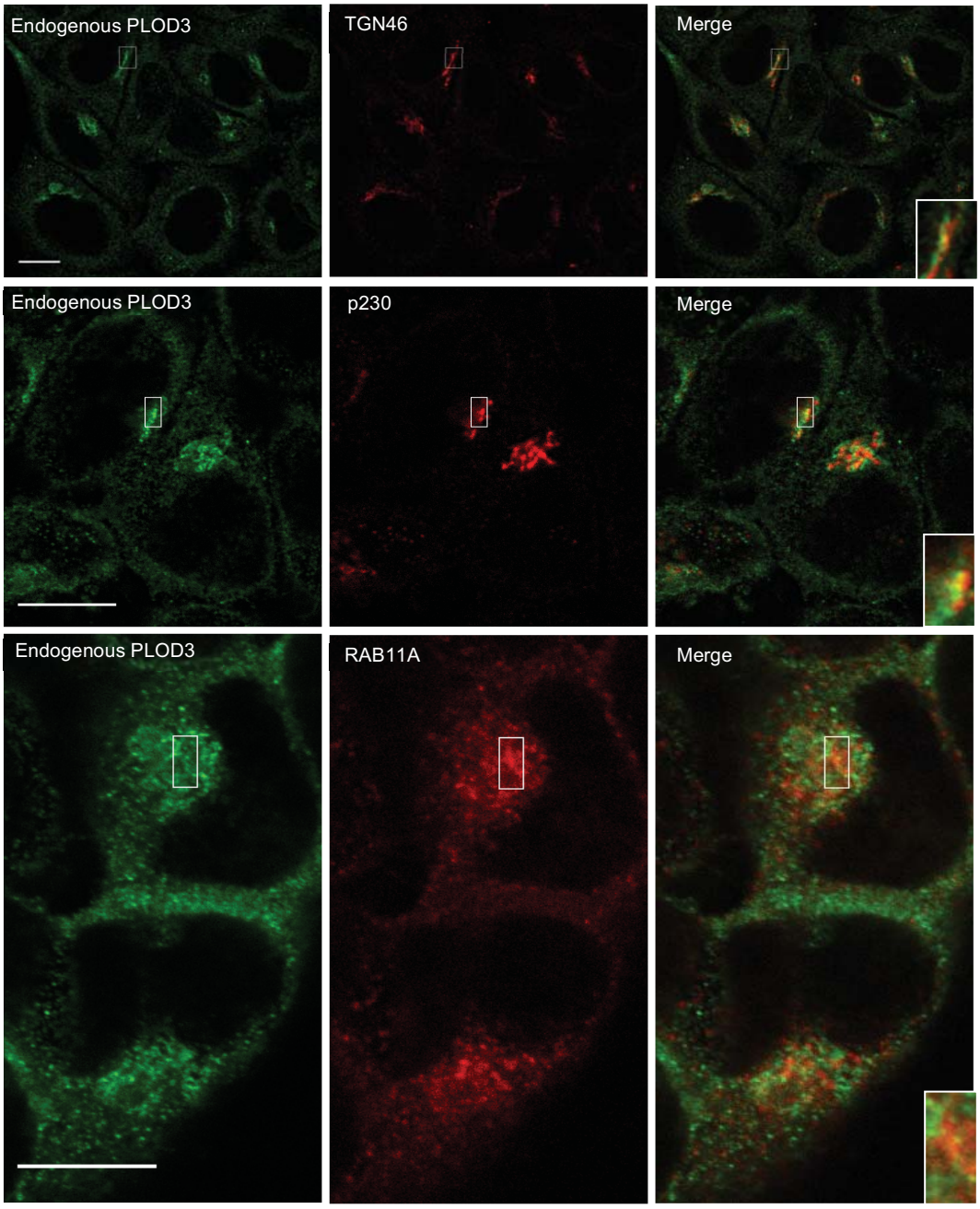


Figure 5.4 - Subcellular localisation of endogenous PLOD3 in HeLa cells using different cell compartment markers. ***(A)****. Co-localisation analysis of endogenously expressed PLOD3 with the lysosomal protein LAMP1 and the ER protein Calreticulin shows no co-localisation between the proteins.* ***(B)****. PLOD3 co-localises with the TGN markers TGN-46 and p230 as well as with endogenously expressed RAB11A. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.*

As mentioned in the previous Chapter (section 4.4.3.3) endogenous PLOD3 co-localises with endogenous VIPAR in the perinuclear area and the aim was to identify the intracellular compartment to which this area corresponds. PLOD3 was localised in the TGN region in fibroblast and HeLa cells (figure 5.3, 5.4) and therefore it is reasonable to assume that VIPAR and PLOD3 co-localise to the TGN area in these cells. Co-localisation analysis of endogenous VIPAR with TGN-46 and RAB11A provides evidence that the intracellular localisation of VIPAR, like PLOD3, is in the TGN area (figure 5.5). A partial overlap between VIPAR and TGN46 was observed even if VIPAR staining clearly resembles the TGN compartment (figure 5.4, 5.5). A different TGN marker, ϒ-Adaptin, was used to co-localise VIPAR and PLOD3 in the TGN (figure 5.6). ϒ-Adaptin is a component of the Golgi adaptor protein complex 1 that links clathrin to transmembrane proteins (e.g. receptors) in clathrin-coated vesicles (section 1.4.1). Endogenously expressed VIPAR and PLOD3 co-localise in the perinuclear area with ϒ-Adaptin (figure 5.6). This observation suggests the involvement of the VPS33B/VIPAR complex in PLOD3 trafficking between the trans-Golgi network (TGN) and endosomes (Nakatsu and Ohno, 2003). This is supported by the fact that AP-1 is involved in the formation of CCVs that exit the TGN route to endosomes (Almomani et al., 2012; Traub and Kornfeld, 1997).

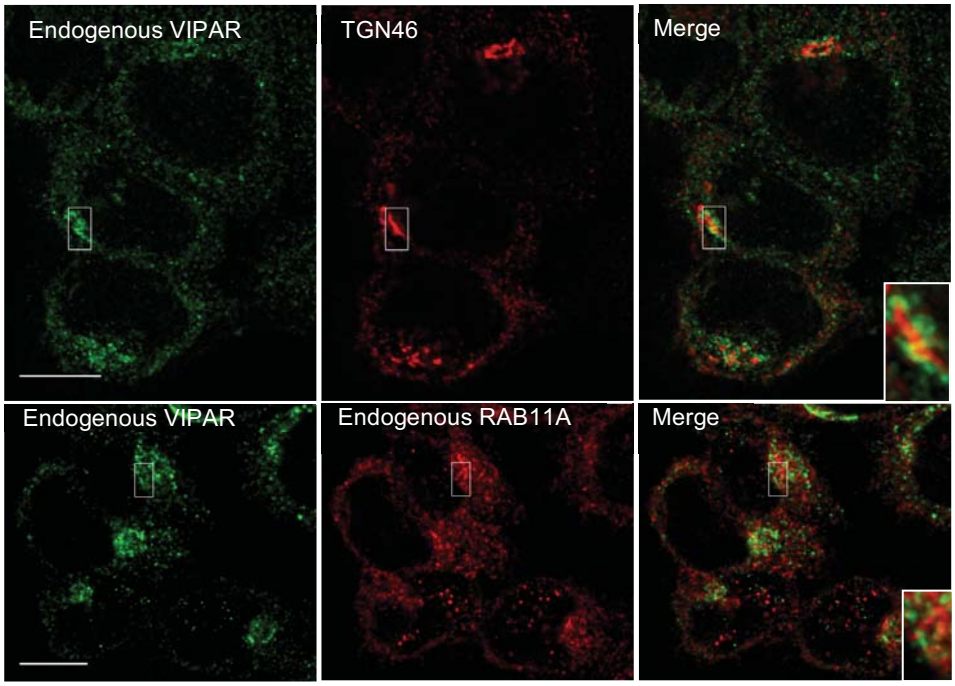


Figure 5.5 - Co-localisation analysis of endogenously expressed VIPAR with TGN46 and RAB11A in HeLa cells. Endogenous VIPAR, TGN46 and RAB11A were detected by immunostaining with an anti-VIPAR, anti-TGN46 and anti-RAB11A antibodies respectively. Projection of 1 optical Z-section at 0.25 µm is shown. The scale bar represents 10 µm.

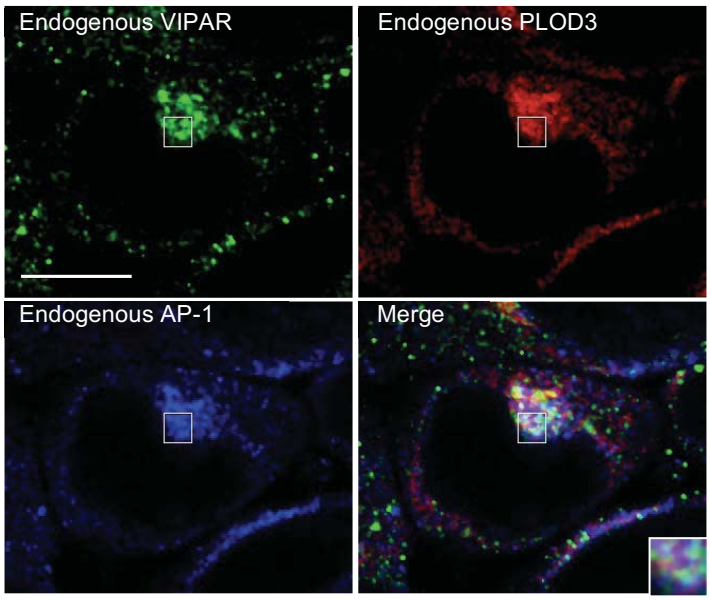


Figure 5.6 - PLOD3 co-localises with VIPAR and AP-1 in HeLa cells. Immunostaining was performed using an anti-VIPAR antibody, a fluorescently labelled anti-PLOD3 antibody and an anti- ϒ-Adaptin antibody for the immunodetection of the endogenous proteins. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.

5.4.3 - PLOD3 is abnormally trafficked in Vps33b/Vipar knockdown mIMCD-3 cells

As mentioned above (section 5.1) PLOD3 is secreted into the extracellular space surrounding some tissues such as kidney, spleen, muscle and liver. It is therefore reasonable to assume that trafficking of PLOD3 from TGN occurs prevalently in some cell types and VPS33B/VIPAR complex is required in one of the PLOD3 trafficking routes in these cells. Mouse inner medullary collecting duct (mIMCD-3), a mouse polarized epithelial cell line, was therefore used to study PLOD3 trafficking. Interestingly the kidney is one of the affected organs in ARC syndrome and Vps33b/Vipar knockdown mIMCD-3 cell lines should mimic the effect of VPS33B/VIPAR absence in ARC syndrome patient's epithelial cells. These knockdown cells are therefore used as a cell model system for the understanding of the pathology of ARC (Cullinane et al., 2010).

Knockdown of VPS33B in these cells lead to down-regulation of VIPAR (figure 5.11) and the same behaviour has been observed in ARC patients' fibroblasts with mutations in the *VPS33B* gene (Cullinane et al., 2010). Several cases of the co-regulation of expression of proteins which are part of the same complexes have been reported in literature (Luo et al., 2010), (Tan et al., 2007). Additionally ARC patients carrying mutations in *VPS33B* or *VIPAR* show no difference between them with respect to clinical symptoms, signs or disease course (section 3.1) and mIMCD-3 cells treated with Vipar-shRNA have the same phenotype as Vps33b-shRNA treated cells (Cullinane et al., 2010). From these data it is possible to conclude that the effect of Vipar or Vps33b stable knockdown mIMCD-3 cells have identical defects and can be both used in the experiments described in this chapter. As mentioned above, knockdown of

Vps33b/Vipar in these cells causes defects in adherens and tight junction formation with loss of polarity and cell-cell contacts. Full polarization of wild-type mIMCD-3 cells is reached by culturing the cells on transwell supports for two weeks. After 5-days of growth, a cavity formation (that resembles the renal tubular structure) is observed in wild-type cells but not in Vps33b-Vipar knockdowns. Endogenous PLOD3 staining was performed in both wild-type and Vps33b knockdown mIMCD-3 cells after they were cultured on transwell supports for two weeks. Additionally a control mIMCD-3 cell line (N/S), obtained with an shRNA containing a scrambled sequence, was used as a further control in this experiment (figure 5.7). A change in localisation of endogenous PLOD3 was observed in Vps33b knockdown mIMCD-3 cells when compared to both wild-type and control (figure 5.7). It is therefore possible that PLOD3 trafficking is disturbed in these cells, which is associated with the absence of VPS33B/VIPAR. This is the first indication that the VPS33B/VIPAR complex may be involved in PLOD3 trafficking in polarized mIMCD-3 cells. As mentioned above, PLOD3 interacts with VIPAR at the TGN compartment, therefore it could be hypothesised that a defect in trafficking from/to TGN may be found in Vps33b/Vipar knockdown cells. To further investigate trafficking of PLOD3 in these cells, immunofluorescence microscopy experiments were carried out by transfecting pre-polarized wild-type and Vipar knockdown mIMCD-3 cells with mCherry-PLOD3 and GFP-TGN38, a marker of TGN in mouse cells (figure 5.8). Interestingly the expression levels of endogenous PLOD3 was observed to be markedly reduced in pre-polarized wild-type mIMCD-3 compared to fully polarized cells, and this was the experimental reason for using overexpressed mCherry-PLOD3 as opposed to the endogenous protein. This observed change in PLOD3 expression with polarization of mIMCD-3 cells might be an indication of the role of PLOD3 in cell

polarity. An accumulation of PLOD3 at the TGN is observed in Vipar knockdown cells but not in the wild-type (figure 5.8). In wild-type mIMCD-3 cells PLOD3 did not co-localise with TGN and therefore further immunofluorescence experiments were performed to investigate the localisation of PLOD3 in these cells. Several markers for different intracellular compartments (including early and late endosomes and lysosomes, endoplasmic reticulum and cis-Golgi and recycling endosomes) were used for this purpose and it was found that transfected mCherry-PLOD3 co-localised with endogenous RAB11A (figure 5.9). This is in line with the previously shown co-localisation between PLOD3 and RAB11A in HeLa cells and fibroblasts (figure 5.4, 5.5). Interestingly mCherry-PLOD3 partially co-localises with endogenous-LAMP1 (a marker of late endosomes and lysosomal compartment) that was not seen in HeLa cells. No co-localisation was observed between mCherry-PLOD3 and Calreticulin (an ER marker), or GFP-RAB4 / GFP-RAB5 (early endosome markers). mCherry-PLOD3 was localised in close proximity to GM130, a marker for cis-golgi network, but no co-localisation between the proteins was observed. Therefore it is possible to conclude that VPS33B/VIPAR complex is involved in PLOD3 trafficking from TGN via a RAB11A dependent vesicular transport and knockdown of this protein complex causes the accumulation of PLOD3 in the TGN compartment (figure 5.8).

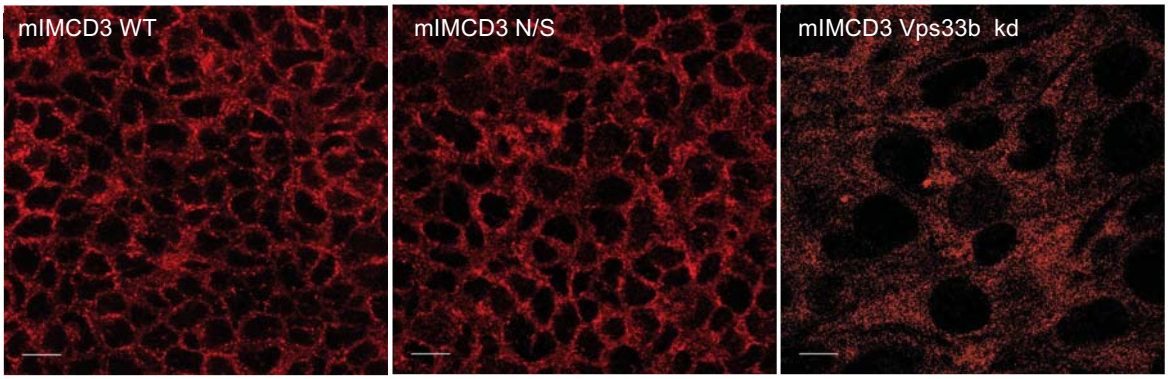


Figure 5.7 - Endogenous PLOD3 staining in wild-type, Control (N/S) and Vps33b stable knockdown mIMCD-3 cells grown on transwell supports for two weeks.
All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.

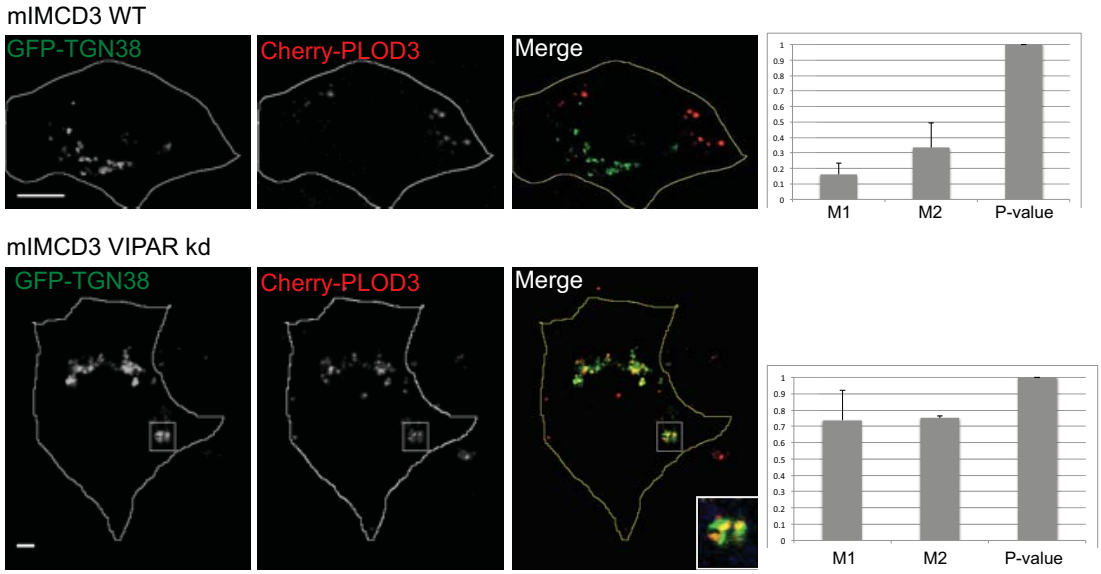


Figure 5.8 - Confocal immunofluorescence of cultured pre-polarized wild-type and VIPAR knockdown mIMCD-3 cells transfected with mCherry-PLOD3 and GFP-TGN38. Co-localisation of mCherry-PLOD3 with GFP-TGN38 is observed only in VIPAR kd mIMCD-3 cells. Endogenous PLOD3 co-localises also with over-expressed YFP-VPS33B. For each co-localisation analysis of the Manders' overlap coefficients M1 and M2, as well as the P-value, are represented as the average of n>3 cells. A rough outline around the cells was drawn with ImageJ. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.

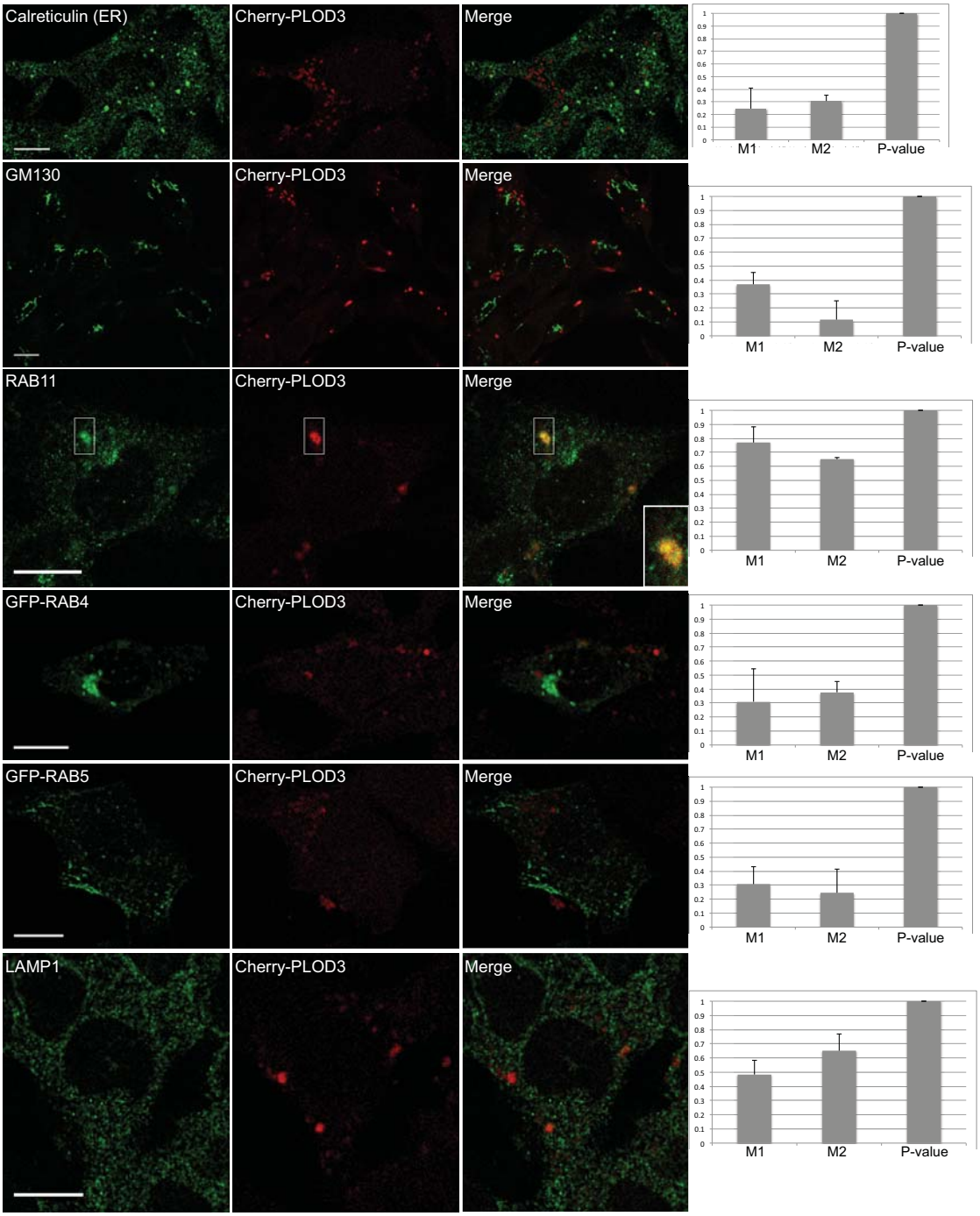


Figure 5.9 - Co-localisation of mCherry-PLOD3 with different intracellular markers in pre-polarized wild-type mIMCD-3 cells. mCherry-PLOD3 co-localises with endogenous RAB11A and partially with endogenous-LAMP1. No co-localisation with Calreticulin or GFP-RAB4 / GFP-RAB5 was seen. mCherry-PLOD3 was localised in close proximity to GM130 but no co-localisation between the proteins was observed. For each co-localisation analysis of the Manders' overlap coefficients M1 and M2, as well as the P-value, are represented as the average of n>3 cells. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.

5.4.4 - VPS33B/VIPAR and PLOD3 express co-ordinately in mIMCD-3 cells

In section 5.4.3 the expression and localisation of PLOD3 in Vps33b/Vipar knockdown mIMCD-3 cells was analysed by immunostaining and mislocalisation of intracellular PLOD3 was found in these cells when compared to controls. In this section the expression of VPS33B/VIPAR was analysed in mIMCD-3 cells with Plod3 knockdown. In order to perform this analysis stable knockdown mIMCD-3 cell lines were created for Plod3, Vps33b and Vipar. The level of knockdown for each cell line was tested by Western blot (figure 5.10). 76.2% knockdown was achieved for VPS33B and 89% for VIPAR (figure 5.10). The survival rate of single clones was very low when stable knockdown of PLOD3 was attempted in these cells. This observation is in agreement with data in literature that show how the depletion of GGT activity of PLOD3 in HT-1080 cells by siRNA leads to arrest of growth followed by lethality (Wang et al., 2009). Among n>30 clones tested, it was possible to isolate only two that were viable and had a decreased intracellular levels in PLOD3. The level of knockdown achieved for each clone were respectively 76.8% and 82.6% (figure 5.10). Both clones were tested for VPS33B and VIPAR expression by Western blot (figure 5.11). VPS33B and VIPAR were both markly down-regulated in Plod3 knockdown mIMCDs (for one Plod3 knockdown clone VPS33B was down-regulated by 81.2 % and VIPAR by 75.3 %, while for the second clone VPS33B was down-regulated by 75.4 % and VIPAR by 88.2% (figure 5.11). A concomitant down-regulation of VIPAR was observed in Vps33b knockdown mIMCD-3 cells (figure 5.11). As mentioned above, the common regulation of proteins that interact mutually as a complex to implement a common biological function is not unusual and several examples have been reported in literature.

Therefore the down-regulation of both VPS33B and VIPAR in PLOD3 knockdown cells is a further indication of the crucial role of VPS33B and VIPAR in PLOD3 trafficking. As shown in section 5.4.3, PLOD3 is mislocalised in Vps33b/Vipar knockdown mIMCD-3 cells but its intracellular expression levels are not changed (as quantified by Western-blot). Thus PLOD3 expression is not dependent on VPS33B/VIPAR interaction.

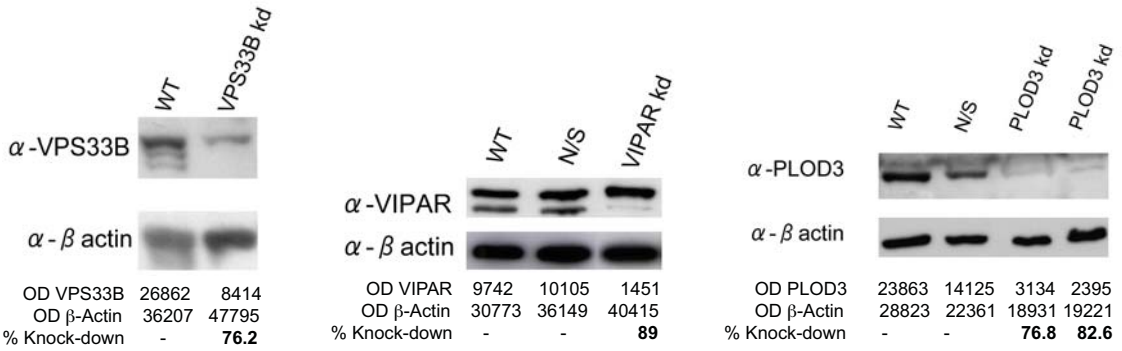


Figure 5.10 - Assessment of Vps33b, Vipar and PLOD3 knockdowns in mIMCD-3 cells. Knockdown efficiency in Vps33b, Vipar and Plod3 shRNA treated mIMCD-3 cells was tested by Western blot and successive densitometry on bands was carried out. Results are normalised to β-actin obtained by immune-blotting the same membrane with an anti-β-actin antibody. Knockdown efficiency was calculated relative to both wild-type (WT) and control (N/S) cells and a mean value is reported. Several clones (n>20) were tested and only the clones with the greatest level of knockdown, used for further experiments, are shown in each blot.
An additional non-specific band is recognised by the anti-VIPAR antibody in WT, N/S and VIPAR knockdown mIMCD-3. Therefore only the band that disappears in the Vipar knockdown clone was considered VIPAR-specific.

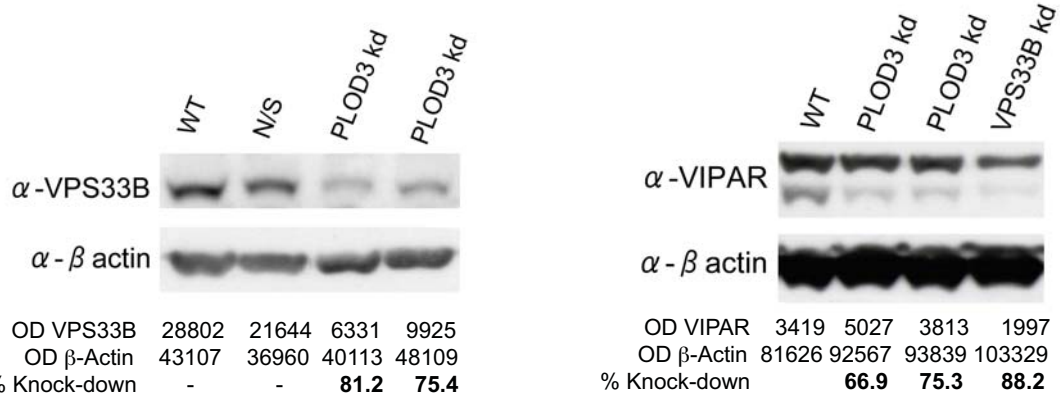


Figure 5.11 - Assessment of VPS33B and VIPAR expression in PLOD3 knockdown mIMCD-3 cells. Immnuoblotting with an anti-Vps33b antibody was performed in Plod-3 knockdown mIMCD-3 and densitometry on bands was carried out. The percentage of VPS33B down-regulation in these cells was calculated after -β-actin normalisation (left blot). Similarly, the levels of VIPAR expression in Plod3 knockdown mIMCD-3 cells were tested by immunoblot with an anti-VIPAR antibody and successive densitometry assessment was carried out. The lower recognized band only is VIPAR specific. Results were normalized to β-actin (right blot). Additionally the levels of VIPAR expression were tested in Vps33b knockdown mIMCD-3 cells and β-actin normalisation was carried out as above.

The close relationship between co-regulation, co-expression and protein-protein interactions between PLOD3, VPS33B and VIPAR was observed also when a cellular fractionation experiment was performed in HEK293 for cell subdivision into four fractions: cytosol, membrane, nucleus and cytoskeleton (section 5.3.2). The use of HEK293 cells was favoured in this experiment due to the high transfection efficiency in this cell type compared to HeLa and mIMCD-3 cells. The expression of endogenous PLOD3 was markedly enhanced when HEK293 cells were transfected with myc-VIPAR compared to untransfected cells (figure 5.12). Additionally endogenous PLOD3 expression was further increased when HEK293 cells were co-transfected with myc-VIPAR and HA-VPS33B, compared to both untransfected and myc-VIPAR only transfected cells (figure 5.12). Only a very faint band for PLOD3 was observed in the membrane fraction in untransfected HEK293 cells. This is in agreement with data in

literature suggesting that PLOD3 is a membrane-associated protein inside the lumen of organelles. The expression of PLOD3 in the membrane fraction is markely increased in Vps33b/Vipar transfected cells. This indicates that the over-expression of VIPAR alone and more markedly of the VPS33B/VIPAR protein complex, enhances the expression of PLOD3 in HEK293 cells. PLOD3 is seen also in the nuclear fraction in transfected cells (figure 5.12). This is attributed to a cross-contamination between membrane and nuclear fraction that is very common in this type of experiments due to the physical communication between nuclei and ER membranes. This was confirmed by immunoblotting the fractions with an anti-PDI antibody whose ER location has been established in literature (figure 5.14). The presence of PDI in both nuclear and membrane fraction confirms the cross-contamination between these two fractions.

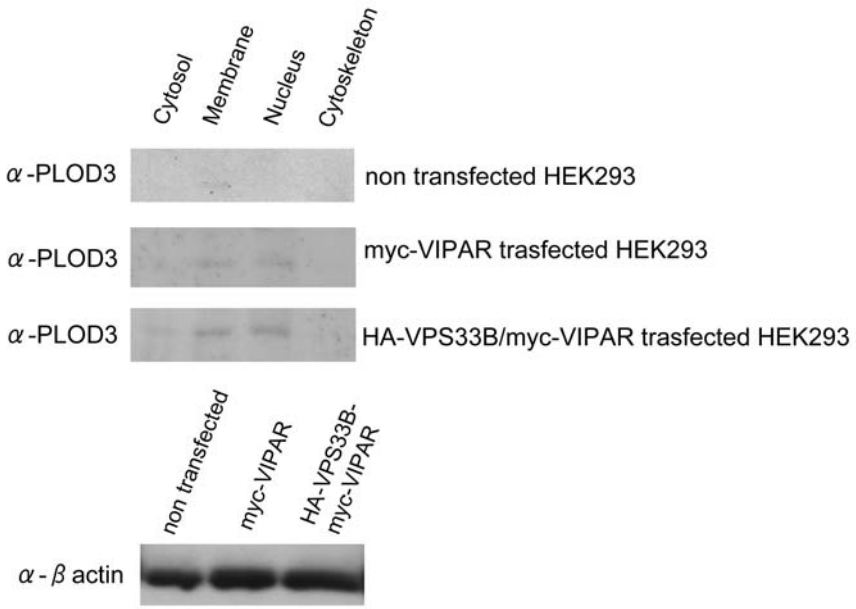


Figure 5.12 - Cell-fractionation of untransfected, myc-VIPAR transfected, and HA-VPS33B/myc-VIPAR transfected HEK293 cells for analysis of endogenous PLOD3 presence in each fraction. Cell fractionation was carried out using the ProteoExtract® Subcellular Proteome Extraction Kit (Merck4Biosciences) for HEK293 cell subdivision into four fractions: cytosol, membrane, nucleus and cytoskeleton. Immunobloting of each fraction was performed with an anti-PLOD3 antibody for the detection of the endogenous protein. The total cell lysate, prior to the fractionation, of each pool of HEK293 cells was blotted with an anti-β-actin antibody.

5.4.5 - The topology of VPS33B/VIPAR/PLOD3 interaction: VIPAR is a transmembrane protein

The fractionation experiments described in the previous section were consistent with the intraluminal localisation of PLOD3 previously reported in literature (Myllyla et al., 2007). To date the topology of VPS33B/VIPAR is not known but cytoplasmic clusters of VPS33B/VIPAR complex have been visualised intracellularly by confocal and electron microscopy (Cullinane et al., 2010; Smith et al., 2012). In this section experiments were performed in order to establish the topology of VPS33B/VIPAR/PLOD3 interaction and the possibility, suggested by FRET-FLIM measurements (section 4.4.3.4), that a direct interaction between PLOD3 and VIPAR can occur. Before proceeding with the experimental work, a bioinformatics analysis was performed for all proteins subject of this study, in order to gain more information regarding their membrane *vs* cytosolic localisation. Several prediction servers were used to predict putative transmembrane regions along VPS33B, VIPAR and PLOD3 protein sequence including: PredictProtein (www.predictprotein.org), Split server (http://split4.pmfst.hr/split/4/), "DAS" - Transmembrane Prediction server (http://www.sbc.su.se/~miklos/DAS/), TopPred2: Topology prediction of membrane proteins file://localhost/(http/::www.sbc.su.se:~erikw:toppred2:), HMMTOP:Prediction of transmembrane helices and topology of proteins (http://www.enzim.hu/hmmtop/). These servers were all in agreement that no transmembrane helices were present in VPS33B and PLOD3, which confirms previous data from literature regarding their respective cytosolic and luminal localisation. For VIPAR a different scenario was presented. PredictProtein, Split server and DAS databases predict two transmembrane helices in VIPAR located around 140-160 amino acid for the first helix and 270-290

amino acid for the second helix (figure 5.13). TopPred2 and HMMTOP predicted only the second transmembrane domain (amino acid 270-290). Additionally, all the above servers predicted the N-terminal region of VIPAR to be located on the luminal side and its C-terminal on the cytosolic side. These predictions nicely match with previous experiments that identified the N-terminal region of VIPAR interacting with PLOD3 and its C-terminal with VPS33B (sections 3.4.4, 4.4.3.5). Based on this analysis, it was possible to predict the topology of VPS33B/VIPAR/PLOD3 interaction: in this scenario VIPAR is a transmembrane protein, with three different hypothesis regarding the location of the transmembrane helices along its sequence (figure 5.13), while VPS33B is located in the cytosol and interacts with the C-terminal cytosolic domain of VIPAR. Thus, according to these predictions PLOD3 interacts with the luminal N-terminal domain of VIPAR in the lumen of the organelle/vesicle (figure 5.13).

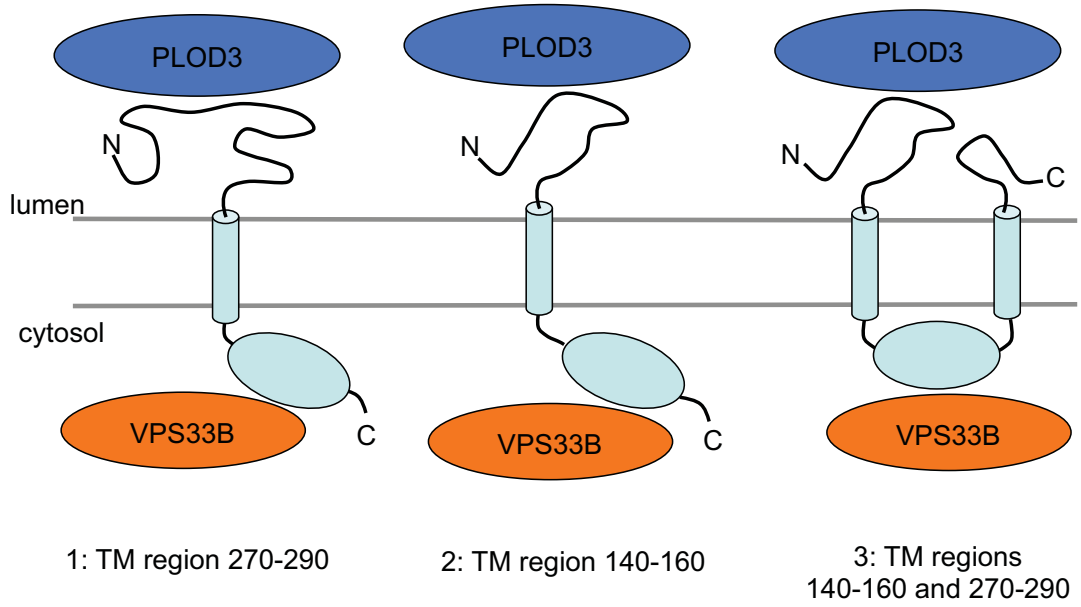


Figure 5.13 - Graphical model of the predicted topology of VPS33B-VIPAR-PLOD3 interaction with three possibilities in respect to the transmembrane domains (TMD) location in VIPAR. Several prediction servers were used to predict putative transmembrane regions along VIPAR protein sequence including: PredictProtein (www.predictprotein.org), Split server (http://split4.pmfst.hr/split/4/), "DAS" - Transmembrane Prediction server (http://www.sbc.su.se/~miklos/DAS/), TopPred2: Topology prediction of membrane proteins file://localhost/(http/::www.sbc.su.se:~erikw:toppred2:), HMMTOP:Prediction of transmembrane helices and topology of proteins (http://www.enzim.hu/hmmtop/). PredictProtein, Split server and DAS databases predict two transmembrane helices in VIPAR located around 140-160 amino acid for the first helix and 270-290 amino acid for the second helix. TopPred2 and HMMTOP predict only the second domain (amino acid 270-290). In all three scenarios VPS33B interacts with the C-terminal cytosolic domain of VIPAR, while PLOD3 with its luminal N-terminal domain.

To further validate the model described for VPS33B/VIPAR/PLOD3 topology, fractionation experiments were carried out with subsequent Western-blot analysis to detect the presence of endogenous VPS33B, VIPAR and PLOD3 in each fraction (section 5.3.1). Non-transfected HEK293 cells used in the fractionation experiment described in the previous section for the detection of endogenous PLOD3, were also used to detect the presence of endogenous VPS33B and VIPAR in the same fractions (figure 5.14). The low quality of the blot is related to the dilution of the protein during

the fractionation procedure that consequently lowers the efficency of recognition by the antibody. VIPAR, similarly to PLOD3 and PDI protein (used as a cross-contamination control protein), is located in the membrane and nuclear fraction and no band was detected in the lane that corresponds to the cytosolic fraction (figure 5.14). Endogenous VPS33B instead was predominantly located in the cytosolic fraction. As shown previously (section 3.4.1) VPS33B interacts directly with VIPAR and therefore the only explanation for VIPAR exclusive membrane localisation is its transmembrane nature and a possible luminal location was therefore excluded.

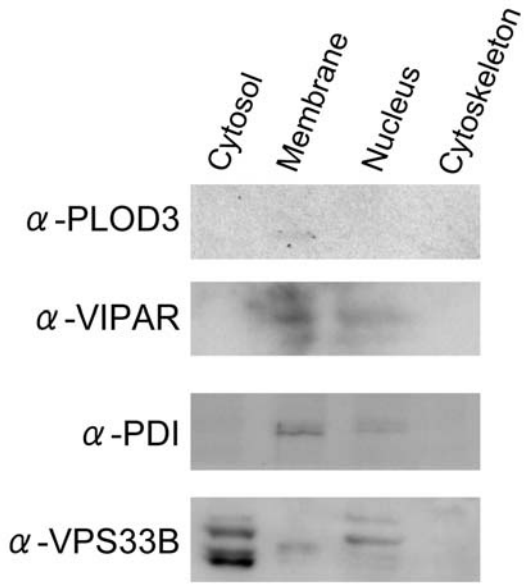


Figure 5.14 - Cell-fractionation of untransfected HEK293 cells and successive Western-blot analysis for the presence of endogenous VPS33B, VIPAR and PLOD3 in each fraction. Cell-fractionation was carried out using the ProteoExtract® Subcellular Proteome Extraction Kit (Merck4Biosciences) for HEK293 cell subdivision into four fractions: cytosol, membrane, nucleus and cytoskeleton. An additional immunoblotting of fractions was carried out with an anti-PDI antibody as a control to assess cross-contamination between fractions.

This data were further confirmed by a coarse fractionation experiment for the separation of the cytosol from crude membrane fractionation of HeLa and mIMCD-3 cells (both wild-type and Vipar knockdown mIMCD-3 cells were used) (section 5.3.2, figure 5.15). Endogenous VIPAR is exclusively located in the membrane fraction of HeLa and mIMCD-3 cells, both in wild-type and Vipar knockdown, confirming previous data (figure 5.15). As shown previously, cells blotted with anti-VIPAR antibody present a typical double band with only the lower molecular weight band being VIPAR-specific (figure 5.10). To exclude misleading results derived from the presence of two bands in the anti-VIPAR blot, unfractionated control and ARC-patient human fibroblast cells were loaded on the same blot for confirmation of VIPAR specific band in the fractionated samples (VIPAR specific band indeed disappears in ARC-patient fibroblast carrying a Vipar missense mutation compared to control cells) (figure 5.15). Additionally, the VIPAR-specific band is reduced in the membrane fraction of Vipar knockdown mIMCD-3 and this is a further confirmation of band-specificity and knockdown efficiency as previously described (figure 5.10). VPS33B is mainly found in the cytosolic fraction of HeLa cells and wild-type mIMCD-3 cells. A small fraction of VPS33B is found in the membrane fraction and this might indicate that VPS33B cycles between soluble and membrane associated (through its interaction with VIPAR), similarly to what observed for the yeast Vps33p that cycles between soluble and particulate forms in an ATP-dependent manner (Gerhardt et al., 2012). VPS33B was not detectable by the anti-VPS33B antibody in Vipar knockdown mIMCD-3 cells.

PLOD3 was predominantly located in the membrane fraction of all cell lines. A small amount of PLOD3 was found in the cytosolic fraction and the same was observed in HeLa cells for PDI, a known luminal ER protein, indicating that a small fraction of the

internal organelle membranes were disrupted by the homogenisation procedure (section 5.3.2) causing the release of their content into the cytosol. This observation is however very indicative in terms of confirming the transmembrane nature of VIPAR. VIPAR, unlike PLOD3 and PDI, is found only in the membrane fraction and possible membrane organelle disruption doesn't cause its release into the cytosol. This further excludes the possibility of VIPAR being a luminal protein as an explanation for its presence in the membrane fraction. The same behaviour was observed for LAMP1, a known transmembrane protein, found exclusively in the membrane pool of HeLa cells (figure 5.15).

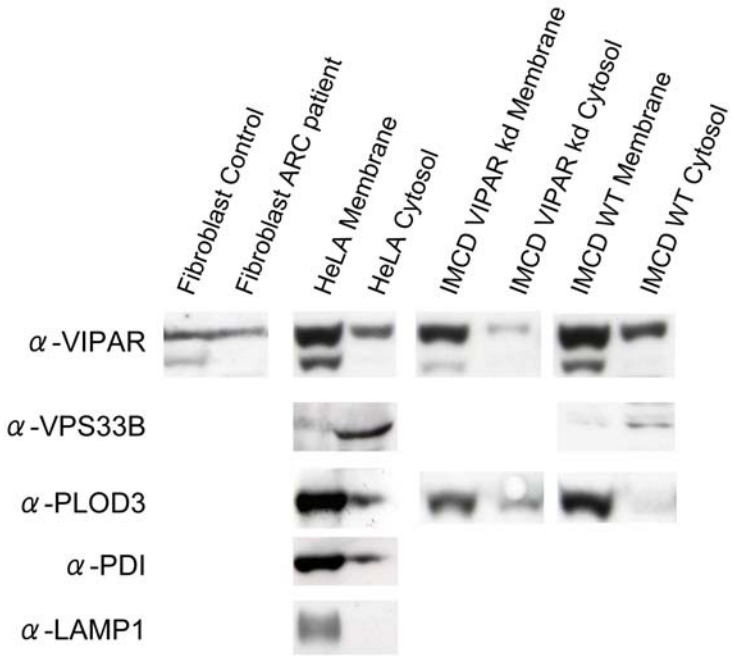


Figure 5.15 - Establishment of the cytosolic versus membrane location of endogenous VIPAR, PLOD3 and VPS33B in HeLa and mIMCD-3 cells. mIMCD-3 wild-type and Vipar knockdown cell lines, as well as HeLa cells, were used in the fractionation experiment for the separation of the cytosol from crude membranes. Both fractions for each cell-line were blotted with anti-VIPAR, anti-VPS33B and anti-PLOD3 antibody in order to determine the cytosolic versus membrane localisation of respective endogenous proteins. Due to the presence of two bands in the samples blotted with anti-VIPAR antibody, control and ARC-patient human fibroblast cells were loaded on the same blot for confirmation of VIPAR specific band (which disappears in ARC-patient fibroblasts carrying a Vipar missense mutation compared to Control cells).
To test for cross-contamination between fractions, cytosolic and membrane fractions from HeLa cells were blotted with antibodies against proteins with known intracellular localisation. Antibodies against PDI (a known luminal ER protein) and LAMP1 (a known transmembrane lysosmal protein) were used for this purpose. The same antibodies were not suitable for the recognition of the same mouse proteins in mIMCD-3 cells.

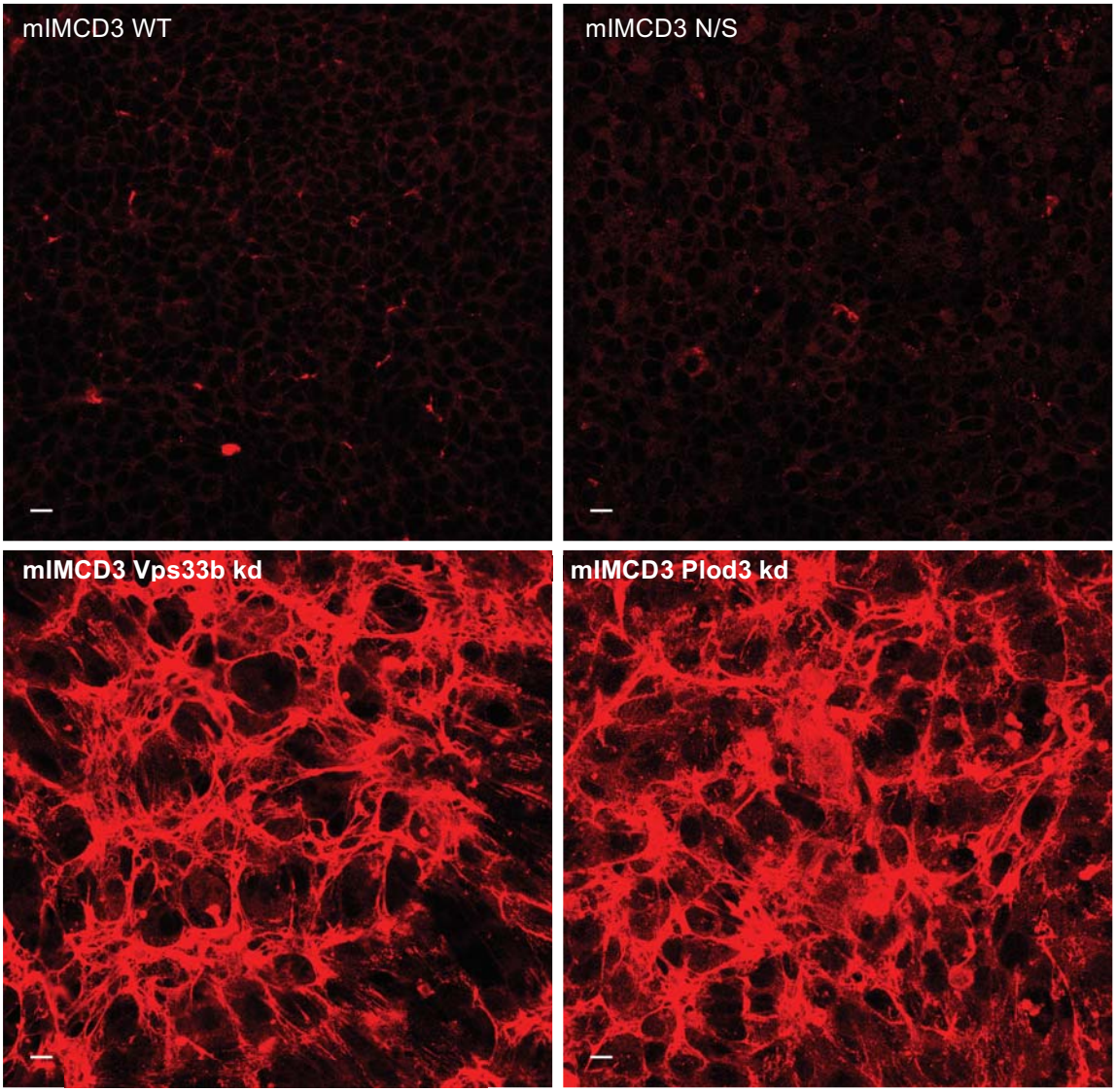
5.4.6 - Collagen type IV and E-cadherin are abnormally expressed in Vps33b/Vipar and Plod3 knockdown cell lines

PLOD3 is an enzyme that catalyses the hydroxylation of lysyl residues in collagen-like peptides as well as the galactosylation of hydroxylysines and glucosylation of galacotosylhydroxylysines. Therefore, to examine possible changes in collagen expression/modification caused by defects in PLOD3 trafficking/expression, immunostaining and immunoblotting of mIMCD-3 cell lines was performed using an anti-collagen type IV antibody. Collagen type IV was chosen among the other types of collagens for its high content in hydroxylated and hydroxyl-glycosylated lysines and also for being the major component of basement membranes of kidneys, remembering that the polarized cells used in this study derive from the inner medullary collecting duct (mIMCD-3) of mouse kidney (section 5.1). The expression of endogenous collagen type IV was analysed in both Vps33b and Plod3 knockdown mIMCD-3 cell lines and was compared to wild-type and control cells. Immunostaining and Western blot analysis demonstrate a abnourmous increase in the intracellular amounts of collagen IV in both Vps33b and Plod3 knockdown mIMCD-3 cells compared to controls (figure 5.16 A and B). This is in line with the data in the literature that demonstrate intracellular accumulation of collagen IV in PLOD3 deficiency, assigning a critical role to hydroxylysine glycosylation in the assembly and secretion of highly glycosylated collagen types (Sipila et al., 2007). From the Western-blot analysis, a slightly enhanced migration on SDS-PAGE of collagen IV chains is observed in Vps33b and Plod3 knockdown mIMCD-3 compared to controls, suggesting a possible difference in their hydroxylation/glycosylation pattern with consequent molecular weight reduction (figure

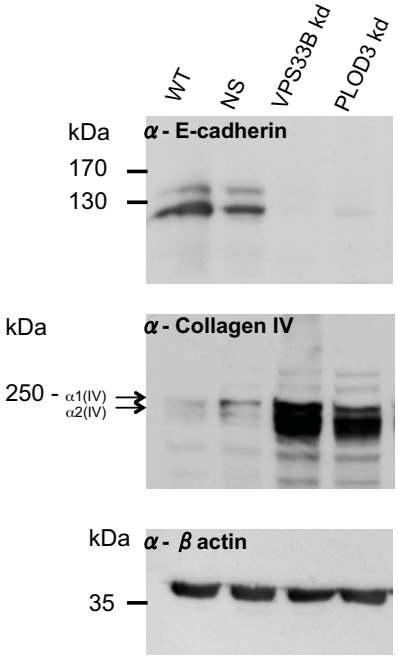
5.16 B). Similar behaviour was observed in E9.5 PLOD3 null embryo lysates compared to wild-type embryo lysates (Rautavuoma et al., 2004).

As mentioned previously (section 5.1) loss of E-cadherin expression and a typical EMT phenotype characterises Vps33b/Vipar knockdown mIMCD-3 cells. The expression of E-cadherin was analysed also in Plod3 knockdown cells and similarly to Vps33b/Vipar knockdowns, a loss in E-cadherin expression was seen (figure 5.16 B, C). This result was quantified by Western-blot (figure 5.16 B) and staining of mIMCD-3 cells cultured for 3 days in 3D type I collagen gels. The staining demonstrated a marked loss of E-cadherin and, to a lesser extent, of Claudin 1 in both Vps33b and Plod3 knockdown cells compared to controls (figure 5.16 C). In 3D cultures, control mIMCD-3 cells formed spheroid structures with a single layer of epithelial cells defining a well-established lumen. Vps33b and Plod3 knockdown cells didn't form the spheroid structures and cells did not grow in a monolayer but instead formed a disorganised mass. These results lead to the conclusion that an EMT phenotype is observed also in Plod3 knockdown mIMCD-3 cells similarly to Vps33b/Vipar knockdowns. Plod3 knockdown mIMCD-3 cells seem to mimic the phenotype of Vps33b/Vipar knockdowns in terms of collagen expression/structure, E-cadherin expression and lumen formation. Additionally, and similarly to what was observed in Vps33b and Vipar knockdown cells (Cullinane et al., 2010), proliferation was significantly increased in Plod3 knockdowns. The maximum level of Trans-Epithelial-Resistance (that measures paracellular permeability to ions), reached by Plod3 knockdown cells were ~40% reduced compared with controls, indicating a defect in junction formation. The similar phenotype of Vps33b, Vipar and Plod3 knockdown mIMCD-3 cell lines is a further indication of the involvement of these proteins in the same intracellular pathway.

A.


mIMCD3 WT
mIMCD3 N/S
mIMCD3 Vps33b kd
mIMCD3 Plod3 kd

B.


WT
NS
VPS33B kd
PLOD3 kd
kDa
α - E-cadherin
170
130
kDa
α - Collagen IV
250 -
α1(IV)
α2(IV)
kDa
α - β actin
35

C.

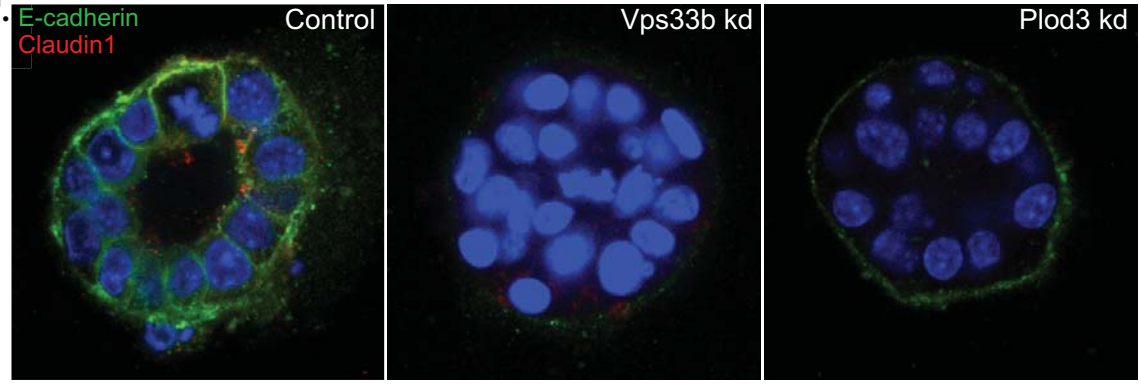


*Figure 5.16 - Endogenous collagen IV and E-cadherin expression in Vps33b and Plod3 knockdown mIMCD-3 cell lines compared to control cells. **(A)**. Immunostaining with anti-collagen IV antibody of mIMCD-3 cell lines (wild-type, control, Vps33b and Plod3 knockdown) grown on transwell supports for two weeks. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.*
*(**B**). Western blot analysis of collagen IV and E-cadherin expression in pre-polarized mIMCD-3 cell lines with a specific anti-collagen IV antibody that recognasies the α1 and α2 chains, and an anti-E-cadherin antibody. The same membrane was blotted with an anti-β-actin antibody as a loading control.*
***(C).** mIMCD-3 cells cultured for 3 days in 3D type I collagen gels before staining with anti-E-cadherin and anti-Claudin antibodies was carried out. Control mIMCD-3 cells formed large spheroid structures with a single layer of epithelial cells that define a well-established lumen. Vps33b and Plod3 knockdown cells don't form the spheroids and cells do not grow in a monolayer. Loss of E-cadherin expression is visible in both Vps33b and Plod3 knockdown cells compared to control. Experiment carried out in collaboration with A Straatman Iwanowska.*

To better understand the relevance of PLOD3 and collagen modifications/expression in ARC syndrome, human skin fibroblast derived from ARC patients and age-matched controls were analysed by Western blot for collagen type I expression. The importance of the GT and GGT activity of PLOD3 has been indeed demonstrated in osteoblasts and an impairment of this LH3 function significantly affects type I collagen fibrillogenesis (Sricholpech et al., 2011). This study shows an accelerated fibrillogenesis in PLOD3 knockdown mouse osteoblastic cell lines (Sricholpech et al., 2011).

A abnourmous intracellular accumulation of collagen I was seen in human skin fibroblast derived from ARC patients compared to age-matched controls (figure 5.17)

similarly to what was previously seen for collagen IV in Vps33b/Vipar and Plod3 knockdown mIMCD-3 cells (figure 5.16B). The consistency of the data between mIMCD-3 cells, the *in-vitro* cell model for ARC, and cell lines derived from patients carrying the disease, indicate plausibility in the observed data and formulated hypothesis.

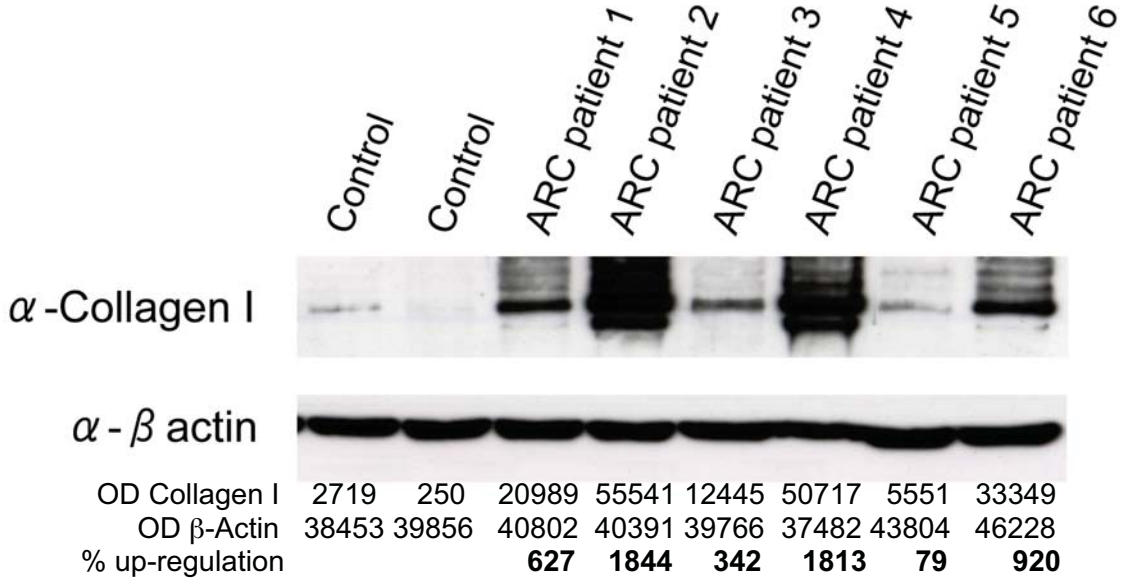


Figure 5.17 - Collagen I expression in human skin fibroblast derived from ARC patients and age-matched controls. Collagen I expression in human skin fibroblast cell lines was tested by Western blot and then densitometry was carried out. Results are normalized to β-actin obtained by immunoblotting the same membrane with an anti-β-actin antibody. Levels of collagen I up-regulation in ARC patient fibroblasts were calculated relative to the first age-matched control in the figure.

One of the major goals of this work is to understand the mechanisms that lead to the E-cadherin down-regulation and consequent EMT phenotype in Vps33b/Vipar mIMCD-3 cell lines. As mentioned above, similarly to Vps33b/Vipar knockdown mIMCD-3 cells, the down-regulation of E-cadherin occurs also in Plod3 knockdown cells. This suggests a common mechanism involving VPS33B, VIPAR and PLOD3 that directly or

indirectly regulates the expression of E-cadherin in these cells. A correlation between E-cadherin down-regulation and abnormal collagen expression has been documented in literature (Espinosa Neira and Salazar, 2012), therefore it is possible to hypothesise that abnormal collagen formation and secretion in Vps33b/Vipar and Plod3 knockdown in mIMCD-3 cells could lead to E-cadherin down-regulation in these cells. In order to understand the possible effect of collagen expression on E-cadherin down-regulation in Vps33b knockdown mIMCD-3 cells, a time course experiment was performed analysing simultaneously the expression of β-actin, E-cadherin and collagen IV on transwell plates growing wild-type and Vps33b knockdown mIMCD-3 cell lines (figure 5.18). From very early stages, at 30 minutes after cells seeding on transwell supports, an increase in collagen IV was seen in Vps33b knockdown cells compared to wild-type (figure 5.18). At this stage, some E-cadherin was still present in the knockdown cells that could not be observed at later stages (figure 5.18). It is therefore possible to assume that the progressive accumulation of collagen IV, that can be seen at later time points in Vps33b knockdown cells, can lead to the progressive loss of E-cadherin expression in these cells. It is important to underline that mIMCD-3 cells with a stable knockdown of Vps33b were used in this experiment and it is therefore not possible to monitor the expression of collagen and E-cadherin at the very first stages of the knockdown establishment. This experiment was useful also for monitoring of the time-course of the establishment of polarity in mIMCD-3 wild-type cells. Predominant intracellular E-cadherin staining was observed at the first stages after cell seeding on transwell supports (from 30 minutes to 6 hours) with progressive junctional staining corresponding to cell-cell contact sites noticeable after 24 hours of growth on transwell supports. The complete polarization was reached after 2 weeks when both E-cadherin and β-actin

were concentrated at the cell-to-cell adhesion regions. Loss of this typical actin cytoskeleton organisation was observed in Vps33b knockdown cells (figure 5.18).

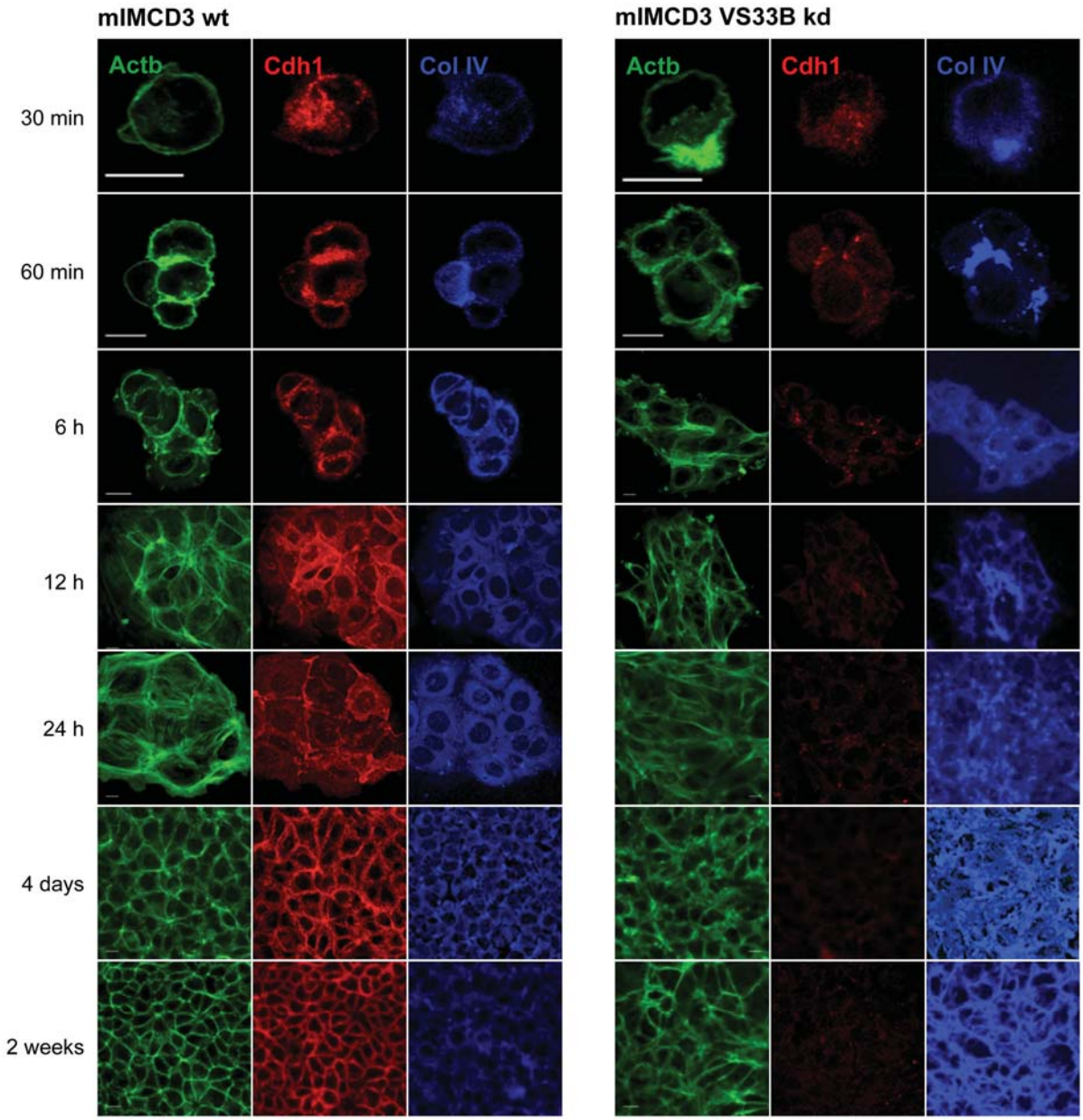
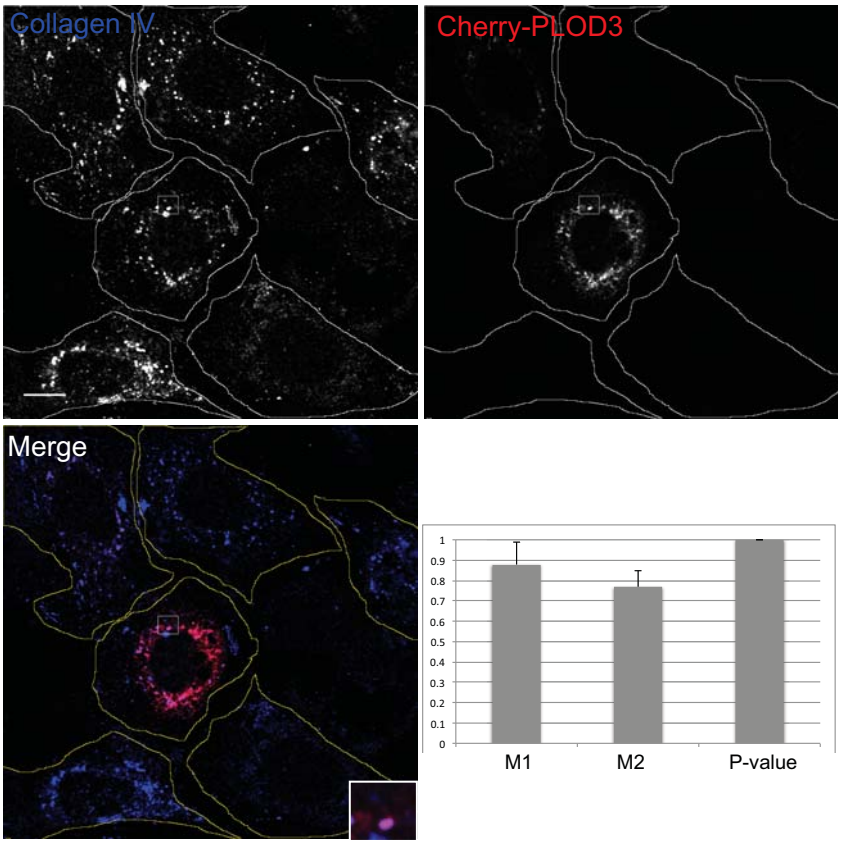


Figure 5.18 - Time course experiment analysing the expression of β-actin, E-cadherin and collagen IV on transwell growing mIMCD-3 cell lines. Wild-type and Vps33b knockdown mIMCD-3 cells were seeded at the same starting point on 7 different transwell supports per cell line. Cells were fixed in 4% PFA at different time points that range from 30min to 2-weeks. After fixation, staining was performed with anti-β-actin, anti-E-cadherin and anti-collagen IV antibodies for each time point of both wild-type and Vps33b knockdown cell lines. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm.

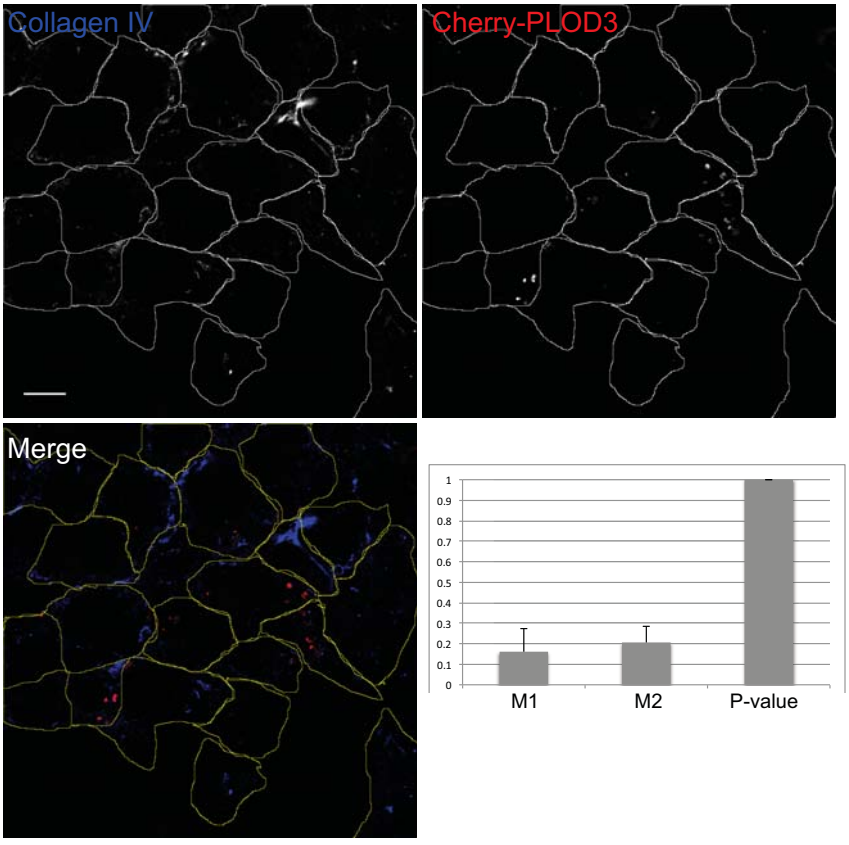
5.4.7 - VPS33B-VIPAR complex is involved in PLOD3 delivery to collagen

As mentioned early in this Chapter, defects in collagen secretion with its subsequent intracellular accumulation, have been associated to loss of GGT activity of PLOD3 (Sipila et al., 2007). In agreement with this data, an increase in the intracellular accumulation of collagen IV was observed in Plod3 knockdown mIMCD-3 cells. The same abnormal increase in intracellular collagen IV was seen also in Vps33b/Vipar knockdown mIMCD-3 cells. It was therefore possible to propose the existence of a novel mechanism for collagen intracellular trafficking involving Vps33b/Vipar/Plod3. To test this hypothesis, co-localisation analysis of endogenous collagen IV with transfected mCherry-PLOD3 was performed in wild-type and Vipar knockdown mIMCD-3 cells. Collagen type IV co-localises with mCherry-PLOD3 in control but not in Vipar knockdown cells (figure 5.19). This experiment therefore indicates that VPS33B/VIPAR complex is crucial for co-localisation and possibly trafficking of PLOD3 to collagen and in the absence of this complex in Vipar knockdown cells, PLOD3 does not modify intracellular collagen IV, which affects its intracellular trafficking (and possibly secretion). Therefore collagen trafficking is affected in PLOD3 knockdown cells (due to the reduced amounts of PLOD3 in these cells), as well as in Vipar knockdown cells (due to a possible failure in the delivery of PLOD3 to collagen). To further validate this data, a rescue experiment was performed by transfecting Vipar knockdown mIMCD-3 cells with YFP-VPS33B/myc-VIPAR. The co-localisation between collagen type IV and mCherry-PLOD3 was achieved in transfected cells confirming the role of VPS33B/VIPAR in the delivery of PLOD3 to collagen (figure 5.19).

mIMCD3 WT



mIMCD3 VIPAR kd



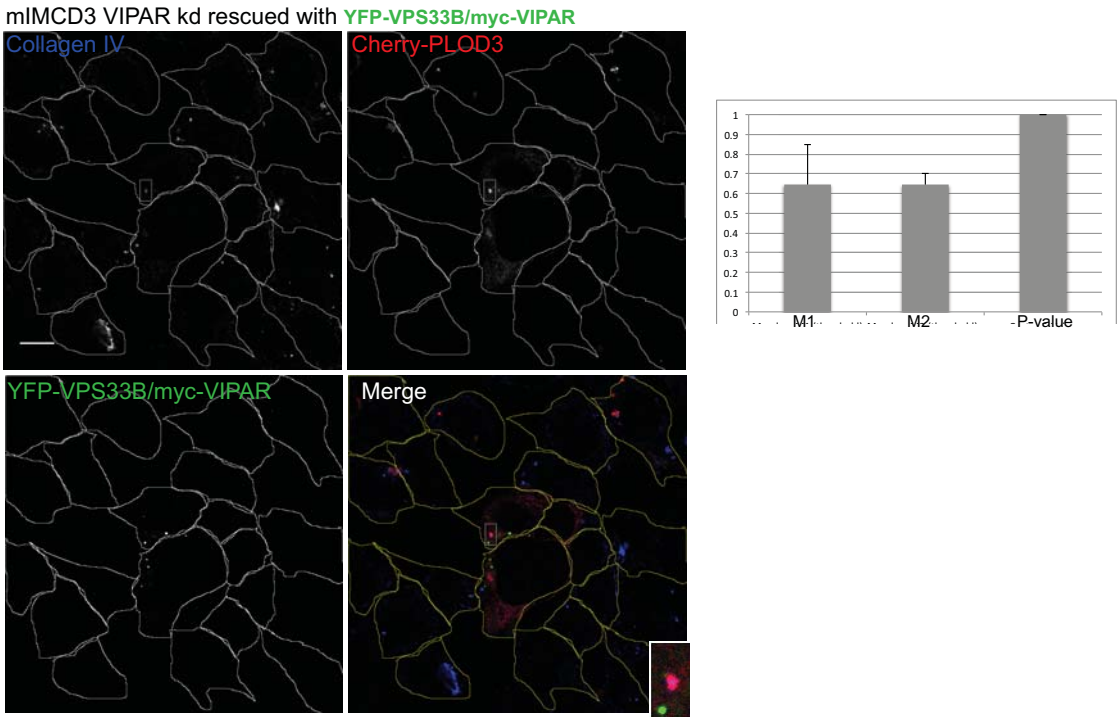


Figure 5.19 - Co-localisation analysis of endogenous collagen IV with mCherry-PLOD3 in wild-type, Vipar knockdown, and Vipar knockdown YFP-VPS33B/myc VIPAR transfected, pre-polarized mIMCD-3 cells. Confocal immunofluorescence of cultured pre-polarized wild-type and VIPAR knockdown mIMCD-3 cells transfected with mCherry-PLOD3 and stained for endogenous collagen IV. Collagen type IV co-localises with mCherry-PLOD3 in control but not in Vipar knockdown cells. The rescue experiment was performed by transfecting Vipar knockdown mIMCD-3 cells with YFP-VPS33B/myc-VIPAR. Collagen type IV co-localises with mCherry-PLOD3 in transfected cells. Endogenous PLOD3 co-localises also with over-expressed YFP-VPS33B. For each co-localisation analysis of the Manders' overlap coefficients M1 and M2, as well as the P-value, are represented as the average of n>3 cells. A rough outline around the cells was drawn with ImageJ. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm. Experiment carried out in collaboration with A.Straatman Iwanowska.

Furthermore, the same cells in the experiment described in figure 5.8 (performed by transfecting cultured pre-polarized wild-type and Vipar knockdown mIMCD-3 cells with mCherry-PLOD3 and GFP-TGN38), were additionally stained for collagen IV. Round granular structures can be seen by co-localising PLOD3 with collagen IV in wild-type cells but the localisation of these granules is not the TGN compartment

(figure 5.20). It is therefore possible to hypothesise that PLOD3 binding to collagen occurs in correspondence of collagen carriers and not at the TGN level. The co-localisation between PLOD3 and collagen IV does not occur in Vipar knockdown mIMCD-3 cells and additionally an accumulation of mCherry-PLOD3 is observed at the TGN compartment of the same cells (figure 5.8, figure 5.20). This further confirms the previous hypothesis stating that PLOD3 binding to collagen does not occur at the TGN compartment.

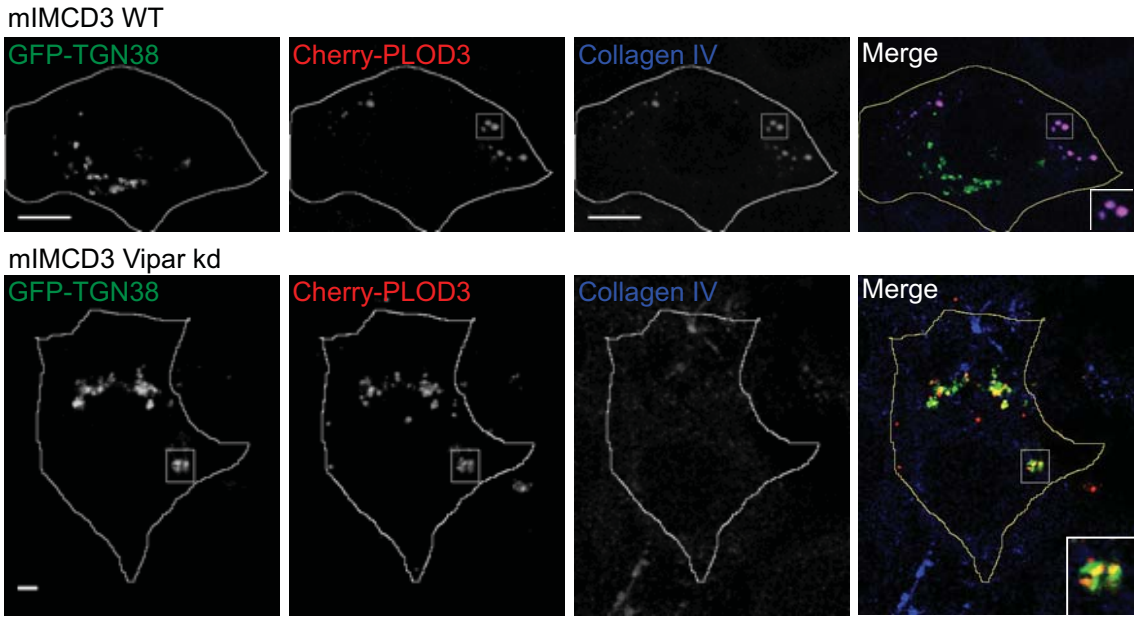


Figure 5.20 - Confocal immunofluorescence of cultured pre-polarized wild-type and Vipar knockdown mIMCD-3 cells transfected with mCherry-PLOD3 and GFP-TGN38 and stained for collagen IV. Collagen IV co-localises with PLOD3 only in wild-type cells but no co-localisation with GFP-TGN-38 was observed. Co-localisation of mCherry-PLOD3 with GFP-TGN38 occurs only in VIPAR knockdown mIMCD-3 cells and no co-localisation between collagen IV and PLOD3 can be seen in these cells. A rough outline around the cells was drawn with ImageJ. All images are projection of one optical section at 0.25 µm Z-spacing. The scale bar represents 10 µm. The co-localisation analysis of the Manders' overlap coefficients M1 and M2, as well as the P-value, are shown in figure 5.8 and 5.19. Experiment carried out in collaboration with A.Straatman Iwanowska.

5.5 - DISCUSSION

The novel PLOD3-VIPAR interaction, unambiguously confirmed in the previous chapter, was here characterised at subcellular level in order to elucidate its potential relevance in the pathogenesis of the ARC syndrome. In order to understand the trafficking pathway involving VPS33B/VIPAR/PLOD3 it was crucial to determine the intracellular localisation of these proteins that is not yet fully characterised. In contrast to the previous reports (Heikkinen et al., 2000; Myllyla et al., 2007; Salo et al., 2006a; Salo et al., 2006b) it was possible to demonstrate that PLOD3 is not an ER resident protein. Unlike the PLOD1 isoform, whose ER-localisation was well-established (Kellokumpu et al., 1994), PLOD3 does not co-localise with the luminal ER protein PDI but is instead found in the Trans-Golgi-Network compartment as shown by co-localisation with TGN-46, p230 and AP1 markers in HeLa cells and fibroblasts. This finding, despite being a contradiction to the previous statements about PLOD3 intracellular localisation, is not completely surprising. As mentioned above, PLOD3 is unique among PLOD isoforms for its localisation and enzymatic activity. Apart from intracellular location, PLOD3 is found also in the extracellular space and at the cell surface. Different trafficking pathways that ensure PLOD3 is delivered to these diverse locations have been suggested. PLOD1 and PLOD2 isoforms, on the other hand are thought to be exclusively ER resident proteins and do not appear to require specific trafficking pathways. Additionally PLOD3 is the only isoform that catalyses the galactosylation of hydroxylysines and glucolsylation of galacotosylhydroxylysines and these post-translational modifications are known to occur in the Golgi complex. The identification of VIPAR at TGN demonstrated by the work described in this thesis, represent the first evidence that establishes a subcellular compartmentalisation for

endogenous VIPAR. This is in agreement with prediction analyses that assign VIPAR to a Golgin subfamily A member 5 domain. Members of Golgin family are coiled-coil motif proteins that mediate tethering of vesicles to Golgi membranes and cisternal membranes to each other (section 1.3).

Additionally, co-localisation experiments established that both VIPAR and PLOD3 co-localise in the vicinity of TGN and a co-localisation of both proteins with RAB11A was also shown. This supports previous data that see VPS33B interacting with RAB11A and their involvement in vesicular trafficking (Cullinane et al., 2010). A trafficking of PLOD3, through VPS33B/VIPAR/RAB11A was postulated and further investigated in mouse Inner Medullary Collecting Duct (mIMCD3) cell lines, the polarized epithelial cell model used for the study of ARC syndrome (Cullinane et al., 2010). Mislocalisation of endogenous PLOD3 was observed in Vps33b/Vipar knockdown mIMCD-3 cells after culturing for two weeks on transwell supports. This allows wild-type cells to become fully polarized but a loss of cell polarity and EMT phenotype is evident in Vps33b/Vipar knockdown cells. The accumulation of mCherry-PLOD3 at the TGN is seen in Vps33b/Vipar knockdown mIMCD-3 cells where PLOD3 is shown to co-localise with GFP-TGN38. The co-localisation of mCherry-PLOD3 with TGN38 was not seen in wild-type pre-polarized mIMCD-3 cells. Additionally PLOD-3 co-localises with RAB11A in wild-type pre-polarized mIMCD-3 cells but not in Vipar knockdown cells, strengthening the hypothesis that PLOD3 is trafficked out of TGN by VIPAR/VPS33B complex via Rab11a positive endosomes. This route appears to be ‘blocked’ in the absence of VPS33B/VIPAR and therefore PLOD3 accumulates in the TGN of Vipar knockdown mIMCD-3 cells.

VPS33B/VIPAR involvement in PLOD3 trafficking was characterized also from the topological point of view. The topology of the VPS33B/VIPAR/PLOD3 interaction that takes place at the TGN, (or vesicles located in the TGN area) was established. Several prediction servers identified two transmembrane domains (TMD) in VIPAR and the transmembrane nature of VIPAR was confirmed by different fractionation experiments. This is a novel finding that for the first time characterises VIPAR topology in the cell. VPS16, the yeast homologue of VIPAR, that constitutes the side of the head of the cytosolic HOPS complex, is not a transmembrane protein. This is in agreement with the prediction analysis for TMD performed for VPS16 in the same way as for VIPAR. These analyses indeed did not identify transmembrane helices in VPS16. As mentioned in the previous Chapter, previously published data indicate that it is unlikely that VIPAR and VPS33B are constituent parts of the HOPS complex in higher organisms. I therefore hypothesised that during evolution VIPAR has acquired a transmembrane topology and new functions that are related to PLOD3 trafficking in the cell. A further indication of this statement derives from TMD prediction analysis on VPS16B, the homologue of VIPAR in *D.melanogaster*. As mentioned in the previous Chapter, the fly database indicates that VPS16B co-immunopreciptates with PLOD, the homologue in fly of PLOD3. The TMD prediction servers identify two TMD in VPS16B at the same regions where TMD in VIPAR are located. Therefore a strong association can be deduced between structure (TMD regions) and function (PLOD3 trafficking) of VIPAR in the course of evolution, giving a stronger support to this theory.

The relevance of PLOD3 in ARC was examined by analyzing the phenotype of mIMCD-3 cell lines knockdown for Vps33b/Vipar that as mentioned above is characterized by a loss of E-cadherin expression and EMT phenotype. A large increase

in the intracellular amounts of collagen IV was observed in these cells. The intracellular accumulation of collagen in PLOD3 deficiency has been previously described in literature, and was evidenced also in Plod3 knockdown mIMCD-3 cells. Additionally, E-cadherin expression was found down-regulated not only in Vps33/Vipar knockdowns but also in Plod3 knockdown mIMCD-3 cells, suggesting the existence of a common mechanism involving VPS33B/VIPAR and PLOD3 that regulates the expression/trafficking of collagen and E-cadherin in these cells. Immunofluorescence experiments showed that collagen type IV co-localises with mCherry-PLOD3 in wild-type mIMCD-3 but not Vipar knockdown cells. This clearly indicates that the VPS33B/VIPAR complex is required for the delivery of PLOD3 to collagen. Given the established importance of PLOD3 in collagen secretion, it was possible to conclude that the intracellular accumulation in collagen IV observed in Vps33b/Vipar and Plod3 knockdown cell lines is caused by the abrogation of PLOD3/collagen binding in these cells. This is caused by the decreased levels of PLOD3 in Plod3 knockdown mIMCD-3 cells, and by the failure in the PLOD3 delivery to collagen in Vps33b/Vipar knockdowns. This crucial role of VPS33B/VIPAR in PLOD3 trafficking could constitute a novel mechanism that regulates collagen trafficking. Additionally, the abnormal expression/modification of collagen, observed in Vps33b/Vipar knockdown mIMCD-3 cells as well as in fibroblasts derived from ARC patients, could explain the EMT phenotype and E-cadherin down-regulation that characterises knockdown mIMCD-3 cell lines. Indeed data in the literature indicate that exogenous collagen can directly lead to E-cadherin down-regulation and EMT (Espinosa Neira and Salazar, 2012). In conclusion, the newly identified PLOD3-VIPAR interaction was fundamental

in elucidating new functions for VPS33B and VIPAR in the cell and getting a better understanding of the pathology of the ARC syndrome.

CHAPTER SIX:

DISCUSSION

6.1 - VPS33B interacts with VIPAR at a protein level forming a stable binary complex

The co-localisation and co-immunoprecipitation of VPS33B along with VIPAR has suggested that these proteins work in a multiprotein complex involved epithelial polarisation and apical junction complex formation (Cullinane et al., 2010). Fundamentally, this project has added to the growing body of evidence showing that VPS33B interacts at a protein level with VIPAR and a direct association between the proteins was described. Additionally the work presented here demonstrated through different biochemical approaches that VPS33B and VIPAR form a stable binary complex *in-vitro.* This was previously shown for the yeast homologues of VPS33B and VIPAR, respectively Vps33 and Vps16 that form a tight sub-complex of HOPS (Brocker et al., 2012). Therefore the data in this thesis represent the first evidence regarding the interaction involving the mammalian homologues VPS33B and VIPAR.

Previously published data suggests that VPS33B and VIPAR are unlikely to form active parts in the conventional HOPS complex and these proteins may be involved in another protein complex (Smith et al., 2012). The work performed in this thesis confirmed that this is the case and a novel interacting partner for VPS33B/VIPAR complex was identified. Despite being a stable complex *in-vitro,* instability phenomena including degradation with time and protein precipitation after tag removal, initially suggested that new binding partners are required to further stabilize the VPS33B/VIPAR complex. The instability was related to the presence of a large disordered region at the N-terminus of VIPAR that is not involved in the interaction with VPS33B. The fact that disordered regions are usually involved *in-vivo* in binding of structured partners, that determine their disordered to ordered transition, suggested that this might be the case for the N-

terminal of VIPAR. Co-immunoprecipitation experiments demonstrated that indeed the N-terminal part of VIPAR is involved in the binding of PLOD3, the novel interactor for VIPAR identified in this work.

6.2 - PLOD3 is a novel interactor for VIPAR

The function of VPS33B and VIPAR in the cell is not yet fully characterised. Therefore the main goal of this work was to clarify the role of VPS33B and VIPAR and understand the pathways and mechanisms in which these proteins are involved. This was achieved by identifying new interacting partners for VPS33B and VIPAR through a pull-down experiment followed my mass-spectrometry identification. Mass spectrometry analysis identified Procollagen-Lysine 2-Oxoglutarate 5-Dioxygenase 1 (PLOD1) and Procollagen-Lysine 2-Oxoglutarate 5-Dioxygenase (PLOD3) as potential interacting partners for VPS33B/VIPAR. The work described in this thesis was focused on PLOD3 due to lack of co-localsiation between PLOD1 and intracellular clusters of VPS33B/VIPAR. It is however important to stress that further investigation is required to exclude the possibility that PLOD1 is also an interactor for VPS33B/VIPAR. All experiments performed in this thesis however demonstrated that PLOD3 is a novel interactor for VIPAR only, while no interaction with VPS33B was demonstrated. These experiments included co-immunoprecipitation, co-localisation of the endogenous proteins and Fluorescence Lifetime Imaging in Förster Resonance Energy Transfer (FRET-FLIM) analysis using fluorescently-tagged proteins. These last measurements provided evidence regarding the intermolecular distance between VPS33B-VIPAR and

VIPAR-PLOD3 from which it was possible to deduce that a direct interaction occurs between PLOD3 and VIPAR in the cell.

6.3 - VIPAR is a transmembrane protein with its luminal N-terminal interacting with PLOD3 and its cytsolic C-terminal interacting with VPS33B

As mentioned above, evidence from this thesis suggests that direct protein-protein interactions exist between VPS33B and VIPAR as well as VIPAR and PLOD3. These data may seem to contradict the accepted knowledge on the topology of these proteins in the cell. PLOD3 is known to be a luminal protein while VPS33B and VIPAR are considered cytosolic proteins because of homology to their yeast counterparts Vps33 and Vps16 that form part of the cytosolic HOPS complex. Bioinformatics analysis performed in this thesis led to the identification of two transmembrane helices along VIPAR protein sequence located approximately at 140-160 amino acid for the first helix and 270-290 amino acid for the second helix. The same TMD were predicted along VPS16B, the fly homologue of VIPAR, that co-immunoprecipitates with PLOD (the fly homologue of PLOD3). Interestingly no transmembrane helices were predicted for the yeast VPS16 that unlike VIPAR, is an integral part of the yeast HOPS complex. The transmembrane nature of VIPAR, predicted by several bioinformatics algorithms, was confirmed by two different fractionation experiments and represents a novel finding in the localisation of VIPAR that was previously uncharacterised. Combined together, the data presented here suggest that the function of VIPAR is not related to that of HOPS complex and depends on the transmembrane nature of the protein that was acquired during evolution of metazoans from single cell organisms. The binomial association between protein structure and function is indeed one of the fundamental principles in

biology.

The same bioinformatics analysis that led to the identification of transmembrane helices in VIPAR, predicted the N-terminal of the protein targeted to the lumen and its C-terminal facing the cytosol. Here a mapping of the interaction along VIPAR sequence was performed for both VPS33B and PLOD3. VPS33B binds the C-terminal region of VIPAR while PLOD3 is involved in the interaction with its N-terminal. These data perfectly match with the prediction analysis and allowed us to unambiguously determine the topology of the VPS33B/VIPAR/PLOD3 interaction: VPS33B, located in the cytosol, interacts with the C-terminal cytosolic domain of VIPAR, while the luminal protein PLOD3 is involved in the interaction with the lumen facing N-terminal domain of VIPAR (figure 6.1 A).

However the exact mapping of the TMD in VIPAR is not yet established and on-going experiments are aimed at determining their localisation along VIPAR protein sequence. As mentioned above, the majority of the prediction servers identify two TMD in VIPAR. Therefore VIPAR might constitute a type II membrane-bound protein. Differently from type I, type II TM proteins are not targeted in the secretory pathway through signal peptides and indeed no signalling sequences have been identified in VIPAR. Instead the targeting to the secretory pathway of type II TM proteins occurs through their first TMD that resembles the signal sequence, but is different from it because it is not cleaved after insertion into the membrane.

The transmembrane nature of VIPAR, together with the two predicted coiled-coil domains, is in agreement with the bioinformatics analysis that identified a large Golgin subfamily A member 5 domain in VIPAR. Golgins indeed are found in different parts of Golgi stacks and they are typically anchored to the membrane by a transmembrane

domain or by binding to a small GTPase (Munro, 2011). We thus concluded that VIPAR is associated with PLOD3, and/or other possible cargos, at the TGN compartment and this binding stabilises the N-terminal disordered region of VIPAR. Fractionation experiment performed in this work showed that VPS33B may be present in a soluble and a membrane associated state (through its interaction with VIPAR). Therefore this experiment, combined with other evidence, suggested that VPS33B binding to VIPAR is associated with formation and trafficking of the vesicles containing PLOD3 and other possible cargos (figure 6.1 A and B).

6.4 - VPS33B/VIPAR complex is involved in PLOD3 delivery to collagen

The main goal of this thesis to clarify the role of VPS33B and VIPAR in the cell and understand the pathways and mechanisms in which these proteins are involved. VPS33B and VIPAR function indeed is not yet fully characterised. VPS33B is known to belong to the SM family of proteins that bind to Soluble N-ethylmalemide sensitive factor Attachment Receptor (SNARE) which in turn mediate all intracellular membrane fusion events (Gallwitz and Jahn, 2003; Peterson and Emr, 2001). As mentioned above, VIPAR belongs to the Golgin family of proteins because a large Golgin subfamily A member 5 domain forms most of its structure. Golgins are coiled-coil motif proteins that mediate tethering of vesicles to Golgi membranes and cisternal membranes to each other (Burguete et al., 2008; Satoh et al., 2003). The discovery of PLOD3 as a novel interacting partner for VIPAR lead to a better understanding of the role of VPS33B/VIPAR complex in the cells. In agreement with the presence of a Golgin domain in VIPAR, the VPS33B/VIPAR complex was localised to the TGN

where it carries out its function in PLOD3 (and possibly other cargos) vesicle mediated transport from this compartment. The mislocalisation of PLOD3 in Vps33b/Vipar knockdown mIMCD-3 cells and its accumulation in the TGN represents a strong evidence for this statement. Additionally the work presented here suggested that RAB11A is involved in the trafficking of PLOD3 from TGN, given the co-localisation of both PLOD3 and VIPAR with this intracellular marker. This confirms previous data that identified a VPS33B interaction with RAB11A (Cullinane et al., 2010). Additionally this project provided insights into the intracellular localisation of PLOD3. The endogenous protein was localised to the TGN compartment in HeLa cells and fibroblasts contradicting the previous data that attribute an ER localisation to all PLOD isoforms (Heikkinen et al., 2000; Salo et al., 2006a). In support of the data shown in this thesis, PLOD3, unlike other PLOD isoforms, is thought to utilise two different trafficking pathways for its delivery to the cell surface and for extracellular secretion (Wang et al., 2012). Additionally PLOD3 enzymatic activity is peculiar among isoforms and as well having the LH activity, common to all three isoforms, PLOD3 is capable of collagen glucosyltransferase (GGT) and collagen galactosyltransferase (GT) activities. Glycosylation (by GT and GGT) is the type of post-translational modification that is known to occur in the Golgi apparatus (Stanley, 2011), which is in agreement with the TGN location of PLOD3 established in this work. The post-translational modifications performed by PLOD3 in collagens are crucial for determining its stability and intracellular trafficking (Sipila et al., 2007). Data in this thesis demonstrate how depletion of PLOD3 activity in Plod3 knockdown mIMCD-3 cells leads to an abnormal intracellular accumulation of collagen IV, similarly to that observed in Vps33b/Vipar knockdown mIMCD-3 cells. Further confirmation of these results was provided by

additional experiments that demonstrated a failure in PLOD3 delivery to collagen in Vps33/Vipar knockdown mIMCD-3 cells. This therefore suggested that VPS33B and VIPAR are required for PLOD3 delivery to collagen and PLOD3-collagen binding is crucial for trafficking of collagen inside the cell (confirming previous data) (figure 6.1). Abnormalities in collagen IV modifications in Plod3 deficiency have been deduced from a 'band shift' in Western-blot of collagen IV chains that is caused by a lack in its post-translational modifications (Norman and Moerman, 2000). Similarly, Western-blot experiments showed a slightly enhanced migration in collage IV chains in Vps33b/Vipar and Plod3 knockdown cells. However, further on-going experiments are required to confirm these data.

The lack of co-localisation between PLOD3 and collagen IV in Vipar knockdown mIMCD-3 cells, that might cause collagen IV under-glycosylation, can lead to the hypothesis that PLOD3 binding to collagen, required for its trafficking, does not occur at the TGN level but in the collagen carriers that are responsible for the transport of collagen in the cell (figure 6.1). This statement is supported by several experiments performed in this thesis. Firstly, the co-localisation between PLOD3 and collagen IV, that is lost in Vipar knockdown mIMCD-3 does not occur at TGN level in wild-type cells. Round granular structures have been evidenced by co-localising PLOD3 with collagen IV and further experiments are required in order to determine their intracellular localisation. Secondly, data in this thesis evidence that mistrafficking of PLOD3 in Vipar knockdown mIMCD-3 is associated with an accumulation of PLOD3 trafficking at the TGN and no co-localisation between PLOD3 and collagen IV occurs in this compartment in knockdown cells. It is therefore possible to conclude that 1) PLOD3 and collagen exit from TGN occurs independently and involves different trafficking

routes and VPS33B/VIPAR are responsible for PLOD3 trafficking only 2) PLOD3 and collagen binding, that is essential for collagen trafficking, occurs at collagen carrier structures (that are either Rab11a or Lamp1 positive) and not at the TGN (figure 6.1). 3) Intermediate organelles might be involved in VPS33B/VIPAR mediated delivery of PLOD3 towards collagen carrier structures (dashed arrow in figure 6.1) as might be suggested by the lack of co-localisation between VIPAR, PLOD3 and collagen in correspondence of the round granular structures. Another possibility is that this process is fast and VPS33B/VIPAR is recycled back for another fusion cycle. Further experiments are required to further investigate the 'delivery' of PLOD3 to collagen carrier structures in the cell.

Trafficking of collagen in the cell is not yet fully understood and controversies are related to procollagen size (>300nm in length) compare to a typical vesicle size (<100 nm) (Stephens, 2012). The evidence provided in this thesis and future planned experiments could lead to a better understanding on the complex machinery involved in collagen trafficking.

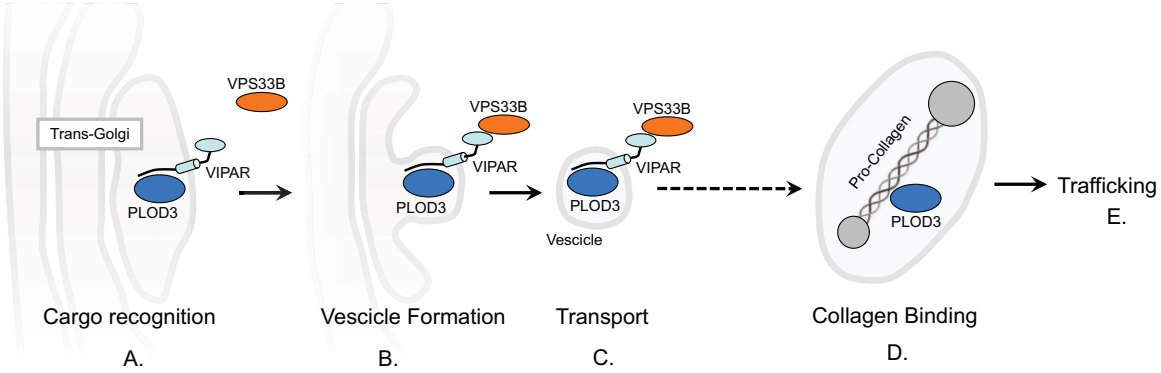


Figure 6.1. Schematic representation of the VPS33B/VIPAR trafficking pathway described in this thesis. ***(A).*** *Transmembrane VIPAR interacts through its N-terminal domain with PLOD3 located at the lumen of the TGN. VPS33B is localised in the cytosol.* ***(B).*** *The VPS33B binding to VIPAR is associated with the formation of the vesicle containing the cargo protein PLOD3.* ***(C)****. VPS33B and VIPAR are required for the transport of the vesicle to the receiver compartment that might be an intermediate organelle (dashed arrow) or collagen carrier structures.* ***(D)****. The fusion of the PLOD3 caring vesicle with the collagen carrier structure determines the binding of PLOD3 to Pro-collagen that is required for collagen trafficking* ***(E)****.*

6.5 - Abnormal collagen in Vps33b/Vipar/Plod3 deficiency and its relevance to ARC

Initial studies by Cullinane at all (2010) described a marked reduction in E-cadherin expression in Vps33b/Vipar knockdown mIMCD-3 cell lines. Similarly E-cadherin was found down-regulated in ARC syndrome patient liver biopsies and Vipar morfolino (MO) knockdown in the zebrafish (Cullinane et al., 2010). Epithelial-to-Mesenchymal transition (EMT) has been observed in knockdown cell lines and this transition is often caused by a reduced expression of E-cadherin. This suggests that an EMT could be occurring in ARC syndrome patient's polarised epithelial cells (Cullinane et al., 2010). The work presented here suggests that abnormal collagen modification and trafficking can lead to E-cadherin down-regulation in Vps33b/Vipar and Plod3 knockdown mIMCD-3 cells. Further experiments are required to confirm these data. However the

mechanism presented here for the first time provides an intracellular pathway for VPS33B/VIPAR action that can link their function with the cell phenotype. The better understanding of the role of this protein complex in the cell will lead in the future to a better characterisation and treatment of the ARC syndrome.

Additionally, an abnormal intracellular accumulation of collagen type I was described in human skin fibroblast derived from ARC patients suggesting an agreement between the findings in the cell model of the disease and patients' phenotype.

Similarities in the phenotype between Vps33b/Vipar knockdown and Plod3 knockdown mIMCD-3 cells suggested that collagen related defects can be observed in ARC patients as well as in patients with mutations in Plod3. Only one compound heterozygous patient with mutations in Plod3 is known and they lead to a markedly reduction in GGT and LH activity of the enzyme (Salo et al., 2008). The common features between the PLOD3 deficient patient and individuals with ARC include: joint contractures, osteopenia, sensorineural deafness, growth retardation (Salo et al., 2008).

The bleeding tendency of patients with ARC might be explained by the disruption of the capillary basement membranes caused by PLOD3 mistrafficking and consequent abnormal collagen formation and secretion. Similarly defects in the glomerular basement membranes (GMB) can explain renal tubular dysfunction in ARC patiensts such as acidosis and renal tubular dysfunction. The non-complete overlap of symptoms between the two conditions suggests that PLOD3 might not be the only cargo of VPS33B/VIPAR and further experiments are required to establish this. These features include the severe degree of liver and kidney disease in ARC and no blood vessel tortuosity in ARC. Indirect secondary effects of PLOD3/VPS33B/VIPAR deficency might explain these symptoms but further investigation is required to elucidate the

mechanisms involved. In addition VPS33B/VIPAR complex was found to be required for platelet α-granule biogenesis (Urban et al., 2012,). The clinical features in the only one patient with PLOD3 deficiency have not been fully investigated and therefore it is difficult to judge the full extent of the phenotype. Nevertheless this available data strongly suggest that PLOD3 is unlikely to be the only cargo for VPS33B/VIPAR. Another hypothesis in that PLOD3 is indirectly involved in these pathways through the post-trasnlational modifications of proteins that are directly involved in the α-granule biogenesis.

Finally, the ubiquitous Vps33b knockout mouse developed in our laboratory is embryonically lethal at E 8.5 while Plod3 knockout mice at E 9.5 (Rautavuoma et al., 2004). These data suggests a crucial role for both VPS33B/VIPAR and PLOD3 in development and disease.

6.6 - Future experiments

1) As mentioned above, it would be interesting to map the location of the transmembrane domains in VIPAR. For this purpose the cloning of new VIPAR constructs, based on predictions of TMD servers, is being carried out.

2) This thesis has established that PLOD3 is a cargo recognised by VIPAR at the TGN level. To date is not known if this recognition is PLOD3-specific or if other cargos are recognised by VIPAR at the TGN. The mass-spectrometry results identified PLOD1 as another possible interacting partner for VPS33B/VIPAR but preliminary co-localisation experiments did not confirm this result. However more experiments are required to exclude the possibility that PLOD1 is not an interactor of VPS33B/VIPAR.

Additionally, different pull-down experiments can be performed by incubating the immobilized VPS33B/VIPAR purified complex with a different cell lysate (e.g derived from mIMCD-3 cells) in order to identify new possible interactors for this complex and answer the question whether PLOD3 is the only cargo recognised by VIPAR in the cell.

3) According to the model proposed in figure 6.1 for VPS33B/VIPAR role in PLOD3 and collagen trafficking, there are still questions that need answering. Primarily, it would be essential to determine the intracellular localisation of PLOD3 and collagen in wild-type mIMCD-3 cells. Secondly, the 'destiny' of VPS33B/VIPAR complex after cargo delivery is not known and would be interesting to establish. Immunofluorescence experiments indeed do not show a simultaneous co-localisation between VPS33B-VIPAR, PLOD3 and collagen IV in wild-type mIMCD-3 cells but VPS33B/VIPAR co-localisation with PLOD3 occurs in a different compartment from PLOD3 co-localisation with collagen. Therefore VPS33B/VIPAR might directly cycle back after PLOD3 delivery to collagen or its involved in the cargo sorting only. Thirdly, the route followed by PLOD3-collagen after binding in collagen carrier structures, and how this is involved in collagen secretion, is not yet known. To answer these questions live cell imaging experiments are currently being performed.

4) On-going mass-spectrometry analysis of collagen IV and collagen I post-translational modifications are aimed at identifying changes in the hydroxylation/glycosylation pattern of collagens in knockdown cells as well as in tissues derived from our mice model for the ARC syndrome. Additionally similar approaches are being performed in order to identify other possible proteins with collagen domains whose glycosylation is altered in Vps33b/Vipar deficiency. This might help at elucidating other features of the ARC syndrome as the mislocalisation of specific proteins in patient's liver and kidney.

5) Mouse model: Experiments are carried out in collaboration with Dr Holly Smith involving the use of the ARC mouse model in order to understand the impact of PLOD3 in the pathophysiology of the ARC syndrome. This indeed is difficult to obtain from human samples and represents the closest model to the disease that can additionally be useful for trials of potential therapeutic agents.

CHAPTER SEVEN: APPENDICES

Appendix 1

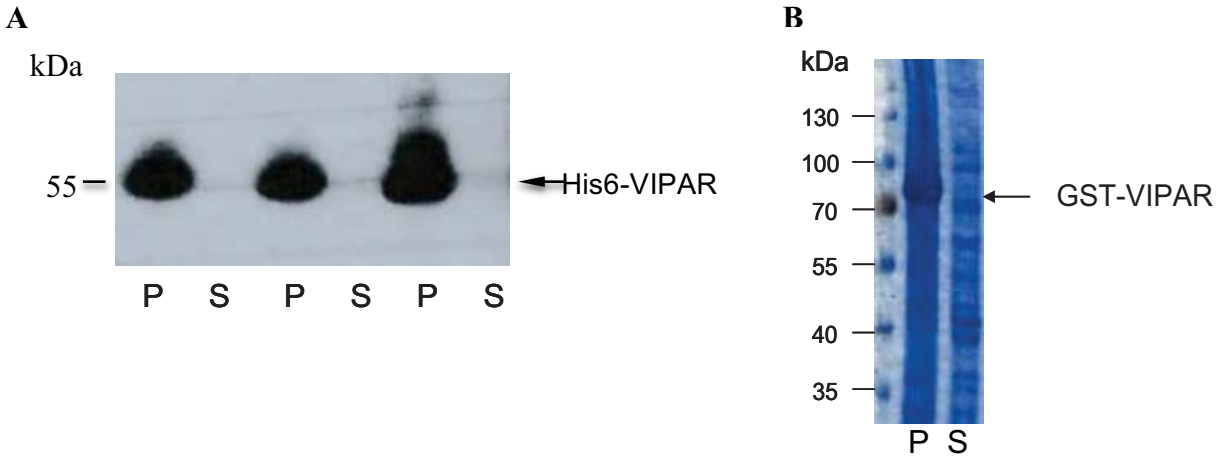


*Figure 7.1 - **(A)** Western blot analysis to identify His6-VIPAR protein soluble (S) and insoluble pellet (P) fractions after protein expression and extraction from SF9 insect cells. Anti-His6 antibody was used for identification. **(B)** SDS-Page analysis to identify GST-tagged VIPAR expression and solubility in BL21 (DE3)pLysS bacterial cells after protein extraction. Soluble (S) and insoluble pellet (P) fractions are shown.*
The molecular weight of VIPAR corresponds to 57 kDa and the molecular weight of the GST is 26 kDa.

Appendix 2

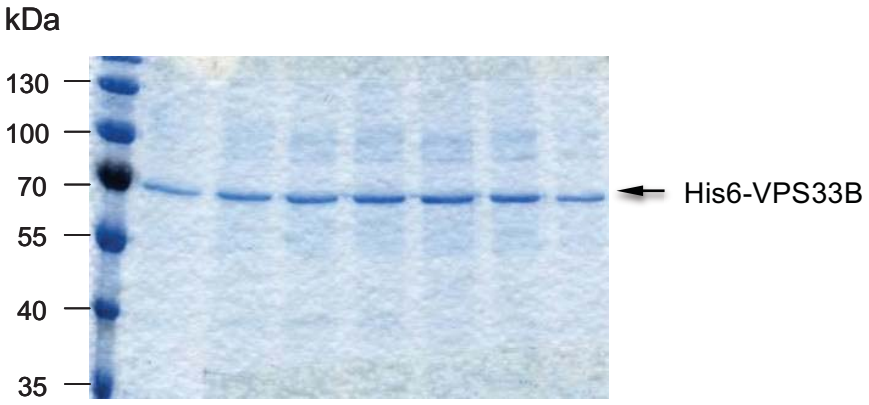
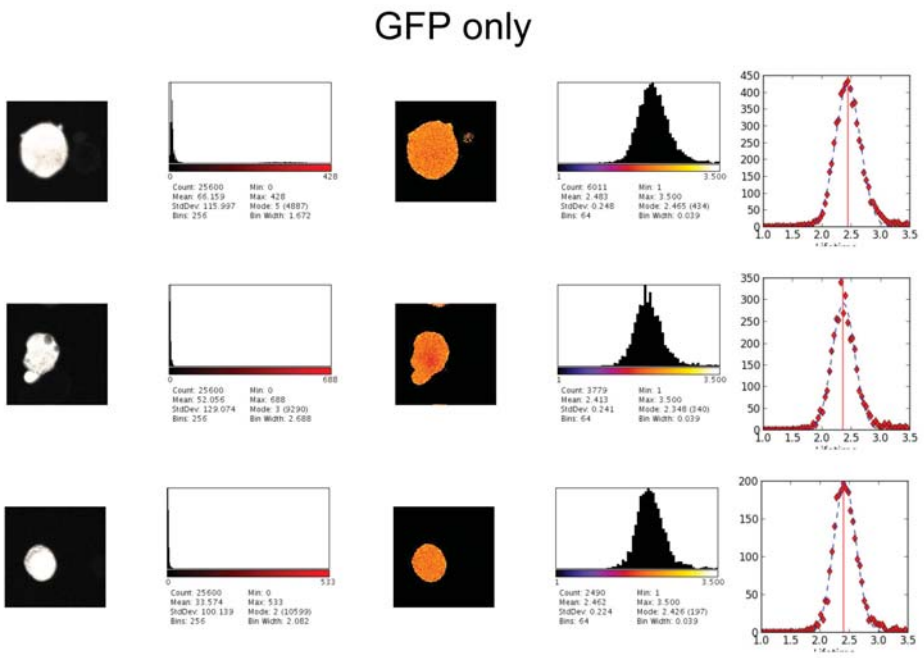
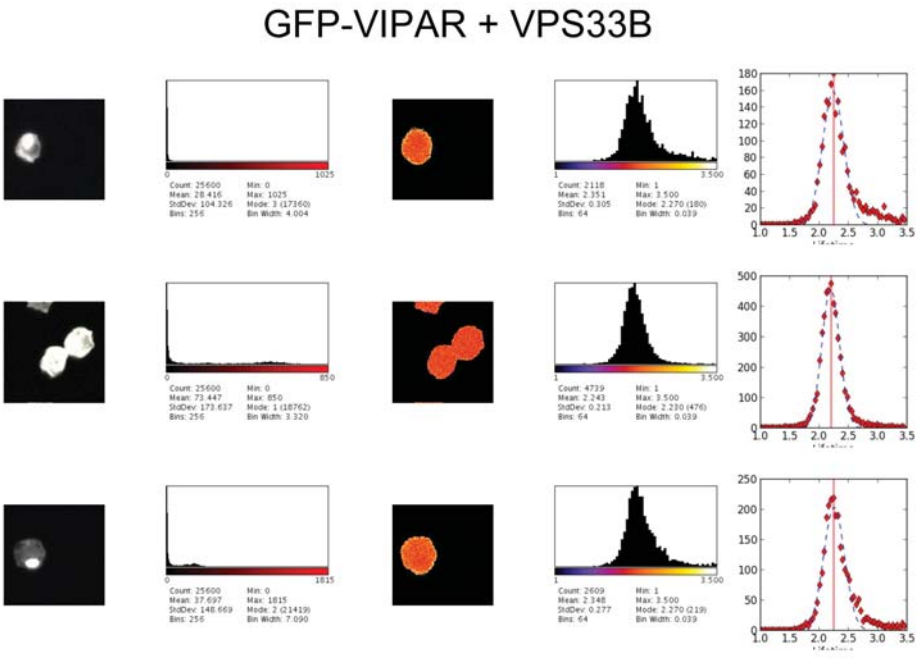


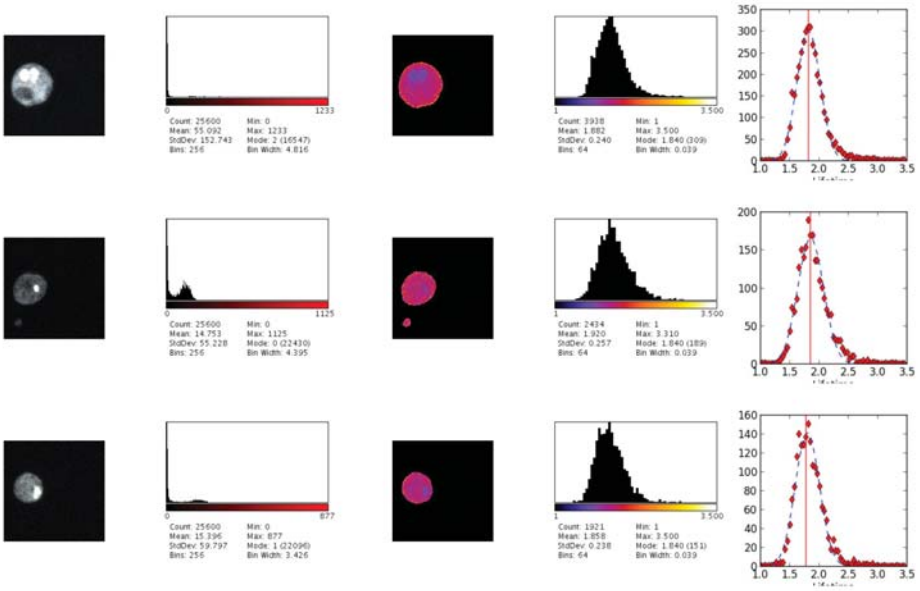
Figure 7.2 - SDS-PAGE of Cobalt chromatography elution fractions. His6-VPS33B was expressed and extracted from SF9 cells before being affinity purified using cobalt (Co^{2+}) resin for the purification of 6x His-tagged VPS33B. The molecular weight of VPS33B corresponds to 70 kDa.

Appendix 3

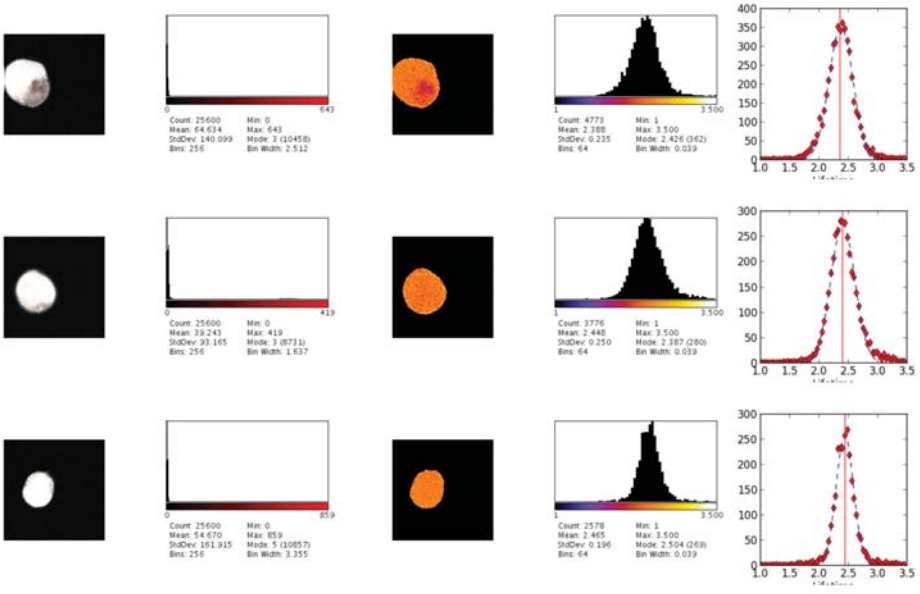




GFP-VIPAR + mCherry-VPS33B



GFP-VIPAR + mCherry



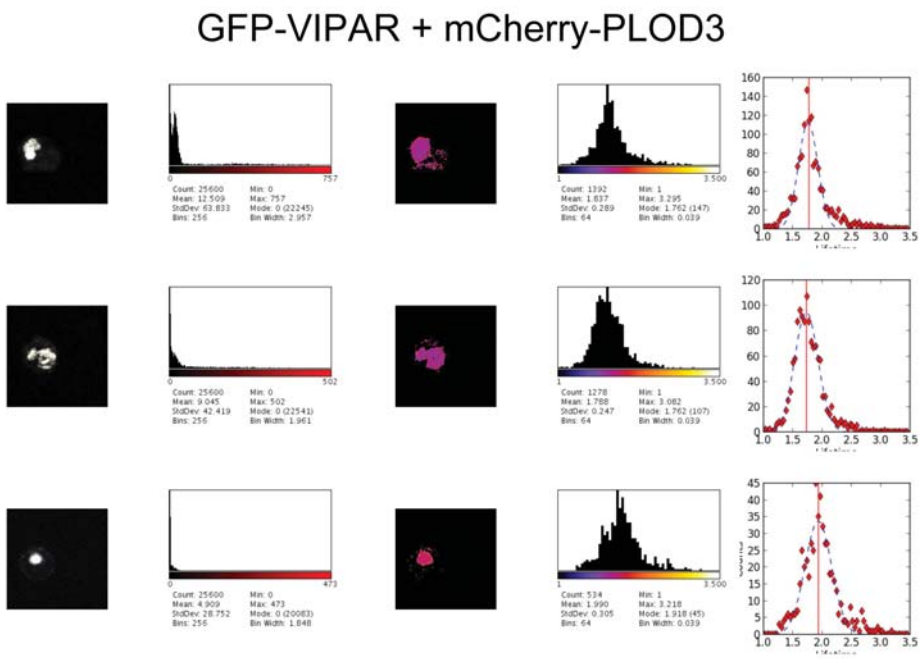


Figure 7.3 - Raw data from FRET data processing. From left to right: intensity image recorded using the donor (GFP) channel; fluorescence intensity histogram; false-color image representing donor fluorescence lifetime, obtained by processing the intensity images using the time-gated phasor plugin for ImageJ; histogram of fluorescence lifetime; gaussian fit of the lifetime histogram, used for the determination of the average fluorescence lifetime from each image.
The panels represent 3 measurements for each of the following samples: GFP only; GFP-VIPAR + VPS33B; GFP-VIPAR + mCherry; GFP-VIPAR + mCherry-VPS33B; GFP-VIPAR + mCherry-PLOD3.

CHAPTER EIGHT: REFERENCES

Aalto, M.K., Ruohonen, L., Hosono, K., and Keranen, S. 1991). Cloning and sequencing of the yeast Saccharomyces cerevisiae SEC1 gene localized on chromosome IV. **Yeast**, *7*, 643-650.

Agarwal, P., and Ballabh, R. (2013). Expression of type IV collagen in different histological grades of oral squamous cell carcinoma: An immunohistochemical study. **J Cancer Res Ther**, *9*, 272-275.

Agarwal, R., Jurisica, I., Mills, G.B., and Cheng, K.W. (2009). The emerging role of the RAB25 small GTPase in cancer. **Traffic**, *10*, 1561-1568.

Akbar, M.A., Tracy, C., Kahr, W.H., and Kramer, H. (2011). The full-of-bacteria gene is required for phagosome maturation during immune defense in Drosophila. **J Cell Biol**, *192*, 383-390.

Almomani, E.Y., King, J.C., Netsawang, J., Yenchitsomanus, P.T., Malasit, P., Limjindaporn, T., Alexander, R.T., and Cordat, E. (2012). Adaptor protein 1 complexes regulate intracellular trafficking of the kidney anion exchanger 1 in epithelial cells. **Am J Physiol Cell Physiol**, *303*, C554-566.

Ang, S.F., and Folsch, H. (2012). The role of secretory and endocytic pathways in the maintenance of cell polarity. **Essays Biochem**, *53*, 29-39.

Antonin, W., Fasshauer, D., Becker, S., Jahn, R., and Schneider, T.R. (2002). Crystal structure of the endosomal SNARE complex reveals common structural principles of all SNAREs. **Nat Struct Biol**, *9*, 107-111.

Aplin, A.E., Howe, A., Alahari, S.K., and Juliano, R.L. (1998). Signal transduction and signal modulation by cell adhesion receptors: the role of integrins, cadherins, immunoglobulin-cell adhesion molecules, and selectins. **Pharmacol Rev**, *50*, 197-263.

Aszodi, A., Legate, K.R., Nakchbandi, I., and Fassler, R. (2006). What mouse mutants teach us about extracellular matrix function. **Annu Rev Cell Dev Biol**, *22*, 591-621.

Aumailley, M., and Gayraud, B. (1998). Structure and biological activity of the extracellular matrix. **J Mol Med (Berl)**, *76*, 253-265.

Bach, H., Papavinasasundaram, K.G., Wong, D., Hmama, Z., and Av-Gay, Y. (2008). Mycobacterium tuberculosis virulence is mediated by PtpA dephosphorylation of human vacuolar protein sorting 33B. **Cell Host Microbe**, *3*, 316-322.

Balch, W.E., Dunphy, W.G., Braell, W.A., and Rothman, J.E. (1984). Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine. **Cell**, *39*, 405-416.

Banta, L.M., Vida, T.A., Herman, P.K., and Emr, S.D. (1990). Characterization of yeast Vps33p, a protein required for vacuolar protein sorting and vacuole biogenesis. **Mol Cell Biol**, *10*, 4638-4649.

Bastaki, M., Braiterman, L.T., Johns, D.C., Chen, Y.H., and Hubbard, A.L. (2002). Absence of direct delivery for single transmembrane apical proteins or their "Secretory" forms in polarized hepatic cells. **Mol Biol Cell**, *13*, 225-237.

Batge, B., Winter, C., Notbohm, H., Acil, Y., Brinckmann, J., and Muller, P.K. (1997). Glycosylation of human bone collagen I in relation to lysylhydroxylation and fibril diameter. **J Biochem**, *122*, 109-115.

Bem, D., Yoshimura, S., Nunes-Bastos, R., Bond, F.C., Kurian, M.A., Rahman, F., Handley, M.T., Hadzhiev, Y., Masood, I., Straatman-Iwanowska, A.A*., et al.* (2011). Loss-of-function mutations in RAB18 cause Warburg micro syndrome. **Am J Hum Genet**, *88*, 499-507.

Bennett, K., Callard, R., Heywood, W., Harper, J., Jayakumar, A., Clayman, G.L., Di, W.L., and Mills, K. (2010). New role for LEKTI in skin barrier formation: label-free quantitative proteomic identification of caspase 14 as a novel target for the protease inhibitor LEKTI. **J Proteome Res**, *9*, 4289-4294.

Bennett, K., Heywood, W., Di, W.L., Harper, J., Clayman, G.L., Jayakumar, A., Callard, R., and Mills, K. (2012). The identification of a new role for LEKTI in the skin: The use of protein 'bait' arrays to detect defective trafficking of dermcidin in the skin of patients with Netherton syndrome. **J Proteomics**, *75*, 3925-3937.

Bock, J.B., Matern, H.T., Peden, A.A., and Scheller, R.H. (2001). A genomic perspective on membrane compartment organization. **Nature**, *409*, 839-841.

Bonifacino, J.S., and Lippincott-Schwartz, J. (2003). Coat proteins: shaping membrane transport. **Nat Rev Mol Cell Biol**, *4*, 409-414.

Brocker, C., Kuhlee, A., Gatsogiannis, C., Balderhaar, H.J., Honscher, C., Engelbrecht-Vandre, S., Ungermann, C., and Raunser, S. (2012). Molecular architecture of the multisubunit homotypic fusion and vacuole protein sorting (HOPS) tethering complex. **Proc Natl Acad Sci U S A**, *109*, 1991-1996.

Bull, L.N., Mahmoodi, V., Baker, A.J., Jones, R., Strautnieks, S.S., Thompson, R.J., and Knisely, A.S. (2006). VPS33B mutation with ichthyosis, cholestasis, and renal dysfunction but without arthrogryposis: incomplete ARC syndrome phenotype. **J Pediatr**, *148*, 269-271.

Burguete, A.S., Fenn, T.D., Brunger, A.T., and Pfeffer, S.R. (2008). Rab and Arl GTPase family members cooperate in the localization of the golgin GCC185. **Cell**, *132*, 286-298.

Butterworth, M.B., Edinger, R.S., Silvis, M.R., Gallo, L.I., Liang, X., Apodaca, G., Frizzell, R.A., and Johnson, J.P. (2012). Rab11b regulates the trafficking and recycling of the epithelial sodium channel (ENaC). **Am J Physiol Renal Physiol**, *302*, F581-590.

Calhoun, B.C., and Goldenring, J.R. (1997). Two Rab proteins, vesicle-associated membrane protein 2 (VAMP-2) and secretory carrier membrane proteins (SCAMPs), are present on immunoisolated parietal cell tubulovesicles. **Biochem J**, *325 (Pt 2)*, 559-564.

Caplan, S., Hartnell, L.M., Aguilar, R.C., Naslavsky, N., and Bonifacino, J.S. (2001). Human Vam6p promotes lysosome clustering and fusion in vivo. **J Cell Biol**, *154*, 109-122.

Carim, L., Sumoy, L., Andreu, N., Estivill, X., and Escarceller, M. (2000). Cloning, mapping and expression analysis of VPS33B, the human orthologue of rat Vps33b. **Cytogenet Cell Genet**, *89*, 92-95.

Carmosino, M., Valenti, G., Caplan, M., and Svelto, M. (2010). Polarized traffic towards the cell surface: how to find the route. **Biol Cell**, *102*, 75-91.

Carr, C.M., and Novick, P.J. (2000). Membrane fusion. Changing partners. **Nature**, *404*, 347-349.

Casanova, J.E., Wang, X., Kumar, R., Bhartur, S.G., Navarre, J., Woodrum, J.E., Altschuler, Y., Ray, G.S., and Goldenring, J.R. (1999). Association of Rab25 and Rab11a with the apical recycling system of polarized Madin-Darby canine kidney cells. **Mol Biol Cell**, *10*, 47-61.

Castro-Sanchez, L., Soto-Guzman, A., Guaderrama-Diaz, M., Cortes-Reynosa, P., and Salazar, E.P. (2011). Role of DDR1 in the gelatinases secretion induced by native type IV collagen in MDA-MB-231 breast cancer cells. **Clin Exp Metastasis**, *28*, 463-477.

Caswell, P.T., Spence, H.J., Parsons, M., White, D.P., Clark, K., Cheng, K.W., Mills, G.B., Humphries, M.J., Messent, A.J., Anderson, K.I*., et al.* (2007). Rab25 associates with alpha5beta1 integrin to promote invasive migration in 3D microenvironments. **Dev Cell**, *13*, 496-510.

Charng, M.J., Zhang, D., Kinnunen, P., and Schneider, M.D. (1998). A novel protein distinguishes between quiescent and activated forms of the type I transforming growth factor beta receptor. **J Biol Chem**, *273*, 9365-9368.

Chen, G., Deng, C., and Li, Y.P. (2012). TGF-beta and BMP signaling in osteoblast differentiation and bone formation. **Int J Biol Sci**, *8*, 272-288.

Chen, W., Feng, Y., Chen, D., and Wandinger-Ness, A. (1998). Rab11 is required for trans-golgi network-to-plasma membrane transport and a preferential target for GDP dissociation inhibitor. **Mol Biol Cell**, *9*, 3241-3257.

Chen, X., Tomchick, D.R., Kovrigin, E., Arac, D., Machius, M., Sudhof, T.C., and Rizo, J. (2002). Three-dimensional structure of the complexin/SNARE complex. **Neuron**, *33*, 397-409.

Chen, Y.A., and Scheller, R.H. (2001). SNARE-mediated membrane fusion. **Nat Rev Mol Cell Biol**, *2*, 98-106.

Chintala, S., Novak, E.K., Spernyak, J.A., Mazurchuk, R., Torres, G., Patel, S., Busch, K., Meeder, B.A., Horowitz, J.M., Vaughan, M.M.*, et al.* (2009). The Vps33a gene regulates behavior and cerebellar Purkinje cell number. **Brain Res**, *1266*, 18-28.

Christoforidis, S., Miaczynska, M., Ashman, K., Wilm, M., Zhao, L., Yip, S.C., Waterfield, M.D., Backer, J.M., and Zerial, M. (1999). Phosphatidylinositol-3-OH kinases are Rab5 effectors. **Nat Cell Biol**, *1*, 249-252.

Clarke, E.P., Cates, G.A., Ball, E.H., and Sanwal, B.D. (1991). A collagen-binding protein in the endoplasmic reticulum of myoblasts exhibits relationship with serine protease inhibitors. **J Biol Chem**, *266*, 17230-17235.

Collins, K.M., Thorngren, N.L., Fratti, R.A., and Wickner, W.T. (2005). Sec17p and HOPS, in distinct SNARE complexes, mediate SNARE complex disruption or assembly for fusion. **EMBO J**, *24*, 1775-1786.

Content, J., Johnston, M.I., De Wit, L., De Maeyer-Guignard, J., and De Clercq, E. (1980). Kinetics and distribution of interferon mRNA in interferon-primed and unprimed mouse L-929 cells. **Biochem Biophys Res Commun**, *96*, 415-424.

Corbeel, L., and Freson, K. (2008). Rab proteins and Rab-associated proteins: major actors in the mechanism of protein-trafficking disorders. **Eur J Pediatr**, *167*, 723-729.

Cox, T.R., and Erler, J.T. (2011). Remodeling and homeostasis of the extracellular matrix: implications for fibrotic diseases and cancer. **Dis Model Mech**, *4*, 165-178.

Cullinane, A.R., Straatman-Iwanowska, A., Seo, J.K., Ko, J.S., Song, K.S., Gizewska, M., Gruszfeld, D., Gliwicz, D., Tuysuz, B., Erdemir, G.*, et al.* (2009). Molecular investigations to improve diagnostic accuracy in patients with ARC syndrome. **Hum Mutat**, *30*, E330-337.

Cullinane, A.R., Straatman-Iwanowska, A., Zaucker, A., Wakabayashi, Y., Bruce, C.K., Luo, G., Rahman, F., Gurakan, F., Utine, E., Ozkan, T.B.*, et al.* (2010). Mutations in VIPAR cause an arthrogryposis, renal dysfunction and cholestasis syndrome phenotype with defects in epithelial polarization. **Nat Genet**, *42*, 303-312.

Dady, A., Blavet, C., and Duband, J.L. (2012). Timing and kinetics of E- to N-cadherin switch during neurulation in the avian embryo. **Dev Dyn**, *241*, 1333-1349.

De Craene, B., and Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. **Nat Rev Cancer**, *13*, 97-110.

De Matteis, M.A., and Luini, A. (2008). Exiting the Golgi complex. **Nat Rev Mol Cell Biol**, *9*, 273-284.

de Vega, S., Iwamoto, T., and Yamada, Y. (2009). Fibulins: multiple roles in matrix structures and tissue functions. **Cell Mol Life Sci**, *66*, 1890-1902.

Desai, R.A., Gao, L., Raghavan, S., Liu, W.F., and Chen, C.S. (2009). Cell polarity triggered by cell-cell adhesion via E-cadherin. **J Cell Sci**, *122*, 905-911.

Desclozeaux, M., Venturato, J., Wylie, F.G., Kay, J.G., Joseph, S.R., Le, H.T., and Stow, J.L. (2008). Active Rab11 and functional recycling endosome are required for E-cadherin trafficking and lumen formation during epithelial morphogenesis. **Am J Physiol Cell Physiol**, *295*, C545-556.

Dulubova, I., Yamaguchi, T., Arac, D., Li, H., Huryeva, I., Min, S.W., Rizo, J., and Sudhof, T.C. (2003). Convergence and divergence in the mechanism of SNARE binding by Sec1/Munc18-like proteins. **Proc Natl Acad Sci U S A**, *100*, 32-37.

Eastham, K.M., McKiernan, P.J., Milford, D.V., Ramani, P., Wyllie, J., van't Hoff, W., Lynch, S.A., and Morris, A.A. (2001). ARC syndrome: an expanding range of phenotypes. **Arch Dis Child**, *85*, 415-420.

Egerton, M., Zueco, J., and Boyd, A. (1993). Molecular characterization of the SEC1 gene of Saccharomyces cerevisiae: subcellular distribution of a protein required for yeast protein secretion. **Yeast**, *9*, 703-713.

Erickson, A.H., and Bocock, J.P. (2007). Targeting to lysosomes in mammalian cells: the biosynthetic and endocytic pathways. **Methods Mol Biol**, *390*, 339-361.

Espinosa Neira, R., and Salazar, E.P. (2012). Native type IV collagen induces an epithelial to mesenchymal transition-like process in mammary epithelial cells MCF10A. **Int J Biochem Cell Biol**, *44*, 2194-2203.

Farsad, K., and De Camilli, P. (2003). Mechanisms of membrane deformation. **Curr Opin Cell Biol**, *15*, 372-381.

Fasshauer, D., Sutton, R.B., Brunger, A.T., and Jahn, R. (1998). Conserved structural features of the synaptic fusion complex: SNARE proteins reclassified as Q- and R-SNAREs. **Proc Natl Acad Sci U S A**, *95*, 15781-15786.

Feres-Filho, E.J., Choi, Y.J., Han, X., Takala, T.E., and Trackman, P.C. (1995). Pre- and post-translational regulation of lysyl oxidase by transforming growth factor-beta 1 in osteoblastic MC3T3-E1 cells. **J Biol Chem**, *270*, 30797-30803.

Fialka, I., Pasquali, C., Kurzbauer, R., Lottspeich, F., and Huber, L.A. (1999). Loss of epithelial polarity is accompanied by differential association of proteins with intracellular membranes. **Electrophoresis**, *20*, 331-343.

Flaumenhaft, R. (2012). alpha-granules: a story in the making. **Blood**, *120*, 4908-4909.
Folsch, H., Ohno, H., Bonifacino, J.S., and Mellman, I. (1999). A novel clathrin adaptor complex mediates basolateral targeting in polarized epithelial cells. **Cell**, *99*, 189-198.

Forman-Kay JD, Mittag T (2013) From sequence and forces to structure, function, and evolution of intrinsically disordered proteins. **Structure** 21:1492-1499.

Franzke, C.W., Bruckner, P., and Bruckner-Tuderman, L. (2005). Collagenous transmembrane proteins: recent insights into biology and pathology. **J Biol Chem**, *280*, 4005-4008.

Friedl, P., and Wolf, K. (2003). Tumour-cell invasion and migration: diversity and escape mechanisms. **Nat Rev Cancer**, *3*, 362-374.

Gallwitz, D., and Jahn, R. (2003). The riddle of the Sec1/Munc-18 proteins - new twists added to their interactions with SNAREs. **Trends Biochem Sci**, *28*, 113-116.

Gelse, K., Poschl, E., and Aigner, T. (2003). Collagens--structure, function, and biosynthesis. **Adv Drug Deliv Rev**, *55*, 1531-1546.

Gerhardt, B., Kordas, T.J., Thompson, C.M., Patel, P., and Vida, T. (1998). The vesicle transport protein Vps33p is an ATP-binding protein that localizes to the cytosol in an energy-dependent manner. **J Biol Chem**, *273*, 15818-15829.

Gissen, P., Johnson, C.A., Gentle, D., Hurst, L.D., Doherty, A.J., O'Kane, C.J., Kelly, D.A., and Maher, E.R. (2005). Comparative evolutionary analysis of VPS33 homologues: genetic and functional insights. **Hum Mol Genet**, *14*, 1261-1270.

Gissen, P., Johnson, C.A., Morgan, N.V., Stapelbroek, J.M., Forshew, T., Cooper, W.N., McKiernan, P.J., Klomp, L.W., Morris, A.A., Wraith, J.E.*, et al.* (2004). Mutations in VPS33B, encoding a regulator of SNARE-dependent membrane fusion, cause arthrogryposis-renal dysfunction-cholestasis (ARC) syndrome. **Nat Genet**, *36*, 400-404.

Gissen, P., and Maher, E.R. (2007). Cargos and genes: insights into vesicular transport from inherited human disease. **J Med Genet**, *44*, 545-555.

Gissen, P., Tee, L., Johnson, C.A., Genin, E., Caliebe, A., Chitayat, D., Clericuzio, C., Denecke, J., Di Rocco, M., Fischler, B.*, et al.* (2006). Clinical and molecular genetic features of ARC syndrome. **Hum Genet**, *120*, 396-409.

Giunta, C., Burer-Chambaz, C., and Steinmann, B. (2009). Novel human pathological mutations. Gene symbol: PLOD1. Disease: Ehlers-Danlos syndrome type VIA, kyphoscoliotic type. **Hum Genet**, *125*, 346.

Goldenring, J.R., Shen, K.R., Vaughan, H.D., and Modlin, I.M. (1993). Identification of a small GTP-binding protein, Rab25, expressed in the gastrointestinal mucosa, kidney, and lung. **J Biol Chem**, *268*, 18419-18422.

Goldenring, J.R., Smith, J., Vaughan, H.D., Cameron, P., Hawkins, W., and Navarre, J. (1996). Rab11 is an apically located small GTP-binding protein in epithelial tissues. **Am J Physiol**, *270*, G515-525.

Grote, E., and Novick, P.J. (1999). Promiscuity in Rab-SNARE interactions. **Mol Biol Cell**, *10*, 4149-4161.

Gruenberg, J., and Maxfield, F.R. (1995). Membrane transport in the endocytic pathway. **Curr Opin Cell Biol**, *7*, 552-563.

Gu, F., Crump, C.M., and Thomas, G. (2001). Trans-Golgi network sorting. **Cell Mol Life Sci**, *58*, 1067-1084.

Guo, W., Sacher, M., Barrowman, J., Ferro-Novick, S., and Novick, P. (2000). Protein complexes in transport vesicle targeting. **Trends Cell Biol**, *10*, 251-255.

Gupta, G.D., and Brent Heath, I. (2002). Predicting the distribution, conservation, and functions of SNAREs and related proteins in fungi. **Fungal Genet Biol**, *36*, 1-21.

Guruharsha KG *et al*. (2011) A protein complex network of Drosophila melanogaster. **Cell** 147:690-703.

Hamano, Y., Zeisberg, M., Sugimoto, H., Lively, J.C., Maeshima, Y., Yang, C., Hynes, R.O., Werb, Z., Sudhakar, A., and Kalluri, R. (2003). Physiological levels of tumstatin, a fragment of collagen IV alpha3 chain, are generated by MMP-9 proteolysis and suppress angiogenesis via alphaV beta3 integrin. **Cancer Cell**, *3*, 589-601.

Hanzlowsky, A., Jelencic, B., Jawdekar, G., Hinkley, C.S., Geiger, J.H., and Henry, R.W. (2006). Co-expression of multiple subunits enables recombinant SNAPC assembly and function for transcription by human RNA polymerases II and III. **Protein Expr Purif**, *48*, 215-223.

Hao, L., Ha, J.R., Kuzel, P., Garcia, E., and Persad, S. (2012). Cadherin switch from E- to N-cadherin in melanoma progression is regulated by the PI3K/PTEN pathway through Twist and Snail. **Br J Dermatol**, *166*, 1184-1197.

Hartsock, A., and Nelson, W.J. (2008). Adherens and tight junctions: structure, function and connections to the actin cytoskeleton. **Biochim Biophys Acta**, *1778*, 660-669.

Heikkinen, J., Risteli, M., Wang, C., Latvala, J., Rossi, M., Valtavaara, M., and Myllyla, R. (2000). Lysyl hydroxylase 3 is a multifunctional protein possessing collagen glucosyltransferase activity. **J Biol Chem**, *275*, 36158-36163.

Hotchin, N.A., Gandarillas, A., and Watt, F.M. (1995). Regulation of cell surface beta 1 integrin levels during keratinocyte terminal differentiation. **J Cell Biol**, *128*, 1209-1219.

Hsu, S.C., Ting, A.E., Hazuka, C.D., Davanger, S., Kenny, J.W., Kee, Y., and Scheller, R.H. (1996). The mammalian brain rsec6/8 complex. **Neuron**, *17*, 1209-1219.

Hudson, B.G., Reeders, S.T., and Tryggvason, K. (1993). Type IV collagen: structure, gene organization, and role in human diseases. Molecular basis of Goodpasture and Alport syndromes and diffuse leiomyomatosis. **J Biol Chem**, *268*, 26033-26036.

Huizing, M., Didier, A., Walenta, J., Anikster, Y., Gahl, W.A., and Kramer, H. (2001). Molecular cloning and characterization of human VPS18, VPS 11, VPS16, and VPS33. **Gene**, *264*, 241-247.

Hutagalung, A.H., and Novick, P.J. (2011). Role of Rab GTPases in membrane traffic and cell physiology. **Physiol Rev**, *91*, 119-149.

Hyry, M., Lantto, J., and Myllyharju, J. (2009). Missense mutations that cause Bruck syndrome affect enzymatic activity, folding, and oligomerization of lysyl hydroxylase 2. **J Biol Chem**, *284*, 30917-30924.

Idrissi, F.Z., Blasco, A., Espinal, A., and Geli, M.I. (2012). Ultrastructural dynamics of proteins involved in endocytic budding. **Proc Natl Acad Sci U S A**, *109*, E2587-2594.

Ishizaka, Y., Ochiai, M., Tahira, T., Sugimura, T., and Nagao, M. (1989). Activation of the ret-II oncogene without a sequence encoding a transmembrane domain and transforming activity of two ret-II oncogene products differing in carboxy-termini due to alternative splicing. **Oncogene**, *4*, 789-794.

Jahn, R., Lang, T., and Sudhof, T.C. (2003). Membrane fusion. **Cell**, *112*, 519-533.

Jahn, R., and Scheller, R.H. (2006). SNAREs--engines for membrane fusion. **Nat Rev Mol Cell Biol**, *7*, 631-643.

Kadler, K.E., Baldock, C., Bella, J., and Boot-Handford, R.P. (2007). Collagens at a glance. **J Cell Sci**, *120*, 1955-1958.

Kadler, K.E., Holmes, D.F., Trotter, J.A., and Chapman, J.A. (1996). Collagen fibril formation. **Biochem J**, *316 (Pt 1)*, 1-11.

Kavalali, E.T. (2002). SNARE interactions in membrane trafficking: a perspective from mammalian central synapses. **Bioessays**, *24*, 926-936.

Keller, P., and Simons, K. (1997). Post-Golgi biosynthetic trafficking. **J Cell Sci**, *110 (Pt 24)*, 3001-3009.

Kellokumpu, S., Sormunen, R., Heikkinen, J., and Myllyla, R. (1994). Lysyl hydroxylase, a collagen processing enzyme, exemplifies a novel class of luminally-oriented peripheral membrane proteins in the endoplasmic reticulum. **J Biol Chem**, *269*, 30524-30529.

Kelly, E.E., Horgan, C.P., and McCaffrey, M.W. (2012). Rab11 proteins in health and disease. **Biochem Soc Trans**, *40*, 1360-1367.
Kerrigan, J.J., Xie, Q., Ames, R.S., and Lu, Q. (2011). Production of protein complexes via co-expression. **Protein Expr Purif**, *75*, 1-14.

Khoshnoodi, J., Pedchenko, V., and Hudson, B.G. (2008). Mammalian collagen IV. **Microsc Res Tech**, *71*, 357-370.

Kinchen, J.M., Doukoumetzidis, K., Almendinger, J., Stergiou, L., Tosello-Trampont, A., Sifri, C.D., Hengartner, M.O., and Ravichandran, K.S. (2008). A pathway for phagosome maturation during engulfment of apoptotic cells. **Nat Cell Biol**, *10*, 556-566.

Kirchhausen, T. (1999). Adaptors for clathrin-mediated traffic. **Annu Rev Cell Dev Biol**, *15*, 705-732.

Kirchhausen, T. (2000). Three ways to make a vesicle. **Nat Rev Mol Cell Biol**, *1*, 187-198.

Kivirikko, K.I., and Myllyla, R. (1983). [The collagen gene family]. **Duodecim**, *99*, 1373-1382.

Kivirikko, K.I., and Pihlajaniemi, T. (1998). Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. **Adv Enzymol Relat Areas Mol Biol**, *72*, 325-398.

Kivirikko, K.I., Ryhanen, L., Anttinen, H., Bornstein, P., and Prockop, D.J. (1973). Further hydroxylation of lysyl residues in collagen by protocollagen lysyl hydroxylase in vitro. **Biochemistry**, *12*, 4966-4971.

Knott, L., and Bailey, A.J. (1998). Collagen cross-links in mineralizing tissues: a review of their chemistry, function, and clinical relevance. **Bone**, *22*, 181-187.

Kokudo, T., Suzuki, Y., Yoshimatsu, Y., Yamazaki, T., Watabe, T., and Miyazono, K. (2008). Snail is required for TGFbeta-induced endothelial-mesenchymal transition of embryonic stem cell-derived endothelial cells. **J Cell Sci**, *121*, 3317-3324.

Kuhn, K., Glanville, R.W., Babel, W., Qian, R.Q., Dieringer, H., Voss, T., Siebold, B., Oberbaumer, I., Schwarz, U., and Yamada, Y. (1985). The structure of type IV collagen. **Ann N Y Acad Sci**, *460*, 14-24.

L'Hernault, S.W., and Faundez, V. (2011). On the endosomal function and gene nomenclature of human SPE-39. **Nat Genet**, *43*, 176.

Lang, K., Schmid, F.X., and Fischer, G. (1987). Catalysis of protein folding by prolyl isomerase. **Nature**, *329*, 268-270.

Lauer-Fields, J.L., Malkar, N.B., Richet, G., Drauz, K., and Fields, G.B. (2003). Melanoma cell CD44 interaction with the alpha 1(IV)1263-1277 region from basement membrane collagen is modulated by ligand glycosylation. **J Biol Chem**, *278*, 14321-14330.

Lian, J.P., Stone, S., Jiang, Y., Lyons, P., and Ferro-Novick, S. (1994). Ypt1p implicated in v-SNARE activation. **Nature**, *372*, 698-701.

Lobingier, B.T., and Merz, A.J. (2012). Sec1/Munc18 protein Vps33 binds to SNARE domains and the quaternary SNARE complex. **Mol Biol Cell**, *23*, 4611-4622.

Lu, P., Takai, K., Weaver, V.M., and Werb, Z. (2011). Extracellular matrix degradation and remodeling in development and disease. **Cold Spring Harb Perspect Biol**, *3*.

Luo, F., Liu, J., and Li, J. (2010). Discovering conditional co-regulated protein complexes by integrating diverse data sources. **BMC Syst Biol**, *4 Suppl 2*, S4.

Malsam, J., Kreye, S., and Sollner, T.H. (2008). Membrane fusion: SNAREs and regulation. **Cell Mol Life Sci**, *65*, 2814-2832.

Matthews, R.P., Plumb-Rudewiez, N., Lorent, K., Gissen, P., Johnson, C.A., Lemaigre, F., and Pack, M. (2005). Zebrafish vps33b, an ortholog of the gene responsible for human arthrogryposis-renal dysfunction-cholestasis syndrome, regulates biliary development downstream of the onecut transcription factor hnf6. **Development**, *132*, 5295-5306.

May, A.P., Whiteheart, S.W., and Weis, W.I. (2001). Unraveling the mechanism of the vesicle transport ATPase NSF, the N-ethylmaleimide-sensitive factor. **J Biol Chem**, *276*, 21991-21994.

Miner, J.H. (1999). Renal basement membrane components. **Kidney Int**, *56*, 2016-2024.

Misura, K.M., Scheller, R.H., and Weis, W.I. (2000). Three-dimensional structure of the neuronal-Sec1-syntaxin 1a complex. **Nature**, *404*, 355-362.

Mundy, G.R. (1996). Regulation of bone formation by bone morphogenetic proteins and other growth factors. **Clin Orthop Relat Res**, 24-28.

Munro, S. (2011). The golgin coiled-coil proteins of the Golgi apparatus. **Cold Spring Harb Perspect Biol**, *3*.

Myers, L.K., Myllyharju, J., Nokelainen, M., Brand, D.D., Cremer, M.A., Stuart, J.M., Bodo, M., Kivirikko, K.I., and Kang, A.H. (2004). Relevance of posttranslational modifications for the arthritogenicity of type II collagen. **J Immunol**, *172*, 2970-2975.

Myllyharju, J., and Kivirikko, K.I. (2004). Collagens, modifying enzymes and their mutations in humans, flies and worms. **Trends Genet**, *20*, 33-43.

Myllyla, R., Pihlajaniemi, T., Pajunen, L., Turpeenniemi-Hujanen, T., and Kivirikko, K.I. (1991). Molecular cloning of chick lysyl hydroxylase. Little homology in primary structure to the two types of subunit of prolyl 4-hydroxylase. **J Biol Chem**, *266*, 2805-2810.

Myllyla, R., Wang, C., Heikkinen, J., Juffer, A., Lampela, O., Risteli, M., Ruotsalainen, H., Salo, A., and Sipila, L. (2007). Expanding the lysyl hydroxylase toolbox: new insights into the localization and activities of lysyl hydroxylase 3 (LH3). **J Cell Physiol**, *212*, 323-329.

Nakamura, N., Hirata, A., Ohsumi, Y., and Wada, Y. (1997). Vam2/Vps41p and Vam6/Vps39p are components of a protein complex on the vacuolar membranes and involved in the vacuolar assembly in the yeast Saccharomyces cerevisiae. **J Biol Chem**, *272*, 11344-11349.

Nakatsu, F., and Ohno, H. (2003). Adaptor protein complexes as the key regulators of protein sorting in the post-Golgi network. **Cell Struct Funct**, *28*, 419-429.

Norman, K.R., and Moerman, D.G. (2000). The let-268 locus of Caenorhabditis elegans encodes a procollagen lysyl hydroxylase that is essential for type IV collagen secretion. **Dev Biol**, *227*, 690-705.

Notbohm, H., Nokelainen, M., Myllyharju, J., Fietzek, P.P., Muller, P.K., and Kivirikko, K.I. (1999). Recombinant human type II collagens with low and high levels of hydroxylysine and its glycosylated forms show marked differences in fibrillogenesis in vitro. **J Biol Chem**, *274*, 8988-8992.

Novick, P., Field, C., and Schekman, R. (1980). Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. **Cell**, *21*, 205-215.

O'Leary, L.E., Fallas, J.A., Bakota, E.L., Kang, M.K., and Hartgerink, J.D. (2011). Multi-hierarchical self-assembly of a collagen mimetic peptide from triple helix to nanofibre and hydrogel. **Nat Chem**, *3*, 821-828.

Oda, H., and Takeichi, M. (2011). Evolution: structural and functional diversity of cadherin at the adherens junction. **J Cell Biol**, *193*, 1137-1146.

Ohlund, D., Franklin, O., Lundberg, E., Lundin, C., and Sund, M. (2013). Type IV collagen stimulates pancreatic cancer cell proliferation, migration, and inhibits apoptosis through an autocrine loop. **BMC Cancer**, *13*, 154.

Ortega, N., and Werb, Z. (2002). New functional roles for non-collagenous domains of basement membrane collagens. **J Cell Sci**, *115*, 4201-4214.

Orzech, E., Livshits, L., Leyt, J., Okhrimenko, H., Reich, V., Cohen, S., Weiss, A., Melamed-Book, N., Lebendiker, M., Altschuler, Y*., et al.* (2001). Interactions between

adaptor protein-1 of the clathrin coat and microtubules via type 1a microtubule-associated proteins. **J Biol Chem**, *276*, 31340-31348.

Ossig, R., Dascher, C., Trepte, H.H., Schmitt, H.D., and Gallwitz, D. (1991). The yeast SLY gene products, suppressors of defects in the essential GTP-binding Ypt1 protein, may act in endoplasmic reticulum-to-Golgi transport. **Mol Cell Biol**, *11*, 2980-2993.

Ozawa, T., Tsuruta, D., Jones, J.C., Ishii, M., Ikeda, K., Harada, T., Aoyama, Y., Kawada, A., and Kobayashi, H. (2010). Dynamic relationship of focal contacts and hemidesmosome protein complexes in live cells. **J Invest Dermatol**, *130*, 1624-1635.

Page, L.J., and Robinson, M.S. (1995). Targeting signals and subunit interactions in coated vesicle adaptor complexes. **J Cell Biol**, *131*, 619-630.

Peissel, B., Geng, L., Kalluri, R., Kashtan, C., Rennke, H.G., Gallo, G.R., Yoshioka, K., Sun, M.J., Hudson, B.G., Neilson, E.G.*, et al.* (1995). Comparative distribution of the alpha 1(IV), alpha 5(IV), and alpha 6(IV) collagen chains in normal human adult and fetal tissues and in kidneys from X-linked Alport syndrome patients. **J Clin Invest**, *96*, 1948-1957.

Pelham, H.R. (2001). SNAREs and the specificity of membrane fusion. **Trends Cell Biol**, *11*, 99-101.

Peplowska, K., Markgraf, D.F., Ostrowicz, C.W., Bange, G., and Ungermann, C. (2007). The CORVET tethering complex interacts with the yeast Rab5 homolog Vps21 and is involved in endo-lysosomal biogenesis. **Dev Cell**, *12*, 739-750.

Peralta, E.R., Martin, B.C., and Edinger, A.L. (2010). Differential effects of TBC1D15 and mammalian Vps39 on Rab7 activation state, lysosomal morphology, and growth factor dependence. **J Biol Chem**, *285*, 16814-16821.

Pereira-Leal, J.B., and Seabra, M.C. (2000). The mammalian Rab family of small GTPases: definition of family and subfamily sequence motifs suggests a mechanism for functional specificity in the Ras superfamily. **J Mol Biol**, *301*, 1077-1087.

Pereira-Leal, J.B., and Seabra, M.C. (2001). Evolution of the Rab family of small GTP-binding proteins. **J Mol Biol**, *313*, 889-901.

Peterson, M.R., and Emr, S.D. (2001). The class C Vps complex functions at multiple stages of the vacuolar transport pathway. **Traffic**, *2*, 476-486.

Pfeffer, S.R. (1999). Transport-vesicle targeting: tethers before SNAREs. **Nat Cell Biol**, *1*, E17-22.

Poirier, M.A., Xiao, W., Macosko, J.C., Chan, C., Shin, Y.K., and Bennett, M.K. (1998). The synaptic SNARE complex is a parallel four-stranded helical bundle. **Nat Struct Biol**, *5*, 765-769.

Pokutta, S., Herrenknecht, K., Kemler, R., and Engel, J. (1994). Conformational changes of the recombinant extracellular domain of E-cadherin upon calcium binding. **Eur J Biochem**, *223*, 1019-1026.

Pols, M.S., van Meel, E., Oorschot, V., ten Brink, C., Fukuda, M., Swetha, M.G., Mayor, S., and Klumperman, J. (2013). hVps41 and VAMP7 function in direct TGN to late endosome transport of lysosomal membrane proteins. **Nat Commun**, *4*, 1361.

Price, A., Seals, D., Wickner, W., and Ungermann, C. (2000). The docking stage of yeast vacuole fusion requires the transfer of proteins from a cis-SNARE complex to a Rab/Ypt protein. **J Cell Biol**, *148*, 1231-1238.

Prockop, D.J., Sieron, A.L., and Li, S.W. (1998). Procollagen N-proteinase and procollagen C-proteinase. Two unusual metalloproteinases that are essential for procollagen processing probably have important roles in development and cell signaling. **Matrix Biol**, *16*, 399-408.

Rabes, H.M. (2001). Gene rearrangements in radiation-induced thyroid carcinogenesis. **Med Pediatr Oncol**, *36*, 574-582.

Rautavuoma, K., Takaluoma, K., Passoja, K., Pirskanen, A., Kvist, A.P., Kivirikko, K.I., and Myllyharju, J. (2002). Characterization of three fragments that constitute the monomers of the human lysyl hydroxylase isoenzymes 1-3. The 30-kDa N-terminal fragment is not required for lysyl hydroxylase activity. **J Biol Chem**, *277*, 23084-23091.

Rautavuoma, K., Takaluoma, K., Sormunen, R., Myllyharju, J., Kivirikko, K.I., and Soininen, R. (2004). Premature aggregation of type IV collagen and early lethality in lysyl hydroxylase 3 null mice. **Proc Natl Acad Sci U S A**, *101*, 14120-14125.

Rehling, P., Darsow, T., Katzmann, D.J., and Emr, S.D. (1999). Formation of AP-3 transport intermediates requires Vps41 function. **Nat Cell Biol**, *1*, 346-353.

Reilly, B.A., Kraynack, B.A., VanRheenen, S.M., and Waters, M.G. (2001). Golgi-to-endoplasmic reticulum (ER) retrograde traffic in yeast requires Dsl1p, a component of the ER target site that interacts with a COPI coat subunit. **Mol Biol Cell**, *12*, 3783-3796.

Reiser, K., McCormick, R.J., and Rucker, R.B. (1992). Enzymatic and nonenzymatic cross-linking of collagen and elastin. **FASEB J**, *6*, 2439-2449.

Rider, L., Oladimeji, P., and Diakonova, M. (2013). PAK1 Regulates Breast Cancer Cell Invasion through Secretion of Matrix Metalloproteinases in Response to Prolactin and Three-Dimensional Collagen IV. **Mol Endocrinol**, *27*, 1048-1064.

Rieder, S.E., and Emr, S.D. (1997). A novel RING finger protein complex essential for a late step in protein transport to the yeast vacuole. **Mol Biol Cell**, *8*, 2307-2327.

Risteli, M., Niemitalo, O., Lankinen, H., Juffer, A.H., and Myllyla, R. (2004). Characterization of collagenous peptides bound to lysyl hydroxylase isoforms. **J Biol Chem**, *279*, 37535-37543.

Risteli, M., Ruotsalainen, H., Salo, A.M., Sormunen, R., Sipila, L., Baker, N.L., Lamande, S.R., Vimpari-Kauppinen, L., and Myllyla, R. (2009). Reduction of lysyl hydroxylase 3 causes deleterious changes in the deposition and organization of extracellular matrix. **J Biol Chem**, *284*, 28204-28211.

Rodriguez-Boulan, E., and Nelson, W.J. (1989). Morphogenesis of the polarized epithelial cell phenotype. **Science**, *245*, 718-725.

Rodriguez-Boulan E, Kreitzer G, Müsch A (2005) Organization of vesicular trafficking in epithelia. *Nat*. **Rev. Mol. Cell Biol**. 6:233-247.

Ruotsalainen, H., Risteli, M., Wang, C., Wang, Y., Karppinen, M., Bergmann, U., Kvist, A.P., Pospiech, H., Herzig, K.H., and Myllyla, R. (2012). The activities of lysyl hydroxylase 3 (LH3) regulate the amount and oligomerization status of adiponectin. **PLoS One**, *7*, e50045.

Ruotsalainen, H., Sipila, L., Kerkela, E., Pospiech, H., and Myllyla, R. (1999). Characterization of cDNAs for mouse lysyl hydroxylase 1, 2 and 3, their phylogenetic analysis and tissue-specific expression in the mouse. **Matrix Biol**, *18*, 325-329.

Ruotsalainen, H., Sipila, L., Vapola, M., Sormunen, R., Salo, A.M., Uitto, L., Mercer, D.K., Robins, S.P., Risteli, M., Aszodi, A.*, et al.* (2006). Glycosylation catalyzed by lysyl hydroxylase 3 is essential for basement membranes. **J Cell Sci**, *119*, 625-635.

Sacher, M., Jiang, Y., Barrowman, J., Scarpa, A., Burston, J., Zhang, L., Schieltz, D., Yates, J.R., 3rd, Abeliovich, H., and Ferro-Novick, S. (1998). TRAPP, a highly conserved novel complex on the cis-Golgi that mediates vesicle docking and fusion. **EMBO J**, *17*, 2494-2503.

Sachse, M., Ramm, G., Strous, G., and Klumperman, J. (2002). Endosomes: multipurpose designs for integrating housekeeping and specialized tasks. **Histochem Cell Biol**, *117*, 91-104.

Salminen, A., and Novick, P.J. (1987). A ras-like protein is required for a post-Golgi event in yeast secretion. **Cell**, *49*, 527-538.

Salo, A.M., Cox, H., Farndon, P., Moss, C., Grindulis, H., Risteli, M., Robins, S.P., and Myllyla, R. (2008). A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. **Am J Hum Genet**, *83*, 495-503.

Salo, A.M., Sipila, L., Sormunen, R., Ruotsalainen, H., Vainio, S., and Myllyla, R. (2006a). The lysyl hydroxylase isoforms are widely expressed during mouse embryogenesis, but obtain tissue- and cell-specific patterns in the adult. **Matrix Biol**, *25*, 475-483.

Salo, A.M., Wang, C., Sipila, L., Sormunen, R., Vapola, M., Kervinen, P., Ruotsalainen, H., Heikkinen, J., and Myllyla, R. (2006b). Lysyl hydroxylase 3 (LH3) modifies proteins in the extracellular space, a novel mechanism for matrix remodeling. **J Cell Physiol**, *207*, 644-653.

Sannerud, R., Saraste, J., and Goud, B. (2003). Retrograde traffic in the biosynthetic-secretory route: pathways and machinery. **Curr Opin Cell Biol**, *15*, 438-445.

Sato, T.K., Rehling, P., Peterson, M.R., and Emr, S.D. (2000). Class C Vps protein complex regulates vacuolar SNARE pairing and is required for vesicle docking/fusion. **Mol Cell**, *6*, 661-671.

Satoh, A., Wang, Y., Malsam, J., Beard, M.B., and Warren, G. (2003). Golgin-84 is a rab1 binding partner involved in Golgi structure. **Traffic**, *4*, 153-161.

Schegg, B., Hulsmeier, A.J., Rutschmann, C., Maag, C., and Hennet, T. (2009). Core glycosylation of collagen is initiated by two beta(1-O)galactosyltransferases. **Mol Cell Biol**, *29*, 943-952.

Schlierf, B., Fey, G.H., Hauber, J., Hocke, G.M., and Rosorius, O. (2000). Rab11b is essential for recycling of transferrin to the plasma membrane. **Exp Cell Res**, *259*, 257-265.

Schneider, V.A., and Granato, M. (2006). The myotomal diwanka (lh3) glycosyltransferase and type XVIII collagen are critical for motor growth cone migration. **Neuron**, *50*, 683-695.

Schneider, V.A., and Granato, M. (2007). Genomic structure and embryonic expression of zebrafish lysyl hydroxylase 1 and lysyl hydroxylase 2. **Matrix Biol**, *26*, 12-19.

Schuck, S., and Simons, K. (2004). Polarized sorting in epithelial cells: raft clustering and the biogenesis of the apical membrane. **J Cell Sci**, *117*, 5955-5964.

Schwartz, S.L., Cao, C., Pylypenko, O., Rak, A., and Wandinger-Ness, A. (2007). Rab GTPases at a glance. **J Cell Sci**, *120*, 3905-3910.

Seals, D.F., Eitzen, G., Margolis, N., Wickner, W.T., and Price, A. (2000). A Ypt/Rab effector complex containing the Sec1 homolog Vps33p is required for homotypic vacuole fusion. **Proc Natl Acad Sci U S A**, *97*, 9402-9407.

Shih, W., Gallusser, A., and Kirchhausen, T. (1995). A clathrin-binding site in the hinge of the beta 2 chain of mammalian AP-2 complexes. **J Biol Chem**, *270*, 31083-31090.

Silver, F.H., Freeman, J.W., and Seehra, G.P. (2003). Collagen self-assembly and the development of tendon mechanical properties. **J Biomech**, *36*, 1529-1553.

Silvis, M.R., Bertrand, C.A., Ameen, N., Golin-Bisello, F., Butterworth, M.B., Frizzell, R.A., and Bradbury, N.A. (2009). Rab11b regulates the apical recycling of the cystic fibrosis transmembrane conductance regulator in polarized intestinal epithelial cells. **Mol Biol Cell**, *20*, 2337-2350.
Simonsen, A., Lippe, R., Christoforidis, S., Gaullier, J.M., Brech, A., Callaghan, J., Toh, B.H., Murphy, C., Zerial, M., and Stenmark, H. (1998). EEA1 links PI(3)K function to Rab5 regulation of endosome fusion. **Nature**, *394*, 494-498.

Sipila, L., Ruotsalainen, H., Sormunen, R., Baker, N.L., Lamande, S.R., Vapola, M., Wang, C., Sado, Y., Aszodi, A., and Myllyla, R. (2007). Secretion and assembly of type IV and VI collagens depend on glycosylation of hydroxylysines. **J Biol Chem**, *282*, 33381-33388.

Smith, H., Galmes, R., Gogolina, E., Straatman-Iwanowska, A., Reay, K., Banushi, B., Bruce, C.K., Cullinane, A.R., Romero, R., Chang, R.*, et al.* (2012). Associations among genotype, clinical phenotype, and intracellular localization of trafficking proteins in ARC syndrome. **Hum Mutat**, *33*, 1656-1664.

Sohda, M., Misumi, Y., Yamamoto, A., Nakamura, N., Ogata, S., Sakisaka, S., Hirose, S., Ikehara, Y., and Oda, K. (2010). Interaction of Golgin-84 with the COG complex mediates the intra-Golgi retrograde transport. **Traffic**, *11*, 1552-1566.

Solinger, J.A., and Spang, A. (2013). Tethering complexes in the endocytic pathway: CORVET and HOPS. **FEBS J**, *280*, 2743-2757.

Sollner, T., Whiteheart, S.W., Brunner, M., Erdjument-Bromage, H., Geromanos, S., Tempst, P., and Rothman, J.E. (1993). SNAP receptors implicated in vesicle targeting and fusion. **Nature**, *362*, 318-324.

Sricholpech, M., Perdivara, I., Nagaoka, H., Yokoyama, M., Tomer, K.B., and Yamauchi, M. (2011). Lysyl hydroxylase 3 glucosylates galactosylhydroxylysine residues in type I collagen in osteoblast culture. **J Biol Chem**, *286*, 8846-8856.

Sricholpech, M., Perdivara, I., Yokoyama, M., Nagaoka, H., Terajima, M., Tomer, K.B., and Yamauchi, M. (2012). Lysyl hydroxylase 3-mediated glucosylation in type I collagen: molecular loci and biological significance. **J Biol Chem**, *287*, 22998-23009.

Stamnes, M. (2002). Regulating the actin cytoskeleton during vesicular transport. **Curr Opin Cell Biol**, *14*, 428-433.

Stanley, P. (2011). Golgi glycosylation. **Cold Spring Harb Perspect Biol**, *3*.

Stein, M.P., Dong, J., and Wandinger-Ness, A. (2003). Rab proteins and endocytic trafficking: potential targets for therapeutic intervention. **Adv Drug Deliv Rev**, *55*, 1421-1437.

Stenmark, H., and Olkkonen, V.M. (2001). The Rab GTPase family. **Genome Biol**, *2*, REVIEWS3007.

Stephens, D.J. (2012). Cell biology: Collagen secretion explained. **Nature**, *482*, 474-475.
Stroupe, C., Collins, K.M., Fratti, R.A., and Wickner, W. (2006). Purification of active HOPS complex reveals its affinities for phosphoinositides and the SNARE Vam7p. **EMBO J**, *25*, 1579-1589.

Subramanian, S., Woolford, C.A., and Jones, E.W. (2004). The Sec1/Munc18 protein, Vps33p, functions at the endosome and the vacuole of Saccharomyces cerevisiae. **Mol Biol Cell**, *15*, 2593-2605.

Sudhof, T.C., and Rothman, J.E. (2009). Membrane fusion: grappling with SNARE and SM proteins. **Science**, *323*, 474-477.

Sutton, R.B., Fasshauer, D., Jahn, R., and Brunger, A.T. (1998). Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. **Nature**, *395*, 347-353.

Takeichi, M. (1991). Cadherin cell adhesion receptors as a morphogenetic regulator. **Science**, *251*, 1451-1455.

Tan, K., Shlomi, T., Feizi, H., Ideker, T., and Sharan, R. (2007). Transcriptional regulation of protein complexes within and across species. **Proc Natl Acad Sci U S A**, *104*, 1283-1288.

Tanaka, H., Nishino, M., Ishida, M., Fukunaga, R., and Sueyoshi, K. (1992). Progression of carotid atherosclerosis in Japanese patients with coronary artery disease. **Stroke**, *23*, 946-951.

Tanos, B., and Rodriguez-Boulan, E. (2008). The epithelial polarity program: machineries involved and their hijacking by cancer. **Oncogene**, *27*, 6939-6957.

Theard, D., Steiner, M., Kalicharan, D., Hoekstra, D., and van Ijzendoorn, S.C. (2007). Cell polarity development and protein trafficking in hepatocytes lacking E-cadherin/beta-catenin-based adherens junctions. **Mol Biol Cell**, *18*, 2313-2321.

Tiainen, P., Pasanen, A., Sormunen, R., and Myllyharju, J. (2008). Characterization of recombinant human prolyl 3-hydroxylase isoenzyme 2, an enzyme modifying the basement membrane collagen IV. **J Biol Chem**, *283*, 19432-19439.

Tong, M., Chan, K.W., Bao, J.Y., Wong, K.Y., Chen, J.N., Kwan, P.S., Tang, K.H., Fu, L., Qin, Y.R., Lok, S.*, et al.* (2012). Rab25 is a tumor suppressor gene with antiangiogenic and anti-invasive activities in esophageal squamous cell carcinoma. **Cancer Res**, *72*, 6024-6035.

Toonen, R.F., and Verhage, M. (2003). Vesicle trafficking: pleasure and pain from SM genes. **Trends Cell Biol**, *13*, 177-186.

Traub, L.M., and Kornfeld, S. (1997). The trans-Golgi network: a late secretory sorting station. **Curr Opin Cell Biol**, *9*, 527-533.
Tryggvason, K. (1996). Mutations in type IV collagen genes and Alport phenotypes. **Contrib Nephrol**, *117*, 154-171.

Tsukita, S., Furuse, M., and Itoh, M. (2001). Multifunctional strands in tight junctions. **Nat Rev Mol Cell Biol**, *2*, 285-293.

Turpeenniemi-Hujanen, T.M., Puistola, U., and Kivirikko, K.I. (1980). Isolation of lysyl hydroxylase, an enzyme of collagen synthesis, from chick embryos as a homogeneous protein. **Biochem J**, *189*, 247-253.

Tuvim, M.J., Adachi, R., Hoffenberg, S., and Dickey, B.F. (2001). Traffic control: Rab GTPases and the regulation of interorganellar transport. **News Physiol Sci**, *16*, 56-61.

Ullrich, O., Reinsch, S., Urbe, S., Zerial, M., and Parton, R.G. (1996). Rab11 regulates recycling through the pericentriolar recycling endosome. **J Cell Biol**, *135*, 913-924.

Ungar, D., and Hughson, F.M. (2003). SNARE protein structure and function. **Annu Rev Cell Dev Biol**, *19*, 493-517.

Ungar, D., Oka, T., Brittle, E.E., Vasile, E., Lupashin, V.V., Chatterton, J.E., Heuser, J.E., Krieger, M., and Waters, M.G. (2002). Characterization of a mammalian Golgi-localized protein complex, COG, that is required for normal Golgi morphology and function. **J Cell Biol**, *157*, 405-415.

Urban, D., Li, L., Christensen, H., Pluthero, F.G., Chen, S.Z., Puhacz, M., Garg, P.M., Lanka, K.K., Cummings, J.J., Kramer, H.*, et al.* (2012). The VPS33B-binding protein VPS16B is required in megakaryocyte and platelet alpha-granule biogenesis. **Blood**, *120*, 5032-5040.

Urbe, S., Huber, L.A., Zerial, M., Tooze, S.A., and Parton, R.G. (1993). Rab11, a small GTPase associated with both constitutive and regulated secretory pathways in PC12 cells. **FEBS Lett**, *334*, 175-182.

Van den Steen, P.E., Proost, P., Brand, D.D., Kang, A.H., Van Damme, J., and Opdenakker, G. (2004). Generation of glycosylated remnant epitopes from human collagen type II by gelatinase B. **Biochemistry**, *43*, 10809-10816.

Wada, Y., Kitamoto, K., Kanbe, T., Tanaka, K., and Anraku, Y. (1990). The SLP1 gene of Saccharomyces cerevisiae is essential for vacuolar morphogenesis and function. **Mol Cell Biol**, *10*, 2214-2223.

Wang, C., Kovanen, V., Raudasoja, P., Eskelinen, S., Pospiech, H., and Myllyla, R. (2009). The glycosyltransferase activities of lysyl hydroxylase 3 (LH3) in the extracellular space are important for cell growth and viability. **J Cell Mol Med**, *13*, 508-521.

Wang, C., Luosujarvi, H., Heikkinen, J., Risteli, M., Uitto, L., and Myllyla, R. (2002a). The third activity for lysyl hydroxylase 3: galactosylation of hydroxylysyl residues in collagens in vitro. **Matrix Biol**, *21*, 559-566.

Wang, C., Risteli, M., Heikkinen, J., Hussa, A.K., Uitto, L., and Myllyla, R. (2002b). Identification of amino acids important for the catalytic activity of the collagen glucosyltransferase associated with the multifunctional lysyl hydroxylase 3 (LH3). **J Biol Chem**, *277*, 18568-18573.

Wang, C., Ristiluoma, M.M., Salo, A.M., Eskelinen, S., and Myllyla, R. (2012). Lysyl hydroxylase 3 is secreted from cells by two pathways. **J Cell Physiol**, *227*, 668-675.

Wang C, Valtavaara M, Myllylä R (2000) Lack of collagen type specificity for lysyl hydroxylase isoforms. **DNA Cell Biol** 19:71-77.

Wang, W., Sacher, M., and Ferro-Novick, S. (2000). TRAPP stimulates guanine nucleotide exchange on Ypt1p. **J Cell Biol**, *151*, 289-296.

Wang, Y., Xu, A., Knight, C., Xu, L.Y., and Cooper, G.J. (2002c). Hydroxylation and glycosylation of the four conserved lysine residues in the collagenous domain of adiponectin. Potential role in the modulation of its insulin-sensitizing activity. **J Biol Chem**, *277*, 19521-19529.

Waters, M.G., Clary, D.O., and Rothman, J.E. (1992). A novel 115-kD peripheral membrane protein is required for intercisternal transport in the Golgi stack. **J Cell Biol**, *118*, 1015-1026.

Waters, M.G., and Hughson, F.M. (2000). Membrane tethering and fusion in the secretory and endocytic pathways. **Traffic**, *1*, 588-597.

Wheelock, M.J., Shintani, Y., Maeda, M., Fukumoto, Y., and Johnson, K.R. (2008). Cadherin switching. **J Cell Sci**, *121*, 727-735.

Whyte, J.R., and Munro, S. (2002). Vesicle tethering complexes in membrane traffic. **J Cell Sci**, *115*, 2627-2637.

Wilcke, M., Johannes, L., Galli, T., Mayau, V., Goud, B., and Salamero, J. (2000). Rab11 regulates the compartmentalization of early endosomes required for efficient transport from early endosomes to the trans-golgi network. **J Cell Biol**, *151*, 1207-1220.

Wolf, K., and Friedl, P. (2006). Molecular mechanisms of cancer cell invasion and plasticity. **Br J Dermatol**, *154 Suppl 1*, 11-15.

Wong, D., Bach, H., Sun, J., Hmama, Z., and Av-Gay, Y. (2011). Mycobacterium tuberculosis protein tyrosine phosphatase (PtpA) excludes host vacuolar-H+-ATPase to inhibit phagosome acidification. **Proc Natl Acad Sci U S A**, *108*, 19371-19376.

Wright, C., Edelmann, M., diGleria, K., Kollnberger, S., Kramer, H., McGowan, S., McHugh, K., Taylor, S., Kessler, B., and Bowness, P. (2009). Ankylosing spondylitis monocytes show upregulation of proteins involved in inflammation and the ubiquitin proteasome pathway. **Ann Rheum Dis**, *68*, 1626-1632.
Wurmser, A.E., Sato, T.K., and Emr, S.D. (2000). New component of the vacuolar class C-Vps complex couples nucleotide exchange on the Ypt7 GTPase to SNARE-dependent docking and fusion. **J Cell Biol**, *151*, 551-562.

Xu, J., Lamouille, S., and Derynck, R. (2009). TGF-beta-induced epithelial to mesenchymal transition. **Cell Res**, *19*, 156-172.

Yeowell, H.N., and Walker, L.C. (2000). Mutations in the lysyl hydroxylase 1 gene that result in enzyme deficiency and the clinical phenotype of Ehlers-Danlos syndrome type VI. **Mol Genet Metab**, *71*, 212-224.

Yeowell, H.N., Walker, L.C., Farmer, B., Heikkinen, J., and Myllyla, R. (2000). Mutational analysis of the lysyl hydroxylase 1 gene (PLOD) in six unrelated patients with Ehlers-Danlos syndrome type VI: prenatal exclusion of this disorder in one family. **Hum Mutat**, *16*, 90.

Yu, I.M., and Hughson, F.M. (2010). Tethering factors as organizers of intracellular vesicular traffic. **Annu Rev Cell Dev Biol**, *26*, 137-156.

Zeevaert, R., Foulquier, F., Jaeken, J., and Matthijs, G. (2008). Deficiencies in subunits of the Conserved Oligomeric Golgi (COG) complex define a novel group of Congenital Disorders of Glycosylation. **Mol Genet Metab**, *93*, 15-21.

Zerial, M., and McBride, H. (2001). Rab proteins as membrane organizers. **Nat Rev Mol Cell Biol**, *2*, 107-117.

Zhao, L., Yang, R., Cheng, L., Wang, M., Jiang, Y., and Wang, S. (2010). Epithelial-mesenchymal transitions of bile duct epithelial cells in primary hepatolithiasis. **J Korean Med Sci**, *25*, 1066-1070.

Zhu, G.D., Salazar, G., Zlatic, S.A., Fiza, B., Doucette, M.M., Heilman, C.J., Levey, A.I., Faundez, V., and L'Hernault S, W. (2009). SPE-39 family proteins interact with the HOPS complex and function in lysosomal delivery. **Mol Biol Cell**, *20*, 1223-1240.

Zlatic, S.A., Tornieri, K., L'Hernault S, W., and Faundez, V. (2011). Metazoan cell biology of the HOPS tethering complex. **Cell Logist**, *1*, 111-117.