LUCIA MAIO

DESIGNED METALLO-TAGS FOR PROBING NATURAL SYSTEMS

Supervisors:

AITOR HIERRO AYUELA
AITZIBER LOPEZ CORTAJARENA

2021

HENRY: Elsa never really believed in the grail. She thought she'd found a prize.
INDIANA: And what did you find, Dad?
HENRY: Me? Illumination.

(Indiana Jones and the Last Crusade)

Acknowledgements

Finally, I am no longer a PhD student.

Getting a PhD is a tortuous and, at the same time, wonderful journey. During last years, I worked hardly, slept badly, over-thought obsessively hypothesis, approaches and experiments. Often, I felt lost and frustrated. Despite all, I think I was exactly where I need to be. Somehow, even where I want to be. My project fed me, and I grow up with it. I am not the same anymore.

Honestly, I am truly convinced that doing research is a wonderful job. I will miss my project. I would like to thank dozens of people who make this experience very remarkable.

Primero, me gustaria agredecer mis supervisores, Dr. Aitor Hierro y Dra. Aitziber L. Cortajarena, por la posibilidad de hacer el doctorado con ellos y de haberme acogido en su laboratorio. Gracias por haberme enseñado a generar ideas y elaborarlas en manera racionable, diseñar experimentos, leer datos con metodo, rigor cientifico y espiritu critico. Voy a echar de menos nuestras reuniones del viernes a la tarde y del lunes a la mañana, donde todo encontraba su sentido y sentias tu mente abrirse.

Ademas, me gustaria agredecer la Dr. Irantzu Llarena de la plataforma de spettroscopia Optica, por su experiencia de medida y las imagines de esta tesis y para haberme enseñado a evaluar las cosas con otro punto de vista mas riguroso y metodico.

Quiero tambien agredecer mis compañeros de laboratorio, tanto en Bilbao tanto en Donosti. Gracias a ellos todo ha sido mucho mas facil, dentro y fuera del laboratorio. A todos os debo algo y mucho mas. Enre ellos, le agradezco especialmente a Carlos Lopez. Puedo decir con bastante seguiridad que esta tesis no hubiera exisistido sin su apoyo. Gracias por haberme enseñado a clonar, purificar, hacer liposomas y, sobretodo, gracias por poder discutir

juntos sobre trafico, ideas, metodos. Gracias por haberme ensenado a andar hasta que no pude andar sola y por haber creido en mi cuanto mas lo necesitaba.

Tambien me gustaria aqui agredecer la Dra Alessandra Girotti, la que fue mi supervisora del master en Valladolid, donde realmente empezò todo. Fue la que me puso por primera vez un plasmido en la mano y me abriò las puertas al mundo de las proteinas y de la nanobiotecnologia. Grazie, per questo e non solo. Per essere aperta di mente, paziente, empatica, sincera e senza peli sulla lingua quando serve. In definitiva una buona amica.

Gracias a Bilbao y sus interminables aventuras. Las noches en el Antxoki, el Klubba, nuestro Mystic, los festivales, la Semana Grande, la Ria, los Garbigunes, los txokos, Bertendona y el Salon de Ilaria y eso de quererse tanto. Seria bonito poder volver atras.

Grazie alla mia famiglia italiana a San Sebastian e per tutta quella vita vissuta insieme per quasi tre anni. Non fate i monelli, vi aspetto qui a Milano da me che mi mancate.

Gracias a Marta Martinez que a parte de ser super graciosa y divertida, ha tenido que aguantar el agobio del papeleo de su tesis y de la mia ☺.

Gracias a Joseba y Arantxa por haberme querido asi como soy, isterica y insoportable cuando me meto en casa. Gracias por haber sido casa y familia durante estos dos años.

Thanks to Tanja to be my pocket great woman and to teach me to be kinder with myself.

Grazie a Ivan e a Carlos, dapprima per la vostra amicizia sincera, per le tonnellate di cose che sapete sulla musica, sui libri, gli spaghetti western e sulla gastronomia. Grazie per avermi dato la possibilità di lavorare insieme e

per quest’anno di lavoro insieme, per tutto quello che mi avete insegnato e per essere stati sempre disponibili a discutere insieme tutto.

Quiero agredecer las LASAI Girls. Me hubiera quedado para siempre en Donosti con vosotras si fuera posible. Gracias por ser como sois, por vuetra comprension, amistad, las ideas traviesas, el amor a la comida y a los vermù y a la vida. Gracias por ser diferente y tan facil de querer. Os echo de menos.

Vorrei inoltre ringraziare le mie cugine e i miei cugini (quelli di sangue e quelli acquisiti). E Marco. Perché tesi o non tesi, ci siete sempre e comunque e non posso fare a meno di voi.

Per ultimo ma non meno importante, vorrei ringraziare i miei genitori e mia sorella, i quali mi sono sempre stati vicini in questo tortuoso percorso che è stato la tesi. Grazie al loro supporto, la loro comprensione, la loro fiducia e i loro consigli, sono sempre riuscita ad affrontare al meglio le varie situazioni e difficoltà che nel tempo ho incontrato sul mio cammino. Nonostante la distanza geografica e le nostre differenze caratteriali, siamo una squadra e niente ci può separare. Sono davvero grata alla mia famiglia, perché è quello fa la differenza.

Table of Contents

List of Abbreviations and Acronyms

AGTm	Modified engineered version of human O^6-alkylguanine-DNA alkyltransferase enzyme
AP	Adaptor protein
ARM	Armadillo repeat
ANK	Ankyrin repeat
BAR	Bin/Amphiphysin/Rvs domain
BG	Benzylguanine derivatives
BC	O^2-benzylcytosine derivates
BSA	Bovine serum albumin
CD	Circular dichroism
cargo	CIMPR $_{(2330\text{-}2491)}$-TEV-$C7_{20Cys}$ (C2342A+C2343A),
cargo-AuNCs	Cys-CIMPR-$C7_{20Cys}$-stabilized AuNCs
cargo-CdS QDs	Cys-CIMPR-$C7_{20Cys}$-stabilized CdS-QDs
cargo-AuNKNCs	Cys-CIMPR-$C7_{20Cys}$-stabilized AuNKNCs
CIMPR	Cation-independent mannose-6-phosphate receptor
CLEM	Correlative light-electron microscopy
CSC	Cargo selection complex
CTPR	Consensus tetratricopeptide repeat
$C7_{20Cys}$	CTPR mutant with 7 repeats and a clamp of 20 Cysteines (Cys, C)
DMT1	Divalent metal transporter 1
DNA	Deoxyribonucleic acid
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine
DOPS	1,2-dioleoyl-sn-glycero-3-phosphoserine
DTT	Threo/1,4/Dimercapto/2,3/butanediol, DL-ithiothreitol

EDTA	Ethylenediaminetetraacetic acid
EE	Early endosomes
EBFP	Enhanced blue fluorescent protein
ECFP	Enhanced cyan fluorescent
EGFP	Enhanced green fluorescent
EGFR	Epidermal growth factor receptor
EM	Electron Microscopy
ER	Endoplasmic reticulum
ESCPE-1	Endosomal SNX-BAR sorting complex for promoting exit 1
ESCRT	Endosomal sorting complex
FPs	Fluorescent proteins
EYFP	Enhanced yellow fluorescent protein
FPLC	Fast Protein Liquid Chromatography
GA	Golgi apparatus
GFP	Green fluorescent protein
GUV	Giant Unilamellar Vesicle
HAADF	High-angle annular dark-field imaging
hAGT	Human O^6-alkylguanine-DNA alkyltransferase
HEPES	4-(2-hydroxyethyl)-1 piperazineethanesulfonic acid
ICP-MS	Inductively coupled plasma mass spectrometry
ILV	Intra-luminal bodies
IEM	Immunoelectron microscopy
IGF1R	Insulin-like growth factor 1 receptor
iRFP	Near- infrared fluorescent protein
IPTG	Isopropyl-B-D-thiogalactopyranoside
LB	Luria Bertani

LE	Late endosomes
LRR	Leucin-rich repeat
LUV	Large unilamellar vesicles
MALDI-TOF	Matrix-Assisted Laser Desorption - Ionisation-Time of Flight Mass SpectrometrY
MCS	Membrane contact sites
miniSOG	Mini Singlet Oxygen Generator
METTEM	Metal-tagging transmission electron microscopy
MLV	Multi lamellar vesicle
MVB	Multi-vesicular bodies
$MT1_{20Cys}$	Cysteine-rich protein metallothionein1 ()
NCs	Nanoclusters
NKNCs	Naked nanoclusters
ncAA	Non-canonical amino acids
OD	Optical density
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PI3P	Phosphatidylinositol 3-phosphate
PM	Plasma membrane
PMSF	Phenylmethylsulfonyl fluoride
POI	Protein of interest
PX	Phox homology
QDs	Quantum dots
RFP	Red fluorescent protein
RE	Recycling endosomes
Rh-PE	Rhodamine-phosphatidylethanolamine
RT	Room temperature

SEC	Size-exclusion chromatography
SDS	Sodium dodecyl sulphate
SENP2	Sentrin specific protease 2
SPL	Self-labelling protein
SNARE	N-ethylmaleimide-sensitive factor attachment protein receptor
SNX	Sorting Nexin
STEM	Scanning Transmission Electron Microscopy
SUMO3	Small ubiquitin-like modifier 3 protein
TAE	Tris-acetate EDTA
TC	Transport Carrier
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TEM	Transmission Electron Microscopy
TEN	Tubular endosomal networks
TEV	Tobacco each virus
TFA	Trifluoroacetic acid
TGN	*Trans* Golgi network
Tm	Melting temperature
TPR	Tetraticopeptide repeat
Tris/HCl	Tris(hydroxymethyl)aminomethane hydrochloride
UV-vis	Ultraviolet-visible
VPS	Vacuolar protein sorting
WASH	Wiskott-Aldrich syndrome protein and scar homolog
WASP	Wiskott-Aldrich syndrome protein

Resumen

Las proteínas son las biomoléculas más importantes de los sistemas biológicos. Por lo tanto, es extremadamente importante investigar la localización, el transporte y la función de las proteínas para comprender mejor sus funciones biológicas.

Con el fin de visualizar las proteínas de interés (POI), se han realizado muchos esfuerzos por mejorar las tecnologías de bioimagen. En las últimas décadas se ha visto el desarrollo de muchas técnicas de microscopía de super-resolución. Dependiendo de la naturaleza de la técnica, la información puede revelar características morfológicas, estructurales, químicas y otras características de la muestra. Actualmente, la microscopía óptica y electrónica se está aplicando principalmente en las investigaciones del campo de la biología celular y la nanotecnología. En concreto, la microscopía de fluorescencia ofrece el potencial de localizar y monitorear específicamente la proteína diana dentro de los sistemas biológicos (a escala micrométrica en el rango de los 0.2 µm). Por otro lado, la microscopía electrónica proporciona información detallada sobre las estructuras subcelulares. Sin embargo, la microscopía electrónica está limitada típicamente a campos de visión pequeños con gran aumento y resolución.

En biología celular, existe una demanda creciente para correlacionar estos niveles individuales de información, que combinan imágenes de resolución nanométrica de microscopía electrónica con la resolución a escala micrométrica de imágenes ópticas. Con este propósito, los investigadores están trabajando extensamente en el desarrollo de nuevas herramientas y metodologías para el etiquetado de proteínas, principalmente enfocadas en imágenes correlativas. Claramente, el desarrollo de nuevas estrategias de etiquetado de proteínas proporcionará una mejora significativa de las técnicas de bioimagen. De hecho, hoy en día existen muchas herramientas y estrategias

disponibles para marcar moléculas y obtener imágenes de microscopía óptica o electrónica. Sin embargo, todavía hay una carencia de herramientas para la microscopía correlativa, en la que se pueda obtener una imagen con una sola etiqueta usando ambas técnicas.

El objetivo de este trabajo es el diseño y desarrollo de nuevas etiquetas mediadas por proteínas para estudiar sistemas biológicamente relevantes. Específicamente, el proyecto se centra en el diseño de andamios moleculares basados en proteínas que promuevan la estabilización de nanoclusters metálicos (NCs) luminiscentes y puntos cuánticos (QDs) como nuevas etiquetas duales para microscopía de fluorescencia y electrónica. Este enfoque aprovecha tanto las propiedades luminiscentes como la densidad electrónica de los nanomateriales estabilizados con proteínas para su aplicación como etiqueta dual para técnicas de imagen correlativa. Finalmente, las nuevas etiquetas duales desarrolladas en este trabajo se utilizan para reproducir el sistema de trasporte retrógrado de proteínas *in vitro* como prueba de concepto para futuros estúdios biológicos.

De forma esquemática, el trabajo se compone de tres capítulos principales.

En primer lugar, en la Introducción se describe el estado del arte de las estrategias de etiquetado mediada por etiquetas proteicas. En particular, presentamos los criterios que deben cumplir las etiquetas ideales basadas en proteínas. Además, proporcionamos la lista y la descripción de las etiquetas disponibles actualmente según las diferentes técnicas de bioimagen. Nos enfocamos en microscopía fluorescente, microscopía electrónica y microscopía electrónica de luz correlativa (CLEM).

En el capítulo titulado “Diseño y síntesis de nanoherramientas de marcaje basadas en proteínas para fluorescencia, crio-TEM y estudios correlativos”, mostramos el diseño y desarrollo de una nueva etiqueta de fusión dual a partir de nanoclusters metálicos (NCs) luminiscentes y puntos cuánticos (QDs)

estabilizados por las proteínas consenso de repetición tetratricopeptídica (acrónimo del inglés, CTPR). Esta etiqueta de fusion dual puede ser utilizada para marcar y estudiar proteínas de interés por fluorescencia y microscopía electrónica. Los nanoclusters metálicos (NCs) luminiscentes y puntos cuánticos (QDs) estabilizadospor péptidos y proteínas representan una herramienta ideal para aplicaciones biológicas ya que pueden realizarse en condiciones suaves de síntesis y las estructuras obtenidas son estables bajo una amplia gama de valores de pH y fuerzas iónicas. Además, poseen gran estabilidad de almacenamiento y trabajo, de fotoestabilidad y biocompatibilidad, condiciones ideales para aplicaciones de bioimagen y bioetiquetado. En este sentido las proteínas consenso de repetición tetratricopeptídica (acrónimo del inglés, CTPR) diseñadas son ideales para estabilizar nanoclusters de oro (AuNC) y puntos cuanticos de sulfuro de cadmio(CdS QD).

Por este motivo clonamos, sobreexpresamos y purificamos con éxito una proteína quimérica (Cys-CIMPR-C720Cys), compuesta por la parte citosolica del receptor de manosa 6-fosfato independente de catión (acrónimo del inglés, CIMPR) y por una CTPR que contiene siete repeticiones y 20 Cys para estabilizar los AuNCs y CdS QDs. De hecho, los AuNCs y los CdS QDs exhiben propiedades ópticas, electrónicas y químicas muy interesantes, incluida una fuerte fotoluminiscencia y un núcleo metálico denso de electrones, aunque tienen un tamaño nanométrico, lo que los convierte en herramientas ideales para microscopía fluorescente y/o electrónica. Usando diferentes protocolos de coordinación de metales, demostramos que Cys-CIMPR-C720Cys puede estabilizar diferentes tipos de AuNCs y CdS QDs con diferentes longitudes de onda de emisión, densidad de electrones y tamaño. Las etiquetas duales aqui desarolladas han sido caracterizadas usando diferentes tecnicas. Para evaluar las propriedades opticas se ha usado spectroscopía a fluorescencia, mientras que para evaluar densidad de

electrones y tamaño se ha usado crio-TEM, MALDI-TOF y ICP-MS. Ademas se ha evaluado la estructura secundaria de las proteinas etiquetadas mediante dicroismo circular. Esta tecnología resulta particularmente atractiva para el etiquetado y para el estudio de sistemas biológicos. Cabe mencionar que es posible ajustar la longitud de onda de emisión y/o el tamaño del metal de la etiqueta dual sin la necesidad de clonar nuevas etiquetas proteicas. En resumen, la capacidad de marcar nanomateriales con una etiqueta con características distintivas, proporciona un enfoque rápido, versátil y que requiere menos tiempo para el etiquetado multimodal.

Finalmente, en el capítulo titulado "Estudios sobre las vías de tráfico retrógrado durante el reclutamento de una proteina modelo (cargo) y la biogénesis de vesículas tubulares mediante nuevas nanoherramientas basadas en proteínas", describimos el desarrollo de un sistema modelo para estudiar las vías de tráfico retrógrado *in vitro* y los estudios de las vías de tráfico retrógrado utilizando las novedosas etiquetas duales, previamente descritas, basados en nanomateriales de proteínas diseñadas.

Después de la introducción general sobre el tráfico de membrana, el objetivo principal de este capítulo es probar la eficacia de las nuevas nanoherramientas basadas en proteínas desarrolladas en este trabajo para estudiar el transporte de proteinas cargo independiente de retrómeros a través de dominios SNX-BAR. Específicamente estudiamos el reciclaje del receptor de manosa-6-fosfato independiente de cationes (CIMPR) por SNX1/SNX5 desde el endosoma al Golgi. Para este propósito, se pueden usar liposomas funcionalizados que imitan la membrana de los endosomas y la proteínas CIMPR etiquetada (Cys-CIMPR-C720Cys), en condiciones fisiológicas, y usada como como modelo *in vitro*. Gracias a la versatilidad de la funcionalización superficial de los liposomas, se pueden utilizar varias estrategias para estudiar las interacciones proteína-membrana. En este trabajo se utilizan dos estrategias. El primer enfoque es inmovilizar la proteína CIMPR modelo etiquetada en la superficie

del liposoma usando acoplamiento tiol-maleimida de acuerdo con protocolos estándarizados, ya que los liposomas están funcionalizados con maleimida y las proteínas etiquetadas tienen una cisteína libre en el extremo N-terminal. El segundo enfoque se basa en la idea de que el reconocimiento de secuencias es un componente importante en la biogénesis de perfiles tubulares y transportadores. Para ello, la proteína CIMPR etiquetada pueden preincubarse con el heterodímero SNX1/SNX5, ya que el dominio PX de SNX5 reconoce directamente CIMPR gracias a su secuencia bipartita, formada por $VSYKYSK_{2349-2455}$ y $WLM_{2369-2371}$. Una vez que se forma complejo CIMPR-SNX, se puede agregar a liposomas enriquecidos con lípido fosfatidilinositol 3-fosfato (PI3P), ya que el heterodímero SNX1/SNX5 se recluta en los endosomas a través de su dominio PX de unión a PI3P y su dominio BAR de sensor de curvatura . Además, estos dos enfoques diferentes para la inmovilización de la proteina CIMPR etiquetada en la superficie del liposoma, brindaron la posibilidad de tener un sistema de modelo *in vitro* más versátil. De hecho, la inmovilización por acoplamiento tiol-maleimida podría usarse para ajustar la concentración de la proteina cargo en la membrana y para estudiar cómo SNX1/SNX5 interactúa con la membrana a diferentes concentraciones de proteina cargo. Por el contrario, el enfoque de reconocimiento de secuencias de SNX1/SNX5 podría usarse para estudios morfológicos, específicamente para estudiar cómo SNX1/SNX5 y las proteínas etiquetadas de proteina cargo se organizan alrededor de las vesículas de transporte.

Nuestros resultados muestran que algunas de las proteínas CIMPR etiquetadas/ desarrolladas en este trabajo son realmente buenas candidatas para estudiar el reciclaje independiente de CIMPR *in vitro* mediante microscopía fluorescente y electrónica. Específicamente, todas las proteínas etiquetadas de proteina cargo pueden reclutarse en la membrana como se muestra en los ensayos de flotación. Por tanto, la interacción directa entre SNX1/SNX5 y CIMPR no se

ve perturbada por parte de la etiqueta. No obstante, las vesiculas tubulares solo se forman con unas de las etiquetas duales desarolladas, las cuales podrían explotarse en trabajos futuros para dilucidar la información estructural utilizando microscopía crioelectrónica de alta resolución (HR cryo-ET) y/o microscopía fluorescente para rastrear diferentes rutas de reciclaje.

1. INTRODUCTION

1. Introduction

1.1. Protein labelling to study biological relevant systems

1.1.1. General aspects

Proteins are one of the most important biomolecules in living systems. Therefore, it is extremely important to investigate protein localization, transport and function for a better understating of mainstream biology. In order to visualize proteins of interest (POIs), many efforts are made to improve bio-imaging technologies. The last decades have seen the development of many super-resolution microscopy techniques. Depending on the nature of the technique, the information may reveal morphological, structural, chemical, and still other sample characteristics. Currently, light microscopy and electron microscopy are mostly applied in the cell biology. Specifically, fluorescence microscopy offers the potential to specifically localize and monitor target protein within biological systems (within micrometric scale 0.2 μm). On the other hand, electron microscopy yields detailed information on subcellular structures. However, electron microscopy is typically restricted to small fields of view at high magnification and resolution. In cell biology, there is an increasing demand to correlate these individual levels of information, that combine nanometric resolution images from electron microscopy with the micrometric scale resolution of optical imaging. For this purpose, researchers are widely working on the development of new tools and methodologies for labelling proteins, mostly focused on correlative imaging. Clearly the development of new protein-labelling strategies will provide significant improvement of bio-imaging techniques[1]. Indeed, nowadays there are many tools and strategies available for labelling molecules for light or electron microscopy imaging. However, there is still a lack of tools for correlative

microscopy, in which a single label can be imaged by both techniques.

As general rule, the main purposes of protein-labelling are the followings: 1) monitoring of biological processes at high resolution; 2) reliable quantification of compounds; 3) specific detection of protein modifications and iso-forms in multiplexed samples; 4) enhancement of detection sensitivity; 5) simplification of detection workflows[2]. At the present the labelling strategies used are multiple and they must be carefully considered and tailored for each application. Currently bio-imaging technologies have several tools and methods already available in the market.

For instance, a useful strategy widely used is the incorporation of non-natural amino acids into proteins which are fluorescent or can be made fluorescent by the addition of fluorescent molecules via click chemistry. The insertion of non-canonical amino acids (ncAA) into proteins provides a minimally invasive site-specific labelling technique. Besides, tag-mediated labelling methods can be used. They consist in genetic tagging of target proteins with a small protein or peptide that can provide a specific signal for microscopy techniques. Tag-mediated labelling provides to researchers derivatizing proteins, which represent powerful tools for studying protein interaction in cells. Indeed, thanks to the tag-mediated labelling, proteins of interest (POIs) can be visualized by bio-imaging techniques, as the tag provides them fluorescent and/or electron signal, showing the versatility of tag-mediated approach[3,4].

1.1.1.2. Tag-mediated Labelling

Scientific community extensively agrees on the idea that an ideal tag should properly work both *in vitro*, in complex biological samples, on the cell surface, within the cell and cellular compartments, and *in vivo* in animal models. An ideal method for tag-based protein labelling should match the following criteria: (1) one-step introduction of the label; (2) specific labelling of target

proteins; (3) the formation of a stable protein-label compound; (4) small label size for not interfering with the protein function; (5) quantitative labelling reaction; (6) no side effects of the reagents used for labelling on protein structure and function. However, none of the existing labelling methods fulfils all these requirements, each presenting inherent strengths and weaknesses according to the sample or bio-imaging techniques planned to use.

In this context, the introduction of fluorescent proteins (FPs) in 1994[5,6] is breakthrough for fluorescent bio-imaging, because FPs are selective, genetic protein tags, which does not interfere with the protein function. Throughout the years, naturally occurring and engineered FPs have been optimized for spectral variation, increased brightness and other properties to provide a wealth of reagents for researchers[7,8]. However, FPs have limitations: none of the FPs can be compared to the best organic fluorophores, above all much less nanoparticles or other emerging chemical probes, in terms of photon output (often measured as brightness and photostability). Additionally, the new emerging chemical probes finally provide new labelling agents for electron microscopy and especially correlative light-electron microscopy (CLEM), which still lacks consistent labels. With chemical tags, rather than tagging the target protein with an FP using standard molecular cloning techniques, the protein is tagged with a polypeptide, which is subsequently modified with chemical probes (i.e. organic fluorophore, nanoparticles). Currently as schematically shown in Figure 1.1 the existing chemical tags can be mainly divided in three groups:

1. Self-labelling Tags

2. Self-labelling Proteins

3. Enzyme-mediated labelling of Tags.

Self-labelling tags and self-labelling proteins are tags that can make a specific, covalent bond with a probe molecule. In self-labelling tags a short peptide tag is used as label while self-labelling proteins are protein-based tag. Meanwhile enzyme-mediated labelling of tags approach enables to combines small tag size with specificity provided by enzymes. Thoroughly the protein of interest (POI) is fused to enzyme derivatives interacting with synthetic substrates. These substrates inhibit enzymes at their active site and most importantly they are modified by fluorophores or tags such as biotin or fluorescein. However, the main drawback for enzyme-mediated labelling of tags relates to the choice of the label that sometimes can interfere with substrate recognition process.[3, 4]

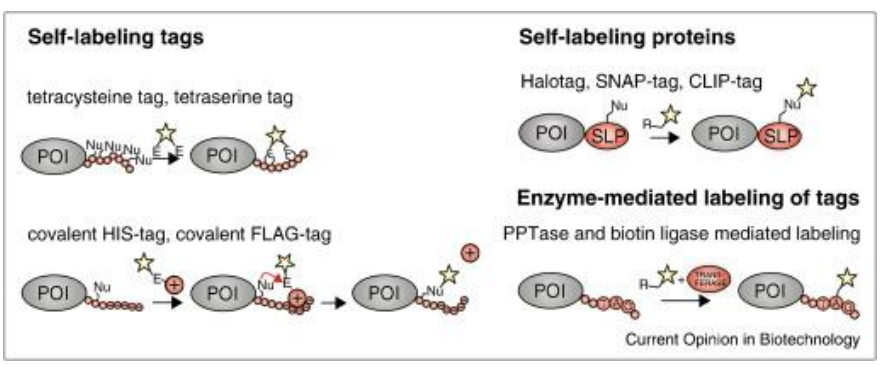

Figure 1.1. Families of chemical tags. POI: Protein of interest; SPL: Self-labelling protein; Nu: Nucleophilic or basic residue; E: Electrophilic moiety; R: moiety required for recognition by the enzyme used for labelling. [3]

1.1.2. Currently bio-imaging techniques for labelled proteins

1.1.2.1. Fluorescent imaging techniques

The advent of modern light microscopy techniques has revolutionized the way scientists image and study biological processes and it has brought researchers in a new era of direct visualisation and characterization of the biological process with a resolution that can reach the single molecule level. In this regard, it is extremely important to have a robust and consistent fluorescent

labelling of protein of interest (POI) for *in vivo* and *in vitro* applications. Hence, the toolbox for fluorescent imagining is continuously being developed, leading to a large variety of genetically encoded fluorescent tags. These tags have fluorescent colour feature spanning the visible spectrum, so that today enables researchers to multicolour imaging of nearly any set of POIs by genetically fusion tags. According to the origin of the fluorophore, technically fluorescent tags can be divided into three broad classes:

1. Intrinsically fluorescent proteins, which emit without addition of any fluorophore.

2. Extrinsically fluorescent proteins, which emit after later reacting with a fluorophore.

Obviously, each class has its own advantages and drawbacks. In many ways intrinsically fluorescent proteins are very easy to use, but their fluorescence properties can be changed only by protein engineering, such as in case of FPs[7–9]. On the other hand, in the case of tags that are fluorescent after binding a fluorophore it is possible to easily vary fluorescence properties as changing the fluorophore. However, in order to properly chose a fluorophore many aspects must be considered, such as its reactivity and interference in biological system monitored[10].

Intrinsically Fluorescent Protein

Green fluorescent protein (GFP) is the first and the most important example of an intrinsically fluorescent protein. GFP is a 238-amino acid protein that folds into an 11-stranded β-barrel. It was rapidly and widely adopted because it can be fused to a broad variety of proteins. Since the gene encoding green fluorescent protein (GFP) was first cloned from jellyfish in 1992[11], fluorescent proteins (FPs) have provided the initial and very powerful instrument for

imaging of proteins of interest (POIs). Over the years, a series of FPs have been developed in order to tune and improve their optical features (i.e., emission wavelength, brightness, photostability, photobleaching, quantum yield), leading for example to the enhanced GFP (eGFP). Besides, blue, cyan, yellow and red variants have been designed by mutation of the GFP chromophore and its surrounding sequence[12]. The engineering GFPs to achieve red fluorescent proteins has been a more challenging scientific problem; currently RFP (red fluorescent proteins) variants consist in engineered bright-red proteins resulting mostly in wild-type tetrameric proteins. The brightness and emission wavelengths are primary concerns, as well as photostability. The new engineered variants share the same β-barrel fold as GFP and the same autocatalytic mechanism of fluorophore formation, but their fluorophores differ in structure, which produces different spectral properties. Additionally, it is noteworthy to mention there are variants of FPs in different oligomeric states. Great efforts have been focused on developing monomeric variants by mutagenesis. It should be emphasized that it is extremely important to use monomeric protein tag to avoid mistargeting and aggregation of fused constructs, which make oligomeric fluorescent proteins generally unsuitable as fusion tags for tracking POIs. In order to engineer monomeric variants, one strategy adopted is to mutate residues at the dimer interface, which allows to reduce significantly the monomer association constant, *i.e.* K_d of dimer formation of EGFP (mEGFP) by 1000-fold respect to eGFP (dimerization Kd ≈ 110 μM). In detail, the hydrophobic patch of Ala^{206}, Leu^{221}, and Phe^{223}, responsible for dimerization of all GFPs and its mutants of any colour, is mutated to positively charged residues in order to prevent dimerization of the β-barrels.[13]However, brightness remains a problem in the far-red monomeric fluorescent proteins, because oligomerization helps stabilize the fluorophore in a conformation that favours bright fluorescence emission [9,14–16]. In summary, over the years, a series of FPs with various colours from blue (λ_{exc} = 400 nm / λ_{em} = 450 nm) to far red (λ_{exc} =600nm

excitation/ λ_{em} =630nm emission) have been developed derived from a range of sources. Besides, recently FPs with longer excitation and emission wavelengths including infrared FPs have been developed and applied to intravital imaging of deep tissue[17]. Protein engineering has provided a large range of fluorescent proteins from which to choose, with properties optimized for specific applications. In Table 1.1. broadly used fluorescent proteins are listed, but many more are available. Additionally, in Figure 1.2. emission spectra of commonly used fluorescent proteins is shown.

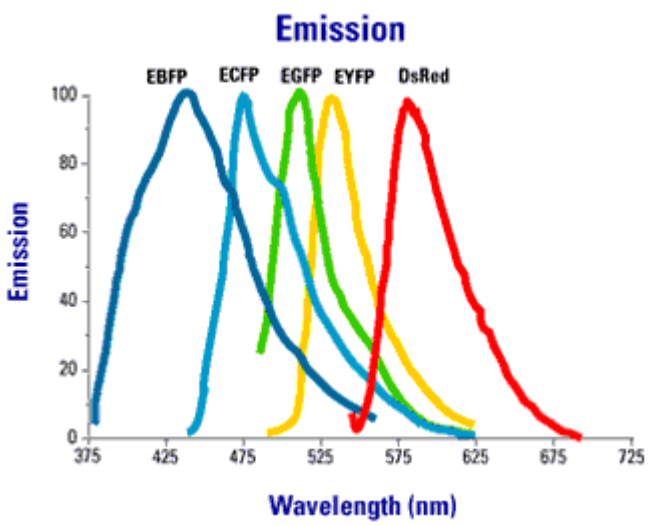

Figure 1.2. Emission spectra of commonly used fluorescent proteins (EBFP: enhanced blue fluorescent protein; ECFP: enhanced cyan fluorescent; EGFP: enhanced green fluorescent; EYFP: enhanced yellow fluorescent protein; DsRed: red fluorescent protein). Spectra reproduced from Clontech web site.

Nevertheless, FPs still have shortcomings, such as low brightness and photostability compared with modern fluorescent dyes, which hamper their use in single particle tracking or long-term live imaging[4]. Moreover, the limitation of the spectral range is still a problem for FPs. For example, in mammals for *in vivo* deep tissue visualization, optimal fluorescent proteins should have both excitation and emission maxima within the near-infrared window (650-900nm)

in order to avoid high absorbance in tissues. As mentioned above, several far-red-shifted fluorescent proteins have been developed for this purpose (such as mKate2, E2-Crimson22, mNeptune23, TagRFP657 and eqFP67), but their spectral properties are still suboptimal. Finally, it should be considered that the labelling using FPs is not applicable to all POI because of their big size (~25 kDa), which can sterically perturb the natural biological function of the protein, thus the labelling by FPs cannot applied to tiny POI.

Table 1.1. Select intrinsically fluorescent proteins.[10]

Name	λ_{ex} (nm)	λ_{em} (nm)	E (mM^{-1} cm^{-1})	QY	E × QY	Type	pK_a	Bleach time (s)
mTagBFP2	399	454	51	0.64	32.4	m	2.7	53
mTurquoise2	434	474	30	0.93	27.9	m	3.1	90
mCerulean3	433	475	40	0.80	32	m	4.7	
EGFP	488	507	56	0.60	33.6	m	6.0	174
mWasabi	493	509	70	0.80	56	m	6.0	93
Superfolder GFP	485	510	83	0.65	54.1	m		
mNeonGreen	506	517	116	0.80	92.8	m	5.7	158
mClover3	506	518	109	0.78	85	m	6.5	80
Venus	515	528	92	0.57	52.5	m	6.0	15
Citrine	516	529	77	0.76	58.5	m	5.7	49
mKOκ	551	563	105	0.61	64	m	4.2	
tdTomato	554	581	138	0.69	95.2	td	4.7	98
TagRFP-T	555	584	81	0.41	33.2	m	4.6	337
mRuby3	558	592	128	0.45	57.6	m	4.8	349
mScarlet	569	594	100	0.70	70.0	m	5.3	277
FusionRed	580	608	95	0.19	18.1	m	4.6	150
mCherry	587	610	72	0.22	15.8	m	4.5	96
mStable	597	633	45	0.17	7.6	m		1002
mKate2	588	633	63	0.40	25	m	5.4	84
mMaroon1	609	657	80	0.11	8.8	m	6.2	178
mCardinal	604	659	87	0.19	16.5	m	5.3	730
T-Sapphire	399	511	44	0.60	26.4	m	4.9	25
mCyRFP1	528	594	27	0.65	18	m	5.6	45
LSSmOrange	437	572	52	0.45	23.4	m	5.7	
mBeRFP	446	611	65	0.27	17.6	m	5.6	

Extrinsically Fluorescent Protein

More recently, in order to overcome FPs drawbacks, many efforts are turned to the development of fluorescent proteins which became fluorescent after binding some fluorescent cofactor. Specifically, fluorescent proteins engineered based on phytochromes and phycobiliproteins represent a good alternative for near-infrared fluorescence applications. For example, a near-infrared emitting FP (iRFP) with a maximal emission wavelength at 713 nm[17] has been obtained based on an engineered phytochrome. In general, phytochromes are proteins that naturally bind the biliverdin, an heme degradation product, or closely related molecules and fluoresce weakly in the infrared. Protein engineering of this family of proteins provided to the scientific community much brighter near-infrared tools and less perturbative in fusions. In addition, even if they use endogenous biliverdin as fluorophores, it needs to be considered that the fluorescence can be increased by supplemental biliverdin or cell-permeant analogues in cells[18] or by co-expressing heme oxygenase in order to increase the intracellular concentration of biliverdin[19]. Among extrinsically fluorescent proteins, other kind of green fluorescent proteins can be counted such as UnaG which uses bilirubin as the fluorophore[20], as well as green fluorescent proteins that use flavins[21]. Among them, it is noteworthy to mention miniSOG (for mini Singlet Oxygen Generator), a fluorescent flavoprotein engineered from Arabidopsis phototropin 2. MiniSOG contains 106 amino acids and it efficiently generates singlet oxygen when illuminated by blue light. The singlet oxygen generated localizes catalyze the polymerization of diaminobenzidine into an osmiophilic reaction product resolvable by EM. It is important to emphasize that even if less bright, miniSOG has some great advantages over EGFP: it is smaller than EGFP and more importantly, it is valuable for imaging anaerobic organisms because they do not require oxygen to become fluorescent[10] and above all, it

could be also used for electron microscopy.[22] Below in table 1.2, common and widely used extrinsically fluorescent proteins are listed.

Table 1.2. Selected extrinsically fluorescent proteins and their cofactors[10].

Name	λ_{ex} (nm)	λ_{em} (nm)	E (mM^{-1} cm^{-1})	QY	E × QY	Type	pK_a	Bleach time (s)	Fluorophore
miniSOG	450	495	12	0,51	6,4	m	3		flavin
UnaG	498	527	77	0,51	39,4	m	4		bilirubin
miRFP670	642	670	71	0,12	8,5	m	4,5	155	biliverdin
TDsmURFP	642	670	170	0,18	30,6	td		190	biliverdin
iRFP670	643	670	114	0,11	12,5	d	4		biliverdin
mIFP	683	704	82	0,08	6,6	m	3,5		biliverdin
iFP2.0	690	711	86	0,08	6,9	m			biliverdin
iRFP720	702	720	96	0,06	5,8	d	4,5		biliverdin

To provide researchers of more completed toolbox for fluorescent microscopy, proteins that bind chemical probes have also been developed. These are commercially available tools with both cell-permeant and cell-impermeant ligands, which give to researchers the possibility to discriminate between intracellular fusion proteins and extracellular fusion proteins. Among this family it is noteworthy to mention fluorogen-activating proteins, single-chain antibodies that bind a non-fluorescent molecule and stabilize it in a fluorescent state[23]. An analogous labelling scheme is to use the six–amino acid tetracysteine motif recognizes arsenic-containing dyes[24,25] named FlAsH/ReAsh. By themselves, the arsenic-containing dyes are not fluorescent, but they become fluorescent when bound to the tetracysteine tag. The great advantages of this labelling scheme are the small size of the tag, while the most important drawback is the nonspecific binding to cysteines which can lead to background fluorescence. This limitation can be overcome by washing with sulfhydryl-containing compounds. Below in Table 1.3 the most widely used tags are listed.

Table 1.3. Selected extrinsically fluorescent proteins with endogenous ligands[10]

Name	λ_{ex} (nm)	λ_{em} (nm)	E (mM^{-1} cm^{-1})	QY	E × QY	Type
FlAsH/peptide	508	528	70	0.85	59.5	m
ReAsH/peptide	593	608	69	0.48	33.1	m
TO1/scFv	509	530	60	0.47	28.2	m
MG/scFv	635	656	105	0.25	26.3	m

An alternative way to fluorescently label a protein of interest is the use of more recently self-labelling tag sequences. This kind of tags covalently react with a small-molecule substrate containing a fluorophore. In table 1.4. the major example is listed.

Table 1.4. Other genetically encoded tagging strategies. [10]

Tag	Description
SNAP tag	20 kDa; covalently labeled by reaction with benzylguanine derivatives
$SNAP_f$ tag	20 kDa; covalently labeled by reaction with benzylguanine derivatives; faster labeling than SNAP tag
CLIP tag	20 kDa; covalently labeled by reaction with benzylcytosine derivatives
$CLIP_f$ tag	20 kDa; covalently labeled by reaction with benzylcytosine derivatives; faster labeling than CLIP tag
Halo tag	33 kDa; covalently labeled by reaction with haloalkane derivatives
TMP tag	Engineered E. coli dihydrofolate reductase that binds trimethoprim-fluorophore conjugates
SunTag	73-kDa tag that recruits up to 24 GFPs
GFP1-10/GFP11 and sfCherry1-10/sfCherry11	19–amino acid peptide from GFP that recruits remaining 222–amino acid GFP sequence;sfCherry is red equivalent; small size enables multimerization or CRISPR knock-in
F30-Broccoli	Green fluorescent RNA aptamer; binds exogenous fluorophore
Mango	Orange or red fluorescent RNA aptamer, depending on exogenous fluorophore
DNB aptamer	Dinitrobenzyl-binding aptamer, enabling light-up labeling of RNA molecules
JX1	Benzylguanine-binding RNA aptamer, allowing use of SNAP-tag reagents for RNA labeling

Among them, SNAP-tag, CLIP-tag, HaloTag and eDHRF are particularly noteworthy. They are enzyme-mediated labelling of tags. SNAP-tag™ is the most versatile of these tags and it is a registered trademark of Covalys Biosciences AG (Covalys Biosciences AG, Witterswil, Switzerland). SNAP-tag is a ~20 kDa modified engineered version of human O^6-alkylguanine-DNA alkyltransferase enzyme (AGTm). The DNA repair protein O^6-alkylguanine-DNA alkyltransferase can react specifically and rapidly with benzylguanine (BG) derivatives of type 1 by transferring an alkyl group to reactive cysteine residues. Therefore, it makes possible that BG derivates bearing fluorescent probes can transfer the label with the benzyl group of BG to the cysteine. The method enables labelling of POIs in bacteria and yeast and even in AGT-deficient mammalian cell lines, thereby avoiding labelling of the endogenous AGT.[26] SNAP-tag is broadly used for cell *in vivo* imaging. Moreover, within self-labelling enzyme tags family, CLIP-tag is noteworthy to mention. Likewise, SNAP-tag, it given rise to human O^6-alkylguanine-DNA alkyltransferase (hAGT) and particularly it is generated from the mutagenesis of eight residues of that. Unlike SPA-tag, CLIP-tag reacts specifically and orthogonally with O^2-benzylcytosine derivates (BC). It is mostly used for study and visualisation of protein-protein interactions[27], additionally in combination with SNAP-tag for simultaneous labelling in live cell. Indeed SNAP/CLIP-tag multi-protein studies using fluorescent imaging is favoured considering that a large variety of fluorescent BC/BG derivates chemically relatively inert are accessible by simple synthetic route. Generally in just two steps from commercially available precursor, it is possible to generate (1) green fluorescent probes using fluorescein, diacetylfluorescein, and dipivaloyl Oregon green, (2) red fluorescent probes using Cy3 and tetramethylrhodamine, and (3) far-red fluorescent Cy5[28]. In addition to SNAP-tag and CLIP-tag, in order to label POIs under physiological conditions, scientist can dispose of Halo-Tag, another type of self-labelling enzyme tag. Halo-tag is a modified haloalkane dehalogenase which catalyses the hydrolytic conversion of

haloalkane (generally chlorine or bromine) to the corresponding alcohol and hydrogen halide by nucleophilic displacement mechanism. Regarding the Halo-tag ligands that react specifically and covalently with Halo-tag, they consist of haloalkane linker and mostly various probing molecules, among them exactly fluorescent dyes. As seen for SNAP-tag and CLIP-tag, cross-reactive interference is negligible with the endogenous mammalian biochemistry and bacterial dehalogenases are relatively small, allowing the specific fluorescent labelling of POI even in living cell systems[29]. Ultimately, it is worth highlighting as labelling tool also the 18kDa stable monomeric *Escherichia coli* protein dihydrofolate reductase (eDHFR). Thanks to the high-affinity interaction between eDHFR and trimethoprim (TMP) derivatives, POI can be fused and studied by fluorescent imaging both in vivo and in vitro systems. Indeed the cell-permeable TMP can be combined with fluorophores and remarkably it shows a higher selectivity for eDHFR than the mammalian DHFRs, hence this labelling system can be effectively used to label and analyse POIs in mammalian cells[30]. As shown above, self-labelling enzyme tags are a effective and versatile systems to overcome FP drawbacks but they may be limited because of reversibility of the chemical attachment, cellular toxicity, and labelling specificity in mammalian systems[4].

1.1.2.2. Electron Microscopy

As largely described above, fluorescence microscopy has revolutionized cell biology, thanks to its extensive toolbox of fluorescent probes. Indeed, fluorescence microscopy is rapid and convenient, as well as compatible with live-cell imaging, which allows dynamic and minimally invasive imaging experiments. However, on the other side, its spatial resolution of ~200–300 nm[31] is an important limiting drawback because most biomolecules are much smaller than these dimensions. In recent years, super-resolution fluorescence microscopy techniques have greatly improved upon the resolution of

conventional light microscopy, but these techniques require specialized fluorophores and equipment, and they do not yet routinely provide spatial resolution in the sub-10-nm regime[32]. Compared to fluorescence microscopy, electron microscopy achieves far superior spatial resolution (~1 nm in biological samples[31]), turning out to be an essential cell biology tool. Electron microscopy has the capability to reveal the entire cellular ultrastructure, including membranes, subcellular organelles, and large proteinaceous complexes. In particular, nowadays electron tomography methods, in combination with technical advances in sample preparation and preservation, already permits 3D reconstruction of intracellular structures to resolutions of 3–8 nm[33,34]. Up to now, the most utilised method for live-cell imaging to track intracellular events is the immunogold labelling. One of most powerful and used technique to understand biological processes is immunocytochemistry. The methodological approaches are simple, and it is based on the fact that antibodies can be labelled directly or indirectly. According to the different selective binding properties of a given antibody, a target protein or epitope can be located with high precision. Immunocytochemistry was born in the era of light microscopy, when Coons *et al.*[35] in 1941 showed that antigens could be identified in tissue sections by antibodies labelled with a fluorescent dye. However only with the advent of electron microscopy and the introduction of particles as markers, such as colloidal gold and alternative electron-dense markers, including ferritin, iron-dextran and uranium, immunocytochemistry became the powerful technique so-called immunoelectron microscopy (IEM). Indeed, electron dense particles of different size can be discriminated with the high resolution offered by the electron microscope. Also, several methods have been developed to intensify the immunogold signal (i.e. silver enhancement by silver salts)[36] opening to the possibility of using also ultra-small gold particles that otherwise will not be visible by electron microscope. However, IEM still presents significant limitations including the limited quantification, and the lack of specificity and precision on the localization of

target protein. Additionally, IEM cannot be applied to all types of samples and systems, because it depends on the availability of antibodies which can interact with the protein of interest (POI) [37,38,39]. In order to overcome these shortcomings, there is a need of a reliable toolbox of genetically encodable protein tags, an approach that could open new possibilities for transmission electron microscopy (TEM), potentially as far-reaching as the applications of fluorescent protein tagging in light microscopy [39]. The ideal tag should meet the following criteria: (1) it should be electron dense to be visualized by electron microscopy; (2) the electron density should be tightly focused and provide good signal-to-noise ratio so the tag is easily distinguishable from background by electron microscopy (EM); (3) genetically encodable so that the biological system can be processed in its native state; and (4) non-toxic and non-disruptive so as not to interfere with biological system[40]. Nowadays, there is an increasing interest on the development of clonable contrast agents for transmission electron microscopy (TEM) which could revolutionize the biology, such as green fluorescent protein did in fluorescence microscopy imaging[8]. For example, proteins with fluorescent tags have been used to generate through photo-conversion electron-dense deposits that could be visualized by TEM. [22,24,41] However, the resulting signals are diffuse, they lack the resolution of particulate probes[42], and sometimes they are too weak to visualize low-abundance proteins[43]. As reported in the scientific literature, several electron-dense clonable tags are inspired from natural proteins, such as bacterial ferritin, or rather its subunit Ftna1. In detail, Ftna1 forms 12 nm cages, composed of 24 monomers and containing up to 4,200 atoms of iron, that are visualized by TEM as 7 nm electron-dense particles. However, the size of ferritin is 18 times that of GFP, which it potentially might affect trafficking and biological function of the target proteins. Besides, because of Ftna1-mediated oligomerization, target protein mislocation and clustering might occur[43]. In addition, the use of metal-binding protein metallothionein (MT) as a tag for EM has been explored[44,45]. Metallothionein proteins are natural

metal-ion chelators and they control the storage and interchange of biologically essential metals such as zinc and copper[46] in eukaryotic cells. Isoform 1 of murine MT, a small 61-aa protein, comprises of 20 cysteine residues that efficiently bind metal atoms. It is been shown that when a single MT tag is fused with a protein of interest and treated *in vitro* with gold it can build an electron-dense gold-thiolate cluster of 1 nm diameter[45]. Besides, it has been shown that intracellular proteins in bacteria can be detected using this methodology, called metal-tagging transmission EM (METTEM)[44]. However, it remained uncertain whether the method could be adapted to work in mammalian cells, which will represent a significantly improvement for mainstream cell biology. Indeed, contrary to bacteria, mammalian cells hardly grow in the presence of heavy metals and in particular they cannot be long exposed to high AuCl concentrations[44]. Also, it was not known whether heavy metals will be transported into these cells with the required efficiency to allow for gold cluster formation, and to what extent resident endogenous MTs might produce background[44]. Therefore, the use of this type of tag will be highly limited in mammalian cells.

1.1.2.3. Correlative Light- Electron Microscopy (CLEM)

Correlative light-electron microscopy (CLEM) represents a powerful technique to explore sub-cellular organization in 3D and locate proteins at high resolution, since it combines the specificity and dynamics of fluorescence light microscopy with the high resolution and cellular context of electron microscopy. More recently, attention has turned to genetically encoded tags for CLEM. The ideal tag for CLEM should combine the criteria already explained for light and electron microscopy, such as being non-toxic and non-disruptive so as not to interfere with normal cellular function, as well as being genetically encodable so that the cell can be processed in its native state without the need for permeabilization. In addition, the tag should be at the

same time fluorescent and electron dense so that it can be visualized by light and electron microscopy. Particular attention is paid to the electron density in order to provide good signal-to-noise ratio of the tag, which can be distinguishable from background by EM. Currently, it is hard to meet all these criteria among existing tags. A widely used approach until now consists of using diaminobenzidine (DAB) for the formation of an electron-dense precipitate either by enzymatic-based polymerization using peroxidase[47,48] or singlet oxygen-based polymerization during photo-oxidation[22,24,49]. The most important shortcoming of this approach is the low labelling resolution by EM due to the diffuse nature of the precipitate. Therefore, this approach is only successful when the proteins to be visualized are located inside organelles or are discretely localized at high densities. Nowadays, nanomaterials are at the leading edge of the rapidly developing field of nanotechnology. Metal nanoclusters are largely used in a variety of biotechnology and material science applications such as optical-based sensors, as chemical catalysts, as biomedical diagnostic and therapeutic tools and for the fabrication of nano-electronics. Metal nanoclusters have unique chemical and physical properties, such as their spectral and electronic structure, which can be tuned varying their composition and size. Furthermore, metal nanoclusters have attracted increasing interest for bioanalytical labelling applications in recent years, since can be used to label biological structures for optical and electron microscopy and they could represent ideal tags for CLEM, sometimes combined with fluorescent proteins. Indeed, metal clusters could improve resolution and be readily distinguishable from background since they contain elements that are able to scatter electrons. Two examples of tags concatenated with fluorescent proteins have been published using metallothionein[40,50] and bacterioferritin[43]. Unfortunately, they present significant drawbacks in their current form. Concerning metallothioneins, it must be said that their use is limited to high abundance proteins and only minimal ultrastructural information is currently possible[50]. Meanwhile, tagging with bacterioferritin is technically limited to

bacteria and it can lead to aggregation and mislocalization of the protein target, because of the large structure of bacterioferritin (>12 nm diameter)[43]. Another technology for labelling proteins in CLEM uses versatile inter-acting peptide (VIP) tags and it is called VIPER technology. VIPER technology consists of a peptide tag that has high specificity in a miniaturized size and it useful to label cellular proteins with several distinct chemical reporters. In detail, VIPER uses a heterodimeric coiled-coil formed between two peptides, a genetically encoded peptide tag (CoilE) and a reporter-conjugated peptide (CoilR). The list of chemical reporters includes BODIPY, Sulfo-Cyanine5 (Cy5), or biotin for detection by streptavidin-Qdot are included[51]. Besides, in literature the use of ferritin for CLEM is described. Indeed, the electron density of ferritin has been exploited for immunoEM for decades[52] and it could be used as an ideal CLEM tag following some modifications. The human ferritin has been adapted for correlative light-electron microscopy is named FerriTag, which is a fluorescent recombinant electrondense ferritin particle that can be attached to a protein-of-interest using rapamycin-induced heterodimerization. Specifically, the human ferritin is a complex of 24 polypeptide subunits of light (FTL) and/or heavy (FTH) chains. Under iron-rich conditions ferritin can store iron and easily visualized by EM and it has a good signal-to-noise ratio and a labelling resolution of approximately 10 nm. It has been demonstrated the utility of FerriTag for correlative light-electron microscopy by labelling proteins associated with various intracellular structures including mitochondria, plasma membrane, and clathrin-coated pits and vesicles[53]. FerriTag represent a great example of the utility and the versatility of CLEM tags to answer biological questions. Unfortunately, except for the example described above, CLEM lacks many tuneable genetically encoded inducible tags. Therefore, the use of CLEM is still limited to few applications, even if it is a highly promising technology for biology.

1.2. Bibliography

1. Yao, J., Yang, M. & Duan, Y. Chemistry, biology, and medicine of fluorescent nanomaterials and related systems: New insights into biosensing, bioimaging, genomics, diagnostics, and therapy. *Chem. Rev.* **114**, 6130–6178 (2014).

2. Obermaier, C., Griebel, A. & Westermeier, R. Chapter 13 Principles of Protein Labeling Techniques. in vol. 1295 153–165 (2015).

3. Hinner, M. J. & Johnsson, K. How to obtain labeled proteins and what to do with them. *Curr. Opin. Biotechnol.* **21**, 766–776 (2010).

4. Liu, J., Liu, J., Cui, Z. & Cui, Z. Fluorescent Labeling of Proteins of Interest in Live Cells: Beyond Fluorescent Proteins. *Bioconjug. Chem.* **31**, 1587–1595 (2020).

5. Heim, R., Prashert, D. C., Tsien, R. Y. & Kandel, E. R. Wavelength mutations and posttranslational autoxidation of green fluorescent protein (Aequorea victoria/blue fluorescent protein/Escherichia coli/imidazolidinone). *Biochemistry* **91**, 12501–12504 (1994).

6. Chalfie, M. *et al.* Expression of the gene and fluorescence characteristics of the recombinant protein Green fluorescent protein as a marker for gene expression. *Fed. Eur. Biochem. Lett.* **341**, 277–280 (1994).

7. Giepmans, B. N. G., Adams, S. R., Ellisman, M. H. & Tsien, R. Y. The fluorescent toolbox for assessing protein location and function. *Science (80-.).* **312**, 217–224 (2006).

8. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing

fluorescent proteins. *Nat. Methods* **2**, 905–909 (2005).

9. Shaner, N. C. *et al.* Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **5**, 545–551 (2008).

10. Thorn, K. Genetically encoded fluorescent tags. *Mol. Biol. Cell* **28**, 848–857 (2017).

11. Prasher, D. C., Eckenrode, V. K., Ward, W. W., Prendergast, F. G. & Cormier, M. J. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* **111**, 229–233 (1992).

12. Goedhart, J. *et al.* Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.* **3**, (2012).

13. Zacharias, D. A., Violin, J. D., Newton, A. C. & Tsien, R. Y. Partitioning of lipid-modified monomeric GFPs into membrane microdomains of live cells. *Science (80-.).* **296**, 913–916 (2002).

14. Merzlyak, E. M. *et al.* Bright monomeric red fluorescent protein with an extended fluorescence lifetime. *Nat. Methods* **4**, 555–557 (2007).

15. ERMAKOVA, G. *et al.* Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.* **418**, 567 (2009).

16. Alford, S. C., Abdelfattah, A. S., Ding, Y. & Campbell, R. E. A fluorogenic red fluorescent protein heterodimer. *Chem. Biol.* **19**, 353–360 (2012).

17. Filonov, G. S. *et al.* Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat. Biotechnol.* **29**, 757–761 (2011).

18. Rodriguez, E. A. *et al.* A far-red fluorescent protein evolved from a

cyanobacterial phycobiliprotein. *Nat. Methods* **13**, 763–769 (2016).

19. Yu, D. *et al.* An improved monomeric infrared fluorescent protein for neuronal and tumour brain imaging. *Nat. Commun.* **5**, 1–7 (2014).

20. Kumagai, A. *et al.* A bilirubin-inducible fluorescent protein from eel muscle. *Cell* **153**, 1602–1611 (2013).

21. Buckley, A. M., Petersen, J., Roe, A. J., Douce, G. R. & Christie, J. M. LOV-based reporters for fluorescence imaging. *Curr. Opin. Chem. Biol.* **27**, 39–45 (2015).

22. Shu, X. *et al.* A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms. *PLoS Biol.* **9**, (2011).

23. Bruchez, M. P. Dark dyes-bright complexes: Fluorogenic protein labeling. *Curr. Opin. Chem. Biol.* **27**, 18–23 (2015).

24. Gaietta, G. *et al.* Multicolor and electron microscopic imaging of connexin trafficking. *Science (80-.).* **296**, 503–507 (2002).

25. Griffin, B. A. *et al.* Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/2896025 REFERENCES Linked references are available on JSTOR for this. *Science (80-.).* **281**, 269–272 (1998).

26. Juillerat, A. *et al.* Directed Evolution of O6-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules In Vivo. *Chem. Biol.* **10**, 313–317 (2003).

27. Mie, M., Naoki, T., Uchida, K. & Kobatake, E. Development of a split

SNAP-tag protein complementation assay for visualization of protein-protein interactions in living cells. *Analyst* **137**, 4760–4765 (2012).

28. Gautier, A. *et al.* An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **15**, 128–136 (2008).

29. Los, G. V *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. *Acs Chem. Biol.* **3**, 373–382 (2008).

30. Miller, L. W., Cai, Y., Sheetz, M. P. & Cornish, V. W. In vivo protein labeling with trimethoprim conjugates: A flexible chemical tag. *Nat. Methods* **2**, 255–257 (2005).

31. Fernández-Suárez, M. & Ting, A. Y. Fluorescent probes for super-resolution imaging in living cells. *Nat. Rev. Mol. Cell Biol.* **9**, 929–943 (2008).

32. Xu, K., Shim, S.-H. & Zhuang, X. Super-Resolution Imaging Through Stochastic Switching and Localization of Single Molecules: An Overview. 27–64 (2013) doi:10.1007/4243_2013_61.

33. Volkmann, N. *Methods for Segmentation and Interpretation of Electron Tomographic Reconstructions*. *Methods in Enzymology* vol. 483 (Elsevier Inc., 2010).

34. Mahamid, J. *et al.* Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science (80-.).* **351**, 969–972 (2016).

35. Coons, A. H., Creech, H. J. & Jones, R. N. Immunological Properties of an Antibody Containing a Fluorescent Group. *Proc. Soc. Exp. Biol. Med.* **47**, 200–202 (1941).

36. Piludu, M., Medda, L., Cugia, F., Monduzzi, M. & Salis, A. Silver Enhancement for Transmission Electron Microscopy Imaging of Antibody Fragment-Gold Nanoparticles Conjugates Immobilized on Ordered Mesoporous Silica. *Langmuir* **31**, 9458–9463 (2015).

37. Waggoner, A. Fluorescent labels for proteomics and genomics. *Curr. Opin. Chem. Biol.* **10**, 62–66 (2006).

38. Hess, M. W. *et al.* Combining high-pressure freezing with pre-embedding immunogold electron microscopy and tomography. *Traffic* **19**, 639–649 (2018).

39. Amiry-Moghaddam, M. & Ottersen, O. P. Immunogold cytochemistry in neuroscience. *Nat. Neurosci.* **16**, 798–804 (2013).

40. Risco, C. *et al.* Specific, sensitive, high-resolution detection of protein molecules in eukaryotic cells using metal-tagging transmission electron microscopy. *Structure* **20**, 759–766 (2012).

41. Grabenbauer, M. *et al.* Correlative microscopy and electron tomography of GFP through photooxidation. *Nat. Methods* **2**, 857–862 (2005).

42. Kireev, I. *et al.* In vivo immunogold labeling confirms large-scale chromatin folding motifs. *Nat. Methods* **5**, 311–313 (2008).

43. Wang, Q., Mercogliano, C. P. & Löwe, J. A ferritin-based label for cellular electron cryotomography. *Structure* **19**, 147–154 (2011).

44. Diestra, E., Fontana, J., Guichard, P., Marco, S. & Risco, C. Visualization of proteins in intact cells with a clonable tag for electron microscopy. *J. Struct. Biol.* **165**, 157–168 (2009).

45. Mercogliano, C. P. & DeRosier, D. J. Gold nanocluster formation using metallothionein: Mass spectrometry and electron microscopy. *J. Mol. Biol.* **355**, 211–223 (2006).

46. Chan, J., Huang, Z., Merrifield, M. E., Salgado, M. T. & Stillman, M. J. Studies of metal binding reactions in metallothioneins by spectroscopic, molecular biology, and molecular modeling techniques. *Coord. Chem. Rev.* **233**–**234**, 319–339 (2002).

47. Martell, J. D. *et al.* Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy. *Nat. Biotechnol.* **30**, 1143–1148 (2012).

48. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2014).

49. Perkovic, M. *et al.* Correlative Light- and Electron Microscopy with chemical tags. *J. Struct. Biol.* **186**, 205–213 (2014).

50. Morphew, M. K. *et al.* Metallothionein as a clonable tag for protein localization by electron microscopy of cells. *J. Microsc.* **260**, 20–29 (2015).

51. Doh, J. K. *et al.* VIPER is a genetically encoded peptide tag for fluorescence and electron microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12961–12966 (2018).

52. Ram, J. S. R. I., Tawde, S. S., Pierce, B., Midgley, A. R. & Depart-, J. R. P. Brief Notes Preparation of Antibodywerritin Conjugates for Immuno-Electron Microscopy. 1–3 (1963).

53. Clarke, N. I. & Royle, S. J. FerriTag is a new genetically-encoded

inducible tag for correlative light-electron microscopy. *Nat. Commun.* **9**, 1–10 (2018).

2. AIM OF WORK

2. Aim of work

The aim of this work is the design and development of novel protein-mediated tags in order to study biologically relevant systems. Specifically, the project focuses on the design of protein scaffolds that stabilise luminescent metal nanoclusters and quantum dots as novel dual tags for both fluorescence microscopy and/or electron microscopy. This approach takes advantage of both the luminescent properties and the electron density of the protein-stabilised nanomaterials for their application as dual tag. Indeed, nowadays correlative imaging techniques, which correlate high resolution protein localization with structural information at nanometric scale, still lack of reliable tools for molecular tagging. Finally, the novel dual tags developed in this work are used to reproduce the retrograde protein recycling system *in vitro* as a biologically relevant proof of concept case study.

For this purpose, the following specific objectives are defined:

1. Design and development of fusion protein tag systems for synthesis of the protein-nanomaterial hybrids for fluorescence and/or electron microscopy imaging.

2. Development of a model system to study the retrograde trafficking pathways *in vitro*.

3. Studies of the retrograde trafficking pathways using the engineered protein-nanomaterial novel dual tags.

3. DESIGN AND SYNTHESIS OF PROTEIN-BASED LABELLING NANOTOOLS FOR FLUORESCENCE, CRYO-TEM AND CORRELATIVE STUDIES

3. Design and synthesis of protein-based labelling nanotools for fluorescence, cryo-TEM and correlative studies

3.1. Introduction

3.1.1. Engineered (repeat) proteins: building blocks for nanotools

Proteins are the building block of living cells, and they are very important for life since they perform numerous essential functions. Among natural proteins, it is noteworthy to mention repeat proteins. Basically, repeat proteins are those proteins which a large fraction of their sequence is composed of repetitions of 3 or more amino acids. They are evolutionary beneficial, since they can expand the repertoire of cellular functions at low energetic cost and, therefore, they play crucial roles in molecular recognition, signalling, protein transport or regulation of gene expression. In nature, there are several families of repeat proteins, which differ in structure, function, and phylogenetic distribution. In general, they have been described and usually classified attending to either the length of their sequence motif or their tertiary structure. In detail, this kind of proteins is composed of multiple tandem copies of a modular structure, that can encode different secondary structure in their constitutional elements, being those elements interconnected to each other forming the base of the structure of the protein. In repeat proteins, the interactions between adjacent units define the shape and curvature of the overall structure[1–3] (Figure 3.1). For example, tetratricopeptide repeat protein (TPR) consisting of 34 amino acids and folding in helix-turn-helix motif, form a right-handed superhelix structure. Whereas Ankyrin (ANK) repeats which contain 33 residues forming a helix-loop-helix motif, they give rise to a lefthanded twist of the stacking. On the contrary, in the case of Leucin-Rich Repeat (LRR) repeats varying between 20 to 30 amino acids and folding in a beta-turn-helix motif, the combination of this structural

motif results in an arch shape with the β-strand and α-helix oriented in antiparallel manner. [4]

Repeat proteins are very attractive and promising for nanotechnological applications. Indeed, the main limitation in the rational protein design is the limited understanding on how protein sequence-structure-function relate. The modular structure of repeat proteins simplifies the design problem since their structure is defined by simple units and interactions between only neighbouring units, which make them local and predictable. Thus, rational computational or experimental design approaches can be more easily used to engineer repeat proteins with different properties in order to expand the sequence and structure space observed in Nature[1]. Specifically, each repeat unit can be used as a building block with individually engineered properties, including stability[5], function[6], and interactions between modules. Those repeat proteins can be used as scaffolds to generate self-assembled functional structures[7–9] and they are ideally suited for protein design and nano-bioengineering through modular approaches [3,10,11].

3.1.2. Consensus Tetratricopeptide Repeat Protein (CTPR) as nanotechnological tool

Repeat proteins are extensively used in protein engineering to build useful protein toolboxes for biotechnological applications, in many cases based on consensus sequence design. Consensus sequence design is a proven and highly effective sequence-based method. It based on the assumption that the respective consensus amino acid at a given position contributes more than average to the stability of the protein than non-conserved amino acids. Consensus amino acid refers to the most common residues at a given position and it is generally determined from multiple sequence alignment of a group of sequences from a given family[12]. Moreover, thanks to consensus sequence design it is possible generating protein with a specific fold and specific

features according to the biological application. The best-studied consensus-designed repeat proteins are the consensus ankyrin repeats (DARPins[13] or CARPs[14]) and consensus tetratricopeptide repeats (CTPRs).

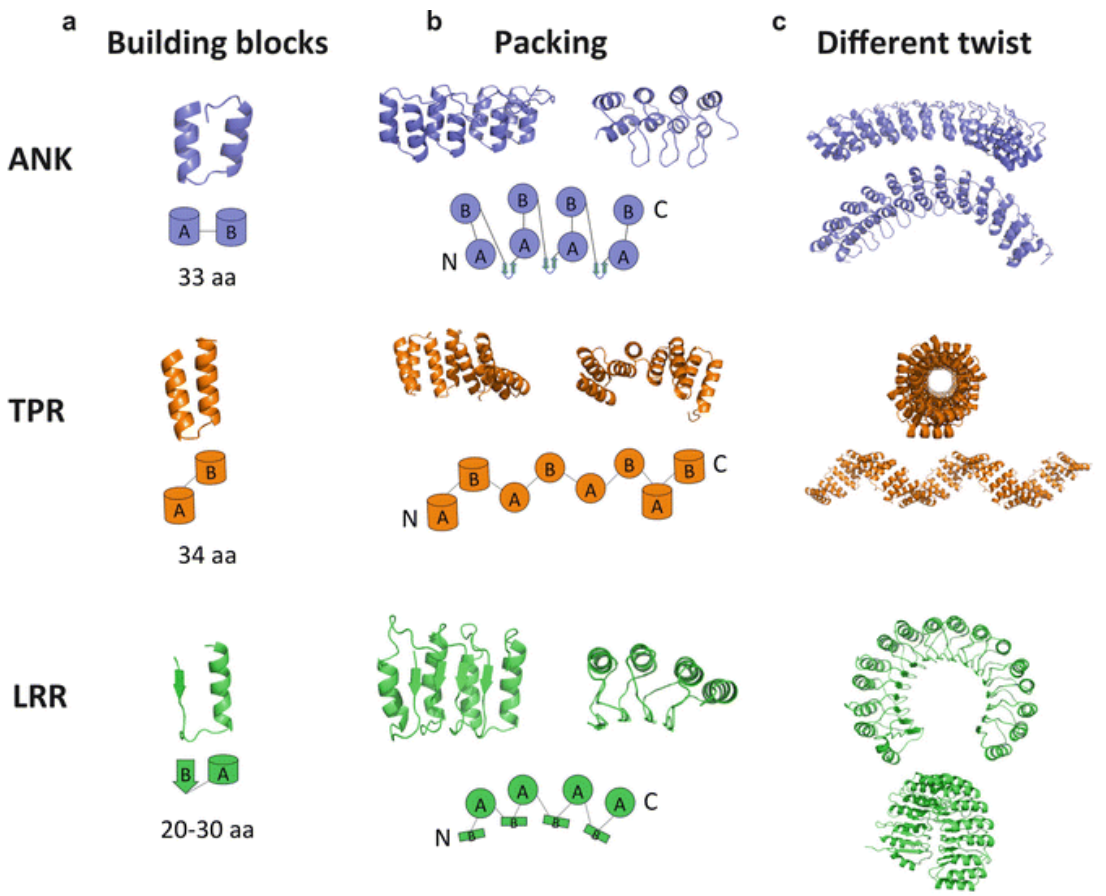

Figure 3.1. Representation of repeat protein family: ANKyrin repeat (ANK) in blue, TetratricoPeptide Repeat (TPR) in orange and Leucin Rich Repeat (LRR) in green. (a) Structure of an individual repeated unit is shown together with a schematic representation of each building block (b) The crystal structures of repeat proteins composed by 4 repeats of each building block (front view on the left side and top view on the right side). The structures illustrate the different packing arrangements between the building blocks as displayed in the schematic representations of the packing from N-terminal to C-terminal of the proteins below the crystal structures. (c) Crystal structures of long repeat arrays; 12 ANK repeats (PDB ID:2XEE); 20 TPR repeats (PDB ID:2AVP); and a 16 LRR repeats (PDB ID: 1A4Y). Figure reproduced from [10]

CTPRs emerges from the statistical analysis of TPR sequences, whose main biological role is associated with molecular recognition and mediate protein-protein interaction[7,15]. CTPR repeat is composed by only eight highly conserved small and large hydrophobic amino acids. These conserved residues are involved in intra- and inter-repeat packing interactions and thus maintain invariant the structure of the protein[15–17]. In detail, CTPR consist of 34 amino acid sequence that folds in a helix-turn-helix motif (Figure 3.2.)[18].

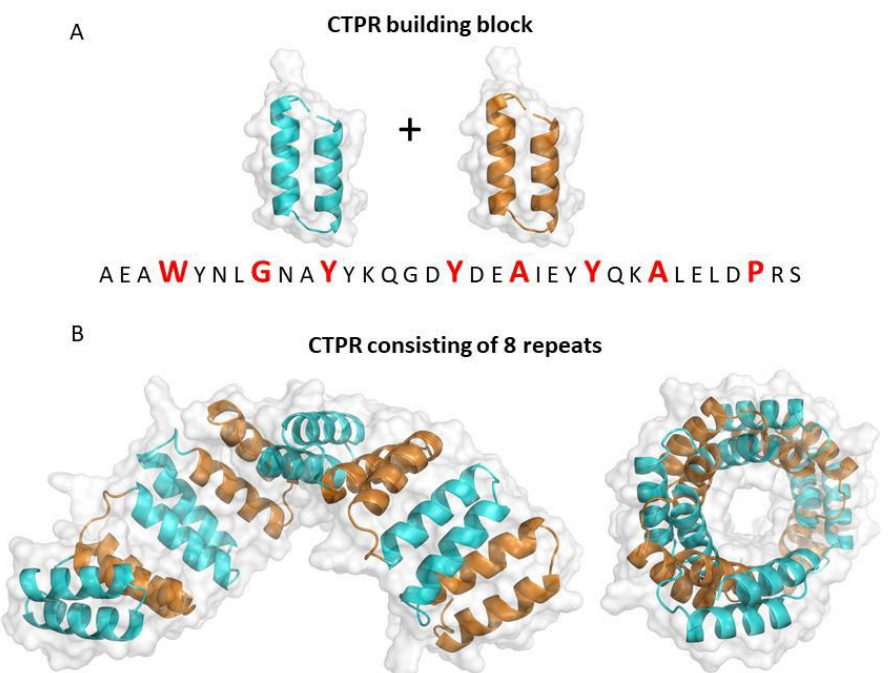

Figure 3.2. (a) CTPR unit structure of the CTPR building blocks. Below: the consensus CTPR sequence with the conserved amino acids highlighted in red (b) The crystal structure of a repeat protein composed of 8 CTPR repeats (PDB ID: 2AVP, 2HYZ). Side and axial views of the CTPR8 are shown. Figure reproduced from [19]

Thus, the CTPR repeat represents the ideal building block to form an idealized structural unit that can be combined in tandem to give rise right-handed super

helical arrays CTPRn with n number of repeats. A full turn of the superhelix (Figure 3.2.) is obtained by an array of eight repeats [16,20].

CTPRs admits sequence variations at many not conserved positions, while still maintaining the TPR-fold[21]. Even though, variations at some positions should follow some constrains and preferences in terms of chemical nature and size, thus some variations may modify the protein packing[21]. Thanks to this strategy, CTPR variants with different ligand-binding and/or novel reactive specificities have already been designed[6,21–23]. Noteworthy, CTPR proteins show higher stability compare to natural TPR domains[18]; the stability of the domain lies in intra- and inter-repeat interactions which can be modulated in a predictable manner[5]. In addition, considering the modularity of the system these new reactivates could be tuned, since different number of varations can be included along the CTPR array. The CTPR proteins can be produced and purified at large scale following simple molecular biology methodologies[7]. For all the above reasons, these features have opened the opportunity to tune the properties of the CTPR building block at will in order to use it as a nanotechnology tool in many applications including sensing[24], biomedicine[25] or bioelectronics[10,25], through organizing different functional elements into the CTPR protein template.

3.1.3. CTPR as designed protein scaffolds to stabilise metal nanostructures

Metal nanostructures are one of the most widely developed and studied systems in nanotechnology over the past decades, because of their optical, magnetic, and catalytic properties. Among them, metal nanoclusters (NC) and quantum dots (QDs) are noteworthy to mention for our purposes.

In particular, metal NCs smaller than 2 nm display different fluorescence emissions depending on their size, and also the molecules used for their stabilization[26]. CTPR proteins have been used as template for metal

nanoclusters[7,11,27]. Metal nanoclusters exhibit very interesting optical, electronical, and chemical properties, including strong photoluminescence, excellent photostability and good biocompatibility, while being sub-nanometre in size. Such properties make metal NCs ideal nanomaterials for applications in biological analysis and imaging, environmental monitoring, industrial catalysis and photoelectronic devices. In particular, the design versatility of CTPR proteins allows to generate protein binding modules that could be employed as novel fusion tags used in fluorescence and electron microscopy tracking strategies. Scientists have made major efforts to understand the role of protein characteristics, including protein size and amino acid content, in the synthesis and resulting properties of metal NCs. Besides, researchers have been focused on the metal coordination environments that promote biomineralization, the NCs structure and formation mechanisms as well as interactions of the metal core with the protein[28–30]. In detail, thiol, amine, and carboxylic acid side-chain functionalities promote metal binding[31,32]. To date, there are only few reported examples on the design artificial and/or natural proteins with an incorporated predefined metal binding site for the synthesis of metal NCs, such as the incorporation of two additional cysteines (Cys) residues in the ferritin cage to enhance the uptake of Au ions[30]. Regarding CTPR proteins, it has been shown their capability to act as template for green synthesis and stabilization of NCs composed of Au, Ag and Cu by the incorporation of a designed cluster coordination site based on cysteines (Cys) [33] or histidines (His)[11].

Also, it is noteworthy to mention the use CTPR proteins capability to stabilize quantum dots. Indeed, QDs have received great attention because of their high emission quantum yields, excellent resistance to photo-bleaching, photo-stability, and large emission Stokes shifts, compared to traditional organic fluorescent dyes. The sustainable synthesis and stabilization of biocompatible CdS-QDs within proteins is extremely advantageous, because it makes possible to skip arising from the conventional chemical synthesis, such as

toxic reagents, organic solvents, extreme reaction conditions. Additionally, after the chemical synthesis, QDs are still not water soluble, and they require post-synthetic processing. In this context, it has been shown CTPRs can stabilized CdS following a green aqueous route at 37 ºC thanks to an engineered metal coordination site based on a His-clamp. It is important to note that apart from histidine, different amino acids, including cysteine, glutamic acid, and aspartic acid bind Cd^{2+} ions and nucleate CdS QDs. The work focused on CdS-QDs, which ~~are~~ is one of the most studied luminescent semiconductors due to their excellent optical and electrochemical properties.[34].

3.1.4. Metal nanostructures stabilized by CTPR and their biological application in imaging

Metal NCs stabilized by peptides and proteins represent ideal tools for biological applications because they can be produced under mild conditions, and the obtained structures are stable under a wide range of pH values and ionic forces. Besides, good storage and work stabilities, good photostability and biocompatibility are required for bio-imaging and bio-labelling applications. CTPR proteins are interesting templates for the synthesis and stabilization of metal NCs. Thanks to their modular structure[2,35,36], it is possible to tune the size and the optical features of the NCs and thus of the newly designed tools. It has been reported that the metal NCs stabilized by CTPRs are stable over a month under storage conditions (PBS at 4°C) and over seven days under physiological conditions (PBS and human plasma at 37°C). Furthermore, the metal NCs stabilized by CTPRs are stable under a temperature range from 20 °C to 70°C, a broad pH range of pH 5–12, an ionic strength range of 0.15–1.0 M of NaCl, to biothiol concentrations up to 1.0 mM of Cys, and in the presence of different metal ions (50–150 μm)[11]. Additionally, the PL of the metal NCs stabilized by CTPRs remained nearly

constant (85–90%) after 10 min under continuous irradiation. Interestingly, metal NCs stabilized by CTPRs present better resistance to photobleaching in comparison to an organic fluorophore (DAPI) and a fluorescent protein (GFP), whose PL decreased to a greater extent (60% and 35%, respectively). Moreover, metal NCs stabilized by CTPRs are successfully cell internalized and they show notable fluorescence properties. These results clearly demonstrate the efficacy of metal NCs stabilized by CTPRs as transfection and labelling agent for live cells. [11,23]. It is noteworthy to mention that the plasma membrane is generally impervious to proteins[37], therefore the internalization of CTPR-metal NCs should be promoted by the presence of the metal NCs coordinated to the CTPR structure. Considering the fluorescein-labelled protein alone cannot be internalized by cells, CTPR-metal NCs are an interesting tool for labelling live cells. Moreover, it should not be overlooked that this kind of tools are very interesting also because metal nanostructures could be easily visualized by electronic microscopy, allow their application in structural biology. Hence metal nanostructures stabilized by CTPRs are ideal fusion protein tag systems for fluorescence, cryo-TEM, and correlative studies.

3.2. Experimental Section

3.2.1. Chemicals

All the commonly used chemicals were purchased from Sigma Aldrich, VWR Chemicals and Fisher Scientific and used without further purification. Milli-Q water was used in all experiments. Milli-Q water used for cloning was previously sterilized at 134 ºC for 20 minutes to avoid presence of nucleases. All DNA was dissolved or eluted in this water.

3.2.2. Buffers

All buffers were prepared as aqueous solutions using milli-Q and when it is necessary, they are filtrated using a nitrocellulose filter with 0.22 μm pore size (Fisher Scientific).

3.2.3. Bacterial strains and culture

Heterologous expression of proteins was performed in *Escherichia coli*. All bacterial strains are shown in Table 3.1. XL1-Blue or DH10B was used for cloning and plasmid amplification. BL21 (DE3), C41 (DE3) and Rosetta (DE3) strains were used for protein expression.

Table 3.1. Bacterial strains used in this study.

Strain	Genotype	Reference
XL1-Blue	*E. coli endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K^- m_K^+)*	Stratagene
DH10B	*E. coli F– mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ (ara-leu)7697 galU galK λ– rpsL(Str^R) nupG*	ThermoFisher
BL21 (DE3)	*E. coli B F - dcm ompT hsdS(r m) gal λ (DE3)*	Stratagene
C41 (DE3)	*E. coli F – ompT hsdSB (rB-mB-) gal dcm (DE3)*	Lucigen
Rosetta (DE3)	*E. coli* B F$^-$ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm (DE3) pRARE (Cam^R)*	Sigma

Bacterial growth and protein expression were carried out in Miller's Luria Bertani medium (Miller's LB, VWR Medicals). Erlenmeyer flasks of 250 ml (for pre-inoculum) or 5 liter (VWR International), in shaking incubators Innova 44. For plasmid production, 10 ml LB culture were grown in 50 ml falcon tubes (Sarstedt). Bacterial growth in solid media were done in Petri dishes (Sigma) with LB supplemented with agar (Pronadisa). All media and material used for bacterial growth were sterilized at 134 ºC for 20 minutes, unless it was bought sterile (e.g., petri dishes or falcon tubes). When needed, kanamycin at 50 mg/ml (Fisher Scientific) were prepared at in milli Q water, filtered through a 0.2 µm filter (Fisher Scientific) and added to the media diluted 1000 times after sterilization process. Chloramphenicol (Panreac) was prepared at 35 mg/ml, dissolved in ethanol 100% and filtered through a 0.2 µm filter (Fisher Scientific).

3.2.4. DNA

$CTPR7_{20Cys}$ was cloned in pProEX-HTa vector inspired by His-variant CTPR6_16His_WT$(4His)_4$WT already published[11]. Cation independent mannose-6-phosphate receptor (CIMPR) was obtained from DNASU plasmid repository (HsCD00403292). Fluorescent tag (mTurquoise2) was kindly donated by Dr. Alvaro Villarroel (UPV/EHU-University of the Basque Country). Metallothionein type I tag (MT1) was obtained from Addgene (Plasmid #32101).

3.2.5. Plasmids

Plasmids used within this project are shown in Table 3.2. Proteins were cloned on expression vectors pET HisSUMO-GFP[38]. pET HisSUMO-GFP is a pET28 derived bacterial expression vector. It has a gene for resistance to kanamycin

as selectable marker, and a poly-histidine tag as affinity tag. This poly-His is flanked downstream with N-terminal (100aa, 11.5 kDa) SUMO3 fusion partner to increase protein solubility. Furthermore, this tag could be removed using a protease specific of the structure of SUMO3, Sentrin specific protease 2 (SENP2). After cleavage, any additional amino acid is left.

3.2.6. Oligonucleotides

Oligonucleotides used in this project are shown in Table 3.3. All oligonucleotides were designed with a melting temperature between 60-70 ºC. They were purchased from Invitrogen, 25 or 50 nanomoles with a purity of ≥98%. Gibson assembly[39] was used for cloning; the primers used for amplifying the vector pET28 HisSumo3 were designed to amplify vector from *AgeI* site (Reverse primer, see table 3.3. and figure 3.3.) and *XhoI* site (Reverse primer, see table 3.3. and figure 3.3.). Moreover, constructs were designed to be inserted instead of the GFP gene (for pET Sumo3 vector), in frame with the corresponding tags and the stop codon (figure 1).

For further details about Gibson assembly, see section 3.2.9.

3.2.7. Enzymes

During the thesis, the following enzymes were used according to the manufacturer instructions:

- Phusion HF II (Fisher Scientific)
- DpnI (Fisher Scientific)
- T4 DNA Ligase (Fisher Scientific)
- T5 Exonuclease (NEB)

3.2.8. SENP2 proteases

SENP2 is cloned in pET28 vector. It was kindly donated by Prof. Karl-Peter Hopfner, at Gene Centre, in Munich. SENP2 was expressed in *E. coli* Rosetta (DE3) cells as described previously[40]

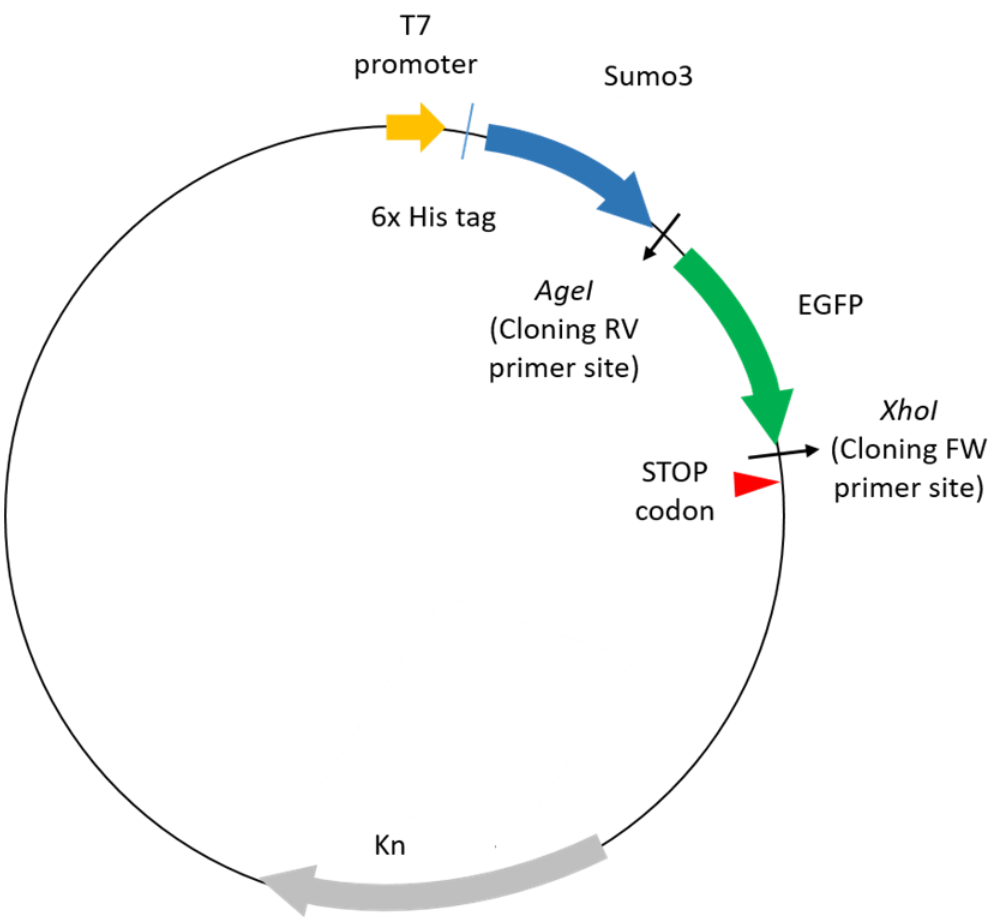

Figure 3.3. Schematic representation of the vectors (pET HisSumo3 GFP) used in this work. It contains a T7 promoter activated by IPTG induction, HisSumo3 as affinity tag and a STOP codon and an antibiotic selectable marker. Gibson assembly primers are designed in the indicated sites in frame with the affinity tag and the stop codon.

Table 3.2. Plasmids and constructions used in the thesis.

Plasmid	Characteristics	Resistance
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$-5H C2342A+C2343A	Human full-length cytosolic tail CIMPR with Cys at N-terminus and 5xHis tag at the C-terminus	Kanamycin
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$ C2342A+C2343A	Human full-length cytosolic tail CIMPR with Cys and His tag at the N-terminus	Kanamycin
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$-TEV-mTURQUOISE2 C2342A+C2343A	Human full-length cytosolic tail CIMPR with Cys and His tag at the N-terminus and mTurquoise2 tag at C-terminus. TEV sequence is used as sequence spacer between CIMPR and mTurquoise2 tag.	Kanamycin
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$-GSG-MT1 C2342A+C2343A	Human full-length cytosolic tail CIMPR with Cys and His tag at the N-terminus and MT1 tag at C-terminus. TEV sequence is used as sequence spacer between CIMPR MT1 tag.	Kanamycin
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$-TEV-C7$_{20Cys}$ C2342A+C2343A	Human full-length cytosolic tail CIMPR with Cys and His tag at the N-terminus and C7_20Cys tag at C-terminus. TEV sequence is used as sequence spacer between CIMPR C7_20Cys tag.	Kanamycin

Table 3.3. Oligonucleotides used in this thesis.

Plasmid	**Oligonucleotide**	**Cloning method**
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$-5H C2342A+C2343A	Fw insert CYS_CIMPR_5H **GTTTAACTTTAAGAAGGAGATATACCATGTGTGAGAGG AGGGAAACAGTGATAAGTAAG** Rv insert CYS_CIMPR_5H **GCAGCCGGATCTCAGTGGTGGTGGTGGTGGATGTGTAA GAGGTCCTCGTCGC** pET Sumo 5H_Cterminal up **CACCACCACCACCACTGAGATCCGGCTGC** pET Sumo NcoI ISO RV **CATGGTATATCTCCTTCTTAAAGTTAAACAAAATT**	Gibson Assembly
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$ C2342A+C2343A	FW SUMO_CYS_CIMPR **CGTGTTCCAGCAGCAGACCGGTGGATGTGAGAGGAGG GAAACAGTGATAAGTAAGCTGACC** RV SUMO_CYS_CIMPR **CAGTGGTGGTGGTGGTGGTGCTCGAGTTAGATGTGTAA GAGGTCCTCGTCGCTGTCG** FW_VECTOR Sumo-isoth-long-up **CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGC** RV VECTOR pETSumo3-AgeI-low **TCCACCGGTCTGCTGCTGGAACACGTCGATGGTGTCCT CG**	Gibson Assembly
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$-TEV-mTURQUOISE2 C2342A+C2343	FW SUMO_CYS_CIMPR_Turquoise **CGTGTTCCAGCAGCAGACCGGTGGATGTGAGAGGAGG GAAACAGTGATAAGTAAGCTGACC** RV SUMO_CYS_CIMPR_Turquoise **CAGTGGTGGTGGTGGTGGTGCTCGAGTTACTTATACAG CTCGTCCATGCCG**	Gibson Assembly

A	FW VECTOR Sumo-isoth-long-up **CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGC** RV VECTOR pETSumo3-AgeI-low **TCCACCGGTCTGCTGCTGGAACACGTCGATGGTGTCCT CG**	
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$-GSG-MT1 C2342A+C2343A	Fw MT1 **CGACGAGGACCTCTTACACATCGGTTCTGGTatggaccccaac tgctcctgc** Rv Mt1 **TCAGTGGTGGTGGTGGTGCTCGAGTTAGGCACAGC ACGTGCACTTGTC** FW VECTOR Sumo-isoth-long-up **CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGC** Rv Cys CIMPR MT1 vector GCGACGAGGACCTCTTACACATC **GATGTGTAAGAGGTCCTCGTCGC**	Gib son Assembly
pET HisSumo3 Cys-CIMPR$_{(2330-2491)}$-TEV-C7$_{20Cys}$ C2342A +C2343A	FW C-CIMPR-CTPR6-16His **GAAAACCTGTATTTTCAGGGCGGCGCCATGGGATCCGC TG** RV C-CIMPR-CTPR6-16His **CAGTGGTGGTGGTGGTGGTGCTCGAGTTAACCCTGTTT CTGTTTAGCGTTACCC** FW VECTOR Sumo-isoth-long-up **CTCGAGCACCACCACCACCACCACTGAGATCCGGCTGC** RV VECTOR -CIMPR-CTPR6-16His **CAGCGGATCCCATGGCGCCGCCCTGAAAATACAGGTTT TCGATGTG**	Gib son Assembly

.

3.2.9. Cloning methods

In general, cloning was performed using Gibson isothermal assembly, described by Daniel G. Gibson[39]. As a general description of the method, DNA of interest was amplified by PCR, as well as the vector where it was going to be introduced. Both PCR products (vector and insert) contained overlapping regions on their flanks. They were digested with DpnI (Fisher Scientific), a restriction enzyme able to digest methylated DNA, avoiding false positive colonies caused by the presence of template DNAs. PCR products were purified using Promega clean-up system, and they were incubated with the so-called "isothermal mix" (Table 3.4. (A)). 50-100 nanograms of vector DNA was mixed with 3-5 molar excess of insert DNA in a total of 5 µl. DNAs were then mixed with 15 µl isothermal mix (see Table 3.4. (B)). Ratios may vary depending on the size of the insert, increasing ratio when insert is smaller than 500 base pairs. The mix was incubated one hour at 50 ºC. The reaction allows the matching of the overlapping regions at any place in the vector, due to the action of the enzymes present in the mixture.

Table 3.4. Isothermal reaction mixture (A), and composition of isothermal mix (B)

A

Reactive	Quantity (picomoles)
DNA insert	0.3-0.5 pmoles
DNA vector	0.1 pmoles
+15 µl Isothermal mix	

Incubate 1 hour at 50ºC

B

Isothermal mix	Isothermal buffer composition
40 µl isothermal buffer	500 mM Tris-Hcl pH 7.5 25% PEG 8000 50 mM MgCl2 50 mM DTT 5 mM NAD 1mM each dNTP Fill up to 6 ml dH_2O
1.5 µl T5 exonuclease 1 U/µl (NEB)	
2.5 µl Phusion polymerase (Thermo)	
10 µl Taq ligase NEB	
96 µl milli Q water	

3.2.9.1. Amplification of DNA by PCR

DNA amplification was accomplished by polymerase chain reaction, or PCR, using Phusion high-fidelity (Thermo Fisher Scientific). Template DNA and oligonucleotides used are described in materials section. Oligonucleotides

were manually designed (see Table 3.3.) and optimized for melting temperature using the webserver OligoCalc (http://biotools.nubic.northwestern.edu/OligoCalc.html). All oligonucleotides were designed to have a melting temperature ≥ 62 ºC, and most annealing temperatures were set at this temperature. For reverse oligonucleotides, a web server was used to obtain the reverse complementary strand (https://www.bioinformatics.org/sms/rev_comp.html). The reaction mixture and conditions of PCR are detailed in Table 3.5.

Table 3.5. PCR reaction mixture for 50 µl reaction (A) and PCR conditions (B)

A

Reactive	Volume (µl)
Milli Q water	37.5 µl
5x Phusion HF buffer	10 µl
dNTPs	200 µM (1 µl of 10 mM)
Oligonucleotides	2.5 µl each (0.5 µM each)
DNA template (1 ng/µl)	1 µl
Phusion HF DNA polymerase	0.5 µl

B

Stage	Temperature	Time (s)	
Initial denaturation	98ºC	30	1
Cycled denaturation	98ºC	10	30
Annealing	60-65ºC	30	
Extension	72ºC	30 s/kb	
Final extension	72ºC	600	1

3.2.9.2. Agarose gel electrophoresis

Amplification of PCR products and evaluation of DNA was carried out using agarose gel electrophoresis. Agarose D1 low EEO (Pronadisa) was dissolved at 1-2% (w/v) in TAE buffer (20 mM Acetic acid, and 1 mM EDTA, 40 mM Tris Base pH 8) and boiled in a microwave until agarose was completely dissolved. When solution was at around 60 ºC, SYBR® safe (Invitrogen) was added at 1:10000 (v/v), and gel was polymerized in a mini sub-cell GT gel caster (Bio-rad). Samples were mixed with 6X DNA loading buffer (Thermo Fisher Scientific), and 5 µl of Generuler 1kb DNA ladder (Thermo sci.) was used as molecular weight marker. Gels were run in a horizontal mini sub® cell

GT chamber (Bio-rad) at 100 V for 45-60 minutes, and DNA was observed using ImageQuant^{TM} LAS 4000 (GE Healthcare).

3.2.9.3. DNA transformation by heat shock

Isothermal or ligation products were directly transformed in *E. coli* XL1 blue or DH10B competent cells. In any case, 5 µl of ligation or Gibson assembly were added to 50 µl of homemade competent cells, and the mixture was incubated on ice for 30 minutes. Cells were then incubated in a heating block at 42 ºC for one minute. Later, cells were incubated on ice again for 5 minutes, and 800 µl LB media was added. Cells were then incubated for 45-60 minutes at 37 ºC. Cells were centrifuged at 9000 g for 1 minute at room temperature, and 600 µl supernatant was removed. The resultant pellet was re-suspended in the excess volume (250 µl) and plated on a Petri dish containing LB agar supplemented with the appropriate antibiotic as selective marker for the plasmid.

3.2.9.4. Colony PCR

Colonies obtained after transformation were checked for the presence of the insert. PCR mixture was prepared with one oligonucleotide of the insert, and one universal primer that hybridized with the vector (e.g., T7 promoter primer). 10 µl of PCR 2x master mix (Thermo Fisher Scientific), was mixed with 0.5 µl of each primer, 10 µl milli Q water and 1 colony. Then, PCR cycles like Table 3.5.(B) were used, but with a denaturation temperature of 95 ºC, and 1 minute of extension per kb. A negative control, using empty vector, was prepared for each test. PCR products were analysed by agarose gel electrophoresis as described before. All colonies tested were plated on fresh LB agar supplemented with antibiotic. One positive colony was selected and grown for plasmid propagation.

3.2.9.5. DNA extraction

Positive colonies were grown in 50 ml sterile falcon tube, with 10 ml LB, supplemented with the appropriate antibiotic, overnight at 37 ºC in an Excella E24 shaking incubator (New Brunswick Sci.). The following day, cells were harvested by centrifuging at 3900 g in Allegra X 22-R centrifuge (Beckman coulter). Plasmid was extracted from the resultant pellet using Wizard® Plus SV Miniprep DNA Purification system (Promega). DNA quality and quantity was assayed in a Nanodrop, as described previously.

3.2.9.6. DNA sequencing

One microgram of DNA was sent for sequencing to STAB VIDA (Caparica, Portugal), mixed with 3 µl of oligonucleotide. For sequencing, universal primers were used (i.e., T7 promoter/terminator). Obtained sequences were aligned with the expected sequence using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Nucleotides) to confirm the presence of the correct insert.

3.2.10. Heterologous protein expression and purification

Heterologous expression of these proteins was achieved in *E. coli* Rosetta (DE3) cells. Pre-inoculum of Rosetta (DE3). 1% of pre-inoculum grew overnight were added to 5 litre flasks containing 1 litre of fresh LB, supplemented with the appropriate antibiotic. Bacteria were grown until optical density at 600 nm (OD_{600}) was 0.6. At this point, temperature of incubators was reduced to 20 ºC, and 30 minutes later 0.5 mM of isopropyl-B-D-thiogalactopyranoside (IPTG) was added to the culture. Expression was carried out overnight at 20 ºC. Cells were harvested in JLA 8.1000 rotor, in an Avanti® J-26 XP centrifuge, at 5000 g for 20 minutes at 18 ºC. Pellet was

directly lysed for protein purification, or frozen in liquid nitrogen and stored at -80 ºC.

General purification strategy is shown in Figure 3.4. All steps were done at 4 ºC, and protein sample was kept on ice as much as possible to avoid protease degradation. It usually followed the same scheme: first purification of protein of interest (POI) by affinity tag, affinity tag's cleavage and consequently second purification by affinity column. For experiments requiring higher purity-grade additionally size exclusion chromatography step was performed using Superdex™ HiLoad 200 (for proteins >50 kDa) or Superdex™ HiLoad 75 (for proteins <50 kDa) 16/60 or 10/300 in an ÄKTA pure system (GE Healthcare) in cold room.

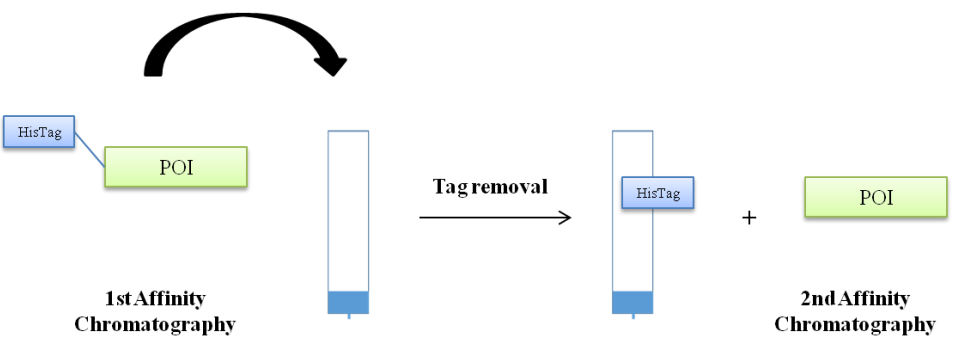

Figure 3.4. General purification strategy followed for all the purifications.

More specifically, proteins tagged with 6xHisSUMO were purified using Protino® Ni-NTA (Nickel-Nitriloacetic acid) agarose, from Macherey-Nagel. Bacterial pellet was dissolved in precooled buffer A (300 mM NaCl, 20 mM Imidazole, 1 mM DTT, Tris-HCl 50 mM pH 7.5), supplemented with 5 mM benzamidine and 0.5 mM phenylmethylsulfonyl fluoride (PMSF). Pellet was stirred at 4 ºC until no chunks of pellets were seen. Later, cells were lysed in a SONICS Vibra cell sonicator with titanium tip of 13 mm, in cycles of 10 seconds ON and 1 minute OFF, in a total time of 5 minutes. Afterwards, lysate was clarified by centrifugation at 60000 g for 45 minutes at 4 ºC using a JA 25.50 rotor in an Avanti® J-26 XP centrifuge. Meanwhile, 10 ml of Ni-NTA

resin was prepared by washing it with 50 ml of distilled water, and then equilibrating it with 50 ml of buffer A. Supernatant was incubated in a roller with pre-equilibrated Ni-NTA resin in 250 ml bottles for 1 hour at 4 ºC. Then, the mixture was loaded into a DWK Konte™ 50 ml column (Thermo Fisher Scientific), and supernatant was discarded. Resin was washed with 30 column volumes (300 ml) buffer A, and protein was then eluted with 30 ml of buffer B (300 mM NaCl, 250 mM Imidazole, 1 mM DTT, 25 mM Tris-HCl pH 7.5). Protein was dialyzed overnight at 4 ºC in buffer C (150 mM NaCl, 1 mM DTT, 25 mM Tris-HCl pH 7.5), in a 10 kDa cut-off SnakeSkin™ dialysis bag, with gentle stirring. 1 mg of SENP2 protease was added inside the dialysis bag for tag removal. The next day, protein was subjected to second purification step by affinity column separating the protein from the 6xHisSUMO tag because of its affinity to the resin. Alternatively, after tag removal POI was loaded into a S200 (proteins > 50 kDa) or S75 (<50 kDa) 16/60 column. In the case the peaks did not resolve well (protein did not separate from degradation products or aggregates), another size exclusion was run. For size exclusion step was used the buffer D (150 mM NaCl, 1 mM DTT, 25 mM Tris-HCl pH 7.5). After the chromatography isolation, fractions were analysed by SDS-PAGE, and the purest samples were pooled together and concentrated. If not used directly, purified proteins were frozen in liquid nitrogen and kept at -80 ºC until further use.

3.2.10.1. Physical-chemical protein characterization

Physico-chemical characteristics of proteins were estimated using Protparam at ExPASy Server (https://web.expasy.org/protparam).

3.2.10.2. Protein concentration estimation

Protein concentration was estimated using Nanodrop N-1000 (Thermo Sci.). Molar extinction coefficient was calculated using ProtParam, and this value was used to calculate protein concentration by measuring absorbance of protein solution at 280 nm, according to Lambert-Beer law (Equation 3.1.):

$$Equation\ 3.1. \qquad A = \varepsilon c l$$

where ε is extinction molar coefficient; c is protein concentration; and l is the optical path length, in centimetres). Absorbance is triggered by aromatic residues (specially Tryptophan, but also Tyrosine and Phenylalanine).

3.2.10.3. Polyacrylamide gel electrophoresis (SDS-PAGE)

Protein quality analysis was followed by polyacrylamide gel electrophoresis (PAGE) in denaturing conditions boiled samples in presence of Sodium Dodecyl Sulphate (SDS). 12-well/15-well gels were prepared at 12 or 15% of polyacrylamide with a size of 8x8 cm length and height, and 1 cm width. Protein samples were prepared by mixing 16 µl of protein with 4 µl with Laemmli loading buffer 5x (Bio-Rad) supplemented with 10 mM DTT. Samples were heated at 95 ºC for 5 minutes. Precision plus protein dual colour standard (Bio-Rad) was used as molecular weight marker. Electrophoresis was carried out in Mini-Protean® Tetra Vertical Electrophoresis Cell (Bio-Rad)/kuroGEL Verti 1010 electrophoresis system. Inner and outer chamber was filled with Tris-glycine "running buffer", (192 mM glycine, 0.1% SDS, 25 mM Tris pH 8.3). Gels were run at 150 – 200 V for 1.5 hours. Gels were stained with Coomassie Brilliant Blue R-250 (0.1% (w/v) Coomassie Brilliant Blue R-250, 45% ethanol, 10% acetic acid and 45% demi-water (v/v)). Gels

were stained for 30 minutes. For protein visualization, gels were unstained using a mixture of 40% ethanol, 10% acetic acid, and 50% demi-water (v/v).

3.2.11. Synthesis of protein-stabilized metal nanoclusters and quantum dots

Cys-CIMPR-$C7_{20Cys}$-stabilized AuNCs and Cys-CIMPR-$C7_{20Cys}$-stabilized CdS-QDs were synthesized following the previously reported procedures [11,26,33,34]. In order to facilitate the reading of the thesis and the discussion of the results, Cys-CIMPR-$C7_{20Cys}$-stabilized AuNCs and Cys-CIMPR-$C7_{20Cys}$-stabilized CdS-QDs are hereinafter named *cargo-AuNCs*, *cargo-AuNKNCs*, and *cargo-CdS QDs*, where cargo is Cys-CIMPR $_{(2330\text{-}2491)}$-TEV-$C7_{20Cys}$ (C2342A+C2343A), and NK=naked nanoclusters.

3.2.11.1. Synthesis of cargo-AuNCs(1)[11,33]

First reducing step of the protein was performed by adding 10mM DTT at least for 15mins, DTT was removed by PD10 desalting column and the buffer was changed to the basic reaction one. In details, 1 mL of 10 μM of protein in sodium phosphate buffer (PBS 150 mM NaCl, 50 mM phosphate pH 10) was mixed with 40 eq. per protein of $HAuCl_4$ for at least 15 min to allow adsorption of gold salts in the protein's stabilizing sites. Then, the reduction of the gold salt to metallic gold was achieved by adding ascorbic acid (100eq per each gold atom). The reaction was incubated at 50°C for 72h. Finally, a centrifugation step (1h, 14000rpm, 4°C) was performed in order to remove protein-metal aggregates and the unreacted salts are removed by PD10 desalting column and the buffer was changed to PBS (150 mM NaCl, 50 mM phosphate pH 7.4).

3.2.11.2. Synthesis of cargo-AuNCs(2)[11,33]

First reducing step of the protein was performed by adding 10 mM DTT at least for 15 mins, DTT was removed by PD10 desalting column and the buffer was changed to the basic reaction one. In details, 1mL of 10 µM of protein in sodium phosphate buffer (PBS: 150 mM NaCl, 50 mM phosphate pH 12) was mixed with 40 eq. per protein of $HAuCl_4$ for at least 15 min to allow adsorption of gold salts in the protein's stabilizing sites. The reaction was incubated at 50 °C for 48h. Finally, a centrifugation step (1h, 14000rpm, 4 °C) was performed in order to remove protein-metal aggregates and the unreacted salts are removed by PD10 desalting column and the buffer was changed to PBS (150mM NaCl, 50 mM phosphate pH 7.4).

3.2.11.3. Synthesis of cargo-AuNKNCs[26]

First Au_7NKNCs for the reaction were synthesized. In a vial, nanopure water (860 mL) was added to NaOH solution (440 mL, 2M), followed by addition of HAuCl4 $\cdot 3H_2O$ (25 mL, 50 mM), and left to react at room temperature stirring for 1 h. Then the synthesis of cargo-Au_7NKNCs was performed. To a 15 mL volumetric vial, reagents were added in the following order: Au_7NKNCs (1.325 mL), HCl aqueous solution (490 mL, 2 M), and protein solution (at least 100 mM). The molar ratio of cysteine/Au was 30. The mixture was stirred with magnetic vigorous stirring at room temperature for 1 day. Before the adding, reducing step of the protein was performed by adding 10mM DTT at least for 15mins, DTT was removed by PD10 desalting column and the buffer was changed to PBS 150 mM NaCl, 50 mM phosphate pH 7.4. Finally, a centrifugation step (1h, 14000rpm, 4°C) was performed in order to remove protein-metal aggregates and the unreacted salts are removed by PD10 desalting column and the buffer was changed to PBS 150mM NaCl, 50 mM phosphate pH 7.4.

3.2.11.4. Synthesis of cargo-CdS QDs(1)[34]

First reducing step of the protein was performed by adding 10mM DTT at least for 15mins, DTT was removed by PD10 desalting column and the buffer was changed to the double distilled water (ddH_2O). In details, 2 mL of 5 µM of protein (ddH_2O) were mixed with 10eq per protein of Cd $(SO_4)_2$ and NH_4OH was added to have basic conditions. The reaction was incubated at 37 °C for at least 1h. Then, in order to reduce Cd^{2+}, Na_2S was added up to obtain a concentration of 0.34 mM. To obtain the final reaction volume, ddH_2O was added up to 2.5 mL. The reaction was incubated at 37 ºC for 48h. Finally, a centrifugation step (1h, 14000 rpm, 4 °C) was performed in order to remove protein-metal aggregates and the unreacted salts are removed by PD10 desalting column and the buffer was changed to PBS 150 mM NaCl, 50 mM phosphate pH 7.4.

3.2.11.5. Synthesis of cargo-CdS QDs(2)[34]

First reducing step of the protein was performed by adding 10mM DTT at least for 15 min., DTT was removed by PD10 desalting column, and the buffer was changed to the double distilled water (ddH_2O). In details, 2 mL of 5 µM of protein (ddH_2O) were mixed with 5 eq. per protein of Cd $(SO_4)_2$ and NH_4OH was added to have basic conditions. The reaction was incubated at 37 °C for at least 1h. Then, in order to reduce Cd^{2+}, Na_2S was added up to obtain a concentration of 0.17 mM. To obtain the final reaction volume, ddH_2O was added up to 2.5 mL. The reaction was incubated at 37 ºC for 48h. Finally, a centrifugation step (1h, 14000 rpm, 4 °C) was performed in order to remove protein-metal aggregates and the unreacted salts are removed by PD10 desalting column and the buffer was changed to PBS (150 mM NaCl, 50 mM phosphate pH 7.4).

3.2.11.6. Synthesis of Cys-CIMPR-MT1$_{20Cys}$-stabilized AuNCs

The metal coordination in Cys-CIMPR-MT1 fusion protein was obtained following the same optimized protocol used for cargo-Au_{10}NCs, as described above in section 3.2.11.1.

3.2.12. Metal Coordinated-Protein Determination

For metal coordinated-protein determination the Bradford Protein Assay[41] was employed. All samples, including a set of standards (BSA: 0.5 mg/mL, 1 mg/mL, 2 mg/mL, 3 mg/mL, 4 mg/mL), were prepared in aqueous solution. For 10 µl of each sample 200 µl of working reagent was added. The samples, including standards and blanks, were incubated at RT for at least 5 min, followed by measurement of absorbance at 595nm on a plate reader.

3.2.13. Biochemical Methods

3.2.13.1. Circular Dichroism

Circular dichroism (CD) is a useful technique used to predict the folding state of a protein in solution[42]. It is based on the differences between the absorption of left- and right- handed circularly polarized light. In the case of proteins, as chiral molecules, the organization of peptide bonds in the molecule gives rise to a specific CD spectrum, which can be correlated with the percentage of α-helixes, β-sheets and random coil structure present in the protein. The protein secondary structure and the metal coordinated-protein was examined by CD using a Jasco J-815 spectrometer (JASCO Corporation, Tokyo, Japan). CD spectra were acquired at 5-10 µM protein concentration in a 1 mm path length cell at 25 ºC using a 1 nm bandwidth with 1 nm increments and 10 s average time, accumulation 5 times (check again/add other parameters). Protein buffer was changed to 10 mM NaCl, 10 mM Tris-HCl pH 7.4, by PD10 desalting

column. Next, proteins were analysed using a Jasco J-815 spectrometer (JASCO Corporation, Tokyo, Japan). Data were obtained at 25 ºC, measuring between 190 – 250 nm, with a quartz cuvette of 0.1 cm path length. Data was blank subtracted.

3.2.13.2. Fluorescence Spectroscopy

Fluorescence spectra were performed on a Varioskan Flash microplate reader (Thermo Scientific). The system was controlled by SkanIt Software 2.4.3. for Varioskan Flash. Protein buffer was changed to PBS (150 mM NaCl, 50 mM phosphate pH 7.4) by PD10 desalting column. Spectra were recorded using 100uL of each sample. Molecular Probes® 96-well microplates were used for the measurement. Data were obtained at 25 ºC, measuring between 400 – 800 nm. Data were plotted using Excel.

3.2.13.3. MALDI-TOF

Matrix assisted laser desorption ionization (MALDI) mass spectrometry Mass spectra were acquired on an Applied Biosystems Voyager Elite MALDI-TOF mass spectrometer with delayed extraction (Applied Biosystems, Framingham, MA, USA) equipped with a pulsed N2 laser (λ= 337 nm). Sinapic acid was used as matrix. An extraction voltage of 20 kV was used. All mass spectra were acquired in positive reflection mode using delayed extraction with an average of 50−100 laser shots. MALDI-TOF sample preparation included 1 μL of the sample mixed with 3 μL of sinapic acid in water:acetonitrile (50/50) with 0.01% TFA. Then, 1 μL of the mixture was deposited onto the MALDI plate and allowed to air-dry.

3.2.13.4. ICP-MS

Sample for inductively coupled plasma mass spectrometry (ICP-MS) were prepared 100 µL of each purified metal coordinated-proteins at approximately at 100 µM concentration were mixed with 300 µL of 37% HCl or HNO_3 the resultant suspension was sonicated for 30 minutes at 40 ºC. Finally, 2700 mL of bi-distilled water was added (in order to have less than 2% of acid in the sample). The Cd and Au concentration was determined by measuring the sample using an iCAP-Q ICPMS (Thermo Scientific, Bremen, Germany) equipped with an autosampler ASX-520 (Cetac Technologies Inc., NE, USA) (n = 3) and QtegraTM v2.6 (Themo Scientific).

3.2.14. Electron Microscopy

3.2.14.1. Transmission electron microscopy (TEM)

TEM samples were prepared by depositing 2 µL of the sample solution on the grid. Before sample deposition, ultrathin carbon films on holey carbon support film, 400 mesh copper grids (TED PELLA INC.) were exposed to glow-discharge treatment. The protein samples in low salt buffer (10 mM Tris pH 7.4, 10 mM NaCl) were deposited at 1 µM protein concentration. The excess solution was air-dried at RT, ON. Micrographs were recorded using a JEOL JEM1200EXII electron microscope with a tungsten filament operated at 100 kV and at 60 K magnifications. Micrographs were digitalized using a Photoscan TD Zeiss-Intergraph scanner (pixel size = 2.33 Å/px).

3.2.14.2. Cryo Electron Microscopy (Cryo-EM)

CryoEM allows the observation of samples in near-native state, and at very low temperatures (-180ºC), protecting them partially from electron radiation

damage. In order to properly see the sample, it needs to be vitrified: it must be fast frozen in order to form vitreous (i.e., amorphous) ice, and not ice crystals. This was achieved using liquid ethane at about -180ºC, cooled in special chamber surrounded by liquid nitrogen[43]. Furthermore, ice must be very thin, not bigger than 500 nm, so electrons can go through the ice and interact with the specimen.

Samples were evaluated on Quantifoil grids R2/2 Cu 300 mesh. Grids were hydrophilized by plasma using a Glow discharge/carbon evaporation MED 020 (BALTEC) and followed by vitrification in liquid ethane in a FEI vitrobot™ Mark III. Samples were stored into a homemade gridbox and kept in a dewar filled with liquid nitrogen. Samples were screened at CIC bioGUNE JEOL 2200FS cryo-electron microscope equipped with an UltraScan 4000 SP (4008×4008 pixels) cooled slow-scan CCD camera (GATAN, UK).

Grids were vitrified in specific conditions that favour homogeneous and thick ice: 30 seconds incubating sample on the grid, 2 seconds blotting, 0 seconds waiting after drying, and offset of -2. Grid were activated using 6 mA of plasma for 30 seconds.

3.3. Results and Discussion

3.3.1. Protein Design and Purification

In the present work, we placed a particular focus on exploring the design of protein scaffolds to stabilise metal nanoclusters and quantum dots as novel dual tags both for fluorescence microscopy and electron microscopy. Indeed, fluorescence microscopy offers the potential to localize specific targets with higher precision, whereas electron microscopy yields detailed information on

subcellular structures. Therefore, correlative imaging techniques, which combine high precise protein localization with high resolution structural information are especially appealing. In this context, new tools such as dual tags, *i.e.* a single tool for the two imaging techniques, are highly demanded for correlative imaging.

As extensively described in the introduction, designed consensus tetratricopeptide repeat (CTPR) proteins are suitable candidates as templates for the stabilization of metal nanoclusters and quantum dots, due to their modular structural and functional properties.

Indeed, metal nanoclusters and quantum dots exhibit very interesting optical, electronical and chemical properties, including strong photoluminescence, electron dense metal core, excellent photostability and good biocompatibility, while having nanometric size. It is thus clear that CTPRs stabilizing metal NCs or QDs represent an ideal protein fusion tags for applications in labelling and imaging of biological systems[33].

Indeed, this technology allows: (1) one-step introduction of the label; (2) specific labelling of target proteins; (3) the formation of a stable protein-label compound; (4) the use of small label size not to interfere with protein function; (5) quick and quantitative labelling reaction.

In the present work $CTPR7_{20Cys}$ was used as fusion tags to label and to study proteins involved in membrane trafficking and protein recycling. Specifically, we focused on the human cation-independent mannose 6-phosphate receptor (CIMPR), a well-characterized cargo in retromer-independent recycling retrograde pathways[44,45], which is extensively described in Chapter 4.

Basically, we designed, cloned, and overexpressed fusion constructs encoding the full-length cytosolic tail of CIMPR (C2342A+C2343A) with $CTPR7_{20Cys}$ at the C-terminus. Fusion protein presents a Cys residue at the N-terminus, which could be exploited for bio-conjugation strategies, such as the thiol-maleimide coupling immobilization strategy. The cysteines at position 2342

and 2342 of CIMPR were mutated to alanine in order to favour the overexpression without having any additional cysteine competing with the cysteine designed at the N-terminus for bio-conjugation strategies.

Additionally, fusion constructs encoding CIMPR and mTurquoise2 and/or Metallothionein1 (MT1) were cloned at the C-terminus of the CIMPR in order to have control tools for the fluorescence microscopy and electron microscopy experiments, respectively. Specifically, mTurquoise2 is the brightest variant of the cyan fluorescent protein (CFP). The enhancement respect to CFP originates from stabilization of the seventh β-strand and the strengthening of the sole chromophore-stabilizing hydrogen bond[46]. It is a suitable control tool to study retrograde trafficking pathways by fluorescence microscopy because it exhibits high-fluorescence quantum yield. On other hand, for electron microscopy the cysteine-rich protein metallothionein1 (MT1) it has been chosen as clonable tag. MT1 belongs to a family of ubiquitous eukaryotic proteins that bind a variety of metals often as a means of detoxification. It has been previously shown that MT1 can bind heavy metal atoms with high affinity forming metal clusters that can be imaged by EM.[47] In Figure 3.5. the tracking methodology used for each tag system designed is summarized schematically.

All the constructs were successfully expressed in *Escherichia coli* Rosetta (DE) and purified according to the protocol described in experimental section. The molecular weight of proteins was confirmed by MALDI-TOF. In the Figure 3.6. size exclusion chromatogram and SDS-PAGE gel show that Cys-CIMPR-$C7_{20Cys}$ can be obtained at high purity grade. The purification yield was around 0.4mg/L of cell culture.

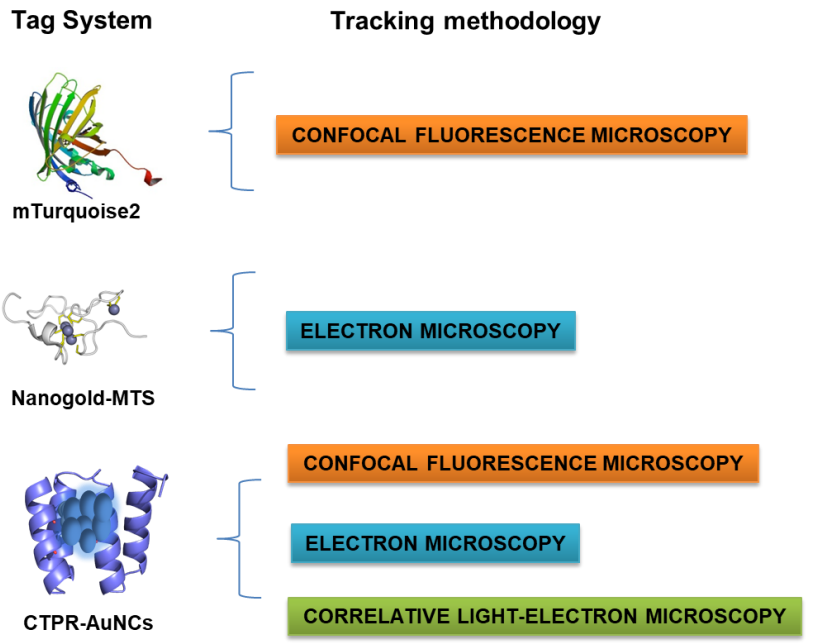

Figure 3.5. Tag system and correspondingly tracking methodology approach.

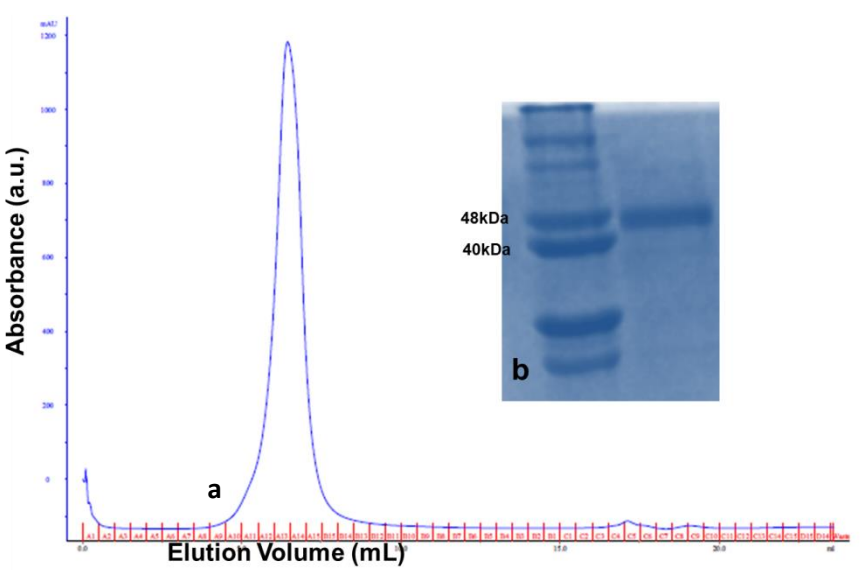

Figure 3.6. Purification of Cys-CIMPR-$C7_{20Cys}$ (a) size exclusion chromatogram of Cys-CIMPR-$C7_{20Cys}$ run over a Superdex 75 column (b) SDS-PAGE gel of purified Cys-CIMPR-$C7_{20Cys}$ (48kDa)

In case of Cys-CIMPR-MT1, the protein was successfully over-expressed, but it was not possible to obtain reproducible purification even using several strategies and attempts. Probably, this is due to the presence of a 20-Cys pocket within the MT1. Indeed the family of MTs are used in nature as natural scavengers for divalent ions[47] and it was observed that during the affinity column purification step some metal nanoclusters were formed inside the column. Probably this phenomenon was favoured by the reducing environment used during the purification. On the other side, no reducing conditions are used disulfide bridges can form and affect severally the purification.

3.3.2. Optimization and characterization of metal-coordinated proteins

Once fusion proteins were obtained at high purity grade, we worked on the optimization of the protocol for the metal coordination to the proteins. Specifically, in this thesis we focused on Cys-CIMPR-$C7_{20Cys}$, which was used to stabilize different nanoclusters and quantum dots thanks to the tunability and versatility of $C7_{20Cys}$ sequence.
Specifically following different procedures and protocols for metal-protein coordination, in the present work we aim at demonstrating how, with a single metal stabilizing scaffold (Cys-CIMPR-$C7_{20Cys}$), a toolbox of different protein-based labelling nanotools can be developed. Indeed Cys-CIMPR-$C7_{20Cys}$ can stabilise different metal clusters which provide different properties, such as emission wavelength, metal core size, and electron density.

The different procedures and protocols for metal-protein coordination can be found detailed in section 3.2.11.

As expected, when Cys-CIMPR-$C7_{20Cys}$ was metal coordinated, it showed the characteristic fluorescence features. As shown in Figure 3.7., the fluorescence

emission spectrum of cargo-AuNCs(1) consisted of a single peak with a maximum at 438 nm, when excited at 370 nm. The fluorescence emission spectrum of cargo-AuNCs(2) consisted of a low intensity peak with a maximum at 441 nm and of a higher intensity peak with a maximum at 660 nm. The fluorescence emission spectrum of for cargo-AuNKNCs consisted of a low intensity peak with a maximum at 432 nm and of a higher intensity peak with a maximum at 645 nm. Also, cargo-CdS QDs fluorescence features, consisting of a peak with a maximum at 472 nm, when excited at 370 nm for the cargo-CdS QDs(2)and a peak with a maximum at 518nm for the cargo-CdS QDs(1).

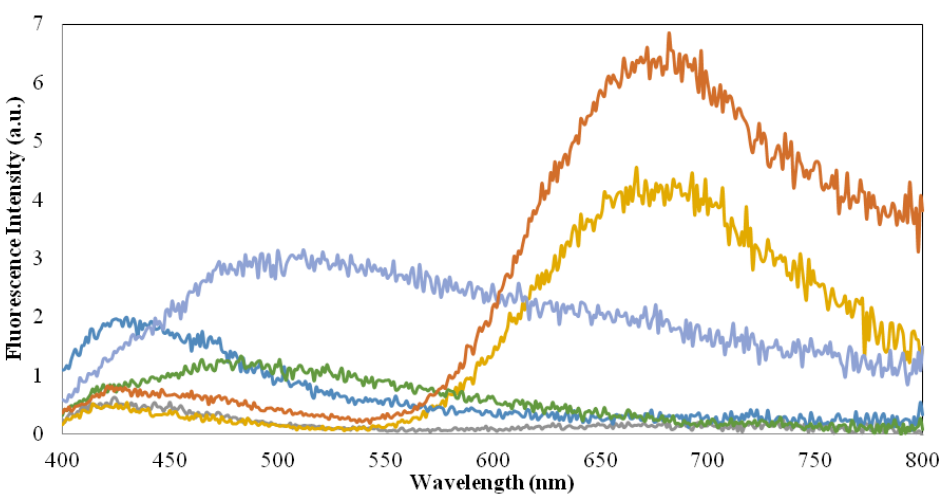

Figure 3.7. Fluorescence emission spectra of cargo-AuNCs(1) (dark blue line), cargo-AuNCs(2) (red line), cargo-AuNKNCs (yellow line), cargo-CdS QDs(2) (light blue line) and cargo-CdS QDs(1) (green line). All the measurements were carried out using the same protein concentration (11 µM) in PBS buffer pH 7.4. All emission spectra are obtained exciting at 370nm.

For Cys-CIMPR-MT1$_{20Cys}$ we did not obtain any fluorescent metal nanoclusters, as expected AuNPs larger than 2 nm. In fact, TEM image (Figure 3.8. (a)) showed that the protein stabilizes larger gold nanoparticles

(AuNPs) with an average diameter of 5 ± 2 nm (Figure 3.8. (b)). These results are not considered robust and consistent, since the purification was not reproducible and it could not be confirmed that the nanoparticles observed are coordinated to the MT1.

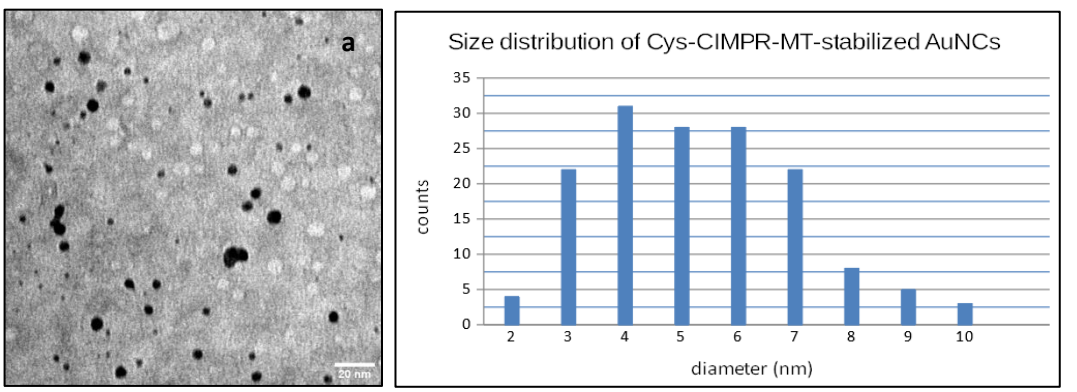

Figure 3.8. (a) TEM image of Cys-CIMPR-MT1$_{20Cys}$ stabilizing gold nanoparticles (scale bar 20nm) and (b) relative size distribution histogram.

MALDI-TOF technique was used to characterize the metal-protein coordination and to determine the size of cargo-AuNCs(1), cargo-AuNCs(2), cargo-AuNKNCs, cargo-CdS QDs(1) and cargo-CdS QDs(2). No results were obtained in case of cargo-AuNCs(1), cargo-AuNCs(2) and cargo-AuNKNCs, because it was not possible to make fly them. In the Figure 3.9., results for cargo-CdS QDs(1) and cargo-CdS QDs(2) are shown. In Figure 3.9.(a) the peak at 48507.660 Da corresponds to the molecular weight of Cys-CIMPR-C7$_{20Cys}$. In Figure 3.9.(b) line 1 the gel electrophoresis showed the high purify grade of Cys-CIMPR-C7$_{20Cys}$. Additionally, SDS-PAGE of cargo-CdS QDs(1) and cargo-CdS QDs(2) are shown in Figure 3.9.(b), respectively in line 2 and line 3. Compared to Cys-CIMPR-C7$_{20Cys}$, the ability in MALDI-TOF of cargo-CdS QDs(1) (Figure 3.9.(c)) and of cargo-CdS QDs(2) (Figure 3.9.(d)) dramatically decreased, as observed previously for metal protein complexes[11].

Nonetheless, the stabilization of the QDs within the Cys-CIMPR-$C7_{20Cys}$ was clear: the peaks are wider and the difference in molecular weight between Cys-CIMPR-$C7_{20Cys}$ and cargo-CdS QDs(1) and cargo-CdS QDs(2) is Cys-CIMPR-$C7_{20Cys}$ is significative (difference of 1168.764u for cargo-CdS QDs(1) and 1378.092 for cargo-CdS QDs (2) .

According to those calculations there may be a slight difference in the composition of the CdS QDs, but considering the error in the mass determination, this difference is not significant. Because of the intrinsic error of technique and above all because cargo-CdS QDs(1) and cargo-CdS QDs(2) did not flying well, it was not possible to have an accurate estimation of the size of quantum dots based in these data.

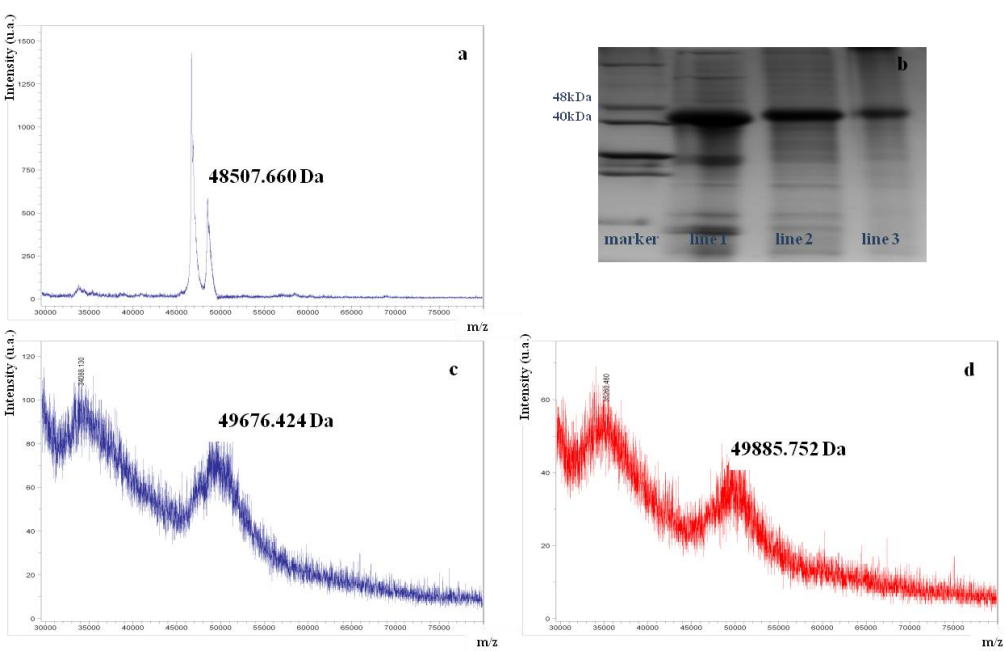

Figure 3.9. (a) MALDI-TOF mass spectrum of Cys-CIMPR-$C7_{20Cys}$ (b) SDS-PAGE of Cys-CIMPR-$C7_{20Cys}$ (line1), cargo-CdS QDs (1) (line2) and of cargo-CdS QDs(2) (line3) (c) MALDI-TOF mass spectrum of cargo-CdS QDs (1) (d) MALDI-TOF mass spectrum of cargo-CdS QDs(2).

In order to implement the data arising from MALDI-TOF, ICP-MS measurements were carried out. Thanks to ICP-MS measurements, we determined the number of atoms which compose the protein-metal complexes. The number of Cd atoms per protein was detected to study cargo-CdS QDs (Table 3.6.) and the number of Au atoms per atoms to study cargo-AuNCs and cargo-AuNKNCs (Table 3.7.).

Table 3.6. Number of Cd atoms for cargo-CdS QDs.

Synthesis	n° atoms of Cd per protein	STD
cargo-CdS (1)	14.1	0.3
cargo-CdS (2)	10.9	0.2

To calculate the number of metal atoms coordinated to the protein, the concentrations of Cd/Au measured by ICP were divided by the protein concentration of cargo-CdS QDs, cargo-AuNCs, and cargo-AuNKNCs quantified by Bradford.

In the case of the cargo-CdS QDs (Table 3.6.) when emission wavelength is larger (red-shifted), the number of Cd coordinated is higher. In the case of gold-protein complexes (Table 3.7.) it was not detected the same behaviour: cargo-AuNCs (1) presents a core with a smaller number of atoms respect to cargo-AuNCs (2). For cargo-AuNKNCs, it was quantified an average of around 7 atoms of Au per protein, even if the emission wavelength is very similar to cargo-AuNCs (2) .

Table 3.7. Number of Au atoms for cargo-AuNCs and cargo-AuNKNCs.

Synthesis	n° atoms of Au per protein	STD
cargo-AuNCs (1)	9.6	0.1
cargo-AuNCs (2)	21.6	1.0
cargo-AuNKNCs	6.7	0.8

Additionally, CryoEM imaging measurements of protein-metal complexes were performed. It is noteworthy to mention that cryoEM was selected since it is a technique friendly for biological, which we used to study the retromer-independent CIMPR recycling using the protein-metal complexes here developed (see Chapter 4). It was not possible to visualize cargo-AuNCs (1), cargo-AuNCs (2) and cargo-AuNKNCs at the resolution of the microscope, probably because of their small size. On the contrary, cargo-CdS QDs (1) and cargo-CdS QDs (2) were clearly visualized, as shown in Figure 3.10. Also, it is noteworthy to underline cargo-CdS QDs (1) is more electron-dense than cargo-CdS QDs (2), confirming the data obtained for ICP-MS measurements which showed a higher number of Cd for cargo-CdS QDs (1) than cargo-CdS QDs (2).

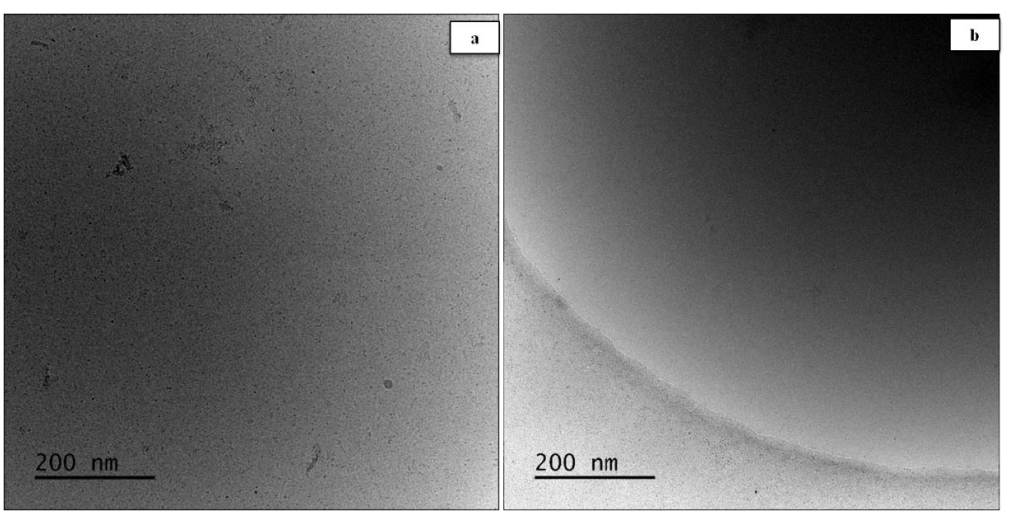

Figure 3.10. CryoEM images of (a) cargo-CdS QDs (1) and (b) cargo-CdS QDs (2)

Additionally, circular dichroism (CD) was carried out in order to check the secondary structure of the protein after metal coordination and nanomaterial stabilization (Figure 3.11.). Compared to the Cys-CIMPR-$C7_{20Cys}$, Cys-CIMPR-$C7_{20Cys}$ stabilizing metal nanoclusters and quantum dots clearly lose

part of their secondary structure. The shape of CD spectrum of Cys-CIMPR-$C7_{20Cys}$ presents the characteristic alpha-helix shape arising from CTPR sequence (CIMPR has normally a mix between random coil and antiparaller β-strands pattern). In detail, there are two minima at 210 nm and 220 nm and a maximum at 198 nm within the CD spectrum of Cys-CIMPR-$C7_{20Cys}$. After metal coordination in all cases, the CD spectra showed the conservation of the signature spectra related to the alpha helical structure, with only a slight change in the shape. These qualitative spectra also showed a significant decrease in CD signal intensity, which may be due to a partial loss of protein in the synthesis procedure. This apparent minor changes on the structure of the fusion protein are not critical for the application of our systems, because the interactions of the CIMPR in retromer- independent retrograde recycling pathway depend on its primary structure, as described in Chapter 4.

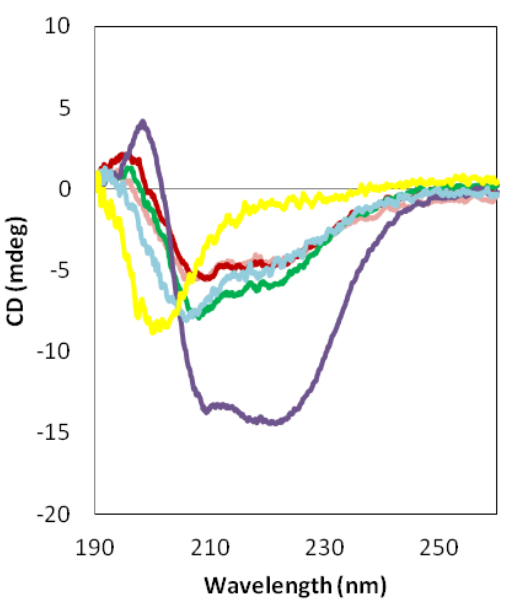

Figure 3.11. CD spectra of CIMPR (yellow line), Cys-CIMPR-$C7_{20Cys}$ (purple line), cargo-CdS QD (1) (pink line), cargo-AuNCs (1) (red line), cargo-AuNCs (2) (green line), and cargo-AuNKNCs (light blue line).

3.4. Conclusion

In this chapter, it has been shown the design and development a novel dual fusion tag which can be used to label and to study proteins of interest by fluorescent and electron microscopy.

Taking into account that CTPR proteins are robust templates to stabilize AuNCs and CdS QDs, we successfully cloned, overexpressed and purified Cys-CIMPR-$C7_{20Cys}$ fusion construct in order to study the retromer-independent recycling of CIMPR.

Indeed AuNCs and CdS QDs exhibit very interesting optical, electronical and chemical properties, including strong photoluminescence and electron dense metal core, while having nanometric size, which make them ideal tools for fluorescent and/or electron microscopy[33]. Using different metal coordination protocols, we showed that Cys-CIMPR-$C7_{20Cys}$ can stabilize different types of AuNCs and CdS QDs with different emission wavelengths, electron density, and size. This technology results particularly appealing for labelling and for studying biological systems, such as in case of the recycling of CIMPR. Indeed Cys-CIMPR-$C7_{20Cys}$ can be used as *in vitro* microscopy model system to track different recycling pathways of CIMPR by fluorescent and/or to provide high resolution information about structure by electron microscopy. It is noteworthy to mention that tuning the emission wavelength and/or the metal size of the dual tag without the need of cloning new tags is particularly appealing, as the capability of labelling a single tag with nanomaterials that provide distinct features it is a fast, versatile, and less-time consuming approach for multimodal labelling.

3.5. Bibliography

1. Brunette, T. J. *et al.* Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).

2. Kajava, A. V. Review: Proteins with repeated sequence - Structural prediction and modeling. *J. Struct. Biol.* **134**, 132–144 (2001).

3. Parmeggiani, F. & Huang, P. S. Designing repeat proteins: a modular approach to protein design. *Curr. Opin. Struct. Biol.* **45**, 116–123 (2017).

4. Andrade, M. A., Perez-Iratxeta, C. & Ponting, C. P. Protein repeats: Structures, functions, and evolution. *J. Struct. Biol.* **134**, 117–131 (2001).

5. Cortajarena, A. L., Mochrie, S. G. J. & Regan, L. Modulating repeat protein stability: The effect of individual helix stability on the collective behavior of the ensemble. *Protein Sci.* **20**, 1042–1047 (2011).

6. Cortajarena, A. L., Liu, T. Y., Hochstrasser, M. & Regan, L. Designed proteins to modulate cellular networks. *ACS Chem. Biol.* **5**, 545–552 (2010).

7. Romera, D., Couleaud, P., Mejias, S. H., Aires, A. & Cortajarena, A. L. Biomolecular templating of functional hybrid nanostructures using repeat protein scaffolds. *Biochem. Soc. Trans.* **43**, 825–831 (2015).

8. Sanchez-deAlcazar, D., Mejias, S. H., Erazo, K., Sot, B. & Cortajarena, A. L. Self-assembly of repeat proteins: Concepts and design of new interfaces. *J. Struct. Biol.* **201**, 118–129 (2018).

9. Sánchez-deAlcázar, D., Velasco-Lozano, S., Zeballos, N., López-Gallego, F. & Cortajarena, A. L. Biocatalytic Protein-Based Materials

for Integration into Energy Devices. *ChemBioChem* **20**, 1977–1985 (2019).

10. Mejías, S. H. *et al.* Repeat protein scaffolds: Ordering photo- and electroactive molecules in solution and solid state. *Chem. Sci.* **7**, 4842–4847 (2016).

11. Aires, A. *et al.* A Simple Approach to Design Proteins for the Sustainable Synthesis of Metal Nanoclusters. *Angew. Chemie - Int. Ed.* **58**, 6214–6219 (2019).

12. Porebski, B. T. & Buckle, A. M. Consensus protein design. *Protein Eng. Des. Sel.* **29**, 245–251 (2016).

13. Binz, H. K., Stumpp, M. T., Forrer, P., Amstutz, P. & Plückthun, A. Designing repeat proteins: Well-expressed, soluble and stable proteins from combinatorial libraries of consensus ankyrin repeat proteins. *J. Mol. Biol.* **332**, 489–503 (2003).

14. Aksel, T. & Barrick, D. Direct observation of parallel folding pathways revealed using a symmetric repeat protein system. *Biophys. J.* **107**, 220–232 (2014).

15. D'Andrea, L. D. & Regan, L. TPR proteins: The versatile helix. *Trends Biochem. Sci.* **28**, 655–662 (2003).

16. Kajander, T., Cortajarena, A. L., Mochrie, S. & Regan, L. Structure and stability of designed TPR protein superhelices: Unusual crystal packing and implications for natural TPR proteins. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **63**, 800–811 (2007).

17. Magliery, T. J. & Regan, L. Beyond consensus: Statistical free energies reveal hidden interactions in the design of a TPR motif. *J. Mol. Biol.* **343**, 731–745 (2004).

18. Main, E. R. G., Xiong, Y., Cocco, M. J., D'Andrea, L. & Regan, L. Design of stable α-helical arrays from an idealized TPR motif. *Structure* **11**, 497–508 (2003).

19. Sanchez de Alcazar Melendo, D. Engineered proteins as molecular templates for functional materials. (2019).

20. Kajander, T., Cortajarena, A. L., Main, E. R. G., Mochrie, S. G. J. & Regan, L. A new folding paradigm for repeat proteins. *J. Am. Chem. Soc.* **127**, 10188–10190 (2005).

21. Cortajarena, A. L., Wang, J. & Regan, L. Crystal structure of a designed tetratricopeptide repeat module in complex with its peptide ligand. *FEBS J.* **277**, 1058–1066 (2010).

22. Cortajarena, A. L., Kajander, T., Pan, W., Cocco, M. J. & Regan, L. Protein design to understand peptide ligand recognition by tetratricopeptide repeat proteins. *Protein Eng. Des. Sel.* **17**, 399–409 (2004).

23. Cortajarena, A. L., Yi, F. & Regan, L. Designed TPR modules as novel anticancer agents. *ACS Chem. Biol.* **3**, 161–166 (2008).

24. Aires, A., Lopez-Martinez, E. & Cortajarena, A. L. Sensors based on metal nanoclusters stabilized on designed proteins. *Biosensors* **8**, (2018).

25. Aires, A., Cadenas, J. F., Guantes, R. & Cortajarena, A. L. An experimental and computational framework for engineering multifunctional nanoparticles: Designing selective anticancer therapies. *Nanoscale* **9**, 13760–13771 (2017).

26. Londoño-Larrea, P., Vanegas, J. P., Cuaran-Acosta, D., Zaballos-García, E. & Pérez-Prieto, J. Water-Soluble Naked Gold Nanoclusters

Are Not Luminescent. *Chem. - A Eur. J.* **23**, 8137–8141 (2017).

27. Aires, A., Fernández-Luna, V., Fernández-Cestau, J., Costa, R. D. & Cortajarena, A. L. White-emitting Protein-Metal Nanocluster Phosphors for Highly Performing Biohybrid Light-Emitting Diodes. *Nano Lett.* **20**, 2710–2716 (2020).

28. Xu, Y. *et al.* The role of protein characteristics in the formation and fluorescence of Au nanoclusters. *Nanoscale* **6**, 1515–1524 (2014).

29. Liu, L., Jiang, H. & Wang, X. Design of dual metal ions/dual amino acids integrated photoluminescent logic gate by high-molecular weight protein-localized Au nanoclusters. *Nano Res.* **11**, 311–322 (2018).

30. Maity, B., Abe, S. & Ueno, T. Observation of gold sub-nanocluster nucleation within a crystalline protein cage. *Nat. Commun.* **8**, 1–9 (2017).

31. Clem Gruen, L. Interaction of amino acids with silver(I) ions. *BBA - Protein Struct.* **386**, 270–274 (1975).

32. Mandal, S., Phadtare, S. & Sastry, M. Interfacing biology with nanoparticles. *Curr. Appl. Phys.* **5**, 118–127 (2005).

33. Couleaud, P. *et al.* Designed Modular Proteins as Scaffolds to Stabilize Fluorescent Nanoclusters. *Biomacromolecules* **16**, 3836–3844 (2015).

34. Aires, A., Möller, M. & Cortajarena, A. L. Protein Design for the Synthesis and Stabilization of Highly Fluorescent Quantum Dots. *Chem. Mater.* **32**, 5729–5738 (2020).

35. Grove, T. Z., Regan, L. & Cortajarena, A. L. Nanostructured functional films from engineered repeat proteins. *J. R. Soc. Interface* **10**, (2013).

36. Mejías, S. H., Sot, B., Guantes, R. & Cortajarena, A. L. Controlled

nanometric fibers of self-assembled designed protein scaffolds. *Nanoscale* **6**, 10982–10988 (2014).

37. Leifert, J. A. & Whitton, J. L. 'Translocatory proteins' and 'protein transduction domains': A critical analysis of their biological effects and the underlying mechanisms. *Mol. Ther.* **8**, 13–20 (2003).

38. Marblestone, J. G. Comparison of SUMO fusion technology with traditional gene fusion systems: Enhanced expression and solubility with SUMO. *Protein Sci.* **15**, 182–189 (2006).

39. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

40. Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J. Struct. Funct. Genomics* **5**, 75–86 (2004).

41. Spector, T. Refinement of the Coomassie blue method of protein quantitation. A simple and linear spectrophotometric assay for ≤0.5 to 50 μg of protein. *Anal. Biochem.* **86**, 142–146 (1978).

42. Martin, S. R. & Schilstra, M. J. Circular Dichroism and Its Application to the Study of Biomolecules. *Methods Cell Biol.* **84**, 263–293 (2008).

43. Russo, C. J., Scotcher, S. & Kyte, M. Europe PMC Funders Group Europe PMC Funders Author Manuscripts A precision cryostat design for manual and semi-automated cryo-plunge instruments. **87**, 1–12 (2017).

44. Simonetti, B. *et al.* Molecular identification of a BAR domain-containing coat complex for endosomal recycling of transmembrane proteins. *Nat. Cell Biol.* **21**, 1219–1233 (2019).

45. Simonetti, B., Danson, C. M., Heesom, K. J. & Cullen, P. J. Sequence-dependent cargo recognition by SNX-BARs mediates retromer-independent transport of CI-MPR. *J. Cell Biol.* **216**, 3695–3712 (2017).

46. Goedhart, J. *et al.* Structure-guided evolution of cyan fluorescent proteins towards a quantum yield of 93%. *Nat. Commun.* **3**, (2012).

47. Mercogliano, C. P. & DeRosier, D. J. Gold nanocluster formation using metallothionein: Mass spectrometry and electron microscopy. *J. Mol. Biol.* **355**, 211–223 (2006).

4. STUDIES ON THE RETROGRADE TRAFFICKING PATHWAYS TRACKING CARGO DURING THE BIOGENESIS OF TUBULE-VESICLES CARRIERS BY NEW PROTEIN-BASED NANOTOOLS

4. Studies on the retrograde trafficking pathways tracking cargo during the biogenesis of tubule-vesicles carriers by new protein-based nanotools

4.1. Protein Homeostasis and Intracellular Protein Trafficking

4.1.1. Overview

Regulation of intracellular protein trafficking in one of the central issues in cell biology, in order to maintain a stable and functional proteome. Alterations in the fine-tuned mechanisms guaranteeing the protein homeostasis underlie the pathogenesis of severe human diseases including, among others, common neurodegenerative disorders such as Alzheimer's or Parkinson's disease[1,2]. The internal membrane system in eukaryotic cells has a pivotal role in the regulation of intracellular protein trafficking. Many cellular processes, from DNA storage to energy production, take place within specialised membrane-separated structures called organelles. Compartmentalisation allows a high degree of specialisation, but it limits the communication within the cell and (in multicellular organisms) with other cells. Hence, the cell needs a transport system regulating the traffic of proteins[3].

When organelles are in close proximity, usually within less than 30 nm, the exchange of protein is driven by membrane contact sites (MCSs), widely spread in endoplasmic reticulum (ER)[4,5]. The endoplasmic reticulum (ER) is the most extensive membrane network in the cell and it interacts with plasma membrane (PM), mitochondria, endosomes, peroxisomes, and lysosomes[6].

Alternatively, the most common protein transport to distal organelles or other cells is mediated by vesicle trafficking (Figure 4.1.). There are many routes of vesicle trafficking in the cell. Basically they can be grouped into two main categories: endocytosis/endocytic pathways concerning the inward flux of

vesicles from the plasma membrane to the cytosol and, on the other hand, exocytosis/secretory pathways regarding the outward flux to the extracellular media[7]. More specifically, in the secretory pathway, proteins destined for secretion or intracellular distribution are synthesized in the endoplasmic reticulum (ER) and pass through the Golgi apparatus and the trans-Golgi network (TGN) where they undergo additional modifications before being delivered to their destination. This forward (anterograde) flow of material is counterbalanced by the internalization of proteins and lipids from the plasma membrane (PM) for delivery to lysosomes (degradation) or to Golgi through retrograde transport routes (recycling).

Therefore, the transport of proteins is mainly driven by vesicular and tubular transport carriers (TCs). Specifically, in any vesicle-trafficking event, there are four distinct events. The first one is vesicle budding, the formation of the vesicle in the original compartment. Many factors participate in this process in order to control the selection of protein (named cargo) and the membrane deformation, but it is noteworthy to mention the crucial role played by coat proteins. The next step is the vesicle displacement to the target organelle, using motor proteins and other elements of the cytoskeleton. Then, the vesicle is recognised and tethered to the target. Finally, the coat proteins, tethering complexes, and N-ethylmaleimide-sensitive factor attachment protein receptors (SNAREs) facilitate the eventual membrane fusion. Many of these processes are thermodynamically unfavourable and are tightly regulated.

4.1.2. The Endolysosomal Trafficking System

The endocytic and secretory pathways intersect at endosomes. Because of the constant exchange of materials, endosomes are highly heterogeneous population that is roughly divided into early, late, and recycling endosomes

according to their protein and lipid composition, morphology, and function. They form, together with the lysosomes, the endolysosomal system.[8]

The endolysosomal system is very important in nutrient uptake, but it also has a key role in the protein homeostasis in the cell. Actually hundreds of integral membrane proteins and their associated proteins (termed “cargo”) gather in endosomes after arriving from the plasma membrane and various other membrane trafficking routes[8–10]. Within endosomes, cargo can be derived into

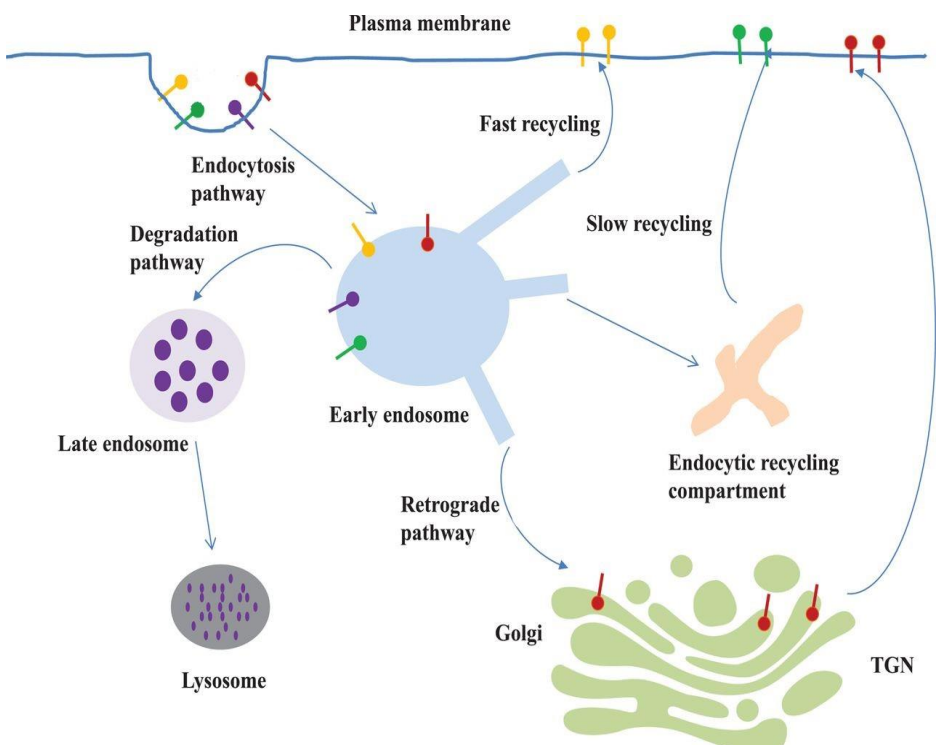

Figure 4.1 Schematic view of protein trafficking pathways.

two principal fates: degradation or recycling. In detail, proteins that continue into the degradation pathway are ubiquitinated and then recognised by the endosomal sorting complex (ESCRT). This is a multi-protein system that recognises ubiquitinated cargoes on the endosome membrane, invaginating the

endosomal membrane to form the intraluminal vesicle (ILV) for degradation during later phases[11]. On the contrary, cargoes that are derived into the recycling pathway are selected for enrichment in endosomal "retrieval" subdomains for being recycled back to the plasma membrane, the trans-Golgi network (TGN), or other specialized organelles[8,9]. There are many protein complexes involved in the cargo retrieval, among which the retromer, retriever, the CCC complex [12]. These are multi-protein complexes that, together with the WASP and SCAR homologue (WASH) and other accessory proteins, can recycle the cargoes and avoid lysosomal degradation [12,13].

4.1.3. Retromer

The retromer is one of the protein complexes that mediates cargo sorting/recycling from endosomes. Retromer malfunction has been linked to severe neurodegenerative diseases[2]. Additionally, retromer machinery is used by several intracellular pathogens. *Legionella Pneumophila*, *Chlamydia trachomatis* and the papillomavirus take advantage of this mechanism to replicate inside the cell, avoiding the immune system[14,15,16]. The complete structure of retromer has been solved by several laboratories, using X-ray crystallography,[17,18] cryo-electron microscopy[19] and cryo-electron tomography[20].

In mammals, retromer is a heterotrimeric protein composed of VPS35, VPS29 and VPS26 subunits, as showed in Figure 3. VPS35 (Figure 3, red) is a 92-kDa protein; its folding resembles that of other proteins involved in coated vesicle trafficking (e.g., AP and clathrin[18]). It is an elongated α-solenoid structure composed of 33 α-helixes grouped in 16 anti-parallel pairs, forming HEAT repeats. VPS35 wraps on its C-terminal around VPS29 (Figure 3, green), a 20-kDa protein with metallophosphoesterase fold. However, VPS29 has no

enzymatic activity, because the catalytic site is occluded by VPS35 and it lacks the critical histidine, which is substituted by a phenylalanine[18]. VPS29 interacts with many retromer effectors regulating the function and localisation of the complex. Finally, VPS26 (Figure 4.2., cyan) is a 38-kDa protein with arrestin-like fold consisting of two β-sandwich domains. Arrestins are a family of proteins important in signal transduction of G protein-coupled receptors and cargo trafficking[21].

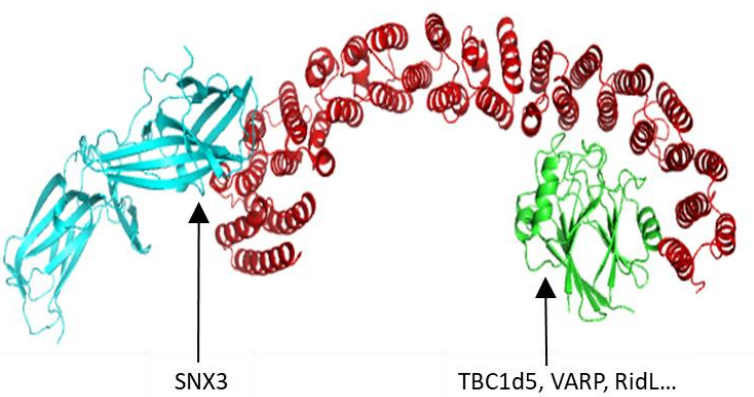

Figure 4.2. Structure of human retromer. VPS26 can be seen at its N-terminal (cyan), VPS35 in the central position (red) and VPS29 (green) at the C-terminal and partially buried in the VPS35 structure. Interaction sites are indicated by arrows.

The main function of retromer is the recycling of cargos back to the plasma membrane (direct recycling) and/or the retrograde transport of cargos to TGN (retrograde trafficking). To carry out this function, retromer needs accessory proteins, which aid in its recruitment to the endosomes and help in cargo recognition and membrane deformation during recycling. Among these accessory proteins can be listed numerous retromer effector proteins such as Rab7a, TBC1d5, VARP and FAM21/WASH complex. Undoubtedly, it is noteworthy to mention among them sorting nexins (SNX), the retromer accessory proteins implicated in membrane localisation of the retromer[22], membrane bending and tubulation[23] and even in the direct protein recycling[24].

The retromer heterotrimer is conserved from yeast to humans. However, despite a high degree of conservation, there are notable differences between yeast and human retromers. In yeast, the Vps26-Vps35-Vps29 heterotrimer, also known as the cargo-selection protein (CSC)[25], binds directly to two auxiliary proteins, Vps5 and Vps17[26]. Vps5 and Vps17 form heterodimers which favours tubule formation and recycling of cargos in the yeast[25]. Although the heteropentameric yeast complex is not stable *in vitro*[27], the structure of the yeast retromer complex (Vps26p–Vps29p–Vps35p) together with an homodimer of Vps5 on membrane tubules has been solved recently[20]. In contrast the human retromer has a larger number of accessory proteins in comparison with the yeast retromer. As such, the mammalian orthologs of Vps5-Vps17 belong to the sorting nexin (SNXs) family proteins, more specifically SNX1/SNX2 and SNX5/SNX6 respectively. In addition, their association is of more transient nature. This increased complexity suggests that human retromer has acquired more sophisticated mechanisms than its yeast counterpart. Indeed, the human retromer has been functionally linked to up to five SNX proteins (SNX1, SNX2, SNX5, SNX6 and SNX32. Furthermore, some of these SNXs can act independently of retromer in cargo sorting.

4.1.3.1. Human retromer and associated SNX proteins

Human sorting nexins (SNXs) are a group of 33 proteins, as shown in Figure 4.3.[28]. SNXs can be classified into three groups. The first group consist of SNXs that, apart from the common Phox (PX) domain, contain a Bin/amphiphysin/Rvs (BAR) domain (SNX-BAR, Figure 4.3., left column). The second group only contains a PX domain (SNX-PX, Figure 4.3., middle column). The SNXs in the last group contain other domains with a variety of functions (other SNXs, Figure 4.3., right column)[28]. The presence of a PX domain can favour the targeting of the protein to membranes via specific

phosphoinositide interactions[29]. On the other hand, thanks to the BAR-domain, SNXs can self-dimerise and sense/induce the membrane curvature[30].

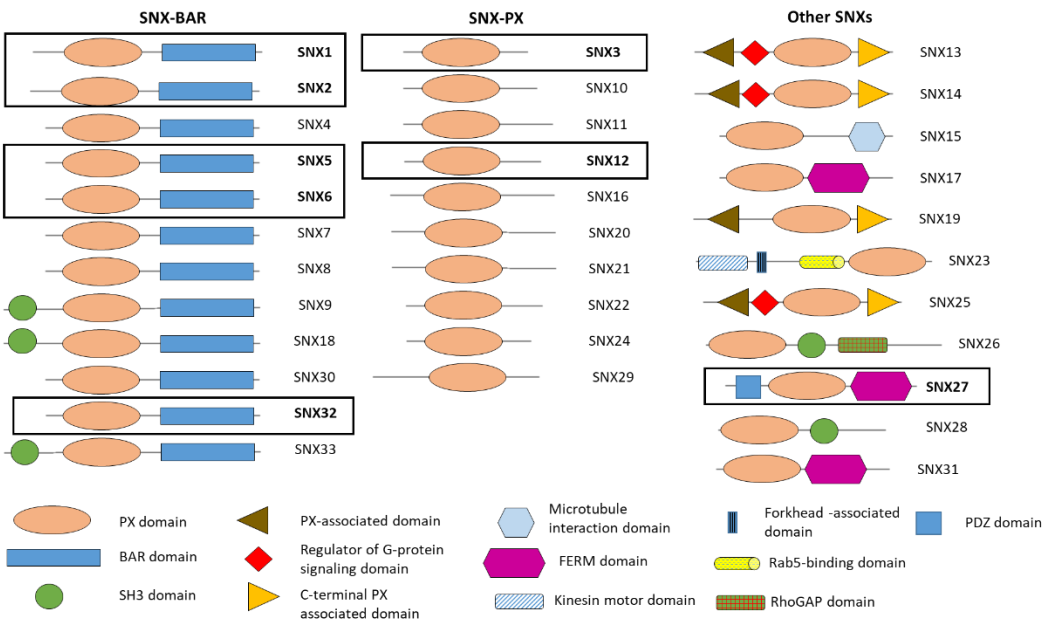

Figure 4.3. Schematic view of known sorting nexins (SNXs). SNXs associated with retromer pathways are shown in the boxes; SNX1, SNX2, SNX5, SNX6 and SNX32 belong to the SNX-BAR group (left column). SNX3 and SNX12 contain only a PX domain (middle column). SNX27 contains, apart from a PX domain, a PDZ domain and a FERM domain, to interact with cargos or other proteins.

Additionally, SNXs can recruit effectors to trigger cargo recycling[17, 31], create a tubular endosomal network (TEN) by membrane bending and tubulation[32] or even directly recycle some cargos[24, 33, 34]. The participation of the different SNXs in cargo recycling is schematically summarized in Figure 4.4.

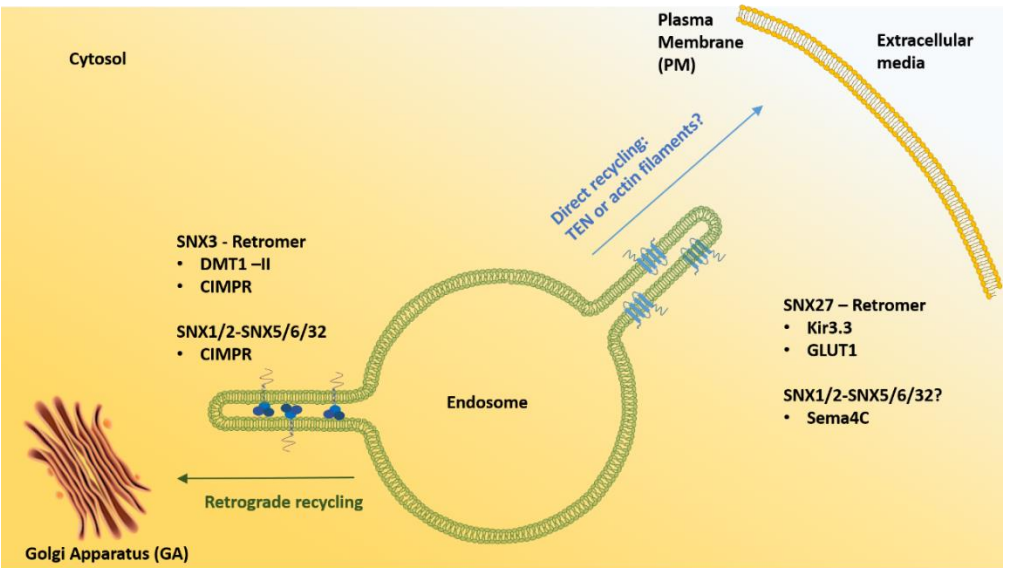

Figure 4.4. Summary of the participation of different SNXs in cargo recycling. SNX3 collaborates with the retromer in retrograde recycling of cargos such as DMT1-II[17]. SNXs-BAR (SNX1, SNX2, SNX5, SNX6 and SNX32) participate in retrograde recycling of cargos (such as CIMPR) to Golgi Apparatus (GA) [24, 35] although it has been associated with direct recycling of cargos like Sema4C[34]. SNX27 can recycle the cargos back to the PM in collaboration with the retromer[36]. Note that the dimensions are not to scale, and the mechanisms are simplified for clarity.

Only 8 SNXs have been associated with the retromer: SNX1, SNX2, SNX3, SNX5, SNX6, SNX12, SNX27 and SNX32[17, 37, 24, 38]. SNX1 and SNX2 contain PX and BAR domains. Although initially thought that these PX domains could interact with PI (3)P-containing membranes, it has been recently demonstrated that bind preferentially PI(3,4)P_2[39]. They can form homodimers between themselves and heterodimers with SNX5, SNX6 or SNX32. In addition, SNX1 and SNX2 have a large unstructured N-terminal region that of unknown function although it has been shown to interact with DENND5 (also called Rab6IP1) [40].

SNX5, SNX6 and SNX32 are the SNX-BAR proteins associated with retromer-dependent recycling pathways. In contrast to SNX1/SNX2 and Vps5,

it is not clear whether SNX5/SNX6/SNX32 are orthologues of the yeast Vps17[41]. They are expected to form homodimers and heterodimers with SNX1 and SNX2 [33, 42]. Although it seems clear that the heterodimers are the functional units of SNX-BAR proteins[24, 35, 42], there is little information about their formation, structure and biology. SNX1/SNX6 heterodimer was purified for the first time in 2018[43]; these complexes are now attracting growing attention[76, 77, 33].

SNX3 and the closely related SNX12 only contain a PX domain; they interact exclusively with PI(3)P[39]. It has been reported that the SNX3 is essential for the recycling of some known retromer cargos, such as DMT1-II[17]. The SNX12 is relatively abundant in neuronal tissues[44], and it has been associated with the recycling of CIMPR from the EE to TGN[45].

The SNX27 is directly linked with retromer function, interacting with retromer through VPS26 to recycle cargos from the endosomes to PM[46, 31]. At the N-terminus, it contains a PDZ domain, which can interact with other proteins and cargos[31]. Downstream from the PDZ domain, there is a PX domain, which directly interacts with PI(3)P-containing membranes[39]. Finally, a FERM domain is found at the C-terminus. It is a highly promiscuous protein–protein interaction domain[46].

To summarize, there are three distinct forms of retromer, SNX-BAR-retromer, SNX27-retromer, and the SNX3-retromer. While SNX27 is responsible for cargo recycling to the plasma membrane, SNX-BAR proteins help promoting the retrograde transport from endosomes to TGN. Besides, it is noteworthy to mention that SNX-BAR proteins can mediate retromer-independent transport of various cargos. [47]

4.1.4. Retromer-independent cargo transport by SNX-BAR domains

The hallmark of SNX-BAR proteins is their ability to bind to membranes and deform them producing the tubular endosomal network (TEN), where the cargo is transported to a specific target in tubulated membranes derived from the endosome[32]. Specifically, the interaction of the PX domain with phosphoinositides[39], and the electrostatic attraction of the negatively charged surface of BAR domains to the membrane[48] make possible the membrane recognition and binding by SNX-BAR proteins. Structurally, BAR domains are formed by coiled-coil α-helices Besides, SNX-BAR proteins can form homodimers or heterodimers (with other BAR domain-containing proteins)[37]. This dimerization allows the formation of a positively charged concave surface (named "banana shape") that can interact directly with negatively charged membranes[49]. After binding to the membrane, SNX-BAR proteins can induce curvature *de novo*. Several factors can affect membrane bending[48], among them it is noteworthy to mention the lipid composition of the membrane[50] and the presence of transmembrane proteins which can have a conical shape that helps in membrane bending, or clustering together, as they increase the membrane curvature.

In yeast, it has been demonstrated that the cargo selection complex (CSC) recruits cargo through Vps35p and Vps26p, while the Vps5 and Vps17 are responsible for the recruitment of the CSC to the endosomes due to the interaction of their PX domain with PIP-containing membanes[51]. In particular recent research has found that the recycling of Vps10 takes places as just described through a bipartite signal present in Vps10[52]. In contrast, mammal variants do not follow the same pattern, as mammal retromer does not form a stable complex with the equivalent SNX proteins.

Several studies have shown that the retrograde transport of the cation independent mannose-6-phosphate receptor (CIMPR) relies on the coordination of retromer. It is well established that numerous other non–retromer associated proteins (i.e. SNX proteins) also contribute to the retrograde trafficking of CIMPR through direct or indirect mechanisms.[53] The retromer-dependent retrograde cargo trafficking pathway of CIMPR requires SNX3 (see Figure 4.4.), whereas a couple of independent studies in 2017 demonstrated that the cation independent mannose-6-phosphate receptor (CIMPR), the human orthologue of Vps10, can be recycled by the SNX-BAR heterodimers of SNX1 or SNX2 with SNX5 or SNX6 in a retromer independent manner from the endosomes to TGN (retrograde recycling)[24,33].

Later on, in 2019 extensive knowledge of what they denominated as "endosomal SNX-BAR sorting complex for promoting exit – 1", or ESCPE-1 [34] was provided. This work showed that the cargo is directly recognised by the PX domain of SNX5 (and SNX6) and the recycling is driven by heterodimers without the intervention of the retromer[34]. Specifically, it was demonstrated that CIMPR contains a bipartite sequence in CIMPR, formed by $VSYKYSK_{2349\text{-}2455}$ and $WLM_{2369\text{-}2371}$. The bipartite sequence generates a dual interaction where VSYKYSK binds to the first alpha-helix following the proline-rich strand called α′ of SNX5, while WLM binds to the second alpha-helix, or α′′. Furthermore, other proteins that are recycled to the TGN e.g., Insulin-like growth factor 1 receptor (IGF1R) also directly bind to SNX-BAR proteins[54]. Using a SILAC-based proteomic approach, over 70 putative SNX-BAR cargos were identified. Therefore based upon these results, it has been proposed a model in which SNX-BARs function as a direct cargo selecting module for a large set of transmembrane proteins transiting the endosome[54,55].

4.2. Experimental Section

4.2.1. Production of Large Unilamellar Vescicles (LUVs)

Production of LUVs was carried out as reported elsewhere[39,22]. Lipids (Avanti Polar Lipids) were pipetted on Durham glass tubes, dried under nitrogen flow for 5-10 minutes, and then dried again in a vacuum chamber for 1 hour. Afterwards, buffer was added to induce Multi Lamellar Vesicle (MLV) hydration. Buffer for MLV was composed of, 100 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5 for electron microscopy samples, while 100 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5, 10% sucrose for co-flotation samples. MLVs were left hydrating for at least 30 minutes, with occasional vortex. After hydration, MLVs were transferred to a 1.5 ml Eppendorf, and freeze-thawed 10 times in liquid nitrogen and a water bath at 50ºC, respectively. Finally, MLVs were extruded 11 times through a 0.4 µm polycarbonate filter (for cryoEM reconstitutions), or 0.2 µm polycarbonate filter for flotation analysis. Liposomes were stored under Argon atmosphere in the fridge up to three days.

4.2.2. Production of Giant Unilamellar Vesicles (GUVs)

GUVs were prepared as described in literature[56]. Lipids (Avanti Polar Lipids) were mixed and dried under nitrogen flow for 5 to 10 minutes, and afterwards they were extensively dried in a desiccator chamber coupled with a vacuum pump for 1 hour. Then, lipid pellet was dissolved in 1 mM HEPES pH 7.5, and left for 30 minutes to 1 hour rehydrating, with occasional vortex. Once the lipids were well dissolved, a 2 µl drop was deposited on Parafilm. Silica beads were washed with 1 mM HEPES pH 7.5 at least 3 times by removing supernatant and spinning at 100g for 1 minute. Then, using a 10 µl tip silica beads were added to the drop of lipids deposited on the Teflon. Beads would

fall by gravity, without pipetting at all. The mixture of lipids + silica beads was dried for at least 30 minutes in a desiccator coupled with a vacuum pump. Meanwhile, a 10 µl tip was cut with a scalpel at narrow side and fresh solution of 1 M trehalose was prepared in 1 mM HEPES pH 7.5. A tip of a glass Pasteur pipette was burned, so a sharped thin and closed glass tip was prepared. Dried MLVs + silica beads were scratched with the glass pipette, and it was deposited very gently inside the cut 10 µl tip which contains the fresh 1 M trehalose solution, entering from the top of the cut tip. The tip, containing the mixture of MLVs + beads, was carefully deposited in a homemade humidity chamber (a falcon tube with milli-Q water, with a hole on the lid, so the tip fitted without falling inside), and incubated in a water bath at 60ºC for 10 minutes. The tip was then taken, and lipids, visible on the bottom of the tip, were dissolved in 100 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5 by just touching the buffer with the tip. Vesicles were left undisturbed for at least 1 hour, so the liposomes can hydrate further an increase its size. Then they were ready to be used for experiments.

4.2.3. Crosslinking of cysteine tagged proteins with Maleimide-functionalized Liposomes

Maleimide-functionalized liposomes were prepared using DSPE-Maleimide lipid (Avanti Polar Lipids) as describe in section 4.2.1. for floatation assay and section 4.2.2. for fluorescent confocal microscopy. In order to perform the maleimide-thiol coupling reaction, cargo-tagged proteins were added to liposomes in 3-fold molar excess over the % of maleimide exposed. We assume that maleimide is equally distributed between both sides, the inner and outer monolayers of the membrane, thus only 50% of total amount of maleimide is exposed on the liposome's surface. For 0% of maleimide sample, the same molar excess of 5% maleimide sample was used. The reaction was

performed overnight at 4ºC in 300 mM NaCl, 0.5 mM TCEP, 25 mM TRIS pH 7.2 buffer, considering the maleimide group reacts specifically with sulfhydryl groups when the pH range is between 6.5 and 7.5. TCEP reducing agents was chosen to exclude extra thiol-containing compounds competing with thiol-tagged cargo protein.

4.2.4. Flotation Assays

Analysis of the interaction of liposomes with proteins were studied using flotation analysis, adapted from previous works[57, 58, 59]. For clarity, a schematic representation of flotation is shown in Figure 4.6. In detail, liposomes were prepared at 1 mM and filtered through 0.2 µm cut-off filter (Whatman®) with the mini extruder, as described in section 4.2.1. Lipid composition used to test cargo immobilization on liposome surface by thiol-maleimide coupling is summarized in Table 4.1. On the contrary, to study the recruitment of cargo-tagged proteins by SNX1/SNX5 the liposome composition used was 45% DOPC, 29% DOPE, 20% DOPS, 1% Rh-PE, and 5% PI3P. SNX1/SNX5 at same molarity of exposed PI3P and 2 fold molar excess of cargo proteins were pre-incubated 15mins on ice, then they were added to liposomes.

Table 4.1. Lipid composition of samples for flotation assay to test cargo immobilization on liposome surface by thiol-maleimide coupling.

	DOPC%	DOPE%	DOPS%	PI3P%	Rh-PE%	DSPE-Mal%
Sample 1	45	29	20	5	1	0
Sample 2	45	24	20	5	1	5
Sample 3	40	24	20	5	1	10
Sample 4	38	18	18	5	1	20

After incubation of proteins with liposomes, the sucrose gradient was manually performed. First, the liposomes and proteins solution were gently mixed with 150 mM NaCl, 80% sucrose, 0.5 mM TCEP, 10 mM HEPES pH 7.5 in order to have a final solution with 30% of sucrose. Later, using a glass Pasteur pipette, a second layer of gradient (150 mM NaCl, 25% sucrose, 0.5 mM TCEP, 10 mM HEPES pH 7.5) was carefully, deposited over the solution at 30% sucrose. Finally, 150 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5 was carefully added over the second layer. Afterwards, samples were centrifuged at 240,000 g for 1 hour at 4ºC, using an Optima L-100 XP Ultracentrifuge (Beckman Coulter), in a SW 55Ti rotor. To run this sample, ultra-Clear^{TM} tubes of 5x41 mm were used (Beckman Coulter, ref. 344090), with adaptors for this rotor (ref. 356860, Beckman Coulter). After centrifugation, tubes were carefully taken out of the rotor and adapter, and the floating fractions of liposomes (seeing due to the presence of Rhodamine B) were taken using a 250 µl Hamilton syringe.

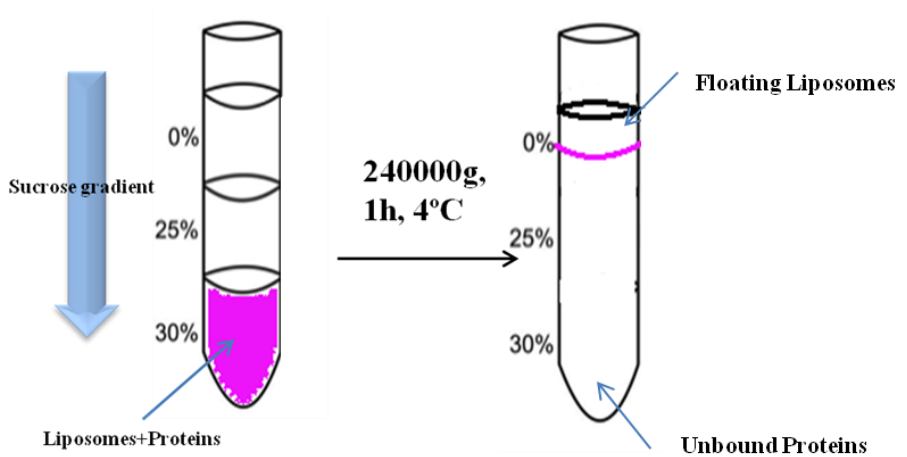

Figure 4.6. Schematic representation of flotation assay.

Samples were evaluated by SDS-PAGE. 20 µl of each sample it was mixed with 5 µl Laemmli buffer 5x, then boiled before running the gel. In order to load in SDS-PAGE the same number of liposomes for each sample,

absorbance at 555 nm was measured and samples were normalized against the sample with lower number of liposomes.

4.2.5. Fluorescence Confocal Microscopy

For fluorescence confocal microscopy GUVs were used, prepared according to the procedure described in section 4.2.2. The liposome composition used was 41% DOPC, 29% DOPE, 20% DOPS, 5% PI3P, 1% RhodamineB-PE, 4% Maleimide-DSPE, mixing lipids (Avanti Polar Lipids) to have a final concentration of 1 mg/ml (1.25 mM). It is noteworthy to mention that GUVS contain 1% of the fluorescent-labelled lipid Rhod-PE (Avanti Polar Lipids) in order to visualise the liposomes, and 4% of DSPE-Maleimide lipid (Avanti Polar Lipids) in order to immobilize cargo on liposome surface using thiol-maleimide coupling reactions (as described in section 4.2.3). Cargo-tagged protein immobilization and later tubule-vesicle genesis were observed using Zeiss LSM 510/ Zeiss LSM 800 Microscopies. Samples were prepared in μ-Slide 8-well plate treated with 0.1 mg/ml BSA (>98% purity, Sigma), washed with buffer and dried under nitrogen flow. Agarose patch low melting point was added to the samples at final concentration 1% (w/v). Cargo-tagged proteins with luminescent nanomaterials (AuNCs or QDs) were excited at 405 nm, whereas the Rhodamine B labelled-GUVs were excited at 560nm. The confocal pinhole was set to 1 AU to optimize z-sectioning. Images and Z-stacks were recorded sequentially using 63x-objective. Image processing were carried out using ImageJ (https://imagej.nih.gov/ij/).

4.2.6. Electron Microscopy

4.2.6.1. Scanning Transmission Electron Microscopy (STEM)

TEM samples were prepared by depositing 2 μL of the sample solution on the grid. Protein samples were deposited at nM protein concentration in low salt

buffer (10mM NaCl, 10mM Tris pH 7.4). The excess solution was air-dried at RT, ON. Before sample deposition, ultrathin carbon films on holey carbon support film, 400 mesh copper grids (TED PELLA INC.) were exposed to glow-discharge treatment. HAADF STEM images were acquired on a JEOL JEM-2100F UHR microscope at 200 kV in scanning mode, with a probe size of 1.5 nm and a choice of the camera length that ensures an inner detector angle of 75 mrad (HAADF) to enhance the contrasts of the metal atoms on the carbon support film.

4.2.6.2. Cryo Electron Microscopy (Cryo-EM)

Reconstitution of proteins on liposomes were evaluated on Quantifoil grids R2/2 Cu 300 mesh. Grids were hydrophilized by plasma using a Glow discharge/carbon evaporation MED 020 (BALTEC) and was then vitrified in liquid ethane in a FEI vitrobot^{TM} Mark III. Samples were stored into a homemade gridbox and kept in a dewar filled with liquid nitrogen. Samples were screened at CIC bioGUNE JEOL 2200FS cryo-electron microscope equipped with an UltraScan 4000 SP (4008×4008 pixels) cooled slow-scan CCD camera (GATAN, UK). Specifically, liposomes were prepared at a size of 0.4 μm, in, 100 mM NaCl, 0.5 mM TCEP, 10 mM HEPES pH 7.5. Liposome composition used was 45% DOPC, 30% DOPE, 20% DOPS and 5% PI (3) P. Screening of conditions for biogenesis of tubule-vesicles by SNX1/SNX5 in presence of cargo was performed by Carlos Lopez in CIC bioGUNE. SNX1/SNX5 at same molarity of exposed PI3P and 2x molar folder excess of cargo proteins were pre-incubated 15mins on ice, then they were added to liposomes. We assume that PI3P is equally distributed between both leaf sides of the membrane, thus only 50% of total amount of PI3P is exposed on the liposome's surface Liposomes + SNX1/SNX5 + cargo proteins were incubated overnight at 4°C after adding Argon to maintain inert atmosphere. The following day, sample was centrifuged at 20,000 xg for 30

minutes and 4 ºC. Supernatant was removed, and half of the volume of fresh buffer was added in order to re-suspend the sample, get rid of unbound protein, and concentrate liposomes. Samples were vitrified in specific conditions that favour thicker ice and favours the presence of more homogeneous ice. Vitrification was performed using the following parameters: 30 seconds incubating sample on the grid, 2 seconds blotting, 0 seconds waiting after drying, and offset of -2. The glow-discharge plasma treatment of grids was performed at 6 mA for 30 seconds.

4.3. Results and Discussion

The aim of this work is to test the effectiveness of the new protein-based nanotools developed in Chapter 3 to study retromer-independent cargo transport by SNX-BAR domains, specifically the recycling of cation-independent mannose-6-phosphate receptor (CIMPR) by SNX1/SNX5 from endosome to Golgi. For this purpose, functionalized liposomes mimicking endosomes and fusion cargo-tagged proteins can be used as *in vitro* model in physiological conditions (Figure 4.7.). As described in Chapter 3, fusion cargo-tagged proteins consist of the Cys-CIMPR$_{(2330\text{-}2491)}$-CTPR7$_{20Cys}$ protein stabilizing different metal nanoclusters and quantum dots. Specifically, it is noteworthy to underline that CIMPR$_{(2330\text{-}2491)}$ corresponds to the cytoplasmatic region of this type-I trans-membrane protein. Thanks to the versatility of the surface functionalization of liposomes, several strategies to study protein-membrane interactions in cargo sorting can be exploited.

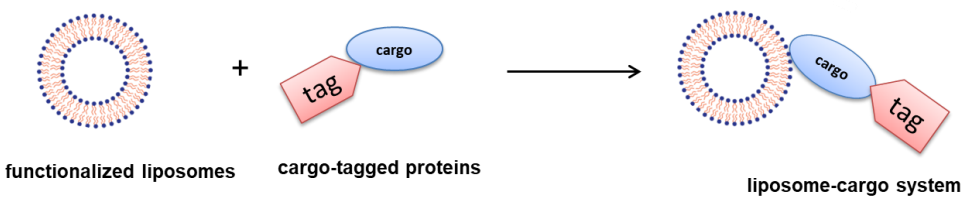

Figure 4.7. Endosome-cargo *in vitro* model

In this work, two strategies were used. The first approach was to immobilize the model cargo-tagged protein on the surface of the liposome using thiol-maleimide coupling according standard protocols[60], since liposomes are maleimide functionalized and cargo-tagged proteins have a free cysteine at N-terminus. The second approach was based on the idea that sequence recognition is an important component in the biogenesis of tubular profiles and transport carriers. To this end, cargo-tagged proteins can be pre-incubated with SNX1/SNX5 heterodimer, because cargo is directly recognised by the PX domain of SNX5 through the bipartite sequence $VSYKYSK_{2349\text{-}2455}$ and $WLM_{2369\text{-}2371}$[34]. Once the complex cargo-SNX heterodimer is formed, it can be added to liposomes enriched with Phosphatidylinositol 3-phosphate (PI3P) lipid, as SNX1/SNX5 heterodimer is recruited to endosomes via its PI3P-binding PX domain and its curvature-sensing BAR domain[16]. Moreover, these two different approaches to cargo immobilization on the liposome's surface, provided the possibility to have a more versatile *in vitro* model system. Indeed, the immobilization by thiol-maleimide coupling could be used to tune the concentration of the cargo on the membrane and to study how SNX1/SNX5 interacts with membrane at different cargo concentrations. On the contrary, the sequence recognition approach by SNX1/SNX5 could be used for morphological studies, specifically to study how SNX1/SNX5 and cargo-tagged proteins arrange around the tubular-vesicular carriers.

Flotation assays were performed in order to test the effectiveness of the immobilization approaches. Liposomes were prepared as described in section

4.2.4. The coupling reaction between maleimide functionalized-liposomes and cargo-tagged proteins were performed as described in section 4.2.3. Specifically, Cys-CIMPR-mTurquoise2 was used for the experiment. As shown in Figure 4.8., cargo-tagged proteins immobilization on liposome's surface approach using thiol-maleimide coupling worked properly. Increasing the % of maleimide in functionalized-liposomes, the highest amount of thiol-tagged cargo proteins was found in bounded fractions..

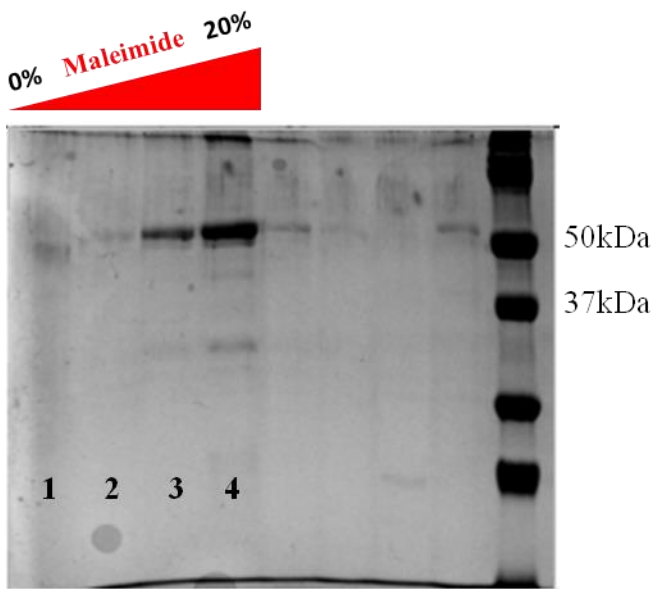

Figure 4.8. SDS-PAGE gel showing the higher amount of protein (Cys-CIMPR-mTurquoise2) immobilized on liposome surface increasing the %of maleimide in functionalized liposomes. Specifically, floating liposome fractions of Cys-mTurquoise2 + LUV 0% maleimide (lane 1), Cys-CIMPR-mTurquoise2 + LUV 5% maleimide (lane 2), Cys-CIMPR-mTurquoise2 + LUV 10% maleimide (lane 3), and Cys-CIMPR-mTurquoise2 + LUV 20% maleimide (lane 4) are shown.

SDS-PAGE gel in Figure 4.9. shows the study of the interactions between cargo proteins, cargo-tagged proteins and SNX1/SNX5 in presence of PI3P-enriched liposomes. The samples were prepared as described in section 4.2.4. Specifically, the heterodimer protein SNX1/SNX5 (MW 59 kDa+47 kDa) was tested with several cargo proteins: Cys-CIMPR-$C7_{20Cys}$ (48 kDa), Cys-

CIMPR-mTurquoise2 (46 kDa) and Cys-CIMPR-5H (18 kDa). Additionally, PI3P-enriched liposomes were tested with only SNX1/SNX5 (lane 1) and with only cargo proteins Cys-CIMPR-$C7_{20Cys}$ (lane 2), Cys-CIMPR-5H (lane 5) and Cys-CIMPR-Turquoise2 (lane 7). In case of SNX1/SNX5+Cys-CIMPR-$C7_{20Cys}$ (lane 3) and SNX1/SNX5+Cys-CIMPR-5H (lane 4) samples, it was possible to find both proteins in bounded fractions, that means the recruitment of these cargo proteins by sequence recognition was an effective strategy. As expected, Cys-CIMPR-5H (lane 6) and Cys-CIMPR-$C7_{20Cys}$ (lane 2) cargoes are not recruited by liposome in absence of SNX1/SNX5. On the contrary, for the sample Cys-CIMPR-mTurquoise2, it was not possible to obtain a clear result. It is not possible to find clearly Cys-CIMPR-mTurquoise2 in the bounded fraction (lane 5) in presence of SNX1/SNX5, whereas Cys-CIMPR-mTurquoise2 was recruited to liposomes in absence of SNX1/SNX5 (lane 7).

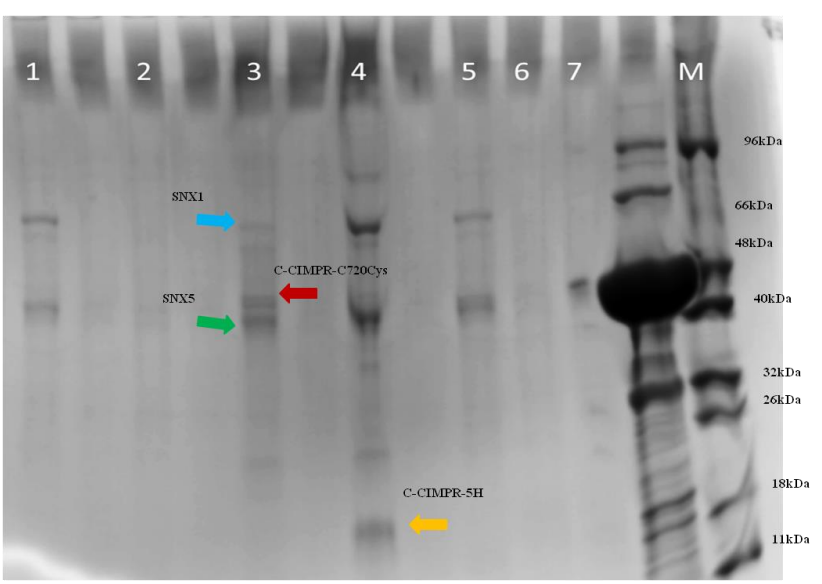

Figure 4.9. SDS-PAGE gel showing liposome flotation assay for studying interactions between SNX1/SNX5 with different construct of CIMPR in presence of PI3P-enriched liposomes. Floating liposome fractions (bounded fractions) of following proteins are shown in the following lanes: (1) LUV+SNX1/SNX5 (2) LUV+Cys-CIMPR-$C7_{20Cys}$ (3)

LUV+SNX1/SNX5+Cys-CIMPR-$C7_{20Cys}$ (4) LUV+SNX1/SNX5+Cys-CIMPR-5H (5) LUV+SNX1/SNX5+Cys-CIMPR-mTurquoise2 (6) LUV+Cys-CIMPR-5H (7) LUV+ Cys-CIMPR-mTurquoise2 (M) Low Molecular Weight (LMW) Protein Marker (Nzytech). The molecular weight for SNX1/SNX5, Cys-CIMPR-$C7_{20Cys}$, and Cys-CIMPR-5H, Cys-CIMPR-mTurquoise2 is respectively 59kDa, 47kDa, 48kDa, 19kDa, and 46kDa.

In order to determine if Cys-CIMPR-mTurquoise2 was recruited by SNX1/SNX5 to the liposomes, floating fractions were also measured by fluorescence microscopy. As shown in figure 4.10., by monitoring the fluorescence signal of the cargo protein, Cys-CIMPR-mTurquoise2 was clearly recruited to liposomes in presence of SNX1/SNX5 (blue line), while in absence of SNX1/SNX5 only a residual amount ofCys-CIMPR-mTurquoise2 (red line) was found, probably due unspecific interaction with liposomes.

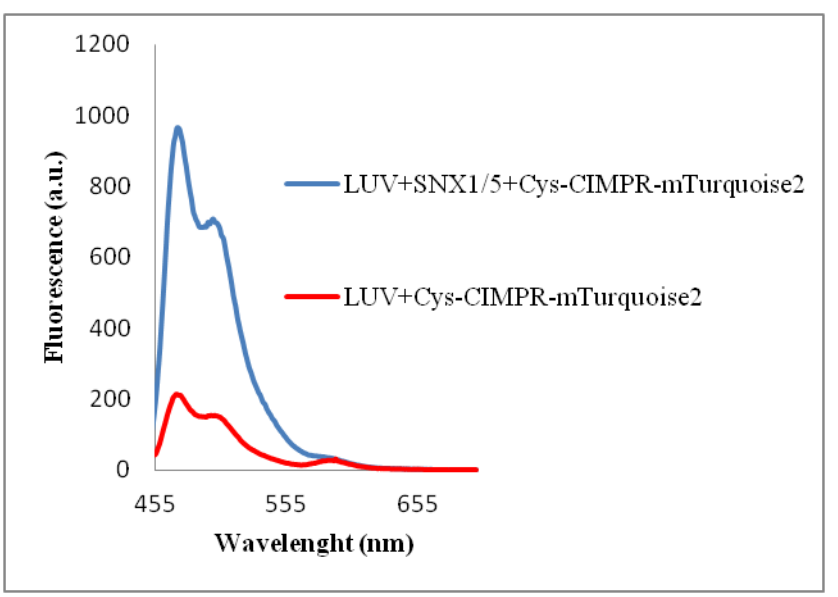

Figure 4.10. Fluorescent spectra showing PI3P-enriched liposome flotation assay for studying interactions between SNX1/SNX5 with different construct Cys-CIMPR-mTurquoise2 in presence (blue line) and in absence (red line) of SNX1/SNX5. Fluorescence spectra were recorded at an excitation wavelength of 434 nm.

Flotation assays were performed also for cargo-tagged proteins stabilizing AuNCs/AuNKNCs and/or CdS-QDs. Unfortunately, it was not possible to

visualise them by SDS-PAGE. Therefore, the same approach used in case of cargo-mTurquoise2 by fluorescence spectroscopy was applied to these samples. Thus, the floating fraction of cargo-stabilized CdS QDs (1) (Figure 4.11.(a)) and cargo-stabilized AuNCs (1) (Figure 4.11.(b)) were analysed. In both cases, in presence of SNX1/SNX5 the cargo-tagged protein was recruited to membrane, as proved from the characteristic emission profile of CdS QDs (1) at 478 nm and AuNCs (1) at 438 nm when they are excited at 370 nm. The higher signal at 595 nm is related to the rhodamine fluorophore of liposomes.

To conclude we have demonstrated that it is possible to immobilize the cargo-tagged protein and, in addition control the amount of cargo present on the liposome's surface. Additionally, the cargo-tagged proteins could be recruited to the membrane only in presence of SNX1/SNX5, which indicates the specificity of the recruitment and that the fusion to the novel tags does not modify the biological function of the cargo.

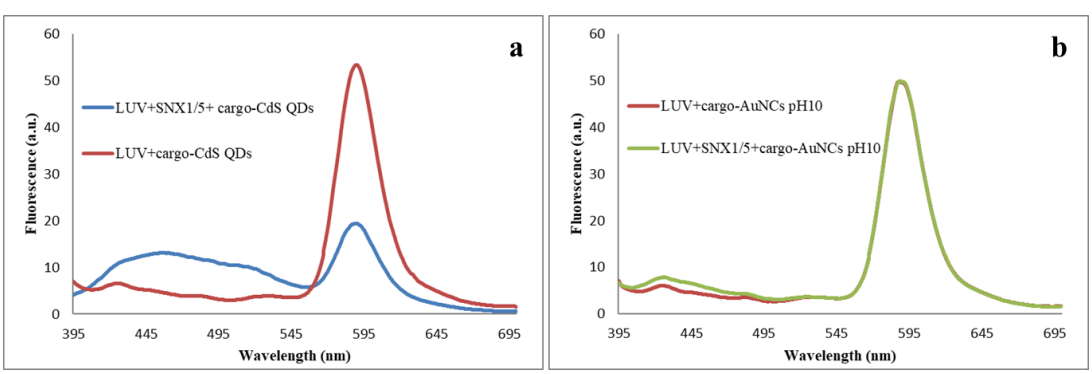

Figure 4.11. Fluorescent characterization of floating fractions of (a) cargo-CdS QDs (1) in presence (blue line) and in absence (red line) of SNX1/SNX5 and (b) cargo-AuNCs (1) in presence (red line) and in absence (green line). Fluorescence spectra were both recorded at excitation wavelength of 370nm.

Since it was possible to successfully immobilize cargo-tagged proteins on liposome surface using thiol-maleimide coupling reaction or sequence recognization approach, the study of the completed *in vitro* system were

performed by microscopy. Samples for cryoEM studies were prepared as described in section 4.2.6.2. Specifically, liposome solution was incubated with a mix solution containing SNX1/SNX5 and different cargo-tagged proteins stabilizing metal clusters (data shown Figure 4.12. are recorded using cargo-CdS QDs (1) (Figure 4.12. (a)), cargo-CdS QDs (2) (Figure 4.12. (b), and cargo-AuNKNCs (2) (Figure 4.12. (c)). Additionally, control-experiments were performed incubating the liposomes only with cargo-tagged proteins (cargo-CdS QDs (1) (Figure 4.12. (d)), cargo-CdS QDs (2) (Figure 4.12. (e), and cargo-AuNKNCs (2) (Figure 4.12. (f)). The model system was not perturbed in presence cargo-CdS QDs (2) and cargo-AuNKNCs, since the tubule-vesicular carriers are formed, and higher electron-density is found on the tube. The same results were obtained using cargo-AuNCs (1) (data not shown). On the contrary, cargo-CdS QDs (1) was clearly be found on the membrane in presence of SNX1/SNX5, but there was not formation of tubule vesicles. This effect could be due to the larger size of the metal clusters in case of cargo-CdS QDs (1), which might be too big for the natural arrangement of the cargo-SNX1-SNX5 coat around the tubule-vesicles. As expected in the control samples (Figure 4.12.(d-f)) there was no recruitment of cargo to the membrane and formation of tubule-vesicles in absence of SNX1-SNX5.

In order to confirm the presence of cargo-CdS QDs (2) and cargo-AuNKNCs on the surface of membrane, the same samples used in cryo-EM were analysed by STEM, as described in section 4.2.6.1. As shown in the Figure 4.13., cargo-AuNKNCs in presence of SNX1/SNX5 was recruited to the membrane. Cargo-AuNKNCs was more concentrated in a specific part of the membrane and followed a pair-wise pattern in some part.

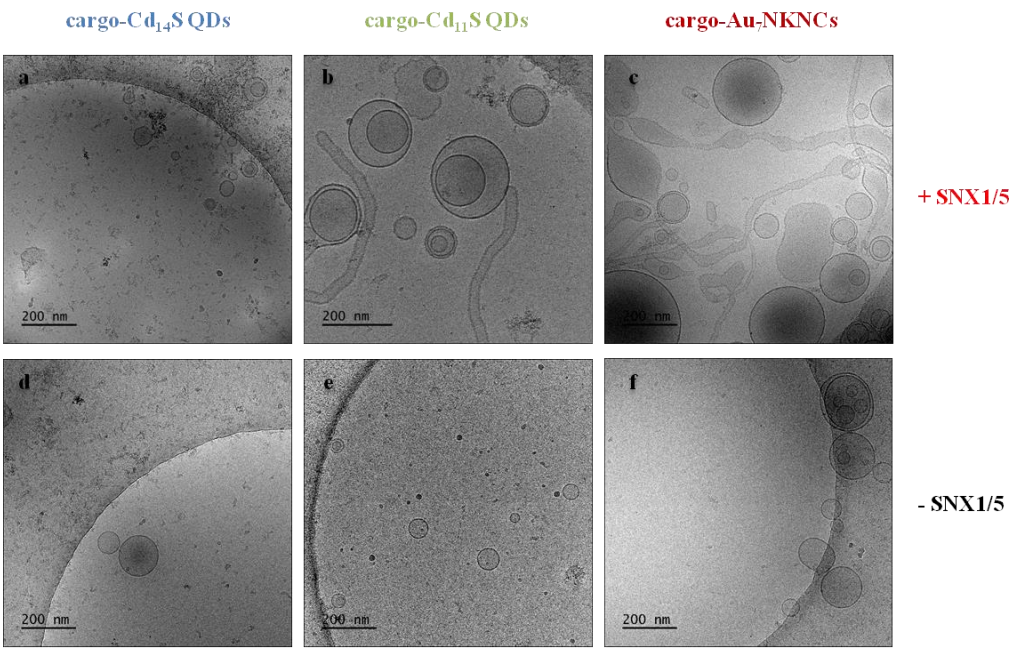

Figure 4.12. Cryo-EM image of recruitment of cargo to the membrane and formation of tubule-vesicular carriers using cargo-tagged proteins and liposomes in presence of SNX1/SNX5 (a-c) and control sample using cargo-tagged proteins and liposomes in absence of SNX1/SNX5 (d-f). (a) LUV+ SNX1/SNX5+ cargo-CdS QDs (1), (b) LUV+ SNX1/SNX5+ cargo-CdS QDs (2) (c) LUV+ SNX1/SNX5+cargo-AuNKNCs (2) (d) (LUV+ cargo-CdS QDs (1), (e) LUV+cargo-CdS QDs (2), (f) LUV+cargo-AuNKNCs (2). Scale bar: 200nm,

It was not clear to determine if there is formation of tubule-vesicular carriers. Indeed, in HAADF STEM samples were simply deposited on the grid, the liposome could be deformed (not allowing a proper morphological study of the sample) or/and destroyed, as shown in case of cargo-CdS QDs (2) in which the shape of liposomes could not be observed at all (Figure 4.14.).

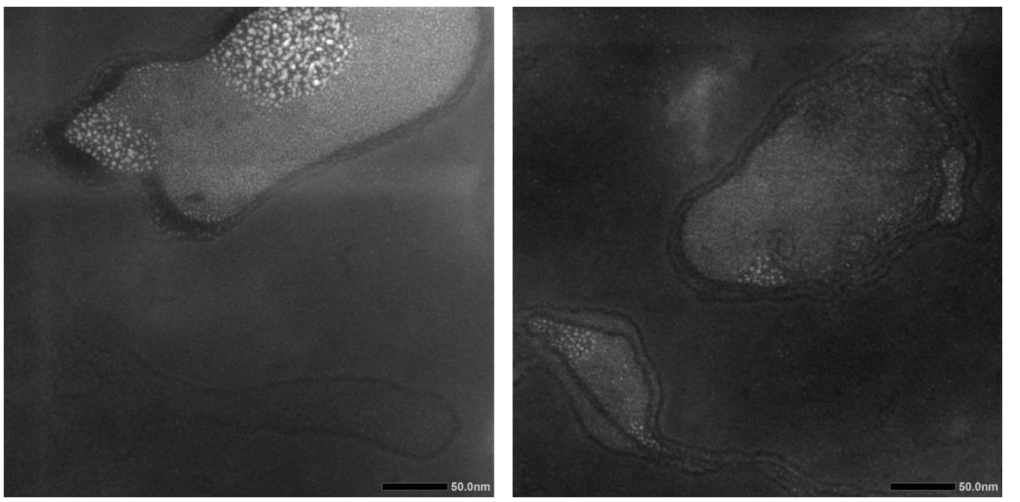

Figure 4.13. HAADF STEM Images of recruitment of cargo-AuNKNCs to the membrane in presence of SNX1/SNX5. The images were obtained mapping Au in the sample..Scale bar: 50nm.

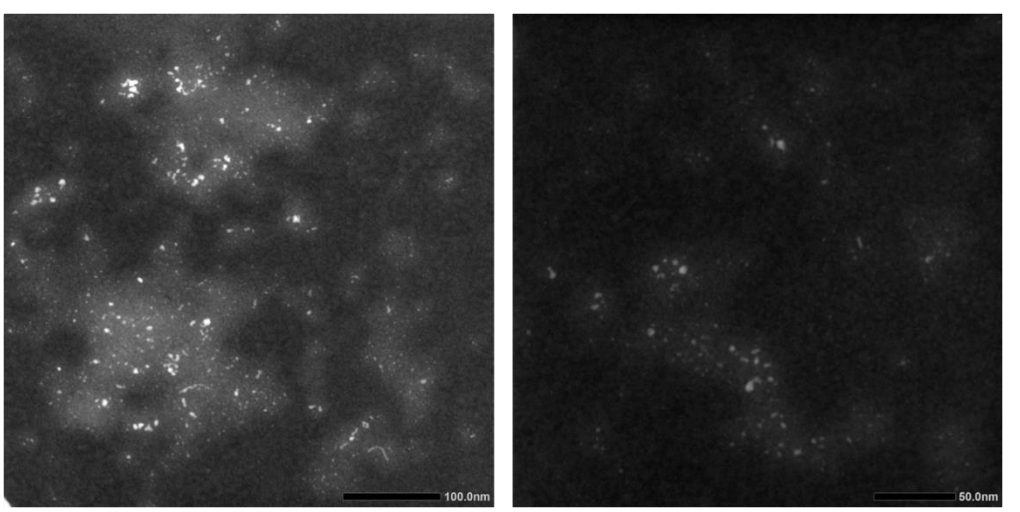

Figure 4.14. HAADF-STEM Images of recruitment of SNX5cargo-CdS QDs (2) to the membrane in presence of SNX1/SNX5. The images were obtained mapping Cd in the sample . Scale bars 100nm.

Regarding the confocal fluorescence microscopy, preliminary studies were performed using Cys-CIMPR-mTurquoise2 in order to test the validity of the model. The immobilization of cargo-tagged proteins was performed using thiol-maleimide coupling strategy. Indeed, we wanted to design a *in vitro*

model system that allow us to determine if the formation of tubule-vesicular carriers is depending of the concentration of cargo. For the experiments Giant Unilamellar Vesicles (GUVs) were used and samples were prepared as described in section 4.2.5.. As shown in Figure 4.15., we could reproduce the formation of tubule-vesicular carriers and we could detect the cargo signal on liposome surface using Cys-CIMPR-mTurquoise2.

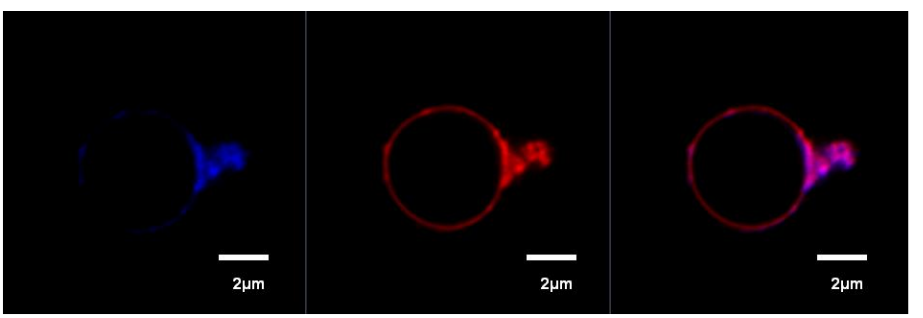

Figure 4.15. Confocal fluorescence microscopy image of the formation of tubule-vesicular carriers using GUVS where Cys-CIMPR-mTurquoise2 was pre-immobilized on liposome surface. Red signal comes from the excitation at 555 nm of RhodamineB-enriched GUVs, blue signal from excitation at 405 nm of Cys-CIMPR-mTurquoise2. Scale bar: 2 µm.

Besides, it was possible to zoom to a single tubule-vesicular carrier. Cys-CIMPR-mTurquoise2 was detected around the tube (Figure 4.16.).

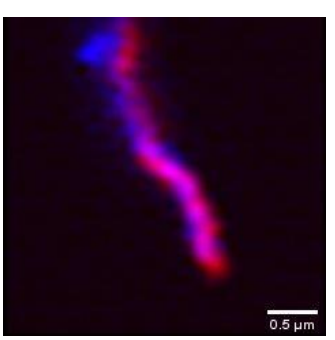

Figure 4.16. Helical arrangement of Cys-CIMPR-mTurquoise2around the tubule-vesicular carriers. Red signal comes from the excitation at 555 nm of RhodamineB-enriched GUVs, blue signal from excitation at 405 nm of Cys-CIMPR-mTurquoise2. Scale bar: 0.5 µm.

The same experiment was repeated using cargo-AuNCs(1). As shown in Figure 4.17., tubule-vesicular carriers were formed and cargo-AuNCs(1) was detected on liposome surface. However the signal arising from cargo-AuNCs(1) was too low to be detected on tubes surface. Increasing the maleimide percentage and consequently the cargo-AuNCs(1) immobilized would provide stronger signal, which it could be detected on surface tubes.

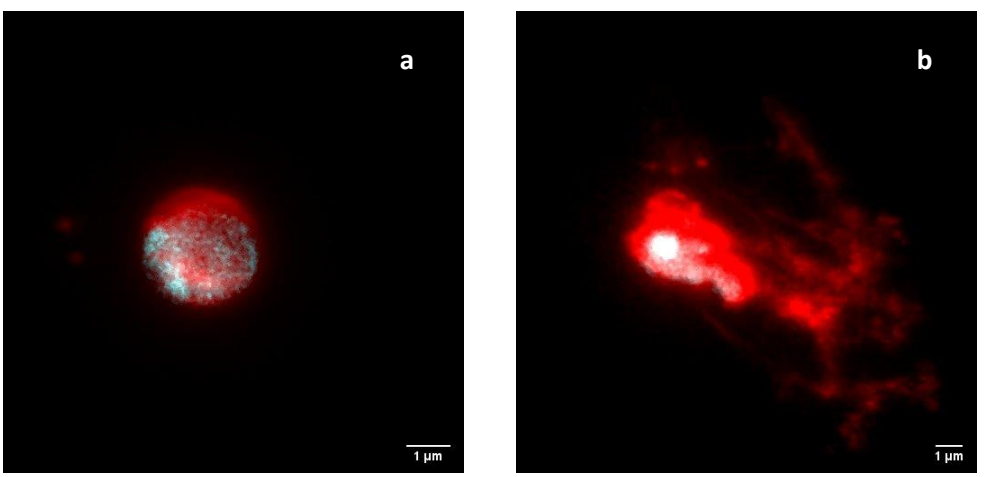

Figure 4.17. (a) Confocal fluorescence microscopy image protein-AuNCs (1) pre-immobilized on liposome surface (b) Confocal fluorescence microscopy image of the formation of tubule-vesicular carriers adding SNX1/SNX5 after the pre-immobilized on liposome surface of protein-AuNCs (1). Red signal comes from the excitation at 555nm of RhodamineB-enriched GUVs, white signal from excitation at 405 nm of cargo-AuNCs (2). Scale bar: 1 µm.

The same results were obtained cargo-CdS QDs (2) and cargo-AuNKNCs (data not showed). Additionally, the same experiments was performed using cargo-AuNCs (2) (Figure 4.18.), but in this case the formation of tubule vesicular carriers was not detected. The liposome surface was deformed. As shown in case of cargo-CdS QDs (1) by cryoEM (Figure 4.12.(a)), when the size of the metal clusters is too big, there is no formation of tubule-vesicular carriers. (Figure 4.18.).

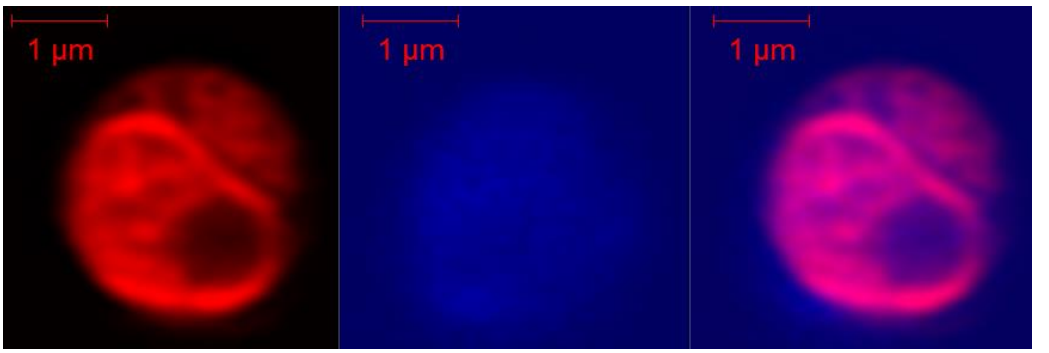

Figure 4.18. (a) Confocal fluorescence microscopy image of the formation of tubule-vesicular carriers adding SNX1/SNX5 after the pre-immobilized on liposome surface of cargo-AuNCs (2). Red signal comes from the excitation at 555nm of RhodamineB-enriched GUVs, blue signal from excitation at 405 nm of cargo-AuNCs (2). Scale bar: 1 µm.

4.4. Conclusions

Our results show that some of cargo-tagged proteins developed in this work are potentially good candidates for studying the retromer-independent recycling of CIMPR *in vitro* systems by fluorescent and electron microscopy. Specifically, all cargo-tagged proteins can be recruited to the membrane as shown in floatation assays. Thus, the direct interaction between SNX1/SNX5 and CIMPR is not perturbed. Tubule-vesicular carriers are formed when cargo-CdS QDs (2), cargo-AuNCs (1) and cargo-AuNKNCs are used. Therefore, cargo-CdS QDs (2), cargo-AuNCs (1) and cargo-AuNKNCs could be exploited for future works to elucidate structural info using high resolution cryo-electron microscopy (HR cryo-ET) or fluorescent microscopy to track different recycling pathways.

4.5. Bibliography

1. Ballabio, A. The awesome lysosome. *EMBO Mol Med* **8**, 73–76 (2016).

2. Petsko, S. A. S. and G. A. Retromer in Alzheimer disease, Parkinson disease and other neurological disorders. *Occup. Environ. Heal.* **16**, 126–132 (2015).

3. Vellai, T. & Vida, G. The origin of eukaryotes: The difference between prokaryotic and eukaryotic cells. *Proc. R. Soc. B Biol. Sci.* **266**, 1571–1577 (1999).

4. Chen S. Valm, A. Lippincot-Schwartz, J. Interacting Organelles. *Curr Opin Cell Biol.* **53**, 84–91 (2018).

5. Helle, S. C. J. *et al.* Organization and function of membrane contact sites. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833**, 2526–2541 (2013).

6. Saheki, Y. & Camilli, P. De. Endoplasmic Reticulum – Plasma Membrane Contact Sites. *Annu. Rev. Biochem.* **86**, 659–684 (2017).

7. Vassilieva, E. V. & Nusrat, A. Vesicular trafficking: Molecular tools and targets. *Methods Mol. Biol.* **440**, 3–14 (2008).

8. Hierro, A., Gershlick, D. C., Rojas, A. L. & Bonifacino, J. S. *Formation of Tubulovesicular Carriers from Endosomes and Their Fusion to the trans-Golgi Network*. *International Review of Cell and Molecular Biology* vol. 318 (Elsevier Ltd, 2015).

9. Burd, C. & Cullen, P. J. Retromer: A master conductor of endosome sorting. *Cold Spring Harb. Perspect. Biol.* **6**, (2014).

10. Huotari, J. & Helenius, A. Endosome maturation. *EMBO J.* **30**, 3481–3500 (2011).

11. Hu, Y., Dammer, E. B., Ren, R. & Wang, G. The endosomal-lysosomal system : from acidification and cargo sorting to neurodegeneration. *Transl. Neurodegener.* 1–10 (2015) doi:10.1186/s40035-015-0041-1.

12. McNally, K. E. & Cullen, P. J. Endosomal Retrieval of Cargo: Retromer Is Not Alone. *Trends Cell Biol.* **28**, 807–822 (2018).

13. Derivery, E., Helfer, E., Henriot, V. & Gautreau, A. Actin polymerization controls the organization of WASH domains at the surface of endosomes. *PLoS One* **7**, (2012).

14. Romano-Moreno, M. *et al.* Molecular mechanism for the subversion of the retromer coat by the Legionella effector RidL. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E11151–E11160 (2017).

15. Elwell, C. A. *et al.* Chlamydia interfere with an interaction between the mannose-6-phosphate receptor and sorting nexins to counteract host restriction. *Elife* **6**, 1–17 (2017).

16. Popa, A. *et al.* Direct Binding of Retromer to Human Papillomavirus Type 16 Minor Capsid Protein L2 Mediates Endosome Exit during Viral Infection. *PLoS Pathog.* **11**, 1–21 (2015).

17. María Lucas, David C. Gershlick, Ander Vidaurrazaga, Adriana L. Rojas, Juan S. Bonifacino, and A. H. Structural Mechanism for Cargo Recognition by the Retromer Complex. *Cell* **167**, 1623–1635 (2016).

18. Hierro, A. *et al.* Functional architecture of the retromer cargo-recognition complex. *Nature* **449**, 1063–1067 (2007).

19. Amy K. Kendall, Boyang Xie, Peng Xu, Jue Wang, Rodger Burcham, Meredith N. Frazier, Elad Binshtein, Hui Wei, Todd R. Graham, Terunaga Nakagawa, and L. P. J. Mammalian retromer is an adaptable scaffold for cargo sorting from endosomes. *bioRxiv* **53**, 1689–1699

(2019).

20. Kovtun, O. *et al.* Structure of the membrane-assembled retromer coat determined by cryo-electron tomography. *Nature* **561**, 561–564 (2018).

21. Moore, C. A. C., Milano, S. K. & Benovic, J. L. Regulation of Receptor Trafficking by GRKs and Arrestins. *Annu. Rev. Physiol.* **69**, 451–482 (2007).

22. Harrison, M. S. *et al.* A mechanism for retromer endosomal coat complex assembly with cargo. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 267–272 (2014).

23. Van Weering, J. R. T. *et al.* Molecular basis for SNX-BAR-mediated assembly of distinct endosomal sorting tubules. *EMBO J.* **31**, 4466–4483 (2012).

24. Simonetti, B., Danson, C. M., Heesom, K. J. & Cullen, P. J. Sequence-dependent cargo recognition by SNX-BARs mediates retromer-independent transport of CI-MPR. *J. Cell Biol.* **216**, 3695–3712 (2017).

25. Purushothaman, L. K. & Ungermann, C. Cargo induces retromer-mediated membrane remodeling on membranes. *Mol. Biol. Cell* **29**, 2709–2719 (2018).

26. Seaman, M. N. J., McCaffery, J. M. & Emr, S. D. A membrane coat complex essential for endosome-to-Golgi retrograde transport in yeast. *J. Cell Biol.* **142**, 665–681 (1998).

27. Purushothaman, L. K., Arlt, H., Kuhlee, A., Raunser, S. & Ungermann, C. Retromer-driven membrane tubulation separates endosomal recycling from Rab7/Ypt7-dependent fusion. *Mol. Biol. Cell* **28**, 783–791 (2017).

28. Cullen, P. J. Endosomal sorting and signalling: An emerging role for sorting nexins. *Nat. Rev. Mol. Cell Biol.* **9**, 574–582 (2008).

29. Zhan, Y., Virbasius, J. V., Song, X., Pomerleau, D. P. & Wayne Zhou, G. The p40phox and p47phox PX domains of NADPH oxidase target cell membranes via direct and indirect recruitment by phosphoinositides. *J. Biol. Chem.* **277**, 4512–4518 (2002).

30. Adam Frost, Rushika Perera, Aurélien Roux, Krasimir Spasov, O. & Destaing, Edward H. Egelman, Pietro De Camilli, and V. M. U. Structural Basis of Membrane Invagination by F-BAR Domains. *Cell* **132**, 807–817 (2008).

31. Gallon, M. *et al.* A unique PDZ domain and arrestin-like fold interaction reveals mechanistic details of endocytic recycling by SNX27-retromer. *Proc. Natl. Acad. Sci. U. S. A.* **111**, (2014).

32. J. Chi, R. Harrison, M. S. & Burd, C. G. Biogenesis of endosome-derived transport carriers. *Cell Mol Life Sci.* **72**, 3441–3455 (2015).

33. Kvainickas, A. *et al.* Cargo-selective SNX-BAR proteins mediate retromer trimer independent retrograde transport. *J. Cell Biol.* **216**, 3677–3693 (2017).

34. Simonetti, B. *et al.* Molecular identification of a BAR domain-containing coat complex for endosomal recycling of transmembrane proteins. *Nat. Cell Biol.* **21**, 1219–1233 (2019).

35. Kvainickas, A. *et al.* Retromer- and WASH-dependent sorting of nutrient transporters requires a multivalent interaction network with ANKRD50. *J. Cell Sci.* **130**, 382–395 (2017).

36. Gallon, M. *et al.* A unique PDZ domain and arrestin-like fold interaction reveals mechanistic details of endocytic recycling by

SNX27-retromer. *PNAS* **111**, 3604–3613 (2014).

37. Wassmer, T. *et al.* A loss-of-function screen reveals SNX5 and SNX6 as potential components of the mammalian retromer. *J. Cell Sci.* **120**, 45–54 (2007).

38. Steinberg, F. *et al.* A global analysis of SNX27 – retromer assembly and cargo specificity reveals a function in glucose and metal ion transport relies on a membrane-curvature-sensing domain in Atg3. *Nat. Publ. Gr.* **16**, 821 (2014).

39. Chandra, M. *et al.* Classification of the human phox homology (PX) domains based on their phosphoinositide binding specificities. *Nat. Commun.* **10**, (2019).

40. Fernandes, H., Franklin, E., Jollivet, F., Bliedtner, K. & Khan, A. R. Mapping the interactions between a RUN domain from DENND5/Rab6IP1 and sorting nexin 1. *PLoS One* **7**, 1–9 (2012).

41. Bonifacino, J. S. & Rojas, R. Retrograde transport from endosomes to the trans-Golgi network. *Nat. Rev. Mol. Cell Biol.* **7**, 568–579 (2006).

42. Wassmer, T. *et al.* The Retromer Coat Complex Coordinates Endosomal Sorting and Dynein-Mediated Transport, with Carrier Recognition by the trans-Golgi Network. *Dev. Cell* **17**, 110–122 (2009).

43. Yong, X. *et al.* Expression and purification of the SNX1/SNX6 complex. *Protein Expr. Purif.* **151**, 93–98 (2018).

44. Mizutani, R. *et al.* Expression of sorting nexin 12 is regulated in developing cerebral cortical neurons. *J. Neurosci. Res.* **90**, 721–731 (2012).

45. Priya, A. *et al.* Essential and selective role of SNX12 in transport of

endocytic and retrograde cargo. *J. Cell Sci.* **130**, 2707–2721 (2017).

46. Ghai, R. *et al.* Structural basis for endosomal trafficking of diverse transmembrane cargos by PX-FERM proteins. *Proc. Natl. Acad. Sci. U. S. A.* **110**, (2013).

47. Hanley, S. E. & Cooper, K. F. Sorting Nexins in Protein Homeostasis. *Cells* **10**, 17 (2020).

48. Gallop, H. T. M. J. L. Membrane curvature and mechanisms of dynamic cell membrane remodelling. *Nature* **438**, 590–596 (2005).

49. Frost, A., Unger, V. M. & De Camilli, P. The BAR Domain Superfamily: Membrane-Molding Macromolecules. *Cell* **137**, 191–196 (2009).

50. Kooijman, E. E. *et al.* Spontaneous Curvature of Phosphatidic Acid and Lysophosphatidic Acid. *Biochemistry* **44**, 2097–2102 (2005).

51. Burda, P., Padilla, S. M., Sarkar, S. & Emr, S. D. Retromer function in endosome-to-Golgi retrograde transport is regulated by the yeast Vps34 Ptdlns 3-kinase. *J. Cell Sci.* **115**, 3889–3900 (2002).

52. Suzuki, S. W., Chuang, Y.-S., Li, M., Seaman, M. N. J. & Emr, S. D. A bipartite sorting signal ensures specificity of retromer complex in membrane protein recycling. *J. Cell Biol.* **218**, 2876–2886 (2019).

53. Cui, Y. *et al.* Retromer has a selective function in cargo sorting via endosome transport carriers. *J. Cell Biol.* **218**, 615–631 (2019).

54. Simonetti, B. *et al.* Molecular identification of a BAR domain-containing coat complex for endosomal recycling of transmembrane proteins. *Nat. Cell Biol.* **21**, 1219–1233 (2019).

55. Yong, X. *et al.* Mechanism of cargo recognition by retromer-linked

SNX-BAR proteins. *PLoS Biol.* **18**, e3000631 (2020).

56. Velasco-Olmo, A., Ormaetxea Gisasola, J., Martinez Galvez, J. M., Vera Lillo, J. & Shnyrova, A. V. Combining patch-clamping and fluorescence microscopy for quantitative reconstitution of cellular membrane processes with Giant Suspended Bilayers. *Sci. Rep.* **9**, 1–12 (2019).

57. Ricarda A. Busse, Andreea Scacioc, Amanda M. Schalk, Roswitha Krick, Michael Thumm, and K. K. Analyzing Protein–Phosphoinositide Interactions with Liposome Flotation Assays. *Lipid Signal. Protoc.* **1376**, 155–162 (2016).

58. Hyland, C., Vuillard, L., Hughes, C. & Koronakis, V. Membrane Interaction of Escherichia coli Hemolysin: Flotation and Insertion-Dependent Labeling by Phospholipid Vesicles. *J. Bacteriol.* **183**, 5364–5370 (2001).

59. Bigay, J. & Antonny, B. Real-time assays for the assembly-disassembly cycle of COP coats on liposomes of defined size. *Methods Enzymol.* **404**, 95–107 (2005).

60. Béduneau, A. *et al.* Design of targeted lipid nanocapsules by conjugation of whole antibodies and antibody Fab' fragments. *Biomaterials* **28**, 4978–4990 (2007).