UNIVERSITÀ DEGLI STUDI DI NAPOLI

FEDERICO II

PhD THESIS IN CHEMICAL SCIENCES

XXXIII COURSE

In vitro and *ex vivo* investigation of functional protein complexes for the elucidation of biological processes

Tutor:

Prof. Maria Monti

Author:

Ilaria Iacobucci

Coordinator:

Prof. Angelina Lombardi

Table of contents

Summary

Functional Proteomics aims to the identification of *in vivo* protein-protein interaction (PPI) in order to piece together protein complexes, and therefore, cell pathways involved in biological processes of interest. Over the years, proteomic approaches used for protein-protein interaction investigation have relied on classical biochemical protocols adapted to a global overview of protein-protein interactions within so-called "interactomics" investigation. In particular, their coupling with advanced mass spectrometry instruments and innovative analytical methods led to make great strides in the PPIs investigation in proteomics. The structural investigation of proteins and protein complexes, as well as the identification of protein-ligand interacting regions, is also a powerful tool to understand the biological roles of protein-protein complexes and develop new and more efficient drugs in the therapy treatment of diseases.

In these perspectives, the present PhD project has been addressed to the investigation of molecular mechanisms at the basis of human diseases, host-virus interaction by functional and structural proteomics approaches. Furthermore, the binding mechanisms of metal-based complexes with protein/peptides have been investigated.

In chapter two, *ex vivo* proteomic strategies have been discussed to investigate the molecular mechanisms of Fabry disease, Huntington's disease, and *in vitro* mass spectrometry-based techniques have been employed to investigate the interaction between Osteopontin (OPN) and ICOS ligand (ICOSL) responsible for tumor metastatization. The role that a specific protein plays in cellular processes is clarified by the identification of its molecular partners. Indeed, the association of an individual protein, whose function is unknown, with protein complexes involved in well definite cellular processes would be strongly suggestive of its biological function. A classical functional proteomics approach consists of

isolating protein complexes involving the target protein (bait) from a cell lysate by immunoprecipitation. Proteins so purified are fractionated by SDS-PAGE, *in situ* digested with trypsin, and identified by nanoLC-MS/MS methodologies integrated with a protein database search. This strategy has been applied to investigate the functional role of α-galactosidase in Fabry disease and ADAM10 in Huntington's disease.

Fabry disease is a genetic disorder caused by a mutation in the *GLA* gene encoding for α-galactosidase (α-GAL) enzyme. α-GAL is a lysosomal hydrolase that degrades some substrates such as Globotriaosylceramide (Gb3). The enzyme deficiency causes an accumulation of Gb3, triggering organ dysfunctions. The investigation of the intracellular pathways involved in the route from the endoplasmic reticulum (ER) to lysosome of the wild-type enzyme and the two recombinant enzymes used in the enzyme replacement therapy (ERT) may elucidate the internalization process of the drugs compared to the physiological traffic of α-GAL. The proteomic investigation of the internalization process of both enzymes revealed that it occurs through endocytosis mediated by caveolae and clathrin vesicles.

Huntington's disease (HD) is a neurodegenerative disorder caused by Huntingtin (HTT) gene mutation. Disintegrin and metalloproteinase domain-containing protein 10 was found to accumulate in HD patient brains. The dysregulation of ADAM10 protein levels has been associated with reduced neurotransmission and cognitive decline in HD mice models. The role of ADAM10 in wild-type and HD mice models has been studied with the functional proteomics approach described above to clarify its role in HD dysregulated processes. Proteomics results revealed, for the first time, that ADAM10 is involved in presynaptic functions, specifically in the regulation of synaptic vesicles (SVs) dynamics at the Active

Zone. In HD density of SVs is reduced for dysregulation of the ADAM10/Piccolo complex.

Previous reports demonstrated the migration of cells high expressing ICOSL in the presence of OPN. In this study, we demonstrated ICOSL as a novel receptor for OPN. The interaction was structurally investigated with cross-link and limited proteolysis approaches coupled to mass spectrometry to understand the functional role. These strategies allowed the definition of a binding site upstream of the RGD motif and the thrombin cleavage site. Furthermore, cross-linking experiments indicate that the region downstream the RGD domain (including K_{170}, K_{172}, and K_{173}) becomes exposed upon ICOSL binding, which in turn indicates that OPN changes its conformation in the complex with ICOSL.

In chapter three, the host-virus interaction processes have been investigated for the SARS-CoV-2 and the *Sulfolobus spindle shape 1* (SSV1) viruses.

The molecular mechanisms of SARS-CoV-2 viral infection have been focused on two fields of application: 1) the interaction between Angiotensin-Converting Enzyme 2 (ACE2), the SARS-CoV-2 main target, and potential virus inhibitors, 2) the study of additional SARS-CoV-2 targets on colon and renal cell surfaces. SARS-CoV-2 is a novel coronavirus discovered because of several pneumonia cases in the Hubei region in China at the end of 2019, but it has rapidly spread worldwide, causing a global pandemic. Besides vaccines, the development of novel and efficient therapies is an urgent issue to be addressed.

Long-chain inorganic polyphosphates (PolyPs) demonstrated to have antiviral activities against HIV-1 infection. To assess the potential antiviral molecular mechanism of PolyPs, we studied their interaction with ACE2 receptor by Size Exclusion Chromatography (SEC) and limited proteolysis-mass spectrometry approaches. SEC data suggested that PolyP120 and Spike S1 have different

binding sites on ACE2 and limited proteolysis experiments confirmed the docking calculation for the prediction of a binding pocket.

The investigation of additional SARS-CoV-2 targets on human cells provided the employment of a pull-down experiment using the S1 subunit of the SARS-CoV-2 S protein to purify its putative binding proteins on the human cell membrane. NanoLC-MS/MS technique was used for protein identification according to a "shotgun" proteomic approach.

Sulfolobus spindle shape 1 (SSV1) virus is an archaeal virus well known as extremophile. SSV1 infects *Sulfolobus solfataricus* bacteria. The viral transcription factor F55 has been found to regulate the viral life cycle through the crosstalk between the host and the virus, but the molecular mechanism at the basis of this process is still unknown. Therefore, an electromobility shift assay was employed to isolate the DNA-bound F55 host protein partners, then identified by a bottom-up proteomic technique. Functional experiments allowed us to propose a model explaining the effect of the F55 interaction with the identified host interactor RadA on the T_6 promoter.

In chapter four, the binding features of platinum complexes and potential anti-aggregation metallodrugs with β-lactoglobulin protein and the beta-amyloid peptide (Aβ), respectively, have been probed by the native-MS technique. The latter is based on a particular approach in which biological analytes are ionized and gas-phase transferred through electrospray ionization (ESI) in a non-denaturing solvent and setting the source parameters as much soft as possible to achieve a good compromise between ionization and complex stability. Native ESI-MS can provide powerful information on protein and protein-ligand complexes binding and stoichiometry.

β-Lactoglobulin is a major globular milk whey carrier with potential applications as an oral drug delivery system thanks to its biochemical and biophysical features.

Cisplatin and oxaliplatin anticancer agents have been tested to define their binding aspects with β-Lactoglobulin by the native ESI-MS technique.

Metallocomplexes have been proposed as potential drugs in amyloidogenic neurodegenerative diseases such as Alzheimer's disease (AD). Indeed, they have some anti-aggregation properties towards fibrils generated by the amyloidogenic peptide Aβ. The amyloid inhibitory activity of the metal-based drugs can be exploited through different mechanisms: 1) coordination chemistry, 2) oxidative, 3) proteolytic reactions for peptide modifications. A native ESI-MS approach aimed to investigate the mechanism of action of several Pt-, Pd-, and Au-based complexes in the aggregation modulation of the C-term of the Aβ peptide spanning from residues 21 to 40 ($A\beta_{21-40}$).

Chapter 1- Introduction

1.1 Why study protein-protein interactions?

In the past two decades, the scientific community has increasingly turned the attention to the investigation of protein-protein complexes, as proved by the massive number of publications in this field and the progressive advances in techniques developed to investigate protein-protein interactions (PPIs). Indeed, since the early 2000s, Kumar and Snyder stated: "no protein is an island entire of itself", reworking a John Donne aphorism[1]. Protein functions are mostly driven by their cooperation in different assemblies or networks, conferring additional activities to the same protein upon different cell conditions, external stimuli, organelle localization, pathological states, and so forth (a phenomenon known as moonlighting)[2]. Therefore, the investigation of a biological process, both in physiological and pathological conditions, is also dependent on the identification of the "interactome" of a specific protein or a large-scale analysis of the macromolecule machineries to find out the molecular mechanisms at the basis of the biological event. Protein complexes isolation coupled to their identification by mass spectrometry is a mainstay branch of proteomics research by now.

Likewise, also the structural investigation of proteins and protein complexes as well as the identification of protein-ligand interacting regions can provide functional readouts of mechanisms of protein biochemical activity and can suggest structural constraints for the set-up of new therapeutic approaches and the optimization of the design of specific drugs.

1.2 Proteomics

The "proteome" can be represented by the overall protein content of a cell that is characterized by their localization, interactions, post-translational modifications, and turnover at a particular time. The term "proteomics" was first used by Marc

Wilkins in 1996 to denote the "PROTein complement of a genOME"[3].

Since it is now known the complete sequence of the genome of various organisms, including the human one, the number of proteins whose function is still partially or completely unknown has grown exponentially inrecent times. This gave rise to the need to consider the knowledge of the human genome as a starting point, and not the end, for the understanding of the cellular processes at a molecular level[4].

The complexity of the proteome is due not only to the high number of components that constitute it (more than 10^6 for Mammalia!) but also by its extreme variability, as it may differ from individual to individual and for a specific organism may change over the time, in response to changes of the external environment or stimuli to which the organism is subjected. Proteomics is the field of biochemical sciences addressed to proteome characterization, including expression, structure, functions, interactions, and modifications of proteins at any stage[5]. Nowadays, proteomics investigations find large employment for early disease diagnosis, prognosis, and monitoring disease development[6]. Furthermore, it also has a vital role in drug development as target molecules[7].

Proteomics application fields can essentially be divided into two main areas: differential and functional proteomics. The first approach aims to evaluate the qualitative and quantitative increase and/or decrease in protein expression levels following the change of microenvironmental conditions, such as cellular stress sources, the presence of pathological states, drug treatments, etc.[8]. On the other hand, functional proteomics intends to define the biological function of proteins whose role is not yet known by identifying protein-protein interactions (PPIs) *in vivo* to piece together protein complexes, and therefore, cell pathways involved in biological processes of interest.

Proteomics and mass spectrometry constitute an indissoluble bionomy. Although the general proteomic workflow can follow two different "bottom-up"

approaches: a “shotgun approach” based on direct tryptic digestion of isolated proteins[9] and “a fractionated approach”, which relies on electrophoretic separation of protein complex components before enzymatic digestion, the peptide mixtures are analyzed by mass spectrometry in both cases. Mass spectrometers are used to measure simply the molecular mass of a polypeptide and to determine additional structural features, including the amino acid sequence or the site of attachment and type of post-translational modifications[10]. In such experiments, referred to as tandem mass spectrometry (MS/MS), after the initial mass determination of specific ions, they are selected and subjected to fragmentation through collision[11]. Detailed structural features of the peptides can be inferred from the analysis of the masses of the resulting fragments. Indeed, the MS and MS/MS acquired data are processed for protein identification through several bioinformatic tools such as Mascot[12], X!Tandem[13], SEQUEST[14], MaxQuant[15], Protein Prospector[16]. They manage bioinformatics algorithms to perform an *in silico* digestion of proteins listed in a specific database and match the theoretical and experimental parents (from MS1) and daughters (from MS/MS) m/z values[17]. Each identified peptide and, consequently, the origin proteins are ranked by a score value, according to the probability that the matches between the candidate peptide/protein’s theoretical and experimental mass data are random events[18]. The protein score is assigned, conferring a p-value calculated considering the true positive and false positive matches, basically identified by decoy and reverse database search[19].

1.3 Mass spectrometry-based approaches for the investigation of functional protein complexes

According to the biological system under investigation, many biochemical strategies have been developed to accomplish protein complex isolation.

Affinity purification-mass spectrometry (AP-MS) techniques as well as high-throughput approaches based on cross-linking (XL), Blue-Native polyacrylamide

gel electrophoresis (BN-PAGE), Size Exclusion Chromatography (SEC) are currently employed in the investigation of *in vivo*/*ex vivo* PPIs. However, while the AP-MS-based approaches are addressed to the isolation of a specific bait's interactome, the latter are potentially suitable to provide a global profiling analysis at the proteomic level of all complexes occurring in a biological system (a cell) in specific conditions.

1.3.1 AP-MS strategies

Affinity-purification mass spectrometry techniques comprise the classical strategies, i.e., pull-down assay, the co-immunoprecipitation (co-IP), and their evolution in proximity-dependent labeling techniques.

The pull-down assay is an *in vitro* methodology largely employed to detect and confirm physical interactions between a bait protein and its partners in non-denaturing conditions. Fishing protein complexes can be achieved by exploiting the bait's intrinsic biological affinity with a ligand, typically immobilized on solid support consisting of different polymers (e.g., sepharose, agarose, polystyrene–divinylbenzene)[20]. Therefore, a classical experimental workflow is reported in Fig. 1.1 and consists of: 1) immobilization of the bait on the insoluble support derivatized with its specific ligand; 2) incubation of the cell extract with the bait to allow interaction with its partners; 3) several washing steps to discard non-specific binding, and finally 4) elution of the purified complexes in denaturing conditions or by competition with a free form of the ligand[21].

In the first step, the bait can be bound onto the insoluble support by either the affinity with its specific ligand (e.g., a drug, oligonucleotide, substrate, etc.)[22] or, more frequently, through the mediation of an affinity tag which the bait can be fused to. In the latter case, the bait is expressed as recombinant, with a tag consisting of a fused protein (e.g., Glutathione S-transferase: GST)[23,24] or a small-molecule (e.g., 6x-histidine tail, biotin)[25,26]. The tags allow the bait's tight and

specific binding onto the resin functionalized with a suitable ligand, i.e., reduced glutathione (GSH), nickel cations (Ni^{2+}) avidin/streptavidin when the bait is GST-fused, 6xHis-tagged, or biotin-tagged, respectively.

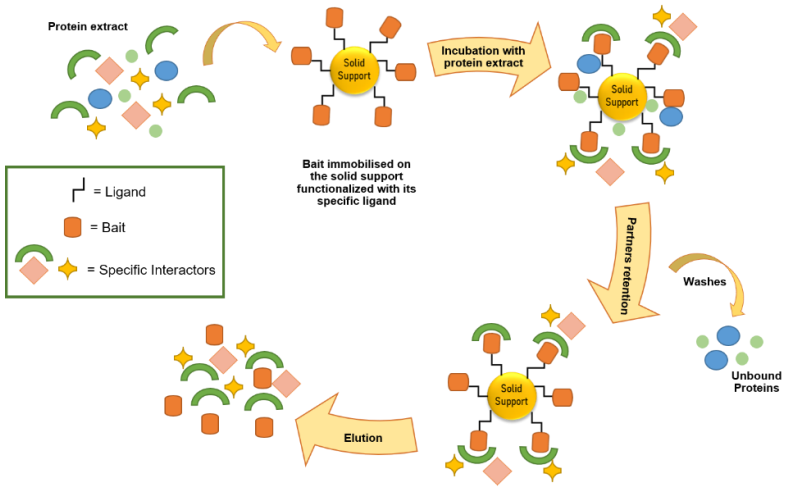

Fig. 1.1 Schematic representation of the pull-down assay workflow[27].

In co-immunoprecipitation experiments, the purification of complexes involving bait is based on antigen-antibody recognition. An anti-bait antibody is bound to protein-A or -G derivatized beads and is employed to capture the bait and its interacting proteins directly from a protein extract (Fig. 1.2). When a suitable anti-bait antibody is not available, this *ex vivo* strategy can be accomplished using an antibody directed toward a tagged form of the bait previously transfected into the cell line of interest. Peptide tags, such as V5, c-Myc, HA, FLAG are the most employed in immunoprecipitation experiments because of 1) their little steric hindrance that does not interfere with the folding process; 2) the availability of commercial high performing anti-tag antibodies covalently conjugated to beads; 3) the possibility to carry out a competitive elution with an excess of free tag reducing the amount of protein contaminants, as it occurs in non-denaturing conditions[28,29,30].

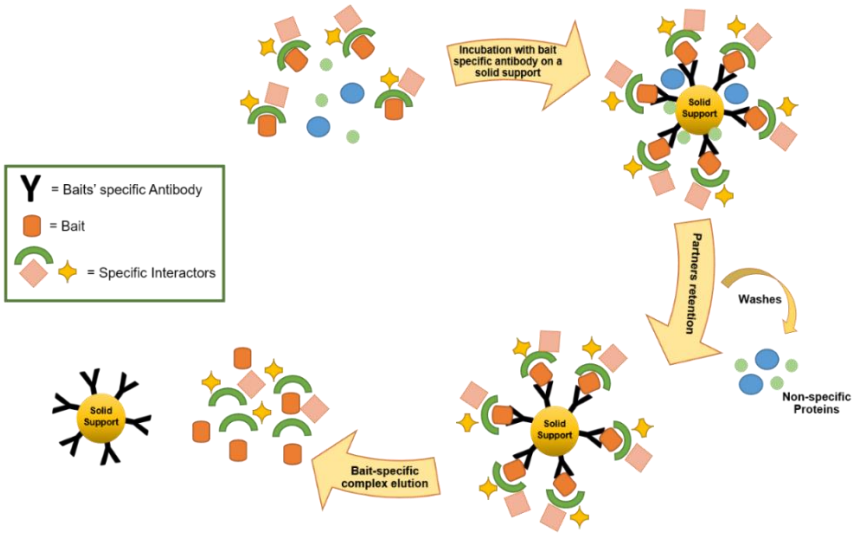

Fig. 1.2 Schematic representation of the co-IP workflow.

Besides the most common and with the highest applicability strategies for protein complexes isolation, based on the pull-down and co-IP assays, complementary approaches have been developed to detect transient interactions[31]. One of them is the Proximity-dependent labeling (PDL) techniques consisting of a pool of strategies including BioID[32], Ascorbate Peroxidase (APEX)[33], selective proteomic proximity labeling assay (SPPLAT)[34], and enzyme-mediated activation of radical sources (EMARS) techniques[35]. PDL techniques aim to label potential interactors, i.e., proteins localized in close spatial proximity (nm of distance) to a protein of interest by mediating the catalytic activity of enzymes fused to the protein of interest and the consequent enrichment of the modified proteins. The first and the most representative methodology of this group is BioID, developed in 2000 when Parrott et al. demonstrated that the endogenous biotin ligase purified from mammalian cells was able to biotinylate other proteins[36]. One year later, the same group showed that the *E. Coli* protein biotin ligase BirA was capable of biotinylate target proteins by applying the technology to the selective modification of secreted

or membrane-bound proteins[37]. A general BioID experiment requires the generation of a stable cell line expressing the protein of interest fused to the *E. Coli* biotin ligase protein. The N- or C- terminus juxtaposition of the enzyme may affect the native conformation of the protein of interest, and functional validation of the fusion protein should be provided at least by confirming the protein localization[38]. Biotinylation reaction is triggered by adding biotin to the cell medium in the presence of ATP: in these conditions, target biotinylation is demonstrated to occur on lysine side chains within a 10 nm range. Following labeling, biotinylated proteins representing the bait interacting partners are purified by cell extract by pulling them down on avidin/streptavidin functionalized beads (Fig. 1.3). Since the biotinylation is a rare modification, the largest part of biotinylated proteins in a BioID experiment can be addressed to BirA-dependent modification. Since the labeling reaction occurs before the lysis step, the protein complexes can be denatured in this phase allowing the detection of the transient PPIs interactions taking a "snapshot" of weak or transient interactions.

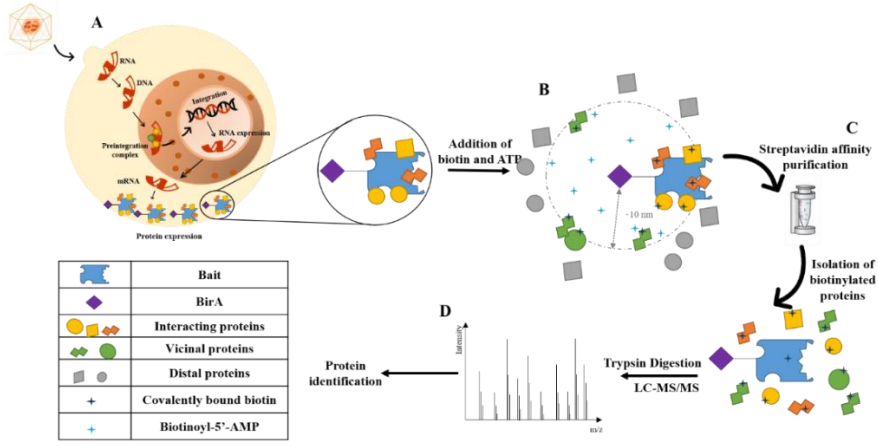

Fig. 1.3 Schematic workflow of a BioID experiment[27].

1.3.2 High-throughput approaches for the "complexome profiling"

Whether the interest is addressed to the investigation of protein complexes to gain global profiling at a proteome level, XL-MS, BN-PAGE, and SEC-MS strategies, coupled to the differential proteomics approach[39], might reveal not only the composition and the physical interacting proteins but also the stoichiometry and apparent molecular weight of stable protein complexes[40], enabling the foundation of the "complexome profiling" approach.

XL-MS is based on the employment of chemical reagents (crosslinkers), which are hetero- or homo-bifunctional organic molecules able to react with specific amino acid residues appropriately. The reactive groups are outdistanced by a spacer arm of defined length (typically dozens of Angstrom)[41]. The cross-linking reaction outcome consists of forming a covalent bond between amino acid side chains properly orientated and spatially close to others.

The extreme reactivity of the cross-linking reagents gives rise to both intra- and intermolecular bonds since the only constraints for reactivity are the nature of the amino acid functional groups, their distance, and orientation. The largest part of cross-linking molecules are designed to react with amino, sulfhydryl, and carbonyl groups, but the last generation of photoinducible reagents displays non-specific reactivity[42]. Moreover, in the absence of two residues with the right distance and spatial orientation, the high reactivity of these compounds frequently leads to single amino acid side chain modification giving rise to dead-end species.

All these reactions occur contemporaneously in a single cross-linking experiment, even when the system comprises only two proteins, leading to extremely heterogeneous samples that are very hard to analyze. The MS profiles of their enzymatic digestions, as well as the intricate patterns of signals in the MS/MS

spectra of cross-linked peptides, are not trivial to be deciphered, often making very hard the identification of cross-linked peptides. The development of suitable software for mass spectrometry-derived data (e.g., XlinkX[43], StavroX [44], MeroX[45], pLink[46]), MS cleavable cross-linking reagents able to penetrate in cell/tissue has broken through the hardness of this methodology moving to the *in vivo* characterization of PPIs at a large scale (Fig. 1.4).

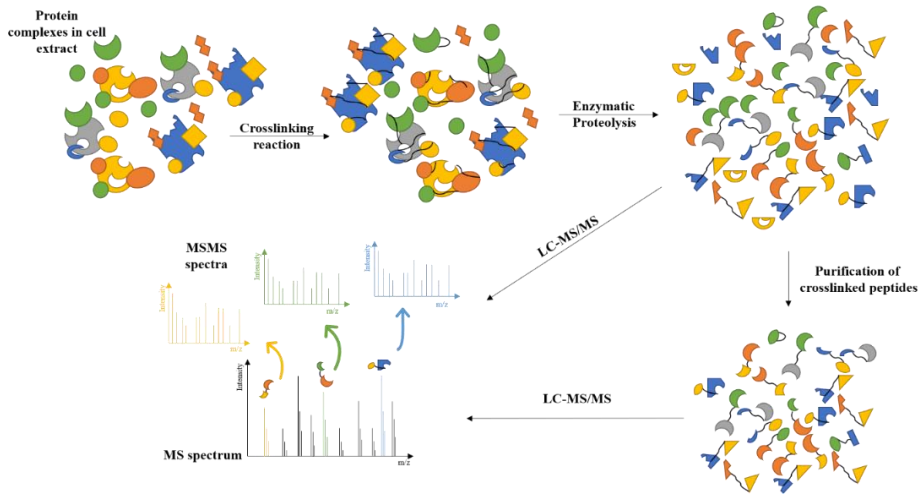

Fig. 1.4 Schematic workflow of an XL-MS experiment[27].

On the other hand, the Blue Native Polyacrylamide Gel Electrophoresis (BN-PAGE) is a technique able to separate protein complexes (from 10kDa up to 10MDa) in native conditions according to their molecular weight and shape. Protein complexes are extracted from cell or tissue samples in non-denaturing conditions using mild non-ionic detergents (e.g., digitonin, dodecylmaltoside, Triton-X100). Fractionation of the complexes is carried out by mixing the extract with the anionic Coomassie blue G-250 dye, which binds protein surfaces, conferring them a negative charge and the consequent ability to move towards the anode within an electric field. Both the sufficient solubility in water and the dye's hydrophobic features allow Coomassie to bind also membrane proteins. In 1D

BN-PAGE, the pore size of a typical gradient gel determines the end-point of complexes migration [for protocol see [47]], and their molecular weights can be retrieved by calibration with standard soluble and membrane-bound proteins or complexes[48]. Identification and quantification of proteins in each gel slice, employing differential proteomics approaches, enables the protein complexes reconstitution and gives information on the complex stoichiometry and molecular weight (Fig. 1.5).

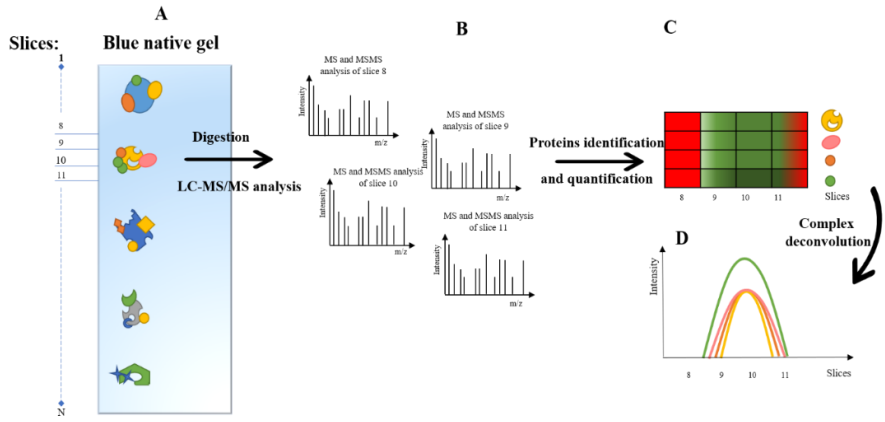

Fig. 1.5 Schematic representation of Blue Native Polyacrylamide Gel Electrophoresis coupled to mass spectrometry strategy (BN-PAGE-MS). Protein complexes are separated by native electrophoresis according to their shape and dimension (A); gel slices are cut and hydrolyzed by trypsin and peptide mixtures analyzed by LC-MS/MS (B). According to single protein abundance estimated based on peptides spectral counts or extracted ion currents, the distribution of each protein among all gel slices is evaluated (C). Protein complexes are rebuilt according to component distribution in each slice (D)[27].

A similar rationale is exploited in the Size Exclusion Chromatography-Mass Spectrometry (SEC-MS) approach. Fractionation of proteins and their assemblies is based on their size, and therefore molecular weight, and shape. The chromatography column is packed with porous gel-like solid phases, and the pore size allows the separation of macromolecules, in particular, higher the molecular weight earlier the retention time. Protein complexes are separated in native

conditions among fractions that are subjected to the classical "bottom-up" procedure for protein identification and quantification (Fig. 1.6).

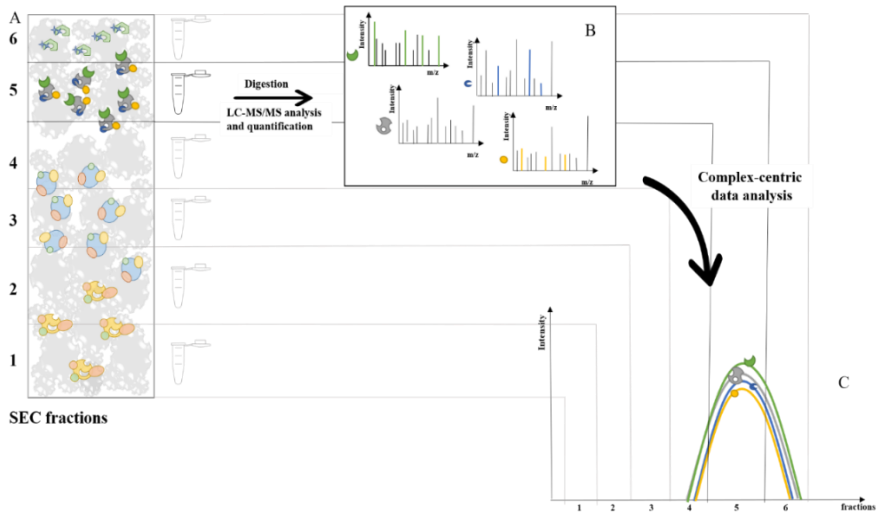

Figure 1.6 Schematic representation of a Size Exclusion Chromatography (SEC)-MS experiment. Protein complexes are separated by Size Exclusion Chromatography (SEC) according to their shape and dimension (A); chromatographic fractions are hydrolyzed by trypsin and peptide mixtures analyzed by LC-MS/MS thus to identify and quantify the eluted proteins (B). The distribution of each protein among all chromatographic fractions is evaluated and then employed to rebuild protein complexes(C)[27].

1.4 *In vitro* structural characterization of protein complexes

The relationship between macromolecule structure and function has been recognized since the earliest years of molecular biology. In 1944 Erwin Shrödinger published the classic book "*What's life?"*[49], inspiring many scientists, particularly physicists, to turn towards biology. In fact, the fifth decades of the last century had seen either the discovery of DNA structure by Francis Crick, James Watson, and Maurice Wilkins and the production of the structural model of monomeric and tetrameric hemoglobin by Max Perutz and John Kendrew[50]. Following these achievements and due to technological improvements, the

structural biology of protein complexes was one of the most flourishing fields in all science.

X-ray crystallography was the first method in structural biology. It is based on the purification of proteins and their complexes to obtain pure enough samples able to crystallize for the acquisition of diffraction patterns with powerful X-ray sources deciphered by computational methods for solving the X-ray diffraction patterns. Even if X-ray crystallography has been, and continues to be, an enormously useful tool to investigate protein/protein complexes, other techniques are widely employed. Among these, cryo-electron microscopy (cryo-EM) is powerfully used to resolve protein complexes above 300 kDa in size. In fact, large complexes are often difficult to crystallize due to the compositional heterogeneity of the samples, which cryo-EM can more easily handle. The two techniques are highly complementary, and many structures are solved at high resolution combining the cryo-EM for a coarse-grained structure, and the X-ray crystallography, for atomic resolution of individual subunits.

Many important biological complexes exist in a dynamic ensemble of conformational states or contain subunits that interact weakly with each other. Such complexes do not lend themselves well to characterization by crystallographic or cryo-EM methods, which can generally resolve one structural state at time. Nuclear magnetic resonance (NMR) is well suitable to study these cases because proteins are analyzed in solution rather than crystallized or frozen. On the other hand, NMR has traditionally struggled to resolve structures beyond 30kDa due to the very efficient relaxation of nuclear spin orientation for large, slowly tumbling molecules. This has the effect of broadening the peaks observed in NMR spectra and, together with the more complex spectra generated by large molecules than smaller ones, ensures that using NMR to study protein complexes is challenging.

Another technique employed mainly in the *in vitro* investigation of protein complexes is mass spectrometry. The arrival of soft ionization MS techniques in the 1980s was of critical importance in protein complexes research, as it allowed delicate non-covalent interactions between proteins to be preserved in gas-phase, making it possible to study intact protein complexes via MS. Combined with the later development of time-of-flight analyzers, this became known as native MS[51]. Since native MS-based protocols have been developed to preserve the main non-covalent PPIs stabilizing complexes subunits, they can be used to study properties such as stoichiometry, compositional heterogeneity, and dynamic processes such as assembly and disassembly.

1.4.1 MS-based strategy for the *in vitro* study of protein complexes

Native-MS term relies on a particular approach in which biological analytes are ionized and gas-phase transferred through electrospray ionization (ESI) in a non-denaturing solvent and setting the source parameters as much soft as possible to achieve a good compromise between ionization and PPIs stability. Of course, the biological status of the samples in solution before the ionization process mostly affects complex stability[51]. Indeed, it is strictly indispensable to control the experimental conditions such as pH and ionic strength to maintain the native folded state of complexes. When the charged state distribution was observed to reflect protein conformation, the field of structural MS came into realizing that non-covalent interactions can be retained in the gas phase[52,53]. This means that the transferring in the mass spectrometer of the entire protein complex and its analysis could be viable in the right set-up. The principal advantages of using MS lies in its ability to probe transient and heterogeneous macromolecular assemblies, with low consumption of sample (pmol)[54].

Above the native-MS, other MS-based strategies have been developed to investigate the complex conformations and dynamics. The aims of these strategies are both the definition of protein-protein/protein-ligand stoichiometry and the identification of the interacting surfaces. Coupled to MS, techniques such as cross-linking (XL), hydroxyl radical footprinting (HRF), and hydrogen/deuterium exchange (HDX) are based on the labeling of amino acidic side chains exposed to solvent[55,56].

Besides their role in the elucidation of protein complexes at the proteome level described above, crosslinkers have a longstanding history in classical protein chemistry for the investigation of two interacting proteins *in vitro*[57,58]. The use of crosslinkers in structural biology has a dual role: 1) enabling the determination of the interacting surfaces by identifying inter-chain cross-linked peptides and 2) distinguishing the surface labeled residues that take into account conformational changes upon protein/ligand interaction[56].

HRF takes advantage of covalent and irreversible reaction starting from hydroxyl radicals on surface-exposed aromatic, aliphatic, and sulfur-containing amino acid side chains[59] (Fig. 1.7 panel A), while HDX relies on the deuterons (deuterium ions) exchange with bulk solvent-exposed amide backbone protons in a time-dependent fashion in the presence of deuterated water (D_2O) in the buffer[60] and a consequent quenching step at pH 2.5 and 0°C to "freeze" the protein conformations because the exchange reaction is decelerated[61] (Fig. 1.7 panel B). The MS identification of the modified peptides allows the topology determination of protein-protein/protein-ligand complexes.

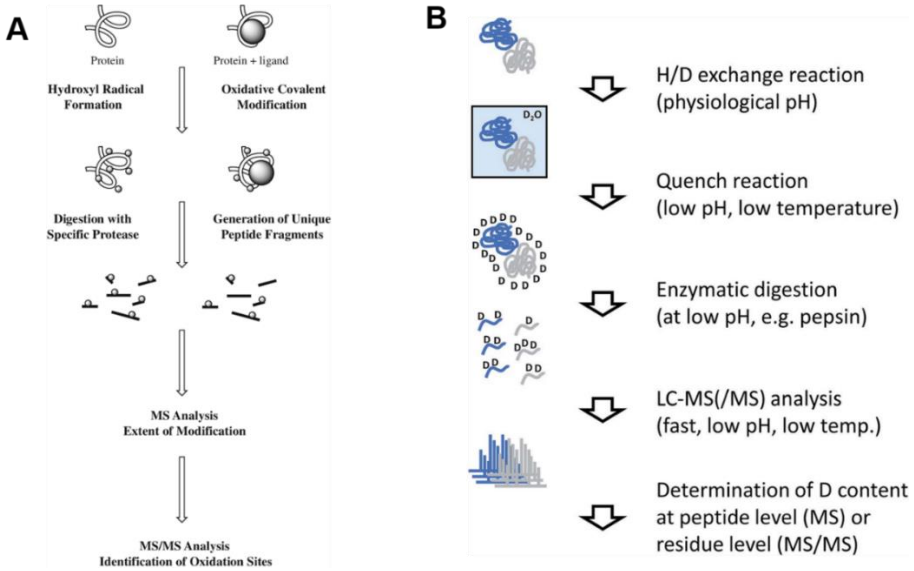

Fig. 1.7 Workflow of hydroxyl radical footprinting (HRF) (panel A)[62] and hydrogen/deuterium exchange (HDX)[63] experiments (panel B).

On the other hand, a non-labeling technique can be coupled to mass spectrometry. Limited proteolysis aims to probe the protein conformational changes upon protein or ligand binding by evaluating the enzymatic proteolysis susceptibility. Indeed, different enzymatic probes can be used to find out the more flexible and solvent-exposed regions, which are preferentially cleaved based on enzyme specificity[64,65,66] (Fig 1.8). Consequently, the MS identification of the different proteolytic patterns of isolated and complexed proteins elicits structural alterations due to complex formation.

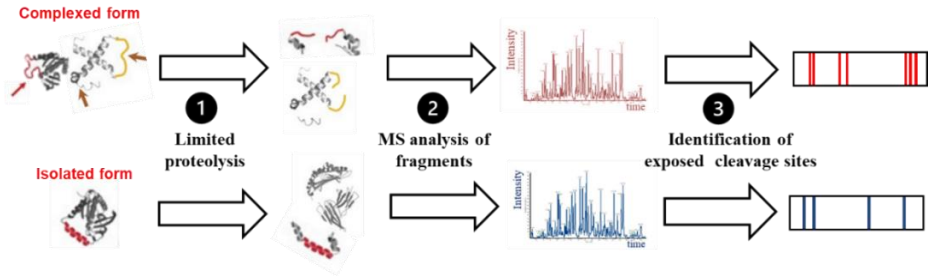

Fig. 1.8 Workflow of a limited proteolysis-mass spectrometry experiment. The isolated and complexed form of the protein is subjected to limited enzymatic digestion to highlight the more exposed and flexible protein regions in the two conditions and the enzymatic cleavage sites are identified by mass spectrometry.

1.5 References

(1) Kumar, A.; Snyder, M. Protein Complexes Take the Bait. *Nature* **2002**, *415* (6868), 123–124. https://doi.org/10.1038/415123a.

(2) Jeffery, C. J. Protein Moonlighting: What Is It, and Why Is It Important? *Phil. Trans. R. Soc. B* **2018**, *373* (1738), 20160523. https://doi.org/10.1098/rstb.2016.0523.

(3) Aslam, B.; Basit, M.; Nisar, M. A.; Khurshid, M.; Rasool, M. H. Proteomics: Technologies and Their Applications. *J Chromatogr Sci* **2017**, *55* (2), 182–196. https://doi.org/10.1093/chromsci/bmw167.

(4) O'Donovan, C.; Apweiler, R.; Bairoch, A. The Human Proteomics Initiative (HPI). *Trends Biotechnol* **2001**, *19* (5), 178–181. https://doi.org/10.1016/s0167-7799(01)01598-0.

(5) Domon, B.; Aebersold, R. Mass Spectrometry and Protein Analysis. *Science* **2006**, *312* (5771), 212–217. https://doi.org/10.1126/science.1124619.

(6) Amiri-Dashatan, N.; Koushki, M.; Abbaszadeh, H.-A.; Rostami-Nejad, M.; Rezaei-Tavirani, M. Proteomics Applications in Health: Biomarker and Drug Discovery and Food Industry. *Iran J Pharm Res* **2018**, *17* (4), 1523–1536.

(7) Frantzi, M.; Latosinska, A.; Mischak, H. Proteomics in Drug Development: The Dawn of a New Era? *Proteomics Clin Appl* **2019**, *13* (2), e1800087. https://doi.org/10.1002/prca.201800087.

(8) Corthals, G. L.; Wasinger, V. C.; Hochstrasser, D. F.; Sanchez, J. C. The Dynamic Range of Protein Expression: A Challenge for Proteomic Research. *Electrophoresis* **2000**, *21* (6), 1104–1115. https://doi.org/10.1002/(SICI)1522-2683(20000401)21:6<1104::AID-ELPS1104>3.0.CO;2-C.

(9) Gilmore, J. M.; Washburn, M. P. Advances in Shotgun Proteomics and the Analysis of Membrane Proteomes. *Journal of Proteomics* **2010**, *73* (11), 2078–2091. https://doi.org/10.1016/j.jprot.2010.08.005.

(10) Glish, G. L.; Burinsky, D. J. Hybrid Mass Spectrometers for Tandem Mass Spectrometry. *J Am Soc Mass Spectrom* **2008**, *19* (2), 161–172. https://doi.org/10.1016/j.jasms.2007.11.013.

(11) Bartlett, K.; Pourfarzam, M. Tandem Mass Spectrometry--the Potential. *J Inherit Metab Dis* **1999**, *22* (4), 568–571. https://doi.org/10.1023/a:1005520726774.

(12) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551–3567. https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2.

(13) Craig, R.; Beavis, R. C. TANDEM: Matching Proteins with Tandem Mass Spectra. *Bioinformatics* **2004**, *20* (9), 1466–1467. https://doi.org/10.1093/bioinformatics/bth092.

(14) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom* **1994**, *5* (11), 976–989. https://doi.org/10.1016/1044-0305(94)80016-2.

(15) Cox, J.; Mann, M. MaxQuant Enables High Peptide Identification Rates, Individualized p.p.b.-Range Mass Accuracies and Proteome-Wide Protein Quantification. *Nat Biotechnol* **2008**, *26* (12), 1367–1372. https://doi.org/10.1038/nbt.1511.

(16) Chalkley, R. J.; Baker, P. R.; Huang, L.; Hansen, K. C.; Allen, N. P.; Rexach, M.; Burlingame, A. L. Comprehensive Analysis of a Multidimensional Liquid Chromatography Mass Spectrometry Dataset Acquired on a Quadrupole Selecting, Quadrupole Collision Cell, Time-of-Flight Mass Spectrometer. *Molecular & Cellular Proteomics* **2005**, *4* (8), 1194–1204. https://doi.org/10.1074/mcp.D500002-MCP200.

(17) Nesvizhskii, A. I. Protein Identification by Tandem Mass Spectrometry and Sequence Database Searching. In *Mass Spectrometry Data Analysis in Proteomics*; Humana Press: New Jersey, 2006; Vol. 367, pp 87–120. https://doi.org/10.1385/1-59745-275-0:87.

(18) McHugh, L.; Arthur, J. W. Computational Methods for Protein Identification from Mass Spectrometry Data. *PLoS Comput Biol* **2008**, *4* (2), e12. https://doi.org/10.1371/journal.pcbi.0040012.

(19) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry. *Nat Methods* **2007**, *4* (3), 207–214. https://doi.org/10.1038/nmeth1019.

(20) Havis, S.; Moree, W. J.; Mali, S.; Bark, S. J. Solid Support Resins and Affinity Purification Mass Spectrometry. *Mol. BioSyst.* **2017**, *13* (3), 456–462. https://doi.org/10.1039/C6MB00735J.

(21) Louche, A.; Salcedo, S. P.; Bigot, S. Protein–Protein Interactions: Pull-Down Assays. In *Bacterial Protein Secretion Systems*; Journet, L., Cascales, E., Eds.; Methods in Molecular Biology; Springer New York: New York, NY, 2017; Vol. 1615, pp 247–255. https://doi.org/10.1007/978-1-4939-7033-9_20.

(22) Higashi, K.; Tomigahara, Y.; Shiraki, H.; Miyata, K.; Mikami, T.; Kimura, T.; Moro, T.; Inagaki, Y.; Kaneko, H. A Novel Small Compound That Promotes Nuclear Translocation of YB-1 Ameliorates Experimental Hepatic

Fibrosis in Mice. *J Biol Chem* **2011**, *286* (6), 4485–4492. https://doi.org/10.1074/jbc.M110.151936.

(23) Shao, G.; Wang, R.; Sun, A.; Wei, J.; Peng, K.; Dai, Q.; Yang, W.; Lin, Q. The E3 Ubiquitin Ligase NEDD4 Mediates Cell Migration Signaling of EGFR in Lung Cancer Cells. *Mol Cancer* **2018**, *17* (1), 24. https://doi.org/10.1186/s12943-018-0784-2.

(24) Hoogenraad, C. C.; Popa, I.; Futai, K.; Sanchez-Martinez, E.; Wulf, P. S.; van Vlijmen, T.; Dortland, B. R.; Oorschot, V.; Govers, R.; Monti, M.; Heck, A. J. R.; Sheng, M.; Klumperman, J.; Rehmann, H.; Jaarsma, D.; Kapitein, L. C.; van der Sluijs, P. Neuron Specific Rab4 Effector GRASP-1 Coordinates Membrane Specialization and Maturation of Recycling Endosomes. *PLoS Biol* **2010**, *8* (1), e1000283. https://doi.org/10.1371/journal.pbio.1000283.

(25) Tomita, T.; Matsushita, H.; Yoshida, A.; Kosono, S.; Yoshida, M.; Kuzuyama, T.; Nishiyama, M. Glutamate Dehydrogenase from *Thermus Thermophilus* Is Activated by AMP and Leucine as a Complex with Catalytically Inactive Adenine Phosphoribosyltransferase Homolog. *J Bacteriol* **2019**, *201* (14), e00710-18, /jb/201/14/JB.00710-18.atom. https://doi.org/10.1128/JB.00710-18.

(26) Meier, N.; Krpic, S.; Rodriguez, P.; Strouboulis, J.; Monti, M.; Krijgsveld, J.; Gering, M.; Patient, R.; Hostert, A.; Grosveld, F. Novel Binding Partners of Ldb1 Are Required for Haematopoietic Development. *Development* **2006**, *133* (24), 4913–4923. https://doi.org/10.1242/dev.02656.

(27) Iacobucci, I.; Monaco, V.; Cozzolino, F.; Monti, M. From Classical to New Generation Approaches: An Excursus of -Omics Methods for Investigation of Protein-Protein Interaction Networks. *J Proteomics* **2021**, *230*, 103990. https://doi.org/10.1016/j.jprot.2020.103990.

(28) Federico, A.; Sepe, R.; Cozzolino, F.; Piccolo, C.; Iannone, C.; Iacobucci, I.; Pucci, P.; Monti, M.; Fusco, A. The Complex CBX7-PRMT1 Has a Critical Role in Regulating E-Cadherin Gene Expression and Cell Migration. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **2019**, *1862* (4), 509–521. https://doi.org/10.1016/j.bbagrm.2019.02.006.

(29) Landriscina, M.; Laudiero, G.; Maddalena, F.; Amoroso, M. R.; Piscazzi, A.; Cozzolino, F.; Monti, M.; Garbi, C.; Fersini, A.; Pucci, P.; Esposito, F. Mitochondrial Chaperone Trap1 and the Calcium Binding Protein Sorcin Interact and Protect Cells against Apoptosis Induced by Antiblastic Agents. *Cancer Res* **2010**, *70* (16), 6577–6586. https://doi.org/10.1158/0008-5472.CAN-10-1256.

(30) Zollo, M.; Ahmed, M.; Ferrucci, V.; Salpietro, V.; Asadzadeh, F.; Carotenuto, M.; Maroofian, R.; Al-Amri, A.; Singh, R.; Scognamiglio, I.; Mojarrad, M.; Musella, L.; Duilio, A.; Di Somma, A.; Karaca, E.; Rajab, A.; Al-Khayat, A.; Mohan Mohapatra, T.; Eslahi, A.; Ashrafzadeh, F.; Rawlins, L. E.; Prasad, R.; Gupta, R.; Kumari, P.; Srivastava, M.; Cozzolino, F.; Kumar Rai, S.; Monti, M.; Harlalka, G. V.; Simpson, M. A.; Rich, P.; Al-Salmi, F.; Patton, M. A.; Chioza, B. A.; Efthymiou, S.; Granata, F.; Di Rosa, G.; Wiethoff, S.; Borgione, E.; Scuderi, C.; Mankad, K.; Hanna, M. G.; Pucci, P.; Houlden, H.; Lupski, J. R.; Crosby, A. H.; Baple, E. L. PRUNE Is Crucial for Normal Brain Development and Mutated in Microcephaly with Neurodevelopmental Impairment. *Brain* **2017**, *140* (4), 940–952. https://doi.org/10.1093/brain/awx014.

(31) Mazina, M. Yu.; Ziganshin, R. H.; Magnitov, M. D.; Golovnin, A. K.; Vorobyeva, N. E. Proximity-Dependent Biotin Labelling Reveals CP190 as an EcR/Usp Molecular Partner. *Sci Rep* **2020**, *10* (1), 4793. https://doi.org/10.1038/s41598-020-61514-0.

(32) Roux, K. J.; Kim, D. I.; Burke, B.; May, D. G. BioID: A Screen for Protein-Protein Interactions. *Current Protocols in Protein Science* **2018**, *91* (1). https://doi.org/10.1002/cpps.51.

(33) Hung, V.; Zou, P.; Rhee, H.-W.; Udeshi, N. D.; Cracan, V.; Svinkina, T.; Carr, S. A.; Mootha, V. K.; Ting, A. Y. Proteomic Mapping of the Human Mitochondrial Intermembrane Space in Live Cells via Ratiometric APEX Tagging. *Molecular Cell* **2014**, *55* (2), 332–341. https://doi.org/10.1016/j.molcel.2014.06.003.

(34) Rees, J. S.; Li, X.; Perrett, S.; Lilley, K. S.; Jackson, A. P. Selective Proteomic Proximity Labeling Assay Using Tyramide (SPPLAT): A Quantitative Method for the Proteomic Analysis of Localized Membrane-Bound Protein Clusters. *Current Protocols in Protein Science* **2017**, *88* (1). https://doi.org/10.1002/cpps.27.

(35) Honke, K.; Kotani, N. Identification of Cell-Surface Molecular Interactions under Living Conditions by Using the Enzyme-Mediated Activation of Radical Sources (EMARS) Method. *Sensors* **2012**, *12* (12), 16037–16045. https://doi.org/10.3390/s121216037.

(36) Parrott, M. B.; Adams, K. E.; Mercier, G. T.; Mok, H.; Campos, S. K.; Barry, M. A. Metabolically Biotinylated Adenovirus for Cell Targeting, Ligand Screening, and Vector Purification. *Mol Ther* **2003**, *8* (4), 688–700. https://doi.org/10.1016/s1525-0016(03)00213-2.

(37) Parrott, M. B.; Barry, M. A. Metabolic Biotinylation of Secreted and Cell Surface Proteins from Mammalian Cells. *Biochem Biophys Res Commun* **2001**, *281* (4), 993–1000. https://doi.org/10.1006/bbrc.2001.4437.

(38) Roux, K. J.; Kim, D. I.; Burke, B.; May, D. G. BioID: A Screen for Protein-Protein Interactions. *Curr Protoc Protein Sci* **2018**, *91*, 19.23.1-19.23.15. https://doi.org/10.1002/cpps.51.

(39) Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative Mass Spectrometry in Proteomics: A Critical Review. *Anal Bioanal Chem* **2007**, *389* (4), 1017–1031. https://doi.org/10.1007/s00216-007-1486-6.

(40) Valerius, O.; Asif, A. R.; Beißbarth, T.; Bohrer, R.; Dihazi, H.; Feussner, K.; Jahn, O.; Majcherczyk, A.; Schmidt, B.; Schmitt, K.; Urlaub, H.; Lenz, C. Mapping Cellular Microenvironments: Proximity Labeling and Complexome Profiling (Seventh Symposium of the Göttingen Proteomics Forum). *Cells* **2019**, *8* (10). https://doi.org/10.3390/cells8101192.

(41) Wong, S. S.; Wong, L.-J. C. Chemical Crosslinking and the Stabilization of Proteins and Enzymes. *Enzyme and Microbial Technology* **1992**, *14* (11), 866–874. https://doi.org/10.1016/0141-0229(92)90049-T.

(42) Arora, B.; Tandon, R.; Attri, P.; Bhatia, R. Chemical Crosslinking: Role in Protein and Peptide Science. *CPPS* **2017**, *18* (9). https://doi.org/10.2174/1389203717666160724202806.

(43) Liu, F.; Rijkers, D. T. S.; Post, H.; Heck, A. J. R. Proteome-Wide Profiling of Protein Assemblies by Cross-Linking Mass Spectrometry. *Nat Methods* **2015**, *12* (12), 1179–1184. https://doi.org/10.1038/nmeth.3603.

(44) Götze, M.; Pettelkau, J.; Schaks, S.; Bosse, K.; Ihling, C. H.; Krauth, F.; Fritzsche, R.; Kühn, U.; Sinz, A. StavroX—A Software for Analyzing Crosslinked Products in Protein Interaction Studies. *J. Am. Soc. Mass Spectrom.* **2012**, *23* (1), 76–87. https://doi.org/10.1007/s13361-011-0261-2.

(45) Iacobucci, C.; Götze, M.; Ihling, C. H.; Piotrowski, C.; Arlt, C.; Schäfer, M.; Hage, C.; Schmidt, R.; Sinz, A. A Cross-Linking/Mass Spectrometry Workflow Based on MS-Cleavable Cross-Linkers and the MeroX Software for Studying Protein Structures and Protein–Protein Interactions. *Nat Protoc* **2018**, *13* (12), 2864–2889. https://doi.org/10.1038/s41596-018-0068-8.

(46) Fan, S.; Meng, J.; Lu, S.; Zhang, K.; Yang, H.; Chi, H.; Sun, R.; Dong, M.; He, S. Using PLink to Analyze Cross-Linked Peptides. *Current Protocols in Bioinformatics* **2015**, *49* (1). https://doi.org/10.1002/0471250953.bi0821s49.

(47) Wittig, I.; Braun, H.-P.; Schägger, H. Blue Native PAGE. *Nat Protoc* **2006**, *1* (1), 418–428. https://doi.org/10.1038/nprot.2006.62.

(48) Wittig, I.; Beckhaus, T.; Wumaier, Z.; Karas, M.; Schägger, H. Mass Estimation of Native Proteins by Blue Native Electrophoresis. *Molecular & Cellular Proteomics* **2010**, *9* (10), 2149–2161. https://doi.org/10.1074/mcp.M900526-MCP200.

(49) Lasker, G. W. What Is Life? The Physical Aspect of the Living Cell. Erwin Schrödinger. The Macmillan Company, New York. 1946. Pp. Viii plus 91 ($1.75). (Illustrated.). *Am. J. Phys. Anthropol.* **1947**, *5* (1), 103–104. https://doi.org/10.1002/ajpa.1330050116.

(50) Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C. Structure of Haemoglobin: A Three-Dimensional Fourier Synthesis at 5.5-A. Resolution, Obtained by X-Ray Analysis. *Nature* **1960**, *185* (4711), 416–422. https://doi.org/10.1038/185416a0.

(51) Leney, A. C.; Heck, A. J. R. Native Mass Spectrometry: What Is in the Name? *J Am Soc Mass Spectrom* **2017**, *28* (1), 5–13. https://doi.org/10.1007/s13361-016-1545-3.

(52) Ganem, B.; Li, Y. T.; Henion, J. D. Detection of Noncovalent Receptor-Ligand Complexes by Mass Spectrometry. *J. Am. Chem. Soc.* **1991**, *113* (16), 6294–6296. https://doi.org/10.1021/ja00016a069.

(53) Ganem, B.; Li, Y. T.; Henion, J. D. Observation of Noncovalent Enzyme-Substrate and Enzyme-Product Complexes by Ion-Spray Mass Spectrometry. *J. Am. Chem. Soc*. **1991**, *113* (20), 7818–7819. https://doi.org/10.1021/ja00020a085.

(54) Sharon, M. How Far Can We Go with Structural Mass Spectrometry of Protein Complexes? *J Am Soc Mass Spectrom* **2010**, *21* (4), 487–500. https://doi.org/10.1016/j.jasms.2009.12.017.

(55) Wang, L.; Chance, M. R. Protein Footprinting Comes of Age: Mass Spectrometry for Biophysical Structure Assessment. *Mol Cell Proteomics* **2017**, *16* (5), 706–716. https://doi.org/10.1074/mcp.O116.064386.

(56) Borch, J.; Jørgensen, T. J.; Roepstorff, P. Mass Spectrometric Analysis of Protein Interactions. *Current Opinion in Chemical Biology* **2005**, *9* (5), 509–516. https://doi.org/10.1016/j.cbpa.2005.08.013.

(57) Birolo, L.; Sacchi, S.; Smaldone, G.; Molla, G.; Leo, G.; Caldinelli, L.; Pirone, L.; Eliometri, P.; Di Gaetano, S.; Orefice, I.; Pedone, E.; Pucci, P.; Pollegioni, L. Regulating Levels of the Neuromodulator D -Serine in Human Brain: Structural Insight into PLG72 and D -Amino Acid Oxidase Interaction. *FEBS J* **2016**, *283* (18), 3353–3370. https://doi.org/10.1111/febs.13809.

(58) Sinz, A. Chemical Cross-Linking and Mass Spectrometry to Map Three-Dimensional Protein Structures and Protein–Protein Interactions. *Mass Spectrom. Rev.* **2006**, *25* (4), 663–682. https://doi.org/10.1002/mas.20082.

(59) Maleknia, S. D.; Downard, K. M. Protein Footprinting with Radical Probe Mass Spectrometry- Two Decades of Achievement. *Protein Pept Lett* **2019**, *26* (1), 4–15. https://doi.org/10.2174/0929866526666181128124241.

(60) Rand, K. D.; Zehl, M.; Jørgensen, T. J. D. Measuring the Hydrogen/Deuterium Exchange of Proteins at High Spatial Resolution by Mass Spectrometry: Overcoming Gas-Phase Hydrogen/Deuterium Scrambling. *Acc Chem Res* **2014**, *47* (10), 3018–3027. https://doi.org/10.1021/ar500194w.

(61) Nguyen, D.; Mayne, L.; Phillips, M. C.; Walter Englander, S. Reference Parameters for Protein Hydrogen Exchange Rates. *J Am Soc Mass Spectrom* **2018**, *29* (9), 1936–1939. https://doi.org/10.1007/s13361-018-2021-z.

(62) Roeser, J.; Bischoff, R.; Bruins, A. P.; Permentier, H. P. Oxidative Protein Labeling in Mass-Spectrometry-Based Proteomics. *Anal Bioanal Chem* **2010**, *397* (8), 3441–3455. https://doi.org/10.1007/s00216-010-3471-8.

(63) Leitner, A. Cross-Linking and Other Structural Proteomics Techniques: How Chemistry Is Enabling Mass Spectrometry Applications in Structural Biology. *Chem. Sci.* **2016**, *7* (8), 4792–4803. https://doi.org/10.1039/C5SC04196A.

(64) Artigues, A.; Nadeau, O. W.; Rimmer, M. A.; Villar, M. T.; Du, X.; Fenton, A. W.; Carlson, G. M. Protein Structural Analysis via Mass Spectrometry-Based Proteomics. *Adv Exp Med Biol* **2016**, *919*, 397–431. https://doi.org/10.1007/978-3-319-41448-5_19.

(65) Suh, M.-J.; Pourshahian, S.; Limbach, P. A. Developing Limited Proteolysis and Mass Spectrometry for the Characterization of Ribosome Topography. *J Am Soc Mass Spectrom* **2007**, *18* (7), 1304–1317. https://doi.org/10.1016/j.jasms.2007.03.028.

(66) Del Giudice, R.; Domingo-Espín, J.; Iacobucci, I.; Nilsson, O.; Monti, M.; Monti, D. M.; Lagerstedt, J. O. Structural Determinants in ApoA-I Amyloidogenic Variants Explain Improved Cholesterol Metabolism despite Low HDL Levels. *Biochim Biophys Acta Mol Basis Dis* **2017**, *1863* (12), 3038–3048. https://doi.org/10.1016/j.bbadis.2017.09.001.

Chapter 2- Structural and functional investigation of protein complexes in human diseases

2.1 A comparative study of the internalization of two recombinant α-galactosidase A, employed in the treatment of Fabry-Anderson disease

2.1.1 Introduction

Lysosomes are membrane-enclosed cytoplasmic organelles that contain lysosomal enzymes, a variety of active hydrolytic enzymes such as glycosidases, sulfatases, phosphatases, lipases, phospholipases, proteases, and nucleases in an acid milieu with a pH of approximately 5[1]. An H^+ pump in the lysosomal membrane uses the energy of ATP hydrolysis to pump H^+ into the lysosome, thereby maintaining the lumen at its acidic pH[2]. Lysosomal hydrolytic enzymes (hydrolases) move from Golgi to lysosomes within specific vesicles, bound to specific receptors for their recognition signal, the mannose-6-phosphate. Extracellular macromolecules (lipids, carbohydrates, proteins) are delivered via endocytic, phagocytic, and autophagic pathways to be degraded in their respective terminal components (fatty acids, monosaccharides, amino acids), which subsequently exit the lysosome. The deficiency of a single hydrolase will create an inability to degrade the macromolecules: the result is a lysosomal storage disease (LSD)[1].

LSDs comprise a group of rare inherited chronic syndromes that cause deficiency of specific native enzymes within the lysosomes. The function impairment of lysosomal enzymes causes intra-lysosomal accumulation of undegraded substrates[1]. The accumulation of macromolecules in lysosomes is responsible for the progressive loss of function in multiple organs. Symptoms may emerge at

variable ages but, the diseases progress and evolve, impairing patients' health and life expectancy[3].

As showed in Table 2.1.1, lysosomal enzyme deficiencies can be categorized based upon the macromolecule that fails to be degraded and is consequently stored. Carbohydrates, for example, are stored in glycogen storage disease type II, or Pompe disease. Glycolipids accumulate in mucolipidoses[1]. Sphingolipids are undegraded in Tay-Sachs disease and Fabry-Anderson disease.

Disease	**Defective protein**	**Materials stored**	**Major organ systems affected**
Sphingolipids			
Fabry	α-Galactosidase A	Globotriasylceramide	Kidney, heart
Gaucher	β-Glucosidase	Glucosylceramide glucosylsphingosine	Spleen, liver, bone marrow
Nieman-Pick A and B	Sphingomyelinase	Sphingomyelin	Spleen, liver, bone marrow, lung (type B)
GM1 Gangliosidosis	β-Galactosidase	GM1 ganglioside	Skeleton, heart
Tay-Sachs	Hexosaminidase A	GM2 ganglioside	Brain nerve cells
Mucopolysaccharides			
MPS1 (Hurler, Scheie, Hurler/Scheie	α-Iduronidase	Dermatan sulfate and heparan sulfate	Organomegaly, skeleton, eye
MPS II (Hunter)	Iduronate-2-sulphatase	Dermatan sulfate and heparan sulfate	Organomegaly, heart

MPS VI (Maroteaux-Lamy)	Acetylgalactosamine-4-sulphatase	Dermatan sulfate	Spleen, liver, heart, macrocephaly
Glycogen			
Pompe	α-1,4-Glucosidase	Glycogen	Liver, heart, and muscles

Table 2.1.1. Most common types of lysosomal storage diseases (LSDs) are classified according to the relevant substrate involved in the pathology.

Fabry-Anderson disease (FAD) is an X-linked disease (Fig. 2.1.1). Heterogenous mothers with each conception have a 50% chance of passing the defective gene to all offspring[4]. Sons who inherit the defective gene will have the disease; daughters can be asymptomatic carriers, or they may develop disease from mild to severe manifestations because of X inactivation. The incidence of the disease is about 1 in 117,000 live births for males; it is not possible to estimate the incidence in the female population according to the reasons previously explained.

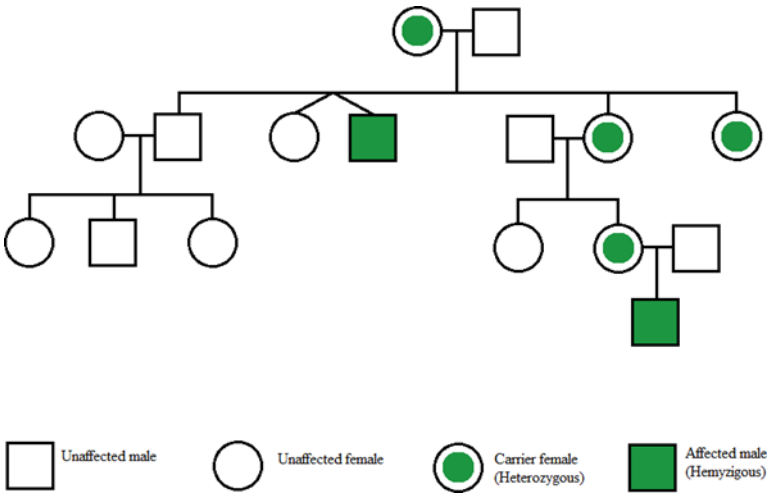

Fig 2.1.1 Genetic inheritance of Fabry disease.

Clinical manifestations include periodic crises of severe pain in the extremities, also known as acroparesthesia, the appearance of vascular cutaneous lesions

called angiokeratomas[4], sweating abnormalities such as anhidrosis, hypohidrosis, and rarely hyperhidrosis, characteristic corneal and lenticular opacities[5]. The kidneys and the cardiovascular system are mainly affected. The kidney disease is usually associated with progressive proteinuria following a decline in glomerular filtration rate, leading over several years to an end-stage renal disease requiring dialysis and kidney transplantation. Regarding the cardiovascular system, complications include progressive hypertrophic cardiomyopathy with diastolic dysfunction, a variety of conduction defects and arrhythmia, and ventricular tachycardia; other complications are atrial fibrillation as well as valvular disease (insufficiency or stenosis) and coronary artery stenosis of large or, more commonly, of small vessels[6].

The molecular events associated with FAD consist in the accumulation of the sphingolipid Globotriaosylceramide (Gb3), caused by mutations within the *GLA* gene that lead to a deficiency of α-galactosidase A enzymatic activity. Although clinical manifestations are well explored, molecular mechanisms linking the intracellular deposition of Gb3 to the cell and tissue dysfunction and finally to the clinical manifestations are still not sufficiently clarified. The most accredited hypothesis is centered on the cascade effects triggered by the lysosomal Gb3 over accumulation on structural cellular change, tissue defects, and organ failure[7].

The hydrolytic lysosomal enzyme α-galactosidase A is encoded by the *GLA* gene (Garman, 2007) localized on the chromosomal region X22.1. The protein is expressed in human skin fibroblasts as a glycosylated precursor of 50 kDa. Following delivery into lysosomes, the α-galactosidase A precursor is processed into the mature 46 kDa form[8]. The X-ray structure revealed that α-galactosidase A is a homodimeric glycoprotein[9] (Fig.2.1.2).

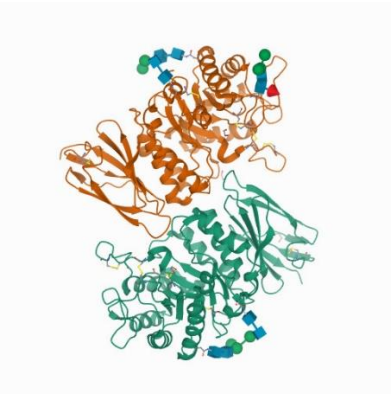

Fig 2.1.2 α-galactosidase A structure by X-Ray diffraction[10]

Each monomer has three potential N-linked carbohydrate attachment sites (at N139, N192, and N215). The adding of mannose 6-phosphate occurs on the N192 and N215, and it is responsible for the targeting of the enzyme to lysosomes[10]. Once the enzyme reaches lysosomes, it degrades its substrates, such as Gb3 (Fig. 2.1.3).

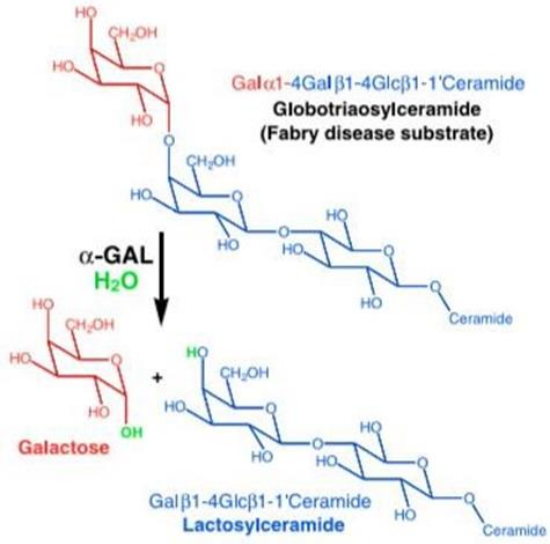

Fig. 2.1.3 The reaction catalyzed by α-galactosidase. Globotriaosylceramide is cleaved by the enzyme to form lactosylceramide and galactose.

Based on the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff (http://www.hgmd.cf.ac.uk/ac/index.php), there are currently 844 mutations described. The largest part is represented by 590 missense/nonsense type mutations, 114 small deletions, 34 large deletions, 41 splice defects, 40 small insertions. Some mutations can cause a total absence of a-galactosidase A activity while, in other cases, a residual enzyme activity ranging from 2% to 25% is preserved[11]. Finally, many mutations do not directly affect the catalytic properties but rather the folding of the protein, destining it to retention in unforeseen cellular districts or, in more severe cases, to proteasomal degradation. To date, several therapies have been developed for FAD cure. Some of them are based on the replacement of the defective gene or enzyme or by directly targeting the organs affected by the accumulation of undegraded metabolites[12].

Furthermore, when the mutations affect the enzyme folding and stability, the mutant enzyme may be stabilized and protected from degradation[3] by using pharmacological chaperones[13] (Fig. 2.1.4). In this regard, 1-deoxygalactonojirimycin (Migalastat), a potent competitive inhibitor of α-galactosidase A, works as a folding template, stabilizing native protein conformation and allowing the normal trafficking towards lysosomes where it dissociates[14].

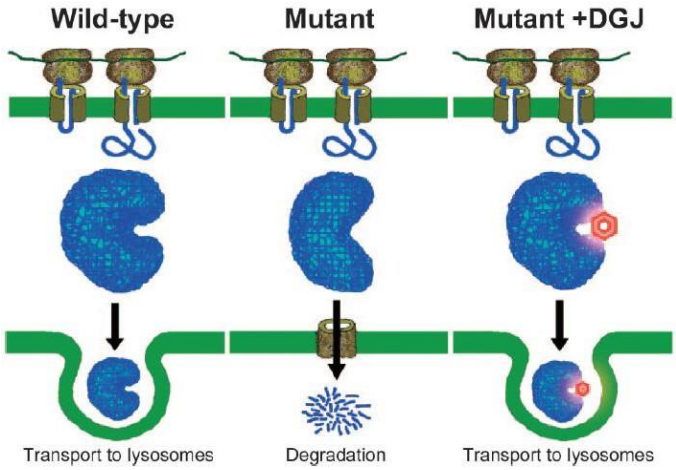

Fig 2.1.4 Schematic representation of the effect of migalastat on mutant α-galactosidase A in the ER. The newly WT synthesized protein folds correctly and reaches the lysosomes. The mutant enzyme does not fold correctly, and it is degraded. In the presence of migalastat, the mutant enzyme folds appropriately and reaches the lysosomes[13].

At present, the most successful treatment available for the Anderson-Fabry disease consists in the Enzyme Replacement Therapy (ERT)[12], in which a recombinant form of the lysosomal enzyme is administered to patients. Two therapeutic forms of the enzyme have been licensed: agalsidase alfa (Replagal®; Shire Human Genetic Therapies, Inc.) and agalsidase beta (Fabrazyme®; Genzyme Corp.), which are given concomitantly with treatments for specific disease-related manifestations[15]. Agalsidase alfa is a glycosylated protein with both sialic acid and mannose-6-phosphate residues. It is produced in a genetically engineered continuous human cell line[16]. The recombinant enzyme agalsidase beta is produced using recombinant DNA technology in mammalian Chinese hamster ovary (CHO) cell culture[17]. The two recombinant proteins share the amino acid sequence of native α-galactosidase A but differ for the glycosylation pattern[18]. Both enzyme preparations are approved in Europe and many other countries, but in the United States, the Food and Drug Administration (FDA) approved only

agalsidase beta. One licensed dose has been recommended for each drug: 0.2mg/kg body weight for Replagal® and 1.0 mg/kg body weight for Fabrazyme®, administration of ERT is through intravenous infusion every other week[19].

The effects of enzyme replacement therapy for FAD have been documented in adult hemizygous patients. In these patients, ERT reduced glycolipid storage in various organs and tissues, decreased pain, improved peripheral nerve function and sweating, and appeared to reduce cardiac hypertrophy[20]. Nevertheless, ERT presents several potential limitations: 1) it has limited tissue penetration; 2) incomplete reversion of FAD pathology and clinical manifestations, 3) it may induce infusion adverse reactions; 4) it may induce the production of anti-drug antibodies with neutralizing effect, reducing the efficacy of the therapy; 5) it is a lifelong therapy requiring intravenous administration every two weeks; and 6) it is associated to a high cost[21].

The present thesis project, carried out in collaboration with Professor Vittoria Cubellis (Department of Biology, University of Naples "Federico II"), was focused on the investigation of the molecular mechanisms involved in the internalization process of the two recombinant enzymes employed for FAD ERT: the agalsidase alfa (Replagal®; Shire Human Genetic Therapies, Inc.) and agalsidase beta (Fabrazyme®; Genzyme Corp.). The strategy employed relies on a functional proteomics approach for the isolation and identification of protein partners of the two recombinant enzymes. Preliminarily, the wild-type GLA interactome was also analyzed in normal fibroblasts to identify the proteins involved in the endogenous enzyme traffic.

2.1.2 Experimental methods

2.1.2.1 Cell lysis and protein quantification

Protein extracts were prepared starting from both fibroblasts with the wild type (WT) gene for the preliminary experiment and from fibroblasts defective (F83) of the α- galactosidase A from a patient with Fabry-Anderson disease, those fibroblasts were incubated for three hours with the recombinant enzymes agalsidase alfa (Replagal®; Shire Human Genetic Therapies, Inc.) and agalsidase beta (Fabrazyme®; Genzyme Corp.) with a concentration of 3µg/ml. The lysis step was carried out by adding a lysis buffer to the cell pellets. The buffer composition was the following:

- 50mM Tris HCl pH=6.5 (Bio-Rad, Hercules, California, US);
- 150mM NaCl (Sigma-Aldrich, Saint Louis, Missouri, US);
- 0.1% Triton X-100 (Bio-Rad);
- 2.5mM KCl (Sigma-Aldrich);
- 1 mini pill of the cocktail of protease inhibitors (Roche, Basilea, SW) for 10 mL of lysis buffer.

A ratio of 1:5 (v/v) pellet/buffer was used for the lysis of each cell pellet. After buffer addition, pellets were left for 10 minutes in ice and successively shacked on the wheel at 4°C for 30 minutes. Finally, the suspensions were centrifuged for 30 minutes at 13000 rpm to separate protein extracts from debris. Bradford assays were carried out for a quantitative evaluation of protein extracts. Bovine serum albumin (BSA, Biorad) was used as a standard for the calibration curve construction.

2.1.2.2 Isolation of protein complexes by immunoprecipitation

Proteins extract from different fibroblast preparations (Fibroblasts WT for the GLA gene, F83 treated with Replagal, F83 treated with Fabrazyme, F83 not treated) were subjected to the immunoprecipitation procedure. As the first step, cell extracts were pre-cleared to remove unspecific proteins adsorbed on beads. Pre-clearing was performed by incubating protein extracts with Dynabeads Protein-G (Thermo Fisher Scientific, Waltham, Massachusetts, US), previously conditioned in lysis buffer, for 2h at 4°C on the wheel. Pre-cleared extracts were then incubated with a specific rabbit polyclonal antibody anti-GLA (Thermo Fisher Scientific), using 2 μg of antibody per mg of protein extract overnight at 4°C on the wheel. Finally, the samples were transferred in new tubes and incubated with Dynabeads Protein-G for 3h at 4°C under gentle agitation on the wheel to capture antigen-antibody complexes. The unbound proteins were then removed, and the beads were repetitively washed with lysis buffer containing either low (150mM) or high (300mM) NaCl concentrations to remove unspecific proteins. Proteins retained were eluted by using Glycine (0,1 M pH=2.5, Bio-rad) for 10 min at 25°C. Beads employed in the pre-cleaning steps were treated in parallel with the same procedure, and the eluted proteins constituted the negative control for the preliminary experiment.

2.1.2.3 Preparative SDS-PAGE

Laemmli Buffer (100mM Tris HCl pH=6,8, 4% SDS (Bio-Rad), 20% glycerol (Sigma-Aldrich), bromophenol blue (Bior-Rad)) with 400mM DTT (dithiothreitol, Sigma-Aldrich) was added to the Glycine elution in a ratio of 1:3 (v/v) buffer/elution, 10 minutes of boiling at 99°C followed. Sample eluates were loaded on a 16x16 cm, 8-15% gradient acrylamide/bis-acrylamide gel for SDS-PAGE fractionation. The electrophoretic runs were performed at an initial voltage of 200 V. The gel was stained with Colloidal Blue Coomassie for a couple of hours, and the excess of dye was removed by extensive washing with deionized

water.

2.1.2.4 In situ hydrolysis

After electrophoresis and staining, the whole lanes were cut in 64 bands for the preliminary experiment and 96 bands for the experiment with both recombinant enzymes by scalpel, chopped, and placed in tubes. The bands were dehydrated by acetonitrile (ACN, Honeywell chemicals, Charlotte, North Carolina, US), removed following agitation with vortex and fast centrifugation.

Protein bands were treated with 80µL of 10mM DTT in 50mM ammonium bicarbonate (NH_4HCO_3, Sigma-Aldrich) for 45 min at 56°C to reduce cysteine residues involved in disulfide bridges. At the end of the reduction, the bands were first dehydrated with acetonitrile and then rehydrated with 80µL of 50mM NH_4HCO_3, containing 55mM iodoacetamide for the cysteine residues alkylation reaction. This latter was carried out in the dark at room temperature for 30 minutes. The excess of reagent was finally removed by washing with acetonitrile and 50mM NH_4HCO_3 alternatively, and the washings were repeated until the complete destaining of gel bands. The dehydrated gel bands were then treated with 20µL of 10ng/µL trypsin (Sigma-Aldrich) solution in 50mM NH_4HCO_3 pH=8 at 4°C, for 1h. Finally, a supplemental volume of 50mM NH_4HCO_3 was added to cover the gel bands, and samples were placed overnight at 37°C. At the end of the hydrolysis reaction, samples were centrifuged at 13200 rpm for a couple of minutes, and the supernatants were collected and placed in a tube. Each sample was acidified by 20% trifluoroacetic acid (TFA, Sigma-Aldrich). Then the remaining gel pieces were shrunk with 50µL of acetonitrile to extract any peptides still present in the gel. The extracts were joined to the respective supernatants, previously removed. A second extraction was performed, adding 20µL of 0,2% formic acid (HCOOH, Chem-Lab, Eernegem, BE) followed by acetonitrile to gel bands, and the supernatants recovered and joined to respective peptide mixtures.

The obtained mixtures were dried by a Speed-Vac system.

2.1.2.5 LC-MS/MS analysis

Each peptide mixture was resuspended in 0.2% HCOOH and analyzed by nanoLC-MS/MS, using a system LTQ Orbitrap XL equipped with a Proxeon-nanoEasy II nano-HPLC. After loading, the peptide mixture was first concentrated and desalinated in the pre-column (C18 EasyColumn L=2cm, 5µm, ID=100µm, Thermo Fisher Scientific). Each peptide sample was then fractionated on a C18 reverse-phase capillary column (C18 Column L=20cm, ID=75µm, 5µm, NanoSeparations, Nieuwkoop, Netherlands) working at a flow rate of 250nL/min, using a step gradient of eluent B (0.2% formic acid, 95% acetonitrile LC-MS Grade) from 10 to 60% over 69 minutes and 60 to 95% over 3 minutes.

Peptide analysis was performed using data-dependent acquisition (DDA) of one MS scan (mass range from 400 to 1800 m/z) followed by MS/MS scans of the five most abundant ions in each MS scan, obtained by CID (Collision Induced Fragmentation) fragmentation in the ion trap.

2.1.2.6 Proteins identification

Raw data from nano-LC-MS/MS analyses were processed and converted into .mgf files to be introduced into the MASCOT software (Matrix Science Boston, USA) to search a non-redundant protein database. Both peptide mass data and the data obtained by fragmentation spectra were included in the peak list for protein identification. The research was done by setting NCBI as protein database, carbamidomethyl (C) as fixed modification, oxidation (M), pyro-Glu (N-termQ), pyro-carbamidomethyl (N-term C) as variable modifications, 10 ppm as MS tolerance, and 0.6 Da as MS/MS cutoff tolerance.

ExPASy Blast Form software was also used to assign to each protein a code that uniquely identifies each sequence in the annotated database UniProt KB.

2.1.3 Results and Discussion

The present study concerning the description of the internalization pathways of two recombinant GLA proteins, active principles of Replagal® (agalsidase alfa) and Fabrazyme® (agalsidase beta), was carried out by employing a functional proteomic approach for the isolation and identification of proteins involved in the traffic of exogenous enzymes. Replagal® and Fabrazyme® are alternatively used in ERT of FAD cure. The same strategy was primarily applied for identifying proteins interacting with endogenous α-galactosidase A in health fibroblasts to describe the route followed by the hydrolase from the endoplasmic reticulum (ER) to lysosomes.

2.1.3.1 Intracellular trafficking of endogenous GLA

The investigation of the interactome of the endogenous α-galactosidase A was investigated in normal human fibroblasts, which were lysed, and the cellular extract was quantified as reported in section 2.1.2.1.

6 mg of total protein extract were immunoprecipitated by using an anti-GLA antibody. A pre-cleaning step was preliminarily performed on the sample to remove proteins with affinity for the Dynabeads-G protein. The pre-cleaned extract was incubated with an anti-GLA antibody overnight and then for 2 hours with the Dynabeads derivatized with the G protein. The supernatant containing the unbound proteins was removed, and beads were washed with lysis buffer.

The proteins bound to the α-galactosidase A and the proteins bound to the pre-cleaning beads, used ad control, were eluted using Glycine at pH=2.5. Laemmli buffer with DTT was added, and an SDS-PAGE was performed to separates the proteins; the gel was finally stained by colloidal blue Coomassie. Thirty-two protein bands for the control (pre-cleaning eluate) and 32 bands for the sample lanes (IP) were excised from the gel (Fig. 2.1.5), *in situ* digested with trypsin, and

the resulting peptide mixtures were analyzed by LC-MS/MS, using an LTQ Orbitrap XL mass spectrometer coupled with a nano-HPLC.

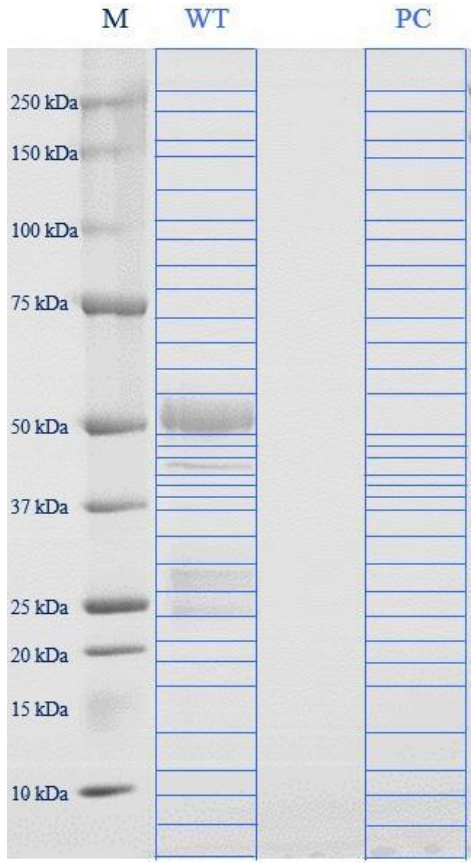

Fig. 2.1.5 Preparative SDS-PAGE gel: the grid shows the excised bands for the proteins of the IP sample (WT) and the pre-cleaning eluate (PC) used as control.

The data obtained by the LC-MS/MS analysis have been processed by Mascot for protein identification.

The lists of proteins identified in the sample (WT) and the control (PC) have been compared, and the proteins present in both have been eliminated. The list of the putative partners of the endogenous α-galactosidase A is reported in the following table 2.1.2, with the total number of peptides identified for each protein.

	Protein Name	Gene	Uniprot code	Peptides
> 250 kDa	Plectin	PLEC	Q15149	76 (49)
250 kDa	Spectrin alpha chain, non-erythrocytic 1	SPTAN1	Q13813	56(40)
	Spectrin beta chain, non-erythrocytic 1	SPTBN1	Q01082	44(35)
250<kDa<150	DNA repair protein RAD50	RAD50	Q92878	15(6)
	LIM domain only protein 7	LMO7	Q8WWI1	12(8)
	Aminopeptidase N	ANPEP	P15144	6(2)
150<kDa<100	Protein transport protein Sec31A	SEC31A	O94979-4	25(13)
	Ankycorbin	RAI14	Q9P0K7	16(14)
	Unconventional myosin-Ib	MYO1B	O43795-2	10(5)
	Unconventional myosin-Ic	MYO1C	O00159	18(8)
	Alpha-actinin-1	ACTN1	P12814	11(6)
	Alpha-actinin-4	ACTN4	O43707	8(4)
	Unconventional myosin-Id	MYO1D	O94832	3(1)
100<kDa<75	Gelsolin isoform 2	GSN	P06396	7(2)
	Endoplasmic reticulum chaperone BiP	HSPA5	P11021	7(2)
75kDa	Stress-70 protein, mitochondrial	HSPA9	P38646	6(5)
	Heat shock cognate 71 kDa protein	HSPA8	P11142	6(4)
75<kDa<50	5'-nucleotidase	NT5E	P21589	10(5)
	RAC-alpha ser/thr-protein kinase	AKT1	P31749	8(4)
	Pyruvate kinase PKM	PKM	P14618	6(2)
	Vimentin	VIM	P08670	51(40)
	Tubulin alpha-1A chain	TUBA1A	Q71U36	6(4)
	Polyadenylate-binding protein 1	PABPC1	P11940-2	8(3)
50kDa	Tubulin beta chain	TUBB	P07437	12(10)
	Tubulin alpha-1B chain	TUBA1B	P68363	7(6)
	Pre-rRNA-processing protein TSR1 homolog	TSR1	Q2NL82	4(3)

	Bystin	BYSL	Q13895	3(2)
50<kDa<37	E3 ubiquitin-protein ligase TRIM21	TRIM21	P19474	9(4)
	Tropomodulin-3	TMOD3	Q9NYL9	10(7)
	Serpin H1	SERPINH1	P50454	8(7)
	Phosphoglycerate kinase 1	PGK1	P00558	7(4)
	Guanine nucleotide-binding protein G(s) subunit alpha isoforms short	GNAS	P63092	6(5)
	Fructose-bisphosphate aldolase A	ALDOA	P04075	7(3)
	Guanine nucleotide-binding protein G(i) subunit alpha-2	GNAI2	P04899	5(3)
	Caveolae-associated protein 3	CAVIN3	Q969G5	7(4)
37kDa	Methylmalonic aciduria type A protein, mitochondrial	MMAA	Q8IVH4	11(5)
	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	P04406	5(2)
	Protein SEC13 homolog	SEC13	P55735	4(3)
	Serine/threonine-protein phosphatase PP1-alpha catalytic subunit	PPP1CA	P62136	2(2)
37<kDa<15	Annexin A2	ANXA2	P07355	15(13)
	L-lactate dehydrogenase A chain	LDHA	P00338	7(4)
	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2	GNB2	P62879	4(4)
	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1	GNB1	P62873	5(5)
	Golgi-associated plant pathogenesis-related protein 1	GLIPR2	Q9H4G4	2(2)
	Myosin light polypeptide 6	MYL6	P60660	6(6)
	Galectin-1	LGALS1	P09382	4(4)

10kDa	Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma	GNG12	Q9UBI6	4(3)
<10kDa	Apolipoprotein D	APOD	P05090	3(3)

Tab. 2.1.2 List of putative interactors of the endogenous α-galactosidase A. The table reports proteins' names, genes' names, the Uniprot code, and the number of peptides (in the brackets those overcoming the Mascot significant statistical threshold). In the brackets, the number of peptides overcoming the Mascot significant statistical threshold is also reported

The 48 putative α-galactosidase A interacting proteins were then classified according to their primary biological function using Gene Ontology, STRING, and/or information reported in the literature (Fig.2.1.6).

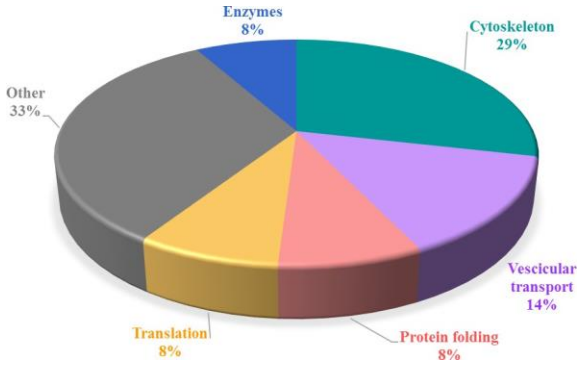

Fig. 2.1.6 Functional classification of WT α-galactosidase A interacting proteins based on protein databases and literature information.

A large number of α- galactosidase A putative partners are associated with the cytoskeleton or are involved in vesicular transport, suggesting that GLA traffic within the cell is connected to these processes.

Among the proteins included in the vesicular transport category, we found several unconventional myosins, such as MYO1B, MYO1C, and MYO1D, a class of single-headed myosin motors, that participate in exocytosis, endocytosis, and trans-Golgi network trafficking by tethering vesicles to the cortical actin filaments[22].

MYO1B is known to be mainly involved in the traffic of cargo along the endocytic pathway. It has also been localized at the plasma membrane in regions enriched for actin filaments and at early endosomes, multivesicular endosomes, and lysosomes. However, a large amount of MYO1B has been found to co-localize partially with Trans-Golgi Network (TGN) and CI-mannose-6-phosphate receptor (CI-MPR) that carries cargos from the TGN to sorting endosomes and recycles back to TGN[23]).

MYO1C has a rigid conformation that is likely to be necessary for the motor to tether exocytic/endocytic vesicles to membranes before the fusion events. Notably, upon Ca^{2+} concentration increases, myosin-Ic undergoes substantial conformational changes and becomes more flexible, as displayed in Fig 2.1.7[22].

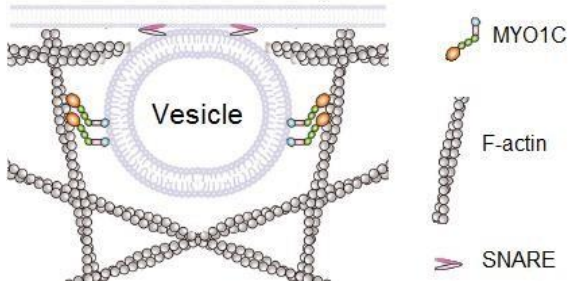

Fig. 2.1.7 Schematic model showing that Myosin-Ic is involved in tethering exocytic vesicles between actin filaments[22].

In addition to cargo loading-induced activity changes, myosins are also regulated by a class of their intimate binding partners, myosin light chains[22]; in this regard, we also found MYL6 (myosin light chain 6).

The functional subset of vesicular transport also included SEC13 and SEC31A. These proteins are involved in the formation of the COPII coated vesicles. SEC31A and SEC13 interact, forming a dimer, which in turn interacts with another SEC31/13 dimer, thus constituting heterotetramers[24]. The COPII coat generates transport vesicles that load newly synthesized proteins and transport

them from the endoplasmic reticulum (ER) to the Golgi apparatus[25].

In the slice “cytoskeleton” of the pie chart, we included the proteins that belong to the intermediate filaments and microtubules, which have also been found. Vimentin, a class III intermediate filament, interacts with actin to contribute to cellular stiffness. Notably, the role of vimentin in regulating intracellular trafficking has also been demonstrated, and several reports document a strong connection between vimentin and late endocytic trafficking[26] and an association between vimentin IFs and clathrin-coated vesicles[27].

Plectin belongs to the cytolinkers and scaffolding proteins. Plectin is a ubiquitous large phosphoprotein that can bind actin on its N-terminus and vimentin on its C-terminus[28]. It cross-links intermediate filaments (IFs), microfilaments, and microtubules with each other and further connects these cytoskeletal networks to the plasma membrane, it also plays an essential role as a regulator of cellular processes linked to actin filament dynamics[29].

Besides actin-interacting proteins, also tubulins (TUBB and TUBA1B) were identified. Tubulin proteins compose microtubules that are hollow tubes[30] on which microtubule motors, dynein, and the kinesins, can move vesicles significantly fast[31].

Gelsolin was first described as a protein able to bind and sever actin filaments, to control polymerization of barbed ends, and to promote actin cytoskeleton remodeling. Its activity is regulated by Ca^{2+} and intracellular pH[32]. Bär *et al*. in 2008 showed that gelsolin activity is required not only to trim actin filaments but also to drive virus export from the nucleus to the outside of infected cells[33]. This transport is mediated by vesicles bearing protein markers of lysosomes and/or late endosomes, suggesting that gelsolin may play a role in the formation, loading, and/or trafficking of these vesicles.

Looking at the functional categories reported in the pie chart, the class of proteins involved in folding is also quite populated. This class includes HSPA5 (known as BiP), HSPA8, and HSPA9 (known as "mortalin"), all belonging to the HSP70 family[34]. These proteins control all aspects of cellular proteostasis, such as nascent protein chain folding, protein export towards different organelles, recovering of proteins from aggregation, and assembly of multi-protein complexes[35]. BiP is located predominantly in the ER and serves as a regulator of the unfolded proteins of ER or secretory proteins, and it helps the assembly of protein complexes. It also serves as an ER stress sensor and targets misfolded proteins for ERAD. HSPA8 was discovered as an uncoating ATPase catalyzing the ATP-dependent uncoating of clathrin-coated pits. It also regulates protein homeostasis and translocation[34]. Both these proteins directly or indirectly drive nascent proteins to their final location, starting from the ER, where they are synthesized.

Based on the function and localization of the proteins identified as GLA interactors, it is possible to speculate about the route followed by the endogenous enzyme when it moves from the endoplasmic reticulum to the lysosome (Fig. 2.1.8).

The newly synthesized protein into the ER meets the chaperone BiP that helps it to fold correctly. In the following step, the α-galactosidase A is packed in COPII vesicles that travel from the ER to the cis-Golgi. The protein leaves the trans-Golgi (TGN) in clathrin-coated vesicles, as suggested by the identification of HSPA8 and MYO1B. The exit of GLA from TGN might either involve Gelsolin, as well as also demonstrated in a previous study, addressed to the identification of protein associated with the α-glucosidase traffic. The vesicles then reach the endosome where, with Vimentin and Plectin help, GLA reaches the lysosome.

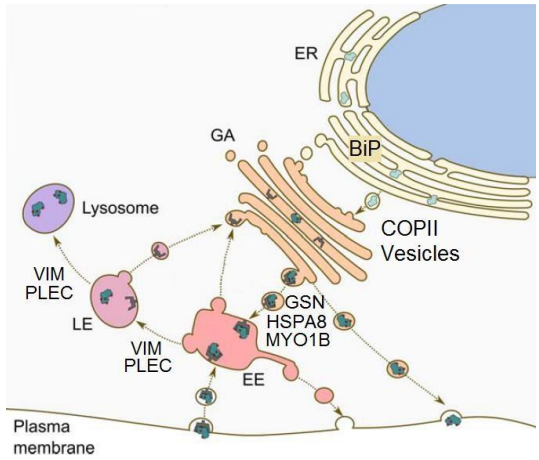

Fig. 2.1.8 The hypothesized pathway of GLA from the endoplasmic reticulum to the lysosome.

2.1.3.2 Investigation of trafficking of exogenous recombinants GLA

The study of the partners involved in the uptake and the trafficking of the two recombinant enzymes Replagal® and Fabrazyme® was carried out in human fibroblasts (F83), not expressing the endogenous enzyme and representing a model of FAD.

Fibroblasts were harvested after being incubated with each recombinant enzyme for three hours at a concentration of 3µg/mL (enzyme/cell medium), and each cellular extract was quantified as previously explained and as reported in section 2.1.2.1. Fibroblasts not treated were employed as control.

5 mg of each protein extracts were immunoprecipitated by using an anti-GLA antibody. A 2 hours pre-cleaning step was performed on the samples to remove unspecific proteins with affinity for Dynabeads- G protein.

The anti-GLA antibody was added to the pre-cleaned extract overnight, followed by 3 hours of incubation with Dynabeads- G protein; the unbound proteins were removed, and beads were washed with lysis buffer. The retained proteins were eluted using Glycine at pH=2.5, separated by SDS-PAGE, and the gel stained by

colloidal blue Coomassie. Thirty-two protein bands for each sample were excised from the gel (Fig. 2.1.9), *in situ* hydrolysis with trypsin followed, and the resulting peptide mixtures were analyzed by LC-MS/MS.

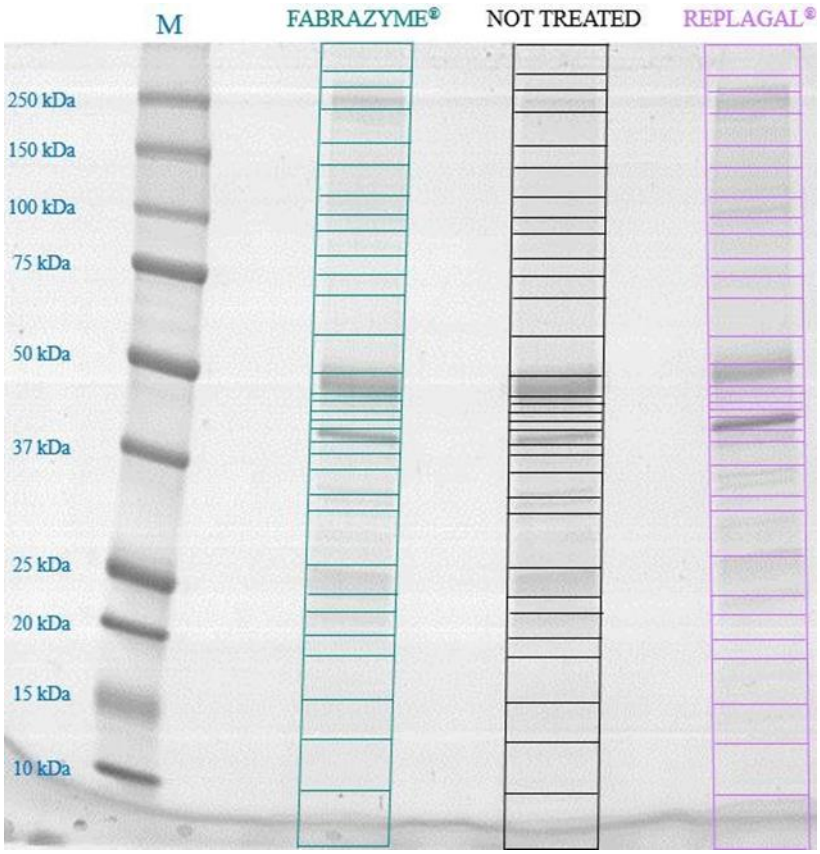

Fig. 2.1.9 Preparative SDS-PAGE gel: the grid shows the excised bands for the proteins of each sample.

The software Mascot was used for protein identification.

The proteins identified in the control were discarded from both sample lists. The remaining proteins, putative interactors of the two recombinant enzymes, are following reported (Table 2.1.3).

	Protein Name	Gene	Uniprot code	Peptides Replagal®	Peptides Fabrazyme®
>250kDa	Plectin	PLEC	Q15149-9	40(4)	28(10)
	Utrophin	UTRN	P46939	8(3)	7(3)
250kDa	Tensin-1	TNS1	Q9HBL0	12(5)	5(2)
	Fibronectin	FN1	P02751	11(3)	14(4)
	Unconventional myosin-Va	MYO5A	Q9Y4I1	7(3)	/
	Poly [ADP-ribose] polymerase 4	PARP4	Q9UKK3	3(2)	5(3)
250<kDa<150	Triple functional domain protein	TRIO	O75962-5	/	4(2)
	Unconventional myosin-XVIIIa	MYO18A	Q92614	7(1)	17(10)
	A-kinase anchor protein 2	AKAP2	Q9Y2D5-4	7(2)	5(3)
150kDa	Tight junction protein ZO-2	TJP2	Q9UDY2	20(6)	13(8)
	Uveal autoantigen with coiled-coil domains and ankyrin repeats	UACA	Q9BZF9	19(5)	15(6)
	Band 4.1-like protein 2	EPB41L2	O43491	14(2)	9(4)
	Coatomer subunit alpha	COPA	P53621	4(2)	/
150<kDa<100	Tight junction protein ZO-1	TJP1	Q07157	28(8)	/
	Unconventional myosin-VI	MYO6	Q9UM54-6	25(9)	17(5)
	Leucine zipper protein 1	LUZP1	Q86V48-2	11(2)	/
	Kinesin-1 heavy chain	KIF5B	P33176	/	6(5)
	Neurabin-2	PPP1R9B	Q96SB3	15(7)	/
100 kDa	Unconventional myosin-Id	MYO1D	O94832	34(16)	27(13)
	Myosin phosphatase Rho-interacting protein	MPRIP	Q6WCQ1	21(8)	8(6)
	Actin filament-associated protein 1	AFAP1	Q8N556	16(6)	/
	AP-2 complex subunit beta	AP2B1	P63010	14(1)	10(2)
	Drebrin	DBN1	Q16643	10(4)	8(4)

	AP-2 complex subunit alpha-1	AP2A1	O95782	11(1)	5(4)
	Phostensin	PPP1R18	Q6NYC8	8(2)	/
	Alanine--tRNA ligase, cytoplasmic	AARS	P49588	/	7(2)
	Catenin alpha-1	CTNNA 1	P35221	/	2(2)
	Protein flightless-1 homolog	FLII	Q13045	5(3)	/
100<kDa<75	Ankycorbin	RAI14	Q9P0K7-4	34(10)	10(1)
	Junctional protein associated with coronary artery disease	JCAD	Q9P266	14(3)	/
	Unconventional myosin-Ie	MYO1E	Q12965	10(2)	/
	Unconventional myosin-Ib	MYO1B	O43795-2	5(1)	/
	Palladin	PLLD	Q8WX93	6(1)	/
	Calnexin	CANX	P27824-2	5(2)	/
	LIM domain and actin-binding protein 1	LIMA1	Q9UHB6	23(11)	12(8)
	LIM domain only protein 7	LMO7	Q8WWI1-3	21(9)	6(3)
	ATP-dependent 6-phosphofructokinase, platelet type	PFKP	Q01813	/	19(9)
	X-ray repair cross-complementing protein 5	XRCC5	P13010	/	10(2)
	Caldesmon	CALD1	Q05682	14(6)	/
	Ezrin	EZR	P15311	12(3)	/
	Nexilin	NEXN	Q0ZGT2	10(1)	5(1)
	Far upstream element-binding protein 2	KHSRP	Q92945	8(4)	6(3)
	Protein transport protein Sec23A	SEC23A	Q15436	8(2)	5(2)

	Unconventional myosin-Ic	MYO1C	O00159-3	8(2)	/
	DNA replication licensing factor MCM7	MCM7	P33993	5(3)	/
	Vacuolar protein sorting-associated protein 35	VPS35	Q96QK1	5(2)	/
	Protein S100-A8	S100A8	P05109	3(2)	/
	Threonine--tRNA ligase, cytoplasmic	TARS	P26639	3(2)	/
	Calpain-1 catalytic subunit	CAPN1	P07384	3(1)	/
75 kDa	Synaptopodin-2	SYNPO2	Q9UMS6	25(12)	10(4)
	TRIO and F-actin-binding protein	TRIOBP	Q9H2D6-3	18(3)	/
	Major vault protein	MVP	Q14764	12(3)	/
	Sorting nexin-18	SNX18	Q96RF0-2	10(5)	7(4)
	Optineurin	OPTN	Q96CV9	/	6(1)
	Calpain-2 catalytic subunit	CAPN2	P17655	/	5(3)
	Fermitin family homolog 2	FERMT2	Q96AC1	/	5(3)
	Glycine--tRNA ligase	GARS	P41250	10(2)	4(2)
	Trifunctional enzyme subunit alpha, mitochondrial	HADHA	P40939	8(1)	/
	ATP-dependent RNA helicase DDX3X	DDX3X	O00571	5(2)	/
	Probable ATP-dependent RNA helicase DDX17	DDX17	Q92841	4(4)	/
	BAG family molecular chaperone regulator 3	BAG3	O95817	3(2)	/
	Synaptopodin	SYNPO	Q8N3V7	3(1)	/

75<kDa<50	Protein phosphatase 1 regulatory subunit 12A	PPP1R12A	O14974	15(5)	/
	Dihydropyrimidinase-related protein 3	DPYSL3	Q14195-2	10(5)	/
	Cytoskeleton-associated protein 4	CKAP4	Q07065	11(4)	8(3)
	Plastin-3	PLS3	P13797	10(3)	10(6)
	EH domain-containing protein 2	EHD2	Q9NZN4	8(3)	9(4)
	5'-nucleotidase	NT5E	P21589	9(2)	4(3)
	1,4-alpha-glucan-branching enzyme	GBE1	Q04446	8(3)	8(4)
	WD repeat-containing protein 1	WDR1	O75083	/	18(7)
	Transketolase	TKT	P29401	/	13(6)
	T-complex protein 1 subunit gamma	CCT3	P49368	/	13(7)
	TRIO and F-actin-binding protein (ISOFORM 1)	TRIOBP	Q9H2D6	/	11(4)
	Protein LTV1 homolog	LTV1	Q96GA3	/	9(5)
	Coatomer subunit delta	ARCN1	P48444	/	10(4)
	Actin, cytoplasmic 1	ACTB	P60709	/	8(5)
	Prolyl 4-hydroxylase subunit alpha-2	P4HA2	O15460	/	7(4)
	Pre-rRNA-processing protein TSR1 homolog	TSR1	Q2NL82	/	6(2)
	Bifunctional purine biosynthesis protein PURH	ATIC	P31939	/	5(2)
	T-complex protein 1 subunit zeta	CCT6A	P40227	/	5(2)

	Peptidyl-prolyl cis-trans isomerase FKBP10	FKBP10	Q96AY3	/	5(2)
	Probable ATP-dependent RNA helicase DDX5	DDX5	P17844	/	4(2)
	Phosphoglucomutase-1	PGM1	P36871	/	5(2)
	26S proteasome non-ATPase regulatory subunit 3	PSMD3	O43242	/	3(2)
	Heterogeneous nuclear ribonucleoprotein K	HNRNPK	P61978	8(2)	13(5)
	Peptidyl-prolyl cis-trans isomerase FKBP9	FKBP9	O95302	5(2)	/
	Serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform	PPP2R1A	P30153	6(2)	7(4)
	Actin-binding LIM protein 3	ABLIM3	O94929	5(2)	12(7)
	Phosphatidylinositol-binding clathrin assembly protein	PICALM	Q13492	3(1)	/
	Vimentin	VIM	P08670	26(13)	/
	Pyruvate kinase PKM	PKM	P14618	20(15)	24(14)
	Beta-2-syntrophin	SNTB2	Q13425	12(4)	/
	RAC-alpha serine/threonine-protein kinase	AKT1	P31749	12(4)	5(4)
	D-3-phosphoglycerate dehydrogenase	PHGDH	O43175	6(4)	6(5)
	Prolyl 4-hydroxylase subunit alpha-1	P4HA1	P13674	7(5)	9(6)
	T-complex protein 1 subunit theta	CCT8	P50990	7(4)	6(3)

	Inosine-5'-monophosphate dehydrogenase 2	IMPDH2	P12268	7(3)	8(5)
	Actin, cytoplasmic 1	ACTB	P60709	10(2)	/
	Beta-1-syntrophin	SNTB1	Q13884	3(2)	/
	T-complex protein 1 subunit beta	CCT2	P78371	8(2)	9(4)
	UDP-glucose 6-dehydrogenase	UGDH	O60701	6(2)	6(3)
	Tryptophan--tRNA ligase, cytoplasmic	WARS	P23381	4(3)	4(3)
	Nucleosome assembly protein 1-like 4	NAPL1L4	Q99733-2	4(3)	/
	Glucose-6-phosphate 1-dehydrogenase	G6PD	P11413	/	5(4)
	Rab GDP dissociation inhibitor alpha	GDI1	P31150	/	3(2)
	tRNA-splicing ligase RtcB homolog	RTCB	Q9Y3I0	/	4(2)
	Atlastin-3	ATL3	Q6DD88	/	4(2)
	Heat shock protein HSP 90-beta	HSP90AB1	P08238	4(2)	/
50kDa	Tubulin alpha-1B chain	TUBA1B	P68363	22(13)	/
	Leucine-rich repeat flightless-interacting protein 2	LRRFIP2	Q9Y608-4	15(8)	/
	PDZ and LIM domain protein 7	PDLIM7	Q9NR12	14(9)	/
	Alpha-actinin-1	ACTN1	P12814	15(8)	/
	Caveolae-associated protein 1	CAVIN1	Q6NZI2	10(8)	3(3)
	Fascin	FSCN1	Q16658	11(4)	/
	Glutamate dehydrogenase 1, mitochondrial	GLUD1	P00367	12(6)	/

	ATP synthase subunit alpha, mitochondrial	ATP5F1A	P25705	8(5)	/
	Plasminogen activator inhibitor 1 RNA-binding protein	SERBP1	Q8NC51	7(3)	/
	Calreticulin	CALR	P27797	8(2)	/
	cAMP-dependent protein kinase type II-alpha regulatory subunit	PRKAR2A	P13861	5(4)	/
	Annexin A11	ANXA11	P50995	5(2)	/
	Death-associated protein kinase 3	DAPK3	O43293	3(2)	/
	Alpha-aminoadipic semialdehyde dehydrogenase	ALDH7A1	P49419-2	4(2)	/
	Nucleosome assembly protein 1-like 1	NAP1L1	P55209	4(2)	/
50<kDa<37	Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	DLST	P36957	8(4)	/
	RuvB-like 2	RUVBL2	Q9Y230	4(3)	/
	Ribonuclease inhibitor	RNH1	P13489	13(4)	/
	Tubulin alpha-1A chain	TUBA1A	Q71U36	6(3)	5(4)
	Elongation factor Tu, mitochondrial	TUFM	P49411	12(7)	/
37 kDa	Alpha-centractin	ACTR1A	P61163	3(2)	/
	Annexin A7	ANXA7	P20073-2	4(1)	/
	Interleukin enhancer-binding factor 2	ILF2	Q12905	6(2)	/
	Reticulocalbin-3	RCN3	Q96D15	3(3)	/

	Isocitrate dehydrogenase [NADP], mitochondrial	IDH2	P48735	/	4(2)
	PRKC apoptosis WT1 regulator protein	PAWR	Q96IZ0	4(1)	/
	cAMP-dependent protein kinase catalytic subunit alpha	PRKACA	P17612	4(1)	4(1)
	Erlin-2	ERLIN2	O94905	3(1)	/
	Reticulocalbin-1	RCN1	Q15293	/	3(2)
	Mitotic checkpoint protein BUB3	BUB3	O43684	/	4(1)
	Transaldolase	TALDO1	P37837	/	12(2)
	Tropomyosin alpha-1 chain	TPM1	P09493	/	13(8)
37<kDa<25	Succinate--CoA ligase [ADP/GDP-forming] subunit alpha, mitochondrial	SUCLG1	P53597	/	3(2)
	Nascent polypeptide-associated complex subunit alpha	NACA	Q13765	/	2(2)
	Ribose-phosphate pyrophosphokinase 1	PRPS1	P60891	/	2(2)
	Cathepsin Z	CTSZ	Q9UBR2	/	2(2)
	Chloride intracellular channel protein 1	CLIC1	O00299	/	9(6)
	Annexin A4	ANXA4	P09525	/	4(2)
	Four and a half LIM domains protein 1	FHL1	Q13642-5	/	5(3)
	S-formylglutathione hydrolase	ESD	P10768	/	4(2)
	Arginase-1	ARG1	P05089	/	2(2)
	Cathepsin B	CTSB	P07858	/	3(2)
	High mobility group protein B1	HMGB1	P09429	/	8(4)

25kDa	Eukaryotic translation initiation factor 4H	EIF4H	Q15056-2	/	5(1)
	Vesicle-associated membrane protein-associated protein A	VAPA	Q9P0L0	/	3(1)
25<kDa<20	Triosephosphate isomerase	TPI1	P60174	/	16(15)
	Rho GDP-dissociation inhibitor1	ARHGDIA	P52565	/	7(6)
	GTP-binding nuclear protein Ran	RAN	P62826	/	9(7)
	Ubiquitin carboxyl-terminal hydrolase isozyme L1	UCHL1	P09936	/	8(7)
	Glutathione S-transferase P	GSTP1	P09211	/	5(5)
	Protein/nucleic acid deglycase DJ-1	PARK7	Q99497	/	6(3)
	Vesicle-trafficking protein SEC22b	SEC22B	O75396	/	5(2)
	28 kDa heat- and acid-stable phosphoprotein	PDAP1	Q13442	/	3(3)
	High mobility group protein B2	HMGB2	P26583	/	5(3)
	Protein-L-isoaspartate(D-aspartate) O-methyltransferase	PCMT1	P22061	/	4(2)
	Peroxiredoxin-6	PRDX6	P30041	/	4(4)
	Reticulon-3	RTN3	O95197-3	/	4(2)
	Membrane-associated progesterone receptor component 1	PGRMC1	O00264	/	3(2)
	Myosin light chain 6B	MYL6B	P14649	/	4(1)
	Ras-related protein Rab-8A	RAB8A	P61006-2	/	4(1)
	Flavin reductase (NADPH)	BLVRB	P30043	/	2(2)

	Peroxiredoxin-1	PRDX1	Q06830	/	9(2)
	Ras-related protein Rab-1A	RAB1A	P62820	/	6(3)
	GTP-binding protein SAR1a	SAR1A	Q9NR31	/	3(2)
20 kDa	Caveolin-1	CAV1	Q03135	/	6(2)
	Ras-related C3 botulinum toxin substrate 1	RAC1	P63000	/	4(1)
	Fibroblast growth factor 2	FGF2	P09038-3	3(1)	/
20<kDa<15	Calmodulin-3	CALM3	P0DP25	/	4(1)
	Cofilin-1	CFL1	P23528	/	4(2)
	Rho-related GTP-binding protein RhoG	RHOG	P84095	/	3(1)
15 kDa	Golgi-associated plant pathogenesis-related protein 1	GLIPR2	Q9H4G4	3(1)	3(2)
10 kDa	Galectin-1	LGALS1	P09382	3(2)	3(3)
	Protein S100-A4	S100A4	P26447	4(1)	6(4)
	40S ribosomal protein S15a	RPS15A	P62244	/	3(2)
	Protein S100-A11	s100A11	P31949	/	4(2)
<10kDa	40S ribosomal protein S29	RPS29	P62273	/	5(3)
	Ubiquitin-60S ribosomal protein L40	UBA52	P62987	/	6(1)
	40S ribosomal protein S28	RPS28	P62857	/	3(3)

Tab. 2.1.3 List of putative interactors of Replagal® and Fabrazyme®. The table reports the name of the protein, the name of the gene, the Uniprot code, and the number of peptides for both conditions. In the brackets, the number of peptides overcoming the Mascot significant statistical threshold is also reported. The symbol / means that a specific protein was not identified in that condition.

Among the 184 identified proteins, 50 were common to both conditions while 61 were unique for the cells treated with Replagal® and 73 the fibroblasts incubated with Fabrazyme®, as reported in Fig. 2.1.10.

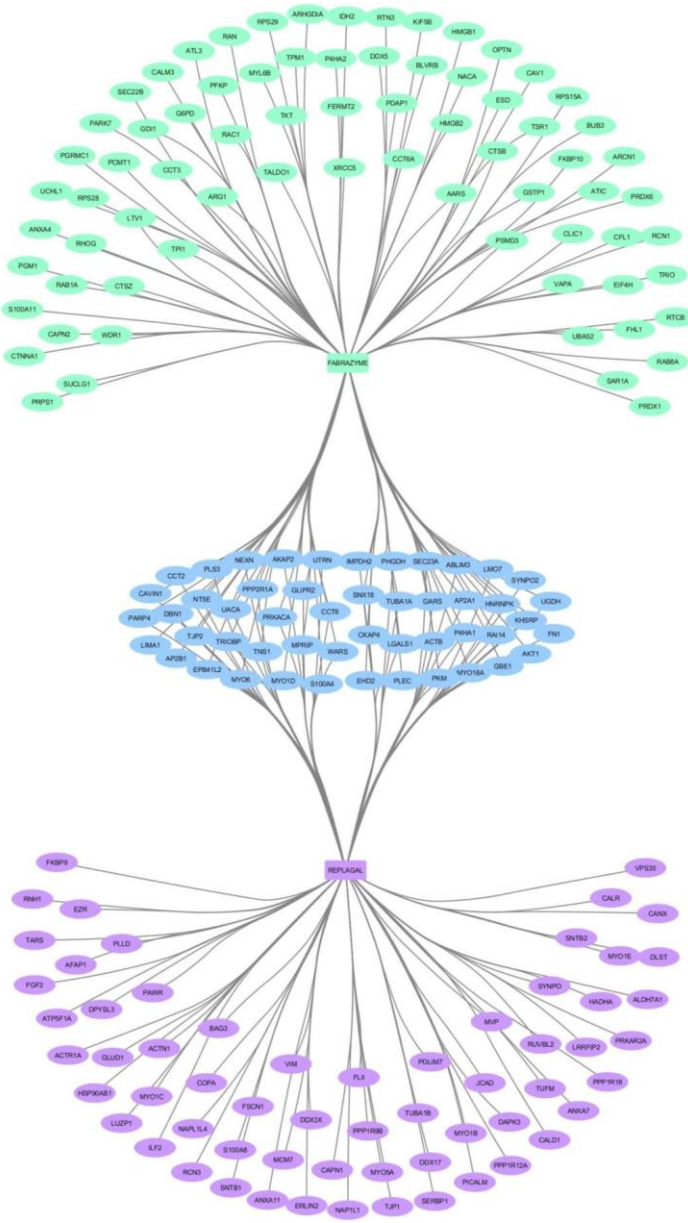

Fig. 2.1.10 Graphical representation of proteins unique for Replagal® in lilac, unique for Fabrazyme® in green, and the ones in common for both conditions in light blue obtained with the software Cytoscape.

The proteins were separately classified according to their main biological functions using bioinformatic tools such as STRING and/or information reported in the literature (Fig. 2.1.11).

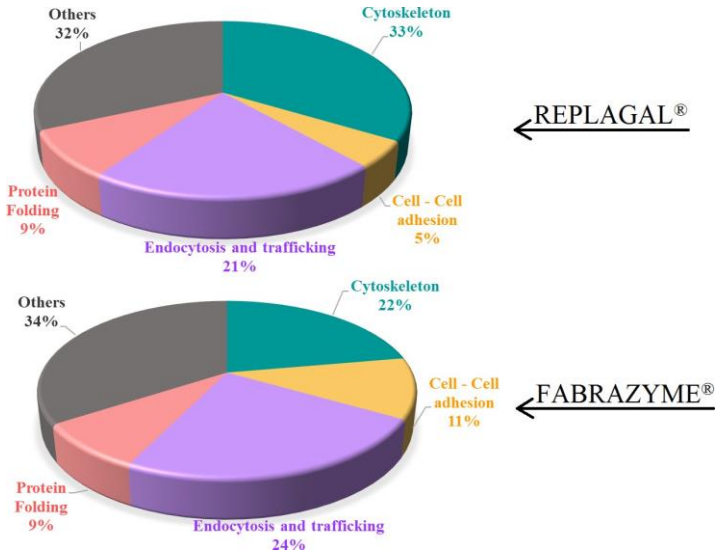

Fig. 2.1.11 Functional classification of the two recombinant α-galactosidase A interacting proteins based on protein databases and literature information. Replagal® pie chart on top and Fabrazyme® pie chart on the bottom.

In both conditions, we found several proteins involved in endocytosis and vesicular trafficking processes and many proteins involved in actin cytoskeleton remodeling and polymerization.

Endocytosis is a key mechanism by which cells collect extracellular material and package it into vesicles that pinch off and enter into the cytosol. Internalized vesicles then fuse with other internal compartments so that their contents can be recycled or degraded[36]. Cells exhibit different endocytic pathways: Clathrin-mediated endocytosis, caveolar-type endocytosis, CLIC/GEEC-type endocytosis, flotillin-dependent endocytosis, phagocytosis, micropinocytosis[37].

Clathrin-mediated endocytosis (CME) is the best characterized of all the

endocytic pathways and involves many actors that concentrate transmembrane cargo proteins and form clathrin-coated vesicles (Fig 2.1.12). In CME, adaptor proteins recognize sorting signals in the cargo proteins and directly link the cargo to clathrin. We found several proteins belonging to this pathway, such as AP2A1, AP2B1, SNX18, and MYO6, in both conditions.

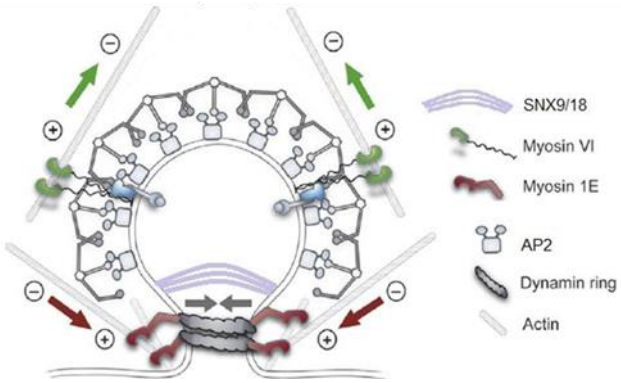

Fig. 2.1.12 Representation of the proteins involved in the formation of clathrin-coated vesicles.

AP2A1 and AP2B1 belong to the adaptor complex AP-2 which localizes at the plasma membrane and facilitates clathrin-mediated endocytosis of a wide range of proteins[38]. AP-2 is involved in the binding of sorting signals in the cytoplasmic tails of cargo proteins, recruiting clathrin and other accessory proteins, and then concentrating the cargo proteins into vesicular carriers. MYO6, a member of the "unconventional" myosins, is an unusual member of its family since it moves along actin filaments towards the minus end, in opposition to other myosins, playing a crucial role in endocytosis. It is involved in vesicle transport through direct interaction with adaptor proteins[39]. Moreover, the presence of MYO6 among identified proteins might be also associated with the process of endosome/lysosome fusion[40], the last step expected in the trip of recombinant proteins towards lysosomes.

SNX18 is a member of sorting nexins, which are concentrated at the membrane,

and induce membrane curvature and form the intermediate tubular membrane structures. Evidence has shown that SNX18 is functionally redundant to SNX9, a sorting nexin that plays a key role in clathrin-mediated endocytosis[41].

In the results of the cells treated with Replagal®, we also found PICALM and MYO1E. PICALM (phosphatidylinositol-binding clathrin assembly protein) is a protein that drives clathrin-coated pit formation together with the AP-2 complex[42]. MYO1E may play a role in both cargo internalization and the transport of the newly internalized clathrin-coated vesicles to their target endosomal compartments[43].

For both conditions, proteins involved in caveolar-mediated endocytosis, such as CAVIN1 and EHD2, were identified. Caveolae are plasma membrane invaginations and are expressed in various tissues and cell types such as fibroblasts. The functions of caveolae are diverse and include endocytosis, transcytosis, potocytosis, calcium signaling, and regulation of various signaling events[44]. CAVIN1, also known as PTFR (polymerase I and transcript release factor), belongs to the Cavins family. Cavins are peripheral membrane proteins that coat the caveolar surface, with caveolins embedded in the interior membrane layer, as shown in Fig. 2.1.13[45].

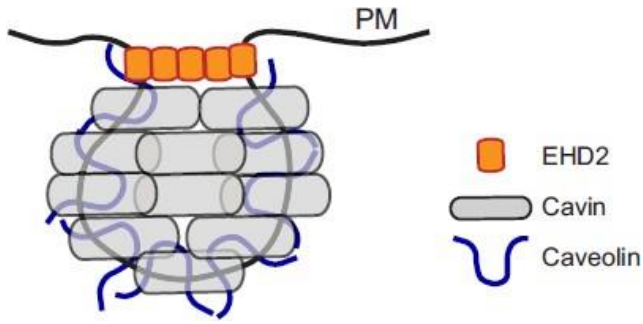

Fig. 2.1.13 Representation of the main proteins involved in the formation of caveolae[45].

EDH2 is one of the EHDs that are dynamin-related proteins involved in nucleotide-dependent membrane remodeling; EHD2 binds and hydrolyzes ATP. It interacts with CAVIN1 and is involved in the remodeling of caveolae[46]. In the Fabrazyme® network, we also found other proteins involved in this endocytic pathway, such as CAV1, RHOG, and ARHGDIA. CAV1 is the principal constituent of caveolae, and it is enough to induce caveolae formation. It is predominantly expressed on the plasma membrane (PM) and is inserted into the inner leaflet of the PM[44]. Prieto-Sánchez *et al.*, in 2006, using dual fluorescence microscopy and other experiments, proved that RHOG, a GTPase member of the Rho/Rac family, is involved in caveolar trafficking[47]. ARHGDIA (Rho GDP-dissociation inhibitor 1) is a regulator of Rho/Rac family GTPases, and it prevents nucleotide exchange and membrane association. It also prevents the GTPases from switching from their inactive GDP-bound state to the active GTP-bound state[48]. Our proteomic findings accord to data already published by Ivanova et al. in 2020, which demonstrated that both CME and caveolar endocytosis mediate the recombinant GLA uptake[49]. In addition, by evaluating all the proteins identified as partners of both recombinant GLA, we can also speculate that Replagal® is up taken in the cell mainly by the CME, while Fabrazyme® might be majorly internalized by caveolae-mediated endocytosis. To support this hypothesis, a larger number of proteins involved in CME was found as Replagal® partners, whereas Fabrazyme® showed an interactome higher enriched in proteins associated with caveolae-mediated endocytosis.

Among the proteins associated with internal trafficking SEC23A, a GTPase-activating protein (GAP) involved in the COPII coat[24] was found. In Fabrazyme®, we identified SAR1A, a small GTPase that when activated, can initiate the assembly of COPII coat proteins involving SEC23A on the ER membrane[50]; whereas in Replagal®, we found SEC22B, a SNARE (Soluble NSF attachment

protein receptor) functioning in the ER-Golgi circuit, that cycles between these compartments and acts in both anterograde and retrograde vesicle fusion events[51].

As suggested by the pie charts (Fig. 2.1.11), we also found many proteins involved in the actin-related cytoskeleton. The rearrangement of the actin cytoskeleton is essential for endocytosis. We found proteins involved in cytoskeleton organization common to both conditions, such as Utrophin that binds to actin and stabilizes actin filaments against depolymerization[52]; Synaptopodin-2 that assists actin polymerization[53]; Plastin-3, which is involved in the formation of filamentous actin (F-actin)[54], in turn, stabilized by TrioBP[55]. In Fabrazyme®'s interactome, Ezrin, a member of the ERM protein family responsible for the plasma membrane linking to the cytoskeleton[56], was present. Analogously, Palladin was identified as a specific Replagal® interactor, and it is known to have a direct role in stabilizing F-actin and/or enhancing actin polymerization[57]. Among the identified cytoskeleton proteins common to Fabrazyme® and Replagal®, another unconventional myosin MYO18A was found. MYO18A links the Golgi complex to the actin cytoskeleton[58]. In Replagal®, we also found MYO5A, an unconventional myosin involved in the ER movement[59]. These identifications strengthen the idea that a fraction of both recombinant proteins, one internalized, travel between ER and Golgi compartments, according to the presence of COPII coating proteins. MYO18A is also involved in the traffic from the Golgi to the plasma membrane[60], suggesting that some of the recombinant proteins might go back outside the cell.

The latter hypothesis is more robust when refers to Fabrazyme® since we also found Optineurin, RAB8A as specific interactors. Optineurin has been identified, as a binding partner for MYO6, both colocalize at the Golgi complex and in vesicles close to the PM. Optineurin also plays a role in colocalizing MYO6 and

RAB8A, a member of the Rab family of G-proteins, which are important regulators of traffic events[61]. These three proteins are involved in exocytosis[62].

MYO1D, already identified among the endogenous GLA interactors, is involved in trafficking from early endosomes to recycling endosomes[63]. We also identified Plectin among the proteins in Fabrazyme® and Replagal®. As expected, the presence of interactors shared with endogenous GLA confirms that, following the internalization, a quote of the recombinant proteins is addressed to the same route of endogenous protein to reach lysosomes.

In light of all results, we can speculate about the pathways involved in the internalization and the traffic of the two recombinant enzymes Replagal® and Fabrazyme® (Fig. 2.1.14).

Replagal® and Fabrazyme® are endocytosed by CME and caveolae. The delivery of both enzymes to the endo/lysosomal system is accomplished by MYO6. However, it seems that a fraction of both molecules are transported to the ER and the Golgi apparatus, as suggested by the presence of SEC23A, SEC22B, SAR1A, MYO5A, and MYO18A. We also hypothesize that Fabrazyme® might be partially exocytosed when associated with Optineurin, RAB8A, and MYO6. The identification of MYO1D suggests that a quote of Replagal® might also be secreted.

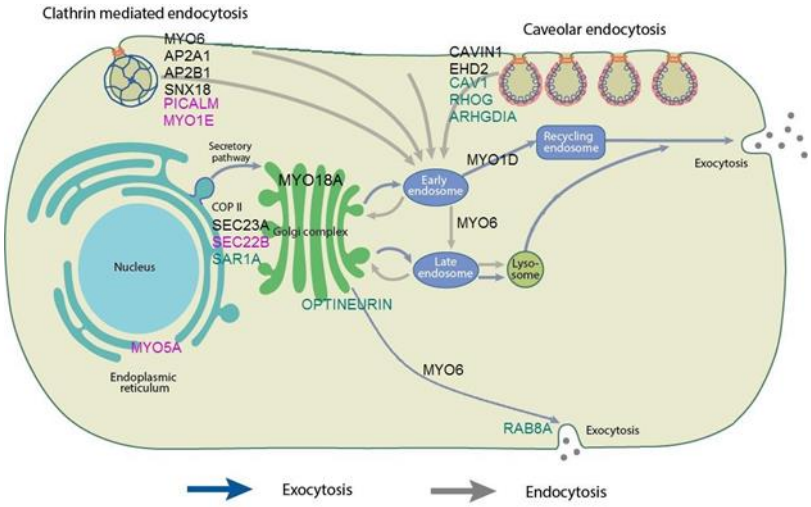

Fig. 2.1.14 The hypothesized pathway of Replagal® and Fabrazyme®. The proteins in black are common, the ones in purple are unique for Replagal® while the ones in green are unique for Fabrazyme®.

Moreover, functional experiments aimed to silence targets of such specific processes might demonstrate the robustness of our hypotheses.

2.1.4 Conclusions

The study of the interactome of the endogenous α-galactosidase A helped us to understand better the route followed by the enzyme when it moves within the cell. The knowledge of the endogenous protein trafficking is important to better understand the results concerning the uptake of the two recombinant enzymes available for ERT to gather information on which molecular mechanisms are involved from the entrance into the cell until lysosomes or other organelles achievement. Using the same approach, the routes followed by the two recombinant enzymes agalsidase alfa (Replagal®) and agalsidase beta (Fabrazyme®) used in ERT of Fabry Disease were investigated, finding the caveolae- and clathrin-mediated endocytosis the common pathways involved in the drug uptake. Algasidase beta has showed a marked dominance of the caveolae-

mediate pathway. Further studies on the differences between the two recombinant proteins could improve the knowledge about the enzymes hence their efficacy. To confirm the hypotheses formulated, other functional experiments such as co-immunoprecipitation and co-localization need to be performed. Moreover, functional experiments aimed to silence targets of such specific processes might demonstrate the robustness of our hypotheses.

2.1.5 References

(1) Ferreira, C. R.; Gahl, W. A. Lysosomal Storage Diseases. *Transl Sci Rare Dis* **2017**, *2* (1–2), 1–71. https://doi.org/10.3233/TRD-160005.

(2) *Molecular Biology of the Cell*, 4th ed.; Alberts, B., Ed.; Garland Science: New York, 2002.

(3) Parenti, G.; Andria, G.; Ballabio, A. Lysosomal Storage Diseases: From Pathophysiology to Therapy. *Annu Rev Med* **2015**, *66*, 471–486. https://doi.org/10.1146/annurev-med-122313-085916.

(4) Mahmud, H. M. Fabry's Disease--a Comprehensive Review on Pathogenesis, Diagnosis and Treatment. *J Pak Med Assoc* **2014**, *64* (2), 189–194.

(5) Mehta, A.; Hughes, D. A. Fabry Disease. In *GeneReviews*®; Adam, M. P., Ardinger, H. H., Pagon, R. A., Wallace, S. E., Bean, L. J., Mirzaa, G., Amemiya, A., Eds.; University of Washington, Seattle: Seattle (WA), 1993.

(6) Schiffmann, R. Fabry Disease. *Pharmacol Ther* **2009**, *122* (1), 65–77. https://doi.org/10.1016/j.pharmthera.2009.01.003.

(7) Seydelmann, N.; Wanner, C.; Störk, S.; Ertl, G.; Weidemann, F. Fabry Disease and the Heart. *Best Pract Res Clin Endocrinol Metab* **2015**, *29* (2), 195–204. https://doi.org/10.1016/j.beem.2014.10.003.

(8) Lemansky, P.; Bishop, D. F.; Desnick, R. J.; Hasilik, A.; von Figura, K. Synthesis and Processing of Alpha-Galactosidase A in Human Fibroblasts. Evidence for Different Mutations in Fabry Disease. *J Biol Chem* **1987**, *262* (5), 2062–2065.

(9) Garman, S. C. Structure-Function Relationships in Alpha-Galactosidase A. *Acta Paediatr* **2007**, *96* (455), 6–16. https://doi.org/10.1111/j.1651-2227.2007.00198.x.

(10) Garman, S. C.; Garboczi, D. N. The Molecular Defect Leading to Fabry Disease: Structure of Human Alpha-Galactosidase. *J Mol Biol* **2004**, *337*

(2), 319–335. https://doi.org/10.1016/j.jmb.2004.01.035.

(11) Schiffmann, R. Fabry Disease. *Handb Clin Neurol* **2015**, *132*, 231–248. https://doi.org/10.1016/B978-0-444-62702-5.00017-2.

(12) Futerman, A. H.; van Meer, G. The Cell Biology of Lysosomal Storage Disorders. *Nat Rev Mol Cell Biol* **2004**, *5* (7), 554–565. https://doi.org/10.1038/nrm1423.

(13) Ishii, S. Pharmacological Chaperone Therapy for Fabry Disease. *Proc Jpn Acad Ser B Phys Biol Sci* **2012**, *88* (1), 18–30. https://doi.org/10.2183/pjab.88.18.

(14) Hamanaka, R.; Shinohara, T.; Yano, S.; Nakamura, M.; Yasuda, A.; Yokoyama, S.; Fan, J.-Q.; Kawasaki, K.; Watanabe, M.; Ishii, S. Rescue of Mutant Alpha-Galactosidase A in the Endoplasmic Reticulum by 1-Deoxygalactonojirimycin Leads to Trafficking to Lysosomes. *Biochim Biophys Acta* **2008**, *1782* (6), 408–413. https://doi.org/10.1016/j.bbadis.2008.03.001.

(15) Deegan, P. B. Fabry Disease, Enzyme Replacement Therapy and the Significance of Antibody Responses. *J Inherit Metab Dis* **2012**, *35* (2), 227–243. https://doi.org/10.1007/s10545-011-9400-y.

(16) Keating, G. M. Agalsidase Alfa: A Review of Its Use in the Management of Fabry Disease. *BioDrugs* **2012**, *26* (5), 335–354. https://doi.org/10.2165/11209690-000000000-00000.

(17) Keating, G. M.; Simpson, D. Agalsidase Beta: A Review of Its Use in the Management of Fabry Disease. *Drugs* **2007**, *67* (3), 435–455. https://doi.org/10.2165/00003495-200767030-00007.

(18) Bekri, S. Importance of Glycosylation in Enzyme Replacement Therapy. In *Fabry Disease: Perspectives from 5 Years of FOS*; Mehta, A., Beck, M., Sunder-Plassmann, G., Eds.; Oxford PharmaGenesis: Oxford, 2006.

(19) Warnock, D. G.; Mauer, M. Fabry Disease: Dose Matters. *J Am Soc Nephrol*

2014, *25* (4), 653–655. https://doi.org/10.1681/ASN.2013121322.

(20) Ries, M.; Clarke, J. T.; Whybra, C.; Mehta, A.; Loveday, K. S.; Brady, R. O.; Beck, M.; Schiffmann, R. Enzyme Replacement in Fabry Disease: Pharmacokinetics and Pharmacodynamics of Agalsidase Alpha in Children and Adolescents. *J Clin Pharmacol* **2007**, *47* (10), 1222–1230. https://doi.org/10.1177/0091270007305299.

(21) Azevedo, O.; Gago, M. F.; Miltenberger-Miltenyi, G.; Sousa, N.; Cunha, D. Fabry Disease Therapy: State-of-the-Art and Current Challenges. *Int J Mol Sci* **2020**, *22* (1). https://doi.org/10.3390/ijms22010206.

(22) Li, J.; Lu, Q.; Zhang, M. Structural Basis of Cargo Recognition by Unconventional Myosins in Cellular Trafficking. *Traffic* **2016**, *17* (8), 822–838. https://doi.org/10.1111/tra.12383.

(23) Almeida, C. G.; Yamada, A.; Tenza, D.; Louvard, D.; Raposo, G.; Coudrier, E. Myosin 1b Promotes the Formation of Post-Golgi Carriers by Regulating Actin Assembly and Membrane Remodelling at the Trans-Golgi Network. *Nat Cell Biol* **2011**, *13* (7), 779–789. https://doi.org/10.1038/ncb2262.

(24) Béthune, J.; Wieland, F. T. Assembly of COPI and COPII Vesicular Coat Proteins on Membranes. *Annu Rev Biophys* **2018**, *47*, 63–83. https://doi.org/10.1146/annurev-biophys-070317-033259.

(25) Nie, C.; Wang, H.; Wang, R.; Ginsburg, D.; Chen, X.-W. Dimeric Sorting Code for Concentrative Cargo Selection by the COPII Coat. *Proc Natl Acad Sci U S A* **2018**, *115* (14), E3155–E3162. https://doi.org/10.1073/pnas.1704639115.

(26) Margiotta, A.; Bucci, C. Role of Intermediate Filaments in Vesicular Traffic. *Cells* **2016**, *5* (2). https://doi.org/10.3390/cells5020020.

(27) Uehara, K.; Uehara, A. Vimentin Intermediate Filaments: The Central Base in Sinus Endothelial Cells of the Rat Spleen. *Anat Rec (Hoboken)* **2010**, *293* (12), 2034–2043. https://doi.org/10.1002/ar.21210.

(28) Favre, B.; Schneider, Y.; Lingasamy, P.; Bouameur, J.-E.; Begré, N.; Gontier, Y.; Steiner-Champliaud, M.-F.; Frias, M. A.; Borradori, L.; Fontao, L. Plectin Interacts with the Rod Domain of Type III Intermediate Filament Proteins Desmin and Vimentin. *Eur J Cell Biol* **2011**, *90* (5), 390–400. https://doi.org/10.1016/j.ejcb.2010.11.013.

(29) Wiche, G. Role of Plectin in Cytoskeleton Organization and Dynamics. *J Cell Sci* **1998**, *111 (Pt 17)*, 2477–2486.

(30) Muroyama, A.; Lechler, T. Microtubule Organization, Dynamics and Functions in Differentiated Cells. *Development* **2017**, *144* (17), 3012–3021. https://doi.org/10.1242/dev.153171.

(31) Fokin, A. I.; Brodsky, I. B.; Burakov, A. V.; Nadezhdina, E. S. Interaction of Early Secretory Pathway and Golgi Membranes with Microtubules and Microtubule Motors. *Biochemistry (Mosc)* **2014**, *79* (9), 879–893. https://doi.org/10.1134/S0006297914090053.

(32) Silacci, P.; Mazzolai, L.; Gauci, C.; Stergiopulos, N.; Yin, H. L.; Hayoz, D. Gelsolin Superfamily Proteins: Key Regulators of Cellular Functions. *Cell Mol Life Sci* **2004**, *61* (19–20), 2614–2623. https://doi.org/10.1007/s00018-004-4225-6.

(33) Bär, S.; Daeffler, L.; Rommelaere, J.; Nüesch, J. P. F. Vesicular Egress of Non-Enveloped Lytic Parvoviruses Depends on Gelsolin Functioning. *PLoS Pathog* **2008**, *4* (8), e1000126. https://doi.org/10.1371/journal.ppat.1000126.

(34) Boswell-Casteel, R. C.; Johnson, J. M.; Duggan, K. D.; Tsutsui, Y.; Hays, F. A. Overproduction and Biophysical Characterization of Human HSP70 Proteins. *Protein Expr Purif* **2015**, *106*, 57–65. https://doi.org/10.1016/j.pep.2014.09.013.

(35) Radons, J. The Human HSP70 Family of Chaperones: Where Do We Stand? *Cell Stress Chaperones* **2016**, *21* (3), 379–404.

https://doi.org/10.1007/s12192-016-0676-6.

(36) Goode, B. L.; Eskin, J. A.; Wendland, B. Actin and Endocytosis in Budding Yeast. *Genetics* **2015**, *199* (2), 315–358. https://doi.org/10.1534/genetics.112.145540.

(37) Doherty, G. J.; McMahon, H. T. Mechanisms of Endocytosis. *Annu Rev Biochem* **2009**, *78*, 857–902. https://doi.org/10.1146/annurev.biochem.78.081307.110540.

(38) Park, S. Y.; Guo, X. Adaptor Protein Complexes and Intracellular Transport. *Biosci Rep* **2014**, *34* (4). https://doi.org/10.1042/BSR20140069.

(39) Dawson, H. J.; Hibbert, A. P.; Chantler, P. D.; Botham, K. M. Myosin VI and Associated Proteins Are Expressed in Human Macrophages but Do Not Play a Role in Foam Cell Formation in THP-1 Cells. *International Journal of Vascular Medicine* **2013**, *2013*, 1–17. https://doi.org/10.1155/2013/516015.

(40) Liao, Y.-W.; Wu, X.-M.; Jia, J.; Wu, X.-L.; Hong, T.; Meng, L.-X.; Wu, X.-Y. Myosin VI Contributes to Maintaining Epithelial Barrier Function. *J Biomed Sci* **2013**, *20*, 68. https://doi.org/10.1186/1423-0127-20-68.

(41) Park, J.; Kim, Y.; Lee, S.; Park, J. J.; Park, Z. Y.; Sun, W.; Kim, H.; Chang, S. SNX18 Shares a Redundant Role with SNX9 and Modulates Endocytic Trafficking at the Plasma Membrane. *J Cell Sci* **2010**, *123* (Pt 10), 1742–1750. https://doi.org/10.1242/jcs.064170.

(42) Suzuki, M.; Tanaka, H.; Tanimura, A.; Tanabe, K.; Oe, N.; Rai, S.; Kon, S.; Fukumoto, M.; Takei, K.; Abe, T.; Matsumura, I.; Kanakura, Y.; Watanabe, T. The Clathrin Assembly Protein PICALM Is Required for Erythroid Maturation and Transferrin Internalization in Mice. *PLoS One* **2012**, *7* (2), e31854. https://doi.org/10.1371/journal.pone.0031854.

(43) Cheng, J.; Grassart, A.; Drubin, D. G. Myosin 1E Coordinates Actin Assembly and Cargo Trafficking during Clathrin-Mediated Endocytosis.

Mol Biol Cell **2012**, *23* (15), 2891–2904. https://doi.org/10.1091/mbc.E11-04-0383.

(44) Lajoie, P.; Nabi, I. R. Lipid Rafts, Caveolae, and Their Endocytosis. *Int Rev Cell Mol Biol* **2010**, *282*, 135–163. https://doi.org/10.1016/S1937-6448(10)82003-9.

(45) Kovtun, O.; Tillu, V. A.; Ariotti, N.; Parton, R. G.; Collins, B. M. Cavin Family Proteins and the Assembly of Caveolae. *J Cell Sci* **2015**, *128* (7), 1269–1278. https://doi.org/10.1242/jcs.167866.

(46) Morén, B.; Shah, C.; Howes, M. T.; Schieber, N. L.; McMahon, H. T.; Parton, R. G.; Daumke, O.; Lundmark, R. EHD2 Regulates Caveolar Dynamics via ATP-Driven Targeting and Oligomerization. *Mol Biol Cell* **2012**, *23* (7), 1316–1329. https://doi.org/10.1091/mbc.E11-09-0787.

(47) Prieto-Sánchez, R. M.; Berenjeno, I. M.; Bustelo, X. R. Involvement of the Rho/Rac Family Member RhoG in Caveolar Endocytosis. *Oncogene* **2006**, *25* (21), 2961–2973. https://doi.org/10.1038/sj.onc.1209333.

(48) Huang, D.; Lu, W.; Zou, S.; Wang, H.; Jiang, Y.; Zhang, X.; Li, P.; Songyang, Z.; Wang, L.; Wang, J.; Huang, J.; Fang, L. Rho GDP-Dissociation Inhibitor α Is a Potential Prognostic Biomarker and Controls Telomere Regulation in Colorectal Cancer. *Cancer Sci* **2017**, *108* (7), 1293–1302. https://doi.org/10.1111/cas.13259.

(49) Ivanova, M. M.; Dao, J.; Kasaci, N.; Adewale, B.; Fikry, J.; Goker-Alpan, O. Rapid Clathrin-Mediated Uptake of Recombinant α-Gal-A to Lysosome Activates Autophagy. *Biomolecules* **2020**, *10* (6). https://doi.org/10.3390/biom10060837.

(50) Urai, Y.; Yamawaki, M.; Watanabe, N.; Seki, Y.; Morimoto, T.; Tago, K.; Homma, K.; Sakagami, H.; Miyamoto, Y.; Yamauchi, J. Pull down Assay for GTP-Bound Form of Sar1a Reveals Its Activation during Morphological Differentiation. *Biochem Biophys Res Commun* **2018**, *503* (3), 2047–2053.

https://doi.org/10.1016/j.bbrc.2018.07.157.

(51) Mancias, J. D.; Goldberg, J. The Transport Signal on Sec22 for Packaging into COPII-Coated Vesicles Is a Conformational Epitope. *Mol Cell* **2007**, *26* (3), 403–414. https://doi.org/10.1016/j.molcel.2007.03.017.

(52) Prochniewicz, E.; Henderson, D.; Ervasti, J. M.; Thomas, D. D. Dystrophin and Utrophin Have Distinct Effects on the Structural Dynamics of Actin. *Proc Natl Acad Sci U S A* **2009**, *106* (19), 7822–7827. https://doi.org/10.1073/pnas.0812007106.

(53) Chalovich, J. M.; Schroeter, M. M. Synaptopodin Family of Natively Unfolded, Actin Binding Proteins: Physical Properties and Potential Biological Functions. *Biophys Rev* **2010**, *2* (4), 181–189. https://doi.org/10.1007/s12551-010-0040-5.

(54) Li, N.; Mruk, D. D.; Wong, C. K. C.; Lee, W. M.; Han, D.; Cheng, C. Y. Actin-Bundling Protein Plastin 3 Is a Regulator of Ectoplasmic Specialization Dynamics during Spermatogenesis in the Rat Testis. *FASEB J* **2015**, *29* (9), 3788–3805. https://doi.org/10.1096/fj.14-267997.

(55) Zaharija, B.; Samardžija, B.; Bradshaw, N. J. The TRIOBP Isoforms and Their Distinct Roles in Actin Stabilization, Deafness, Mental Illness, and Cancer. *Molecules* **2020**, *25* (21), 4967. https://doi.org/10.3390/molecules25214967.

(56) Biri-Kovács, B.; Kiss, B.; Vadászi, H.; Gógl, G.; Pálfy, G.; Török, G.; Homolya, L.; Bodor, A.; Nyitray, L. Ezrin Interacts with S100A4 via Both Its N- and C-Terminal Domains. *PLoS ONE* **2017**, *12* (5), e0177489. https://doi.org/10.1371/journal.pone.0177489.

(57) Gurung, R.; Yadav, R.; Brungardt, J. G.; Orlova, A.; Egelman, E. H.; Beck, M. R. Actin Polymerization Is Stimulated by Actin Cross-Linking Protein Palladin. *Biochem J* **2016**, *473* (4), 383–396. https://doi.org/10.1042/BJ20151050.

(58) Makowski, S. L.; Tran, T. T.; Field, S. J. Emerging Themes of Regulation at the Golgi. *Curr Opin Cell Biol* **2017**, *45*, 17–23. https://doi.org/10.1016/j.ceb.2017.01.004.

(59) Wagner, W.; Brenowitz, S. D.; Hammer, J. A. Myosin-Va Transports the Endoplasmic Reticulum into the Dendritic Spines of Purkinje Neurons. *Nat Cell Biol* **2011**, *13* (1), 40–48. https://doi.org/10.1038/ncb2132.

(60) Buschman, M. D.; Field, S. J. MYO18A: An Unusual Myosin. *Adv Biol Regul* **2018**, *67*, 84–92. https://doi.org/10.1016/j.jbior.2017.09.005.

(61) Sahlender, D. A.; Roberts, R. C.; Arden, S. D.; Spudich, G.; Taylor, M. J.; Luzio, J. P.; Kendrick-Jones, J.; Buss, F. Optineurin Links Myosin VI to the Golgi Complex and Is Involved in Golgi Organization and Exocytosis. *J Cell Biol* **2005**, *169* (2), 285–295. https://doi.org/10.1083/jcb.200501162.

(62) Chibalina, M. V.; Roberts, R. C.; Arden, S. D.; Kendrick-Jones, J.; Buss, F. Rab8-Optineurin-Myosin VI: Analysis of Interactions and Functions in the Secretory Pathway. *Methods Enzymol* **2008**, *438*, 11–24. https://doi.org/10.1016/S0076-6879(07)38002-6.

(63) Huber, L. A.; Fialka, I.; Paiha, K.; Hunziker, W.; Sacks, D. B.; Bähler, M.; Way, M.; Gagescu, R.; Gruenberg, J. Both Calmodulin and the Unconventional Myosin Myr4 Regulate Membrane Trafficking along the Recycling Pathway of MDCK Cells. *Traffic* **2000**, *1* (6), 494–503. https://doi.org/10.1034/j.1600-0854.2000.010607.x.

2.2 ADAM10 interacts with Piccolo at pre-synapsis and affects synaptic vesicle stores in Huntington's disease

2.2.1 Introduction

Disintegrin and metalloproteinase domain-containing protein 10 (ADAM10) is one of the 22 members of the ADAM family of transmembrane proteases found in humans and one of the most widely expressed in the brain[1]. Nascent ADAM10 is produced as an approximately 95 kDa zymogen that matures into a 60 kDa active protease (m-ADAM10) after cleavage of its prodomain by furin or proconvertase 7 (PC7) in the trans-Golgi secretory pathway[1]. Within neurons, ADAM10 localizes specifically at the synapse and is associated with the postsynaptic compartment, together with the scaffold protein synapse-associated protein 97 (SAP97)[2], which is required for insertion of ADAM10 into the synaptic membrane, as shown in Fig. 2.2.1. This enzyme exerts the main physiological neuronal alpha-secretase activity, responsible for non-amyloidogenic cleavage of β-amyloid precursor protein (APP)[1]. Besides APP, ADAM10 has a large number of synaptic substrates, indicating extensive flexibility of the protease concerning substrate recognition and cleavage activity in the brain[3] (Fig. 2.2.1).

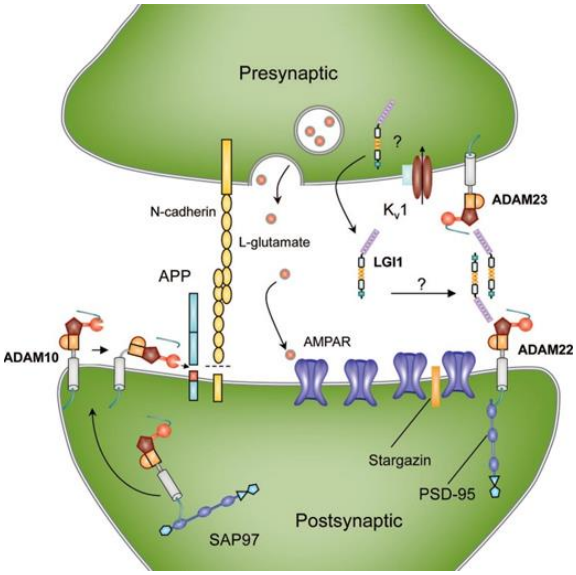

Fig. 2.2.1 ADAM10 has metalloproteinase activity and localizes at postsynaptic sites via its association with synapse-associated protein 97 (SAP97). ADAM10 acts on several postsynaptic substrates, including N-cadherin and amyloid precursor protein (APP)[4].

Consequently, the biological effect of ADAM10 activity may change over time and in different conditions, as it relies on the presence and function of the target substrates and cleavage products. Accordingly, ADAM10 plays fundamental roles during brain development and in the homeostasis of adult neuronal networks[1,5], as it is involved in dendritic spine formation, maturation, and stabilization; axon guidance and extension; neuronal migration; myelination; and regulation of the molecular organization of the glutamatergic synapse[1,5]. ADAM10-dependent cleavage of N-cadherin, Nectin 1, and Neurologin 1 also represents a new modality for regulating the strength and activity of the glutamatergic synapse[1,5]. Glutamatergic synapses are the main excitatory synapses in the brain. These synapses contain glutamate localized inside presynaptic vesicles and glutamate receptors on the postsynaptic membrane (Fig. 2.2.2).

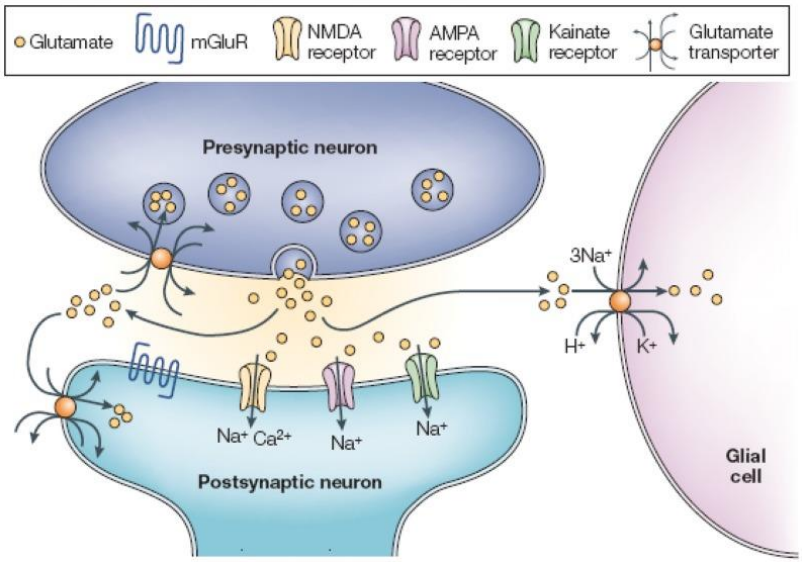

Fig. 2.2.2 Schematic drawing of a glutamatergic synapse[6].

An excess of ADAM10-mediated N-cadherin proteolysis in the brain has been reported to compromise adhesion between the presynaptic and postsynaptic membrane and reduces the overall long-term stability and neurotransmission of the glutamatergic synapse[7,8,9]. Defects in this circuitry have been implicated in Huntington's disease (HD), an autosomal dominant neurodegenerative disorder caused by an unstable CAG triplet repeat expansion in exon 1 of the huntingtin gene (HTT)[10]. An alteration in presynaptic and/or postsynaptic compartments has been demonstrated in several mouse models of the disease since the early stages and documented in HD patients in cross-sectional and neuroimaging studies[11,12,13]. ADAM10 has been linked to HTT and dysfunction of the glutamatergic synapse in HD. Previous studies in embryonic stem cells depleted of HTT demonstrated that the homotypic interaction between neuroepithelial cells is dependent on the presence of normal HTT, which acts physiologically to inhibit ADAM10 and the maturation of its substrate N-cadherin, allowing cell-cell contacts[14]. It was also reported in the adult brain, HTT similarly affects synapse remodeling through ADAM10 activity on N-cadherin[14]. In HD, this pathway is hyper-activated, with

deleterious consequences on the morphology and functionality of the glutamatergic synapse[15]. In particular, ADAM10 accumulates at postsynaptic densities (PSDs) and causes excessive N-cadherin proteolysis[15], leading to loss of excitatory synaptic contacts, synapse deficiencies, and cognitive decline in HD mice. Previous studies also demonstrated that pharmacological, molecular, and genetic approaches aimed at normalizing the level of the active enzyme in the HD mouse brain prevent electrophysiological synaptic defects at the excitatory synapse and cognitive impairment in the mice[15]. Taken together, these previous studies point to a critical role of ADAM10 at the glutamatergic synapse in the HD brain. To unravel the molecular mechanisms through which ADAM10 is associated with glutamatergic synaptic dysfunction in HD, in collaboration with the research group of the Professor Chiara Zuccato of the Department of Biomedical Sciences for Health at the University of Milan, we performed a system-level study of ADAM10 interactors in the brains of HD mice. A functional proteomic experiment based on immunoaffinity purification-mass spectrometry identification (IP-MS) of endogenous ADAM10 interactors in the brains of wild-type and HD mice model (zQ175) revealed a never described role at pre-synapsis for ADAM10, as well as its involvement in unexpected synaptic processes.

2.2.2 Experimental methods

2.2.2.1 Mice brain tissue lysis and protein extract quantification

Total protein lysates from n=4 wild-type and n=4 homozygous 50-week-old zQ175 mice were obtained by pestling tissues in an N_2 pre-chill mortar to grind small frozen tissue samples to a fine powder. Mouse tissue homogenization was carried out by 20 strokes in a Teflon-glass homogenizer (clearance 0.25 mm, 700 rpm) in RIPA buffer (50 mM Tris HCl pH 8, 150 mM NaCl, 0.1% SDS, 1% NP40 (Sigma-Aldrich) with 1 mM PMSF (Sigma-Aldrich) and 1× protease inhibitor cocktail. Tissue lysates were then passed through a 25 gauge needle attached to a

1 ml syringe. After 20 minutes on ice, lysates were put 15 minutes on a rotating wheel at 4°C. Lysates were then cleared by centrifugation for 30 minutes at 12,000 g and 4°C. The resulting supernatants were collected and protein concentrations were determined by Bradford.

2.2.2.2 Isolation of protein complexes by immunoprecipitation

Tissue lysates from n=4 mice/genotypes were pooled and 6 mg of protein lysates for wild-type and zQ175 mice were pre-cleared by incubation with Dynabeads Protein G (Thermo Fisher Scientific, Waltham, Massachusetts, US) for 2h at 4°C on a rotating wheel. The pre-cleared extracts were immunoprecipitated by the addition of 24 µl of rabbit polyclonal anti-ADAM10 provided by P. Saftig (University of Kiel) and followed by the overnight incubation at 4°C on a rotating wheel. The pre-cleared-antibody mixture was incubated with Dynabeads Protein G (40 µl of slurry resin each milligram of protein extract). The elution from pre-clearing (control) and immunoprecipitated beads was carried out in Sample Buffer 2× (200 mM Tris HCl pH 6.8, 40% glycerol, 8% SDS, 200 mM DTT, 0.01% bromophenol blue) at 99°C for 10 minutes.

2.2.2.3 Preparative SDS-PAGE

Eluted samples were fractionated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). The 8-15% gradient SDS-PAGE gel was run starting from 200 V and stained with GelCode™ Blue Safe Protein Stain (Thermo Fisher Scientific). Each lane was cut into 1 mm slices, and the gel bands digested *in situ* with trypsin as previously reported in section 2.1.2.4.

2.2.2.4 LC-MS/MS analysis and protein identification

Peptide mixtures were analyzed by nano LC-MS/MS using LTQ Orbitrap XL mass spectrometer coupled to Proxeon nanoEasy II capillary HPLC (Thermo Fisher Scientific Inc., Waltham, MA). Peptides were fractionated onto a C18 reverse-phase capillary column (3 µm biosphere, 7.5 µm internal diameter, 100

mm length) working at 300 nl/min flow rate and adopting a linear gradient from 5% to 100% of eluent B (0.2% formic acid, 95% acetonitrile LC-MS Grade) over 49 min. Mass spectrometric analyses were carried out in Data Depending Acquisition (DDA): from each MS scan, spanning from 400 to 1800 m/z, the five most abundant ions were selected and fragmented. Output data were further processed into mgf files suitable for protein identification procedure by Mascot licensed software (Matrix Science Boston, USA) searching for proteins in the NCBI database. Proteins identification was carried out by using 10 ppm as peptides mass tolerance for MS and 0.6 Da for MS/MS search; Mus Musculus as taxonomy, carbamidomethyl (Cys) as fixed modification and Gln->pyro-Glu (N-term Gln), Oxidation (Met), Pyro-carbamidomethyl (N-term Cys) as variable modifications.

2.2.2.5 Bioinformatic filtering of immunoprecipitation contaminants

For the definition of the proteins that are non-specific contaminants of immunoprecipitation, we relied on the Contaminant Repository for Affinity Purification (CRAPome) 2.0 web tool (https://reprint-apms.org), querying the human experiment collection. To this end, starting from the list of Mouse Genome Informatics (MGI) symbols for ADAM10 interactors in wild-type and zQ175 mice, human homolog genes were retrieved by biomaRt R package (Ensembl version 98)[16]. CRAPome was then interrogated using 'Workflow1' and 'H. sapiens - All' parameter. Contaminants were defined as proteins reported in at least 50% of the experiments.

2.2.2.6 Bioinformatic analysis of ADAM10 protein networks

A network visualizing ADAM10 interactors in wild-type and zQ175 striata was generated in Cytoscape software (version 3.7.2). Interactors are visualized as nodes connected to ADAM10 in WT and/or MUT conditions. Moreover, ADAM10 interactors identified for wild-type and zQ175 mice were subjected to

gene ontology enrichment analysis. Gene ontology enrichment analysis was performed in R with TopGO package considering the Cellular Component domain of the ontology and using the Fisher statistics and the Weight01 algorithm. Terms were considered significant with a P-Value < 0.01 and an enrichment value of at least 2. The top-three categories (ranked by p-value) in both wild-type and zQ175 mice were: (I) glutamatergic synapse; (II) post-synaptic density; (III) pre-synaptic active-zone cytoplasmic component. Starting from the interactome generated from the proteomics data, the genes associated with the "glutamatergic synapse" category were highlighted in the network as blue nodes.

2.2.2.7 Co-immunoprecipitation and western blot analysis

Co-IP was performed on total protein lysates and synaptosomes prepared from cortical tissues of wild-type and R6/2 mice at 12 weeks of age. Cortical tissues were pooled in an N_2 pre-chill mortar and pestle to grind small frozen tissue samples to a fine powder, and homogenization was carried out in RIPA buffer (50 mM Tris HCl pH 8, 150 mM NaCl, 0.1% SDS, 1% NP40 with 1 mM PMSF and 1× protease inhibitor cocktail (Roche Scientific). Tissue lysates were then passed through a 25gauge needle attached to a 1 ml syringe. After 20 minutes in ice, lysates were put 15 minutes on a rotating wheel at 4°C. Lysates were then cleared by centrifugation for 30 minutes at 12000 g and 4°C. The resulting supernatants were collected, and protein concentrations were determined by Bradford assay with Protein Assay Dye Reagent Concentrate (Bio-Rad). 1 mg protein lysate was pre-cleared using Dynabeads Protein G according to the manufacturer's instructions and then incubated with 50 µl of rabbit polyclonal anti-PCLO antibody (Synaptic Systems, Göttingen, Germany) overnight at 4°C on a rotation wheel and then loaded onto Dynabeads Protein G according to the manufactured instruction, and incubated for 3 hours at 4°C on a rotation wheel. After washing, the beads were eluted in Sample Buffer 2× (200 mM Tris HCl pH 6.8, 40%

glycerol, 8% SDS, 200 mM DTT, 0.01% bromophenol blue) at 99°C for 10 minutes. For the Western blot analysis, Co-IP samples were denatured at 99°C and loaded on a 6% and 10% SDS-PAGE gel. Input samples (5% of the amount of the corresponding lysate used for IP) were loaded as controls. Separated proteins were transferred onto a nitrocellulose membrane (Bio-Rad) using Trans-blot Turbo Transfer System (Bio-Rad) (standard protocol: 1.5 A constant; up to 25 V; 10 minutes). Membranes were blocked with 5% nonfat milk (Bio-Rad) in PBS for 1 hour and were incubated overnight at 4°C with rabbit polyclonal anti-ADAM10 antibody provided by P. Saftig (1:1000 in PBS and 0.2% Tween20), with rabbit polyclonal anti-PCLO antibody (1:1000 in 5% nonfat milk; Synaptic System), and with mouse monoclonal anti-α-Tubulin (1:5000; MilliporeSigma, Massachusetts, US). After washing, filters were incubated for 45 minutes at room temperature with 1:200 Very Blot for IP Detection Reagent (HRP) (Abcam, Cambridge, UK) in 5% nonfat milk in PBS and 0.2% Tween20. Immunoreactive bands were developed by Pierce™ ECL Western Blotting Substrate (Thermo Fisher Scientific), and the proteins were visualized by exposing the membrane to autoradiography films or by using the ChemiDoc MP Imaging System from Bio-Rad. Densitometric analyses were performed by using Image Lab version 6.0.1 or Quantity One 4.6.8 (Bio-Rad).

To verify the fractionated lysis of pre- and post-synaptic cell lysates, the protein extracts were loaded onto a 10% SDS-PAGE and transferred onto a nitrocellulose membrane as described above. The membranes were incubated with rabbit polyclonal anti-ADAM10 provided by P. Saftig (1:1000 in TBST), mouse monoclonal anti-PSD95 antibody (1:2000 in TBST; Thermo Fisher Scientific), and mouse monoclonal anti-SYP antibody (1:500 in BSA 5%; Abcam) at room temperature for 3 hours or at 4°C overnight. After washing, filters were incubated for 1 hour at room temperature with a peroxidase-conjugated secondary antibody (1:3000 in 5% BSA; Bio-Rad, goat anti-rabbit horseradish peroxidase (HRP) and

goat anti-mouse HRP) and then washed 3 times with TBST. The Clarity Western ECL Substrate (Bio-Rad) was used to visualize immunoreactive bands. Blot visualization was performed by using the ChemiDoc MP Imaging System from Bio-Rad. Densitometric analyses were performed by using Image Lab version 6.0.1.

2.2.3 Results and Discussion

2.2.3.1 The ADAM10 interactome is enriched in synaptic proteins

The ADAM10 interactome was analyzed in the striatum wild-type mice and compared to that of homozygous (zQ175) knock-in mice, a model for HD. ADAM10 protein complexes were isolated by immunoprecipitation, and each component was identified by mass spectrometry. Proteins retained onto dynabeads- G Protein during pre-clearing of wild-type or zQ175 tissue lysates in the absence of specific antibodies were eluted and used as controls. Both samples and controls were fractionated by SDS-PAGE, and the lanes were cut into 33 gel bands (Fig. 2.2.3), which were digested *in situ* with trypsin[17].

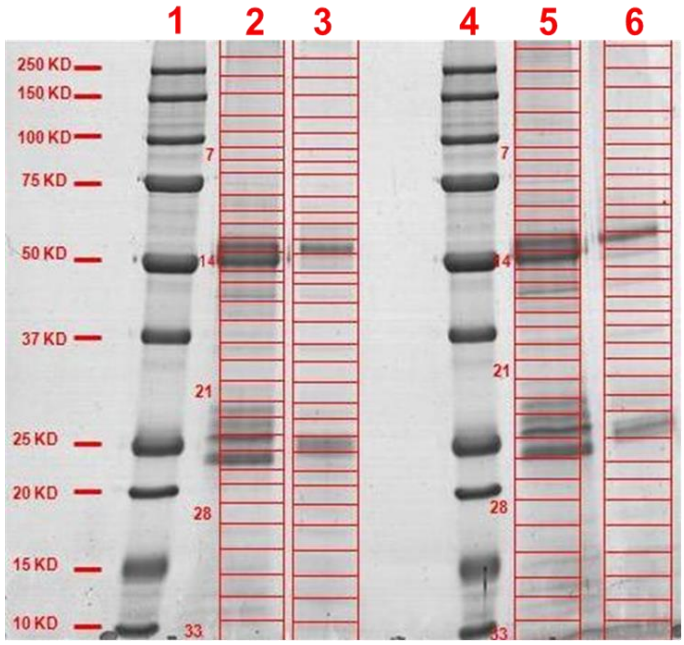

Fig 2.2.3 Preparative SDS-PAGE gel. Samples loaded are IP from zQ175 (lane 2), pre-clearing zQ175 (lane 3), IP WT(lane 5), pre-clearing WT (lane 6). The molecular-weight size markers were loaded in lanes 1 and 4.

The resulting peptide mixtures were directly analyzed by nanoLC-MS/MS. Mass spectral data were used to search a non-redundant protein database using Mascot software. Proteins detected in the samples and in the corresponding pre-clearing controls were discarded, and only those solely present in the sample lanes were considered as candidates for ADAM10 interactors. Two lists of 178 and 146 proteins were found in the ADAM10 interactome for WT and zQ175 mice, respectively. Among them, 101 proteins were shared (Table 2.2.1).

Protein Name	Gene name	Uniprot code	Sequence coverage (WT)	Peptides WT	Sequence coverage (zQ175)	Peptides zQ175
AP2-associated protein kinase 1	Aak1	Q3UHJ0	9.91	6	22.49	8

Alanine--tRNA ligase, cytoplasmic	Aars	Q8BGQ7	4.44	4	10.33	9
4-aminobutyrate aminotransferase, mitochondrial	Abat	P61922	18.80	8	22.37	6
3-ketoacyl-CoA thiolase, mitochondrial	Acaa2	Q8BWT1	12.09	4	1.15	4
Long-chain specific acyl-CoA dehydrogenase, mitochondrial	Acadl	P51174	16.28	6	11.57	11
Acetyl-CoA acetyltransferase, cytosolic	Acat2	Q8CAY6	7.81	3	7.51	5
ATP-citrate synthase	Acly	Q91V92	6.23	7	4.40	6
Peroxisomal acyl-coenzyme A oxidase 1	Acox1	Q9R0H0	34.04	20	11.63	10
Long-chain-fatty-acid--CoA ligase 6	Acsl6	Q91WC3	8.46	6	6.83	14
Alpha-actinin-2	Actn2	Q9JI91	/	/	22.59	11
Actin-related protein 3	Actr3	Q99JY9	25.84	10	32.10	18
Disintegrin and metalloproteinase domain-containing protein 10	Adam10	O35598	40.19	31	23.65	12
Arf-GAP with GTPase, ANK repeat and PH domain-containing protein 3	Agap3	Q8VHH5	7.47	5	19.28	11
Adenosylhomocysteinase	Ahcy	P50247	16.20	6	5.61	7

Jouberin	Ahi1	Q8K3E5	4.39	4	6.98	3
Alpha-aminoadipic semialdehyde dehydrogenase	Aldh7a1	Q9DBF1	/	/	5.25	4
Ankyrin-2	Ank2	Q8C8R3	1.69	6	11.32	9
Acidic leucine-rich nuclear phosphoprotein 32 family member A	Anp32a	O35381	/	/	6.79	5
Annexin A4	Anxa4	P97429	/	/	10.77	7
Annexin A7	Anxa7	Q07076	47.52	23	4.08	2
AP-2 complex subunit alpha-1	Ap2a1	P17426	7.78	8	7.49	32
Bifunctional purine biosynthesis protein PURH	Atic	Q9CWJ9	/	/	5.83	6
V-type proton ATPase subunit C 1	Atp6v1c1	Q9Z1G3	11.52	6	13.22	7
V-type proton ATPase subunit D	Atp6v1d	P57746	/	/	32.04	25
V-type proton ATPase subunit G 2	Atp6v1g2	Q9WTT4	25.42	3	7.62	6
Phospholipid-transporting ATPase IA	Atp8a1	P70704	/	/	7.32	5
Methylglutaconyl-CoA hydratase, mitochondrial	Auh	Q9JLZ3	14.97	5	7.52	3
D-beta-hydroxybutyrate dehydrogenase, mitochondrial	Bdh1	Q80XN0	/	/	8.64	4
Protein bassoon	Bsn	O88737	14.16	47	4.92	6

BTB/POZ domain-containing protein 17	Btbd17	Q9DB72	22.59	7	17.91	7
CAD protein	Cad	B2RQC6	9.98	19	15.27	12
Calcium-dependent secretion activator 1	Cadps	Q80TJ1	7.60	9	29.73	8
Calcium/calmodulin-dependent protein kinase type II subunit alpha	Camk2a	P11798	43.93	21	7.34	6
Calcium/calmodulin-dependent protein kinase type II subunit beta	Camk2b	P28652	34.87	16	17.87	20
Calcium/calmodulin-dependent protein kinase type II subunit delta	Camk2d	Q6PHZ2	17.64	9	5.85	2
Calcium/calmodulin-dependent protein kinase type II subunit gamma	Camk2g	Q923T9	25.52	14	18.05	9
Calnexin	Canx	P35564	10.83	7	13.91	10
Caprin-1	Caprin1	Q60865	7.36	5	21.81	6
F-actin-capping protein subunit alpha-1	Capza1	P47753	18.88	6	10.97	6
F-actin-capping protein subunit beta	Capzb	P47757	22.74	5	9.63	7
Caskin-1	Caskin1	Q6P9K8	10.76	12	10.66	4
Carbonyl reductase [NADPH] 1	Cbr1	P48758	/	/	12.00	3
T-complex protein 1 subunit beta	Cct2	P80314	/	/	3.85	5

T-complex protein 1 subunit delta	Cct4	P80315	/	/	9.66	4
T-complex protein 1 subunit epsilon	Cct5	P80316	14.97	8	10.67	10
T-complex protein 1 subunit zeta	Cct6a	P80317	/	/	8.22	3
T-complex protein 1 subunit eta	Cct7	P80313	3.49	2	5.54	2
T-complex protein 1 subunit theta	Cct8	P42932	8.58	4	4.86	5
Charged multivesicular body protein 4b	Chmp4b	Q9D8B3	/	/	8.14	3
Citron Rho-interacting kinase	Cit	P49025	3.55	7	23.25	7
Chloride intracellular channel protein 4	Clic4	Q9QYB1	/	/	25.06	15
CAP-Gly domain-containing linker protein 2	Clip2	Q9Z0H8	7.45	6	8.10	3
COP9 signalosome complex subunit 7a	Cops7a	Q9CZ04	/	/	13.42	5
C-terminal-binding protein 1	Ctbp1	O88712	6.58	3	14.78	22
Catenin alpha-2	Ctnna2	Q61301	/	/	15.38	4
Catenin beta-1	Ctnnb1	Q02248	6.02	5	5.41	11
Cytochrome c1, heme protein, mitochondrial	Cyc1	Q9D0M3	/	/	4.96	6
Cytoplasmic FMR1-interacting protein 2	Cyfip2	Q5SQX6	6.86	7	11.79	15
Drebrin	Dbn1	Q9QXS6	7.93	4	7.34	9
Dynactin subunit 1	Dctn1	O08788	8.12	10	11.22	5
Dynactin subunit 2	Dctn2	Q99KJ8	15.67	5	2.92	7

ATP-dependent RNA helicase DDX1	Ddx1	Q91VR5	7.30	5	6.12	5
Disks large homolog 1	Dlg1	Q811D0	3.09	3	28.14	18
Disks large homolog 4	Dlg4	Q62108	10.36	7	25.18	24
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial	Dlst	Q9D2G2	13.00	6	19.16	5
DnaJ homolog subfamily A member 2	Dnaja2	Q9QYJ0	9.22	4	16.59	6
Dynamin-1-like protein	Dnm1l	Q8K1M6	26.55	16	4.14	2
Cytoplasmic dynein 1 heavy chain 1	Dync1h1	Q9JHU4	5.84	27	32.35	13
Elongation factor 2	Eef2	P58252	2.80	2	29.95	14
EF-hand domain-containing protein D2	Efhd2	Q9D8Y0	/	/	11.59	38
Eukaryotic translation initiation factor 3 subunit I	Eif3i	Q9QZD9	18.77	6	4.10	18
Eukaryotic initiation factor 4A-II	Eif4a2	P10630	13.02	5	27.74	6
Eukaryotic initiation factor 4A-III	Eif4a3	Q91VC3	/	/	7.64	8
Echinoderm microtubule-	Eml3	Q8VC03	9.36	7	11.58	6

associated protein-like 3						
Band 4.1-like protein 3	Epb41l3	Q9WV92	11.73	9	20.00	9
EPM2A-interacting protein 1	Epm2aip1	Q8VEH5	5.12	3	32.16	18
ERC protein 2	Erc2	Q6PH08	19.54	22	29.25	13
Erlin-2	Erlin2	Q8BFZ9	10.59	3	9.81	10
Ermin	Ermn	Q5EBJ4	14.23	4	3.17	5
Electron transfer flavoprotein subunit alpha, mitochondrial	Etfa	Q99LC5	28.53	7	14.18	6
Ezrin	Ezr	P26040	16.38	10	17.97	9
Phenylalanine--tRNA ligase beta subunit	Farsb	Q9WUA2	9.17	5	5.97	14
Fatty acid synthase	Fasn	P19096	/	/	21.31	40
Fragile X mental retardation protein 1 homolog	Fmr1	P35922	6.84	4	30.90	17
Fascin	Fscn1	Q61553	23.53	13	30.00	31
Fragile X mental retardation syndrome-related protein 1	Fxr1	Q61584	3.99	3	17.60	26
Ras GTPase-activating protein-binding protein 1	G3bp1	P97855	34.19	13	17.93	9
Ras GTPase-activating protein-binding protein 2	G3bp2	P97379	19.50	9	16.73	10
Rab GDP dissociation inhibitor beta	Gdi2	Q61598	20.90	9	26.09	14
Glyoxalase domain-containing protein 4	Glod4	Q9CPV4	15.44	4	13.70	4

GMP synthase [glutamine-hydrolyzing]	Gmps	Q3THK7	9.52	6	22.79	7
Glycerol-3-phosphate dehydrogenase 1-like protein	Gpd1l	Q3ULJ0	18.52	6	21.66	7
Gephyrin	Gphn	Q8BUV3	7.15	5	11.31	8
Hyaluronan and proteoglycan link protein 4	Hapln4	Q80WM4	9.00	3	3.35	6
Potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 1	Hcn1	O88704	11.21	8	14.34	4
Heterogeneous nuclear ribonucleoprotein A3	Hnrnpa3	Q8BG05	12.14	4	27.31	5
Heterogeneous nuclear ribonucleoprotein A/B	Hnrnpab	Q99020	12.98	3	19.75	5
Heterogeneous nuclear ribonucleoprotein K	Hnrnpk	P61979	/	/	5.16	5
Heterogeneous nuclear ribonucleoprotein M	Hnrnpm	Q9D0E1	16.32	11	7.59	3
Heterogeneous nuclear ribonucleoprotein U	Hnrnpu	Q8VEK3	10.50	8	25.81	15
Homer protein homolog 1	Homer1	Q9Z2Y3	18.58	6	3.69	5

3-hydroxyacyl-CoA dehydrogenase type-2	Hsd17b10	O08756	11.11	2	/	/
Isocitrate dehydrogenase [NADP] cytoplasmic	Idh1	O88844	11.59	5	/	/
IQ motif and SEC7 domain-containing protein 3	Iqsec3	Q3TES0	6.03	7	/	/
Integral membrane protein 2C	Itm2c	Q91VK4	12.27	2	/	/
Kinesin-1 heavy chain	Kif5b	Q61768	/	/	/	/
Kinesin heavy chain isoform 5C	Kif5c	P28738	13.49	12	/	/
Importin subunit beta-1	Kpnb1	P70168	/	/	/	/
LanC-like protein 2	Lancl2	Q9JJK2	11.11	4	/	/
Protein lin-7 homolog C	Lin7c	O88952	/	/	/	/
Protein LSM12 homolog	Lsm12	Q9D0R8	15.90	3	/	/
Mitogen-activated protein kinase kinase kinase 5	Map3k5	O35099	22.68	35	/	/
Microtubule-associated protein 6	Map6	Q7TSJ2	44.04	28	/	/
Mitogen-activated protein kinase 1	Mapk1	P63085	18.99	7	/	/
Mitogen-activated protein kinase 3	Mapk3	Q63844	19.21	7	/	/
Myeloid leukemia factor 2	Mlf2	Q99KX1	11.74	3	/	/
Cytochrome c oxidase subunit 2	Mtco2	P00405	18.94	4	/	/

Myosin-10	Myh10	Q61879	20.85	42	/	/
Myosin light polypeptide 6	Myl6	Q60605	21.85	3	/	/
Unconventional myosin-Va	Myo5a	Q99104	9.61	17	/	/
Alpha-soluble NSF attachment protein	Napa	Q9DB05	18.31	6	/	/
Neurocan core protein	Ncan	P55066	/	/	/	/
Nck-associated protein 1	Nckap1	P28660	7.71	9	/	/
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 12	Ndufa12	Q7TMF3	25.52	4	/	/
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13	Ndufa13	Q9ERS2	22.22	3	/	/
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 7	Ndufa7	Q9Z1P6	38.94	4	/	/
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 9, mitochondrial	Ndufa9	Q9DC69	/	/	/	/
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 4	Ndufb4	Q9CQC7	27.13	4	/	/

NADH dehydrogenase [ubiquinone] iron-sulfur protein 2, mitochondrial	Ndufs2	Q91WD5	15.12	7	/	/
NADH dehydrogenase [ubiquinone] flavoprotein 2, mitochondrial	Ndufv2	Q9D6J6	20.56	3	/	/
Adaptin ear-binding coat-associated protein 1	Necap1	Q9CR95	/	/	/	/
Neogenin	Neo1	P97798	18.35	22	/	/
Neurofascin	Nfasc	Q810U3	8.23	11	/	/
Cysteine desulfurase, mitochondrial	Nfs1	Q9Z1J3	10.46	4	/	/
Nuclear receptor-binding protein	Nrbp1	Q99J45	10.47	6	/	/
Neuronal cell adhesion molecule	Nrcam	Q810U4	/	/	/	/
2-oxoglutarate dehydrogenase, mitochondrial	Ogdh	Q60597	/	/	/	/
Dynamin-like 120 kDa protein, mitochondrial	Opa1	P58281	/	/	/	/
Ubiquitin thioesterase OTUB1	Otub1	Q7TQI3	19.93	5	/	/
Oxidation resistance protein 1	Oxr1	Q4KMM3	6.24	5	/	/
Polyadenylate-binding protein 1	Pabpc1	P29341	29.09	21	/	/

Phosphofurin acidic cluster sorting protein 1	Pacs1	Q8K212	42.66	39	/	/
Platelet-activating factor acetylhydrolase IB subunit alpha	Pafah1b1	P63005	17.80	7	/	/
Multifunctional protein ADE2	Paics	Q9DCL9	10.12	5	/	/
Pyruvate carboxylase, mitochondrial	Pc	Q05920	8.49	9	/	/
Propionyl-CoA carboxylase alpha chain, mitochondrial	Pcca	Q91ZA3	6.63	4	/	/
Protein piccolo	Pclo	Q9QYX7	5.84	25	/	/
Pyruvate dehydrogenase protein X component, mitochondrial	Pdhx	Q8BKZ9	8.18	4	/	/
Pyridoxal kinase	Pdxk	Q8K183	/	/	/	/
Pyridoxal phosphate phosphatase	Pdxp	P60487	20.89	5	/	/
Profilin-1	Pfn1	P62962	42.14	4	/	/
Glucose 1,6-bisphosphate synthase	Pgm2l1	Q8CAA7	14.98	8	/	/
Plastin-3	Pls3	Q99K51	7.94	4	/	/
Liprin-alpha-2	Ppfia2	Q8BSS9	5.57	6	/	/
Liprin-alpha-3	Ppfia3	P60469	8.96	9	/	/
Protein phosphatase 1 regulatory subunit 7	Ppp1r7	Q3UM45	14.96	6	/	/
Serine/threonine-protein phosphatase 2A 65 kDa regulatory	Ppp2r1a	Q76MZ3	/	/	/	/

subunit A alpha isoform						
Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	Ppp2r2a	Q6P1F6	10.74	4	/	/
Serine/threonine-protein phosphatase 2B catalytic subunit alpha isoform	Ppp3ca	P63328	22.07	12	/	/
cAMP-dependent protein kinase type II-alpha regulatory subunit	Prkar2a	P12367	10.72	4	/	/
Protein arginine N-methyltransferase 1	Prmt1	Q9JIF0	37.47	15	/	/
Protein arginine N-methyltransferase 8	Prmt8	Q6PAK3	29.95	14	/	/
Proteasome subunit alpha type-4	Psma4	Q9R1P0	/	/	/	/
26S protease regulatory subunit 8	Psmc5	P62196	/	/	/	/
26S proteasome non-ATPase regulatory subunit 3	Psmd3	P14685	5.66	3	/	/
Tyrosine-protein phosphatase non-receptor type 11	Ptpn11	P35235	9.27	5	/	/
Receptor-type tyrosine-protein phosphatase zeta	Ptprz1	B9EKR1	/	/	/	/
Glutaminyl-peptide cyclotransferase	Qpct	Q9CYK2	9.39	3	/	/

Ras-related protein Rab-1A	Rab1A	P62821	48.78	9	/	/
Ras-related protein Rab-7a	Rab7a	P51150	28.50	5	/	/
Rap guanine nucleotide exchange factor 4	Rapgef4	Q9EQZ6	/	/	/	/
RNA binding motif protein, X-linked-like-1	Rbmxl1	Q91VM5	19.33	7	/	/
Rho-associated protein kinase 2	Rock2	P70336	4.61	6	/	/
Rabphilin-3A	Rph3a	P47708	52.57	36	/	/
60S ribosomal protein L12	Rpl12	P35979	28.48	4	/	/
60S ribosomal protein L23	Rpl23	P62830	27.14	3	/	/
60S ribosomal protein L6	Rpl6	P47911	9.12	3	/	/
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	Rpn1	Q91YQ5	12.17	6	/	/
40S ribosomal protein S3	Rps3	P62908	20.58	4	13.55	5
Reticulon-4	Rtn4	Q99P72	11.79	12	8.11	8
UPF0568 protein C14orf166 homolog	RTRAF	Q9CQE8	20.08	5	21.05	6
Septin-6	Sept6	Q9R1T4	34.33	14	9.09	2
Septin-9	Sept9	Q80UG5	/	/	11.52	9

Plasminogen activator inhibitor 1 RNA-binding protein	Serbp1	Q9CY58	9.83	3	6.86	3
Leukocyte elastase inhibitor A	Serpinb1a	Q9D154	14.25	5	17.24	5
Serpin B6	Serpinb6	Q60854	12.96	5	11.99	5
Splicing factor 3A subunit 1	Sf3a1	Q8K4Z5	/	/	24.67	16
Splicing factor 3A subunit 2	Sf3a2	Q62203	9.68	5	22.02	6
SH3 and multiple ankyrin repeat domains protein 3	Shank3	Q4ACU6	2.14	3	11.54	11
Tyrosine-protein phosphatase non-receptor type substrate 1	Sirpa	P97797	12.09	4	11.16	2
Sickle tail protein	Skt	A2AQ25	11.66	21	19.76	4
Solute carrier family 12 member 2	Slc12a2	P55012	/	/	18.18	5
Calcium-binding mitochondrial carrier protein Aralar1	Slc25a12	Q8BH59	/	/	16.00	4
Clathrin coat assembly protein AP180	Snap91	Q61548	8.99	7	17.49	5
Protein SOGA3	Soga3	Q6NZL0	5.19	4	9.48	10
SRC kinase signaling inhibitor 1	Srcin1	Q9QWI6	35.60	42	33.75	7
Serine-threonine kinase receptor-associated protein	Strap	Q9Z1Z2	7.14	2	7.54	3

Succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	Suclg2	Q9Z2I8	11.55	4	2.64	6
Synapsin-3	Syn3	Q8JZP2	28.32	12	18.79	9
Ras/Rap GTPase-activating protein SynGAP	Syngap1	F6SEU4	11.64	13	3.42	3
Synaptojanin-1	Synj1	Q8CHC4	19.00	30	12.25	12
Transaldolase	Taldo1	Q93092	31.75	11	12.20	5
Tubulin-folding cofactor B	Tbcb	Q9D1E6	/	/	3.71	5
Nucleolysin TIAR	Tial1	P70318	15.05	7	3.66	4
Mitochondrial import inner membrane translocase subunit TIM50	Timm50	Q9D880	9.35	3	4.47	4
Tropomyosin alpha-1 chain	Tpm1	P58771	/	/	27.20	8
Tropomyosin alpha-3 chain	Tpm3	P21107	/	/	18.78	4
Nucleoprotein TPR	Tpr	F6ZDS4	5.80	12	12.50	4
TSC22 domain family protein 1	Tsc22d1	P62500	3.90	4	7.91	9
Translin	Tsn	Q62348	/	/	2.03	4
Tetraspanin-14	Tspan14	Q8QZY6	10.74	3	11.32	5
Tetraspanin-15	Tspan15	F7BWT7	22.79	7	11.71	7
Tetraspanin-7	Tspan7	Q62283	/	/	6.97	7
Ubiquitin carboxyl-terminal hydrolase 5	Usp5	P56399	4.90	3	4.55	3

Vesicle-associated membrane protein-associated protein A	Vapa	Q9WV55	11.24	2	24.86	13
Vesicle-associated membrane protein-associated protein B	Vapb	Q9QY76	19.75	4	13.91	7
Vinculin	Vcl	Q64727	8.26	8	13.18	6
Wiskott-Aldrich syndrome protein family member 3	Wasf3	Q8VHI6	6.59	3	12.85	3
WD repeat-containing protein 37	Wdr37	Q8CBE3	28.63	15	13.16	3
WD repeat-containing protein 7	Wdr7	Q920I9	3.02	4	27.82	9
Serine/threonine-protein kinase WNK1	Wnk1	P83741	9.63	22	26.67	9
Serine/threonine-protein kinase WNK2	Wnk2	Q3UH66	26.24	50	12.30	3
Zinc finger C2HC domain-containing protein 1A	Zc2hc1a	Q8BJH1	14.51	4	23.48	4

Table 2.2.1 Table of the identified ADAM10 putative interactors in WT and zQ175 mice. Protein name, gene name, Uniprot code, sequence coverage, and peptides are reported for each protein. In the brackets, the number of peptides overcoming Mascot significant statistical threshold is also reported. The symbol / means that a specific protein was not identified in that condition.

Both lists were then filtered according to the Contaminant Repository for Affinity Purification database (CRAPome), a web-accessible resource that stores and annotates negative control experiments generated by the proteomics research community[18]. Considering non-specific contaminant proteins frequently occurring in at least 50% of the experiments collected by the database, we excluded 25 proteins, thus obtaining a final list of 197 putative ADAM10's interactors: 91 shared among the two conditions, the 68 and 38 proteins found

exclusively in the wild-type or HD brain, respectively.

By analyzing ADAM10 WT and MUT interactomes by Cytoscape, the pools of shared and specific interactors were obtained and reported in Fig. 2.2.4, where in blue, “Glutamatergic synapse”-related proteins are highlighted.

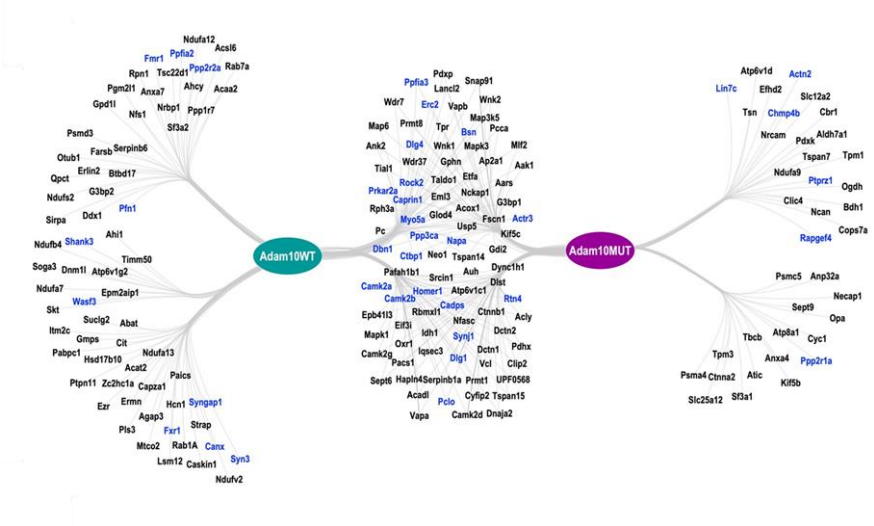

Fig. 2.2.4 Cytoscape network representing putative ADAM10 interactors. Nodes (proteins) identified selectively in wild-type (ADAM10WT) and zQ175 (ADAM10 MUT), and interactors identified in both the wild-type and zQ175 striatum are shown. Genes associated according to Gene Ontology enrichment analysis with ‘Glutamatergic synapse’-related proteins are highlighted in blue.

Next, we analyzed proteomics data by a Gene Ontology enrichment functional analysis. By interrogating the Cellular Component domain and imposing the thresholds $P < 0.01$ and enrichment ≥ 2, 38 enriched ontology categories were found for the wild-type and 34 for the zQ175 system. Fig 3.2.5 displays the top-12 categories (ranked by P-value); as expected, most of them were shared between the two conditions, with similar enrichment and P-values. Notably, this holds true for the top-three terms: glutamatergic synapse, post-synaptic density (PSD), and presynaptic active zone (AZ). The latter category suggests for the first time an ADAM10 localization and biological activity at the level of pre-synapsis, never

described before.

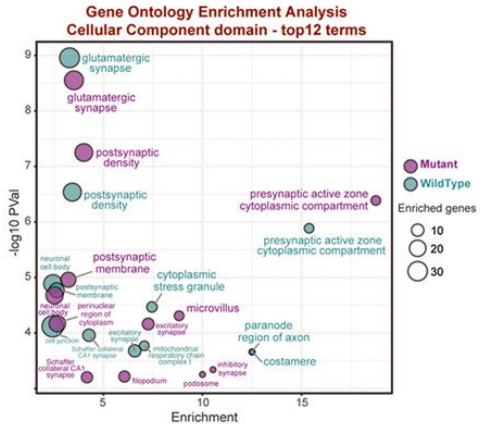

Fig. 2.2.5 Gene Ontology enrichment analysis of the ADAM10 interactome in wild-type (green) or zQ175 (purple). The plot depicts the top-12 terms for each condition (ranked by their P-value); the x-axis reports the enrichment folds and the y-axis the significance (as -log10 P values), whereas the size of each dot is proportional to the number of genes assigned to the category.

The ADAM10 distribution between pre- and post-synapsis was then investigated by immunoblotting of separated fractions and by using synaptophysin and PSD95 as pre- and post-synaptic markers, respectively. As shown in Fig. 2.2.6, ADAM10 is present in both synaptic compartments, and its levels in R6/2 is always higher than that estimated for wild-type littermates.

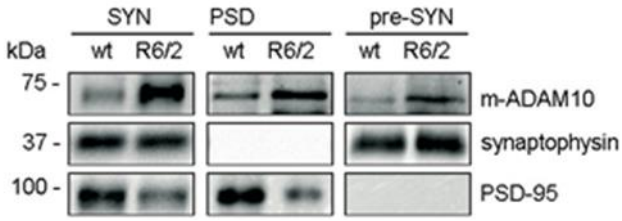

Fig. 2.2.6 Western blot for m-ADAM10 on PSDs and PRE-SYN from a pool of three cortices from 12-week-old wild-type and R6/2 mice. Synaptophysin, PRE-SYN marker. PSD95, PSD marker.

2.2.3.2 Active ADAM10 and the presynaptic protein Piccolo form a complex that is altered in the HD brain

To validate the predicted interactions of ADAM10 with proteins at the AZ, we performed co-immunoprecipitation experiments in cortices from wild-type and R6/2 mice, another certified model for HD. We focused on Piccolo, because of its recognized role in the structural and functional assembly of presynaptic AZ, in guiding synaptic vesicles (SVs) from the backfield of the synapse to the AZ for release, and in the recycling and maintenance of SVs[19,20]. We found that the active form of ADAM10 co-immunoprecipitated with Piccolo in total protein lysates prepared from the wild-type mouse cortex, as showed in the western blot assay reported in Fig. 2.2.7.

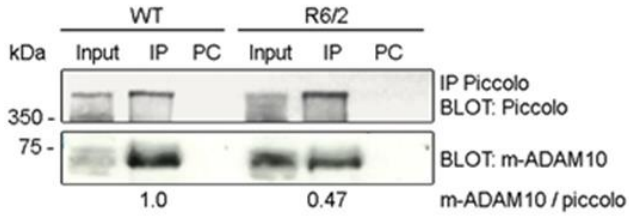

Fig. 2.2.7 Co-IP experiments for the ADAM10/Piccolo complex in cortical protein lysates from a pool wild-type and R6/2 mice, respectively.

Interestingly, western blot assay showed that the amount of ADAM10 co-immunoprecipitated with Piccolo was largely lower in the R6/2 than in wt mice cortices. Indeed, although ADAM10 is overexpressed in HD brains, estimating an increase of approximately about 60%[15] (Fig. 2.2.8, panel A and panel B), the quote interacting with Piccolo was decreased about 50% in the cortices of symptomatic R6/2 mice (Fig. 2.2.7).

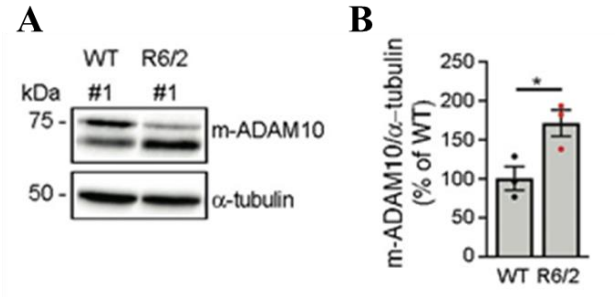

Fig. 2.2.8 Representative Western blot and quantification of m-ADAM10 in cortices from 12-week-old wild-type (n =3) and R6/2 mice (n = 3). α-Tubulin was the loading control. Data are presented as mean ± SEM; *p < 0.05, unpaired t-test.

To exclude the possibility that the observed changes in interaction abundance were due to changes in the Piccolo protein level in HD, we performed Western blot for Piccolo in 3 biological replicates for wt and R6/2 transgenic mice. A shown in Fig. 2.2.9, protein levels of Piccolo were similar between all genotypes and replicates.

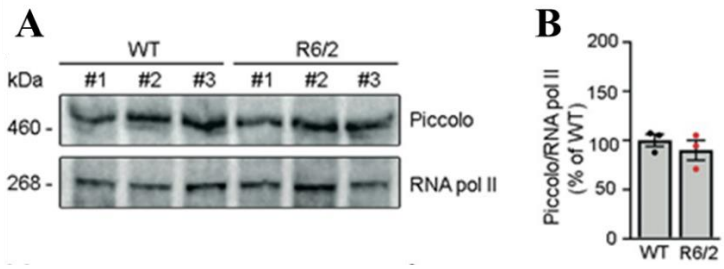

Fig. 2.2.9 Representative Western blot and quantification of Piccolo in cortices from 12-week-old wild-type (n = 3) and R6/2 mice (n = 3). RNA polymerase II was the loading control. Data are presented as mean ± SEM.

As already mentioned, Piccolo exerts multiple functions at the AZ at the level of endosome formation and the maintenance of the SV reserve pool[20]. Piccolo also regulates the translocation of SVs from the reserve to the readily releasable pool[21] and safeguards a fraction of them for delayed action potential-induced release. This enables the synapse to sustain high-frequency synaptic transmission over long periods[22].

We used transmission electron microscopy (TEM) to test SV density and distribution in WT, R6/2, and R6/2-ADAM10cKO cortical neurons. TEM analyses revealed a reduction in vesicle density in R6/2 cortical presynaptic boutons compared to the wild-type mice with similar synaptic area, active zone (AZ) length, and vesicle diameter (Fig. 2.2.10, panel A). Next, we compared the cumulative % of SV distribution 0 to 500 nm from the AZ in the wild-type and R6/2 presynaptic terminal. SVs whose centers were located within 25 nm from the presynaptic membrane (the whole vesicle is contained within 50 nm) were considered docked, whereas the non-docked SVs were classified as reserve (vesicle centers located 50–300 nm from membrane) or resting (vesicle centers located > 300 nm from membrane) pools[23] (Fig. 2.2.10, panel B).

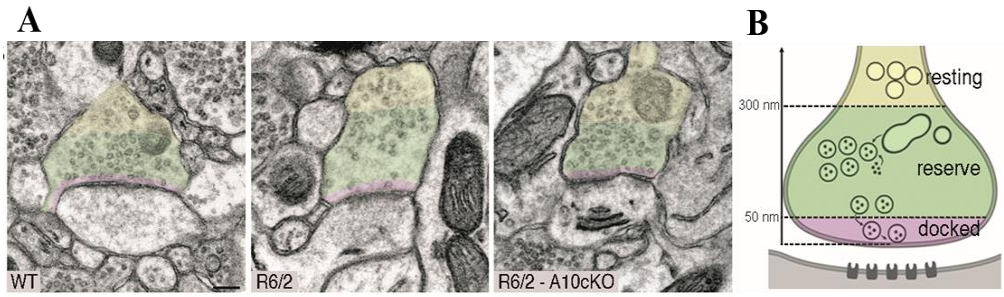

Fig 2.2.10 (A) Representative TEM images from wild-type, R6/2, and R6/2-A10cKO cortical synapses. Scale bar: 100 nm. (B) Diagram showing synaptic vesicle classification based on distance from the presynaptic membrane.

In R6/2 mice, the cumulative SV distribution had a markedly disturbed profile, with a significant reduction in the cumulative % of SVs located within 50 nm (docked vesicles) and between 50-300 nm from the AZ (reserve vesicles). In R6/2 KO (Fig. 2.2.10, panel B), a rescue of the physiological profile has been restored, confirming that the unbalance of ADAM10 and Piccolo interaction induces a perturbation in vesicles density and distribution at pre-synapsis.

2.2.4 Conclusions

Altogether these data have revealed that ADAM10 is involved in presynaptic functions, specifically in the regulation of SV dynamics at the AZ. We also demonstrated the presence of morphological presynaptic defects, particularly an altered distribution of SVs at the AZ, in the HD mice *in vivo*. Mutant HTT reduces the density of SVs in the reserve and docked pools, and we showed that, despite the presence of increased m-ADAM10, the formation of m-ADAM10/Piccolo complex is reduced in HD. We also demonstrated that the observed presynaptic defects in the HD cortex correlate with reduced formation of the m-ADAM10/Piccolo complex. These findings indicate that ADAM10 and Piccolo at the presynapse can be a relevant component of HD pathogenesis and reinforces the role of the cortex and the presynaptic compartment in HD, which have to be considered in future therapies. This study suggests a novel aspect of ADAM10 biology and creates a paradigm for future investigation of the ADAM10 physiological role at the presynaptic terminal. Additional studies are also needed to understand the contribution of presynaptic and postsynaptic ADAM10 to HD cortico-striatal dysfunction. Hyperactive ADAM10 in the postsynaptic terminal causes excessive proteolysis of the synaptic cell adhesion protein N-cadherin. Notably, postsynaptic N-cadherin proteolysis has been proposed to influence SV dynamics and release from the presynaptic terminal[24,25,26,27]. Targeting ADAM10 selectively in the presynaptic or postsynaptic terminal will help elucidate the relevance of the action of the enzyme in the two synaptic compartments and could reveal which site of the HD cortico-striatal synapse could be the primary target for future ADAM10-modulating therapies.

The results discussed in this chapter (including figures and tables) have been published in the following article: Cozzolino, F.; Vezzoli, E.; Cheroni, C.; Besusso, D.; Conforti, P.; Valenza, M.; **Iacobucci, I.**; Monaco, V.; Birolini, G.;

Bombaci, M.; Falqui, A.; Saftig, P.; Rossi, R. L.; Monti, M.; Cattaneo, E.; Zuccato, C. ADAM10 Hyperactivation Acts on Piccolo to Deplete Synaptic Vesicle Stores in Huntington's Disease. *Hum Mol Genet* **2021**. https://doi.org/10.1093/hmg/ddab047.

2.2.5 References

(1) Saftig, P.; Lichtenthaler, S. F. The Alpha Secretase ADAM10: A Metalloprotease with Multiple Functions in the Brain. *Prog Neurobiol* **2015**, *135*, 1–20. https://doi.org/10.1016/j.pneurobio.2015.10.003.

(2) Marcello, E.; Gardoni, F.; Mauceri, D.; Romorini, S.; Jeromin, A.; Epis, R.; Borroni, B.; Cattabeni, F.; Sala, C.; Padovani, A.; Di Luca, M. Synapse-Associated Protein-97 Mediates -Secretase ADAM10 Trafficking and Promotes Its Activity. *Journal of Neuroscience* **2007**, *27* (7), 1682–1691. https://doi.org/10.1523/JNEUROSCI.3439-06.2007.

(3) Kuhn, P.-H.; Colombo, A. V.; Schusser, B.; Dreymueller, D.; Wetzel, S.; Schepers, U.; Herber, J.; Ludwig, A.; Kremmer, E.; Montag, D.; Müller, U.; Schweizer, M.; Saftig, P.; Bräse, S.; Lichtenthaler, S. F. Systematic Substrate Identification Indicates a Central Role for the Metalloprotease ADAM10 in Axon Targeting and Synapse Function. *Elife* **2016**, *5*. https://doi.org/10.7554/eLife.12748.

(4) Benarroch, E. E. ADAM Proteins, Their Ligands, and Clinical Implications. *Neurology* **2012**, *78* (12), 914–920. https://doi.org/10.1212/WNL.0b013e31824c4728.

(5) Endres, K.; Deller, T. Regulation of Alpha-Secretase ADAM10 In Vitro and In Vivo: Genetic, Epigenetic, and Protein-Based Mechanisms. *Front Mol Neurosci* **2017**, *10*, 56. https://doi.org/10.3389/fnmol.2017.00056.

(6) Attwell, D.; Gibb, A. Neuroenergetics and the Kinetic Design of Excitatory Synapses. *Nat Rev Neurosci* **2005**, *6* (11), 841–849. https://doi.org/10.1038/nrn1784.

(7) Malinverno, M.; Carta, M.; Epis, R.; Marcello, E.; Verpelli, C.; Cattabeni, F.; Sala, C.; Mulle, C.; Di Luca, M.; Gardoni, F. Synaptic Localization and Activity of ADAM10 Regulate Excitatory Synapses through N-Cadherin

Cleavage. *J Neurosci* **2010**, *30* (48), 16343–16355. https://doi.org/10.1523/JNEUROSCI.1984-10.2010.

(8) Barthet, G.; Georgakopoulos, A.; Robakis, N. K. Cellular Mechanisms of γ-Secretase Substrate Selection, Processing and Toxicity. *Prog Neurobiol* **2012**, *98* (2), 166–175. https://doi.org/10.1016/j.pneurobio.2012.05.006.

(9) Andreyeva, A.; Nieweg, K.; Horstmann, K.; Klapper, S.; Müller-Schiffmann, A.; Korth, C.; Gottmann, K. C-Terminal Fragment of N-Cadherin Accelerates Synapse Destabilization by Amyloid-β. *Brain* **2012**, *135* (Pt 7), 2140–2154. https://doi.org/10.1093/brain/aws120.

(10) A Novel Gene Containing a Trinucleotide Repeat That Is Expanded and Unstable on Huntington's Disease Chromosomes. The Huntington's Disease Collaborative Research Group. *Cell* **1993**, *72* (6), 971–983. https://doi.org/10.1016/0092-8674(93)90585-e.

(11) Zuccato, C.; Valenza, M.; Cattaneo, E. Molecular Mechanisms and Potential Therapeutical Targets in Huntington's Disease. *Physiol Rev* **2010**, *90* (3), 905–981. https://doi.org/10.1152/physrev.00041.2009.

(12) Saudou, F.; Humbert, S. The Biology of Huntingtin. *Neuron* **2016**, *89* (5), 910–926. https://doi.org/10.1016/j.neuron.2016.02.003.

(13) Veldman, M. B.; Yang, X. W. Molecular Insights into Cortico-Striatal Miscommunications in Huntington's Disease. *Curr Opin Neurobiol* **2018**, *48*, 79–89. https://doi.org/10.1016/j.conb.2017.10.019.

(14) Lo Sardo, V.; Zuccato, C.; Gaudenzi, G.; Vitali, B.; Ramos, C.; Tartari, M.; Myre, M. A.; Walker, J. A.; Pistocchi, A.; Conti, L.; Valenza, M.; Drung, B.; Schmidt, B.; Gusella, J.; Zeitlin, S.; Cotelli, F.; Cattaneo, E. An Evolutionary Recent Neuroepithelial Cell Adhesion Function of Huntingtin Implicates ADAM10-Ncadherin. *Nat Neurosci* **2012**, *15* (5), 713–721. https://doi.org/10.1038/nn.3080.

(15) Vezzoli, E.; Caron, I.; Talpo, F.; Besusso, D.; Conforti, P.; Battaglia, E.;

Sogne, E.; Falqui, A.; Petricca, L.; Verani, M.; Martufi, P.; Caricasole, A.; Bresciani, A.; Cecchetti, O.; Rivetti di Val Cervo, P.; Sancini, G.; Riess, O.; Nguyen, H.; Seipold, L.; Saftig, P.; Biella, G.; Cattaneo, E.; Zuccato, C. Inhibiting Pathologically Active ADAM10 Rescues Synaptic and Cognitive Decline in Huntington's Disease. *J Clin Invest* **2019**, *129* (6), 2390–2403. https://doi.org/10.1172/JCI120616.

(16) Gur, T. Biobtree: A Tool to Search, Map and Visualize Bioinformatics Identifiers and Special Keywords. *F1000Res* **2019**, *8*. https://doi.org/10.12688/f1000research.17927.4.

(17) Medugno, L.; Costanzo, P.; Lupo, A.; Monti, M.; Florio, F.; Pucci, P.; Izzo, P. A Novel Zinc Finger Transcriptional Repressor, ZNF224, Interacts with the Negative Regulatory Element (AldA-NRE) and Inhibits Gene Expression. *FEBS Lett* **2003**, *534* (1–3), 93–100. https://doi.org/10.1016/s0014-5793(02)03783-3.

(18) Mellacheruvu, D.; Wright, Z.; Couzens, A. L.; Lambert, J.-P.; St-Denis, N. A.; Li, T.; Miteva, Y. V.; Hauri, S.; Sardiu, M. E.; Low, T. Y.; Halim, V. A.; Bagshaw, R. D.; Hubner, N. C.; Al-Hakim, A.; Bouchard, A.; Faubert, D.; Fermin, D.; Dunham, W. H.; Goudreault, M.; Lin, Z.-Y.; Badillo, B. G.; Pawson, T.; Durocher, D.; Coulombe, B.; Aebersold, R.; Superti-Furga, G.; Colinge, J.; Heck, A. J. R.; Choi, H.; Gstaiger, M.; Mohammed, S.; Cristea, I. M.; Bennett, K. L.; Washburn, M. P.; Raught, B.; Ewing, R. M.; Gingras, A.-C.; Nesvizhskii, A. I. The CRAPome: A Contaminant Repository for Affinity Purification-Mass Spectrometry Data. *Nat Methods* **2013**, *10* (8), 730–736. https://doi.org/10.1038/nmeth.2557.

(19) Gundelfinger, E. D.; Reissner, C.; Garner, C. C. Role of Bassoon and Piccolo in Assembly and Molecular Organization of the Active Zone. *Front Synaptic Neurosci* **2015**, *7*, 19. https://doi.org/10.3389/fnsyn.2015.00019.

(20) Ackermann, F.; Schink, K. O.; Bruns, C.; Izsvák, Z.; Hamra, F. K.;

Rosenmund, C.; Garner, C. C. Critical Role for Piccolo in Synaptic Vesicle Retrieval. *Elife* **2019**, *8*. https://doi.org/10.7554/eLife.46629.

(21) Leal-Ortiz, S.; Waites, C. L.; Terry-Lorenzo, R.; Zamorano, P.; Gundelfinger, E. D.; Garner, C. C. Piccolo Modulation of Synapsin1a Dynamics Regulates Synaptic Vesicle Exocytosis. *J Cell Biol* **2008**, *181* (5), 831–846. https://doi.org/10.1083/jcb.200711167.

(22) Parthier, D.; Kuner, T.; Körber, C. The Presynaptic Scaffolding Protein Piccolo Organizes the Readily Releasable Pool at the Calyx of Held. *J Physiol* **2018**, *596* (8), 1485–1499. https://doi.org/10.1113/JP274885.

(23) Miranda, R.; Nudel, U.; Laroche, S.; Vaillend, C. Altered Presynaptic Ultrastructure in Excitatory Hippocampal Synapses of Mice Lacking Dystrophins Dp427 or Dp71. *Neurobiol Dis* **2011**, *43* (1), 134–141. https://doi.org/10.1016/j.nbd.2011.02.017.

(24) Bozdagi, O.; Valcin, M.; Poskanzer, K.; Tanaka, H.; Benson, D. L. Temporally Distinct Demands for Classic Cadherins in Synapse Formation and Maturation. *Mol Cell Neurosci* **2004**, *27* (4), 509–521. https://doi.org/10.1016/j.mcn.2004.08.008.

(25) Jüngling, K.; Eulenburg, V.; Moore, R.; Kemler, R.; Lessmann, V.; Gottmann, K. N-Cadherin Transsynaptically Regulates Short-Term Plasticity at Glutamatergic Synapses in Embryonic Stem Cell-Derived Neurons. *J Neurosci* **2006**, *26* (26), 6968–6978. https://doi.org/10.1523/JNEUROSCI.1013-06.2006.

(26) Vitureira, N.; Letellier, M.; White, I. J.; Goda, Y. Differential Control of Presynaptic Efficacy by Postsynaptic N-Cadherin and β-Catenin. *Nat Neurosci* **2011**, *15* (1), 81–89. https://doi.org/10.1038/nn.2995.

(27) Gottmann, K. Transsynaptic Modulation of the Synaptic Vesicle Cycle by Cell-Adhesion Molecules. *J Neurosci Res* **2008**, *86* (2), 223–232. https://doi.org/10.1002/jnr.21484.

2.3 Mapping of OPN-ICOSL interacting regions

2.3.1 Introduction

ICOSL (B7-H2, CD275) belongs to the B7 family and regulates the immune response by delivering costimulatory signals through ICOS, a surface receptor mainly expressed by activated T cells[1,2,3]. ICOSL is constitutively expressed by antigen-presenting cells (APCs) as well as some non-lymphoid cells, including several tumor cell types[4,5]. The expression of ICOSL in non-lymphoid tissues, such as the brain, lungs, heart, kidney, liver, and gut, suggests that it regulates the activation of antigen-experienced effector/memory T cells. However, ICOSL is expressed at high levels in T helper follicular (TFH) cells, and ICOS deficiency has been associated with the defective formation of lymphoid follicles in mice and the development of common variable immunodeficiency in humans[6].

The main known function of ICOSL is triggering ICOS, which serves as a costimulatory molecule for T cells and supports cytokine-driven polarization of T helper (Th) cells. Conversely, when ICOS binds, ICOSL triggers "reverse signaling" into the ICOSL expressing cell. In particular, following the interaction between ICOSL and a recombinant soluble form of ICOS (ICOS-Fc), adhesiveness and migration of human umbilical vein endothelial cells (HUVECs) and several tumor cell lines were inhibited, as well as the development of lung metastases in the B16 melanoma model[7,8]. Other effects of ICOSL triggering have been detected in dendritic cells (DC), whose treatment with ICOS-Fc not only impairs cell migration and adhesion but also modulates cytokine secretion and antigen cross-presentation in class I MHC molecules[9,10,11].

Osteopontin (OPN) is a phosphoprotein secreted by several cell types, such as macrophages, DC, and Th cells; it can function both as a matricellular protein and a cytokine mediating several biological functions. These include migration,

adhesion, activation of inflammatory cells, and modulation of T cell activation supporting differentiation of proinflammatory type 1 (Th1) and type 17 (Th17) Th cells[12]. It is highly expressed and secreted in several types of tumors, promoting invasion and metastatic dissemination[13]. Importantly, current research indicates that OPN inhibition would be a good therapeutic approach to counteract metastatic diseases. In preclinical models, knocking down OPN by RNAi, aptamers, or antibodies lead to promising results in cancer containment. Due to its autocrine and paracrine activities, OPN appears to be a crucial mediator of cellular cross-talk and an influential factor in the tumor microenvironment.

OPN's pleiotropic activities are partly due to its capacity to interact with multiple ligands, including several cell surface receptors, namely several integrins and CD44, calcium, and heparin. OPN's biological functions are also influenced by post-translational modifications, such as phosphorylation, glycosylation, and protein cleavage mediated by thrombin and metalloproteinases[14].

OPN is an intrinsically disordered protein (IDP), and this feature confers to OPN the ability to adopt different functional structures by folding upon binding and to interact with multiple binding partners[15]. Thrombin cleaves OPN in the middle of the molecule, near to an RGD motif, and generates two fragments with slightly different functional activities[12].

Recent findings depict a functional network between ICOS, ICOSL, and OPN. Firstly, all these molecules support Th17 cell responses. Moreover, they are involved in cell migration tuning since it is activated by OPN and inhibited by ICOS-mediated ICOSL triggering[10]. Finally, they are involved in bone metabolism since OPN is a key bone component produced by osteoblasts, whereas ICOS-mediated triggering of ICOSL expressed by osteoclasts (OC) inhibits OC differentiation from monocytes and bone resorption activity of mature OC *in vitro*, and the development of osteoporosis *in vivo*[16].

A prior study showed that HUVECs treatment with ICOS-Fc inhibits ERK phosphorylation induced by OPN, but not that induced by ATP[8]. In a more recent work, it has been reported that ICOS-Fc inhibits HUVEC tubulogenesis induced by OPN, but not that induced by VEGFα.

The structural and functional crosstalk between ICOS/ICOSL and OPN was further *in vitro* and *in vivo* investigated in collaboration with the research group of the Professor Annalisa Chiocchetti of the Department of Health Sciences of the University of "Piemonte Orientale". In detail, I have dealt with conformational studies to define the interacting regions between OPN and ICOSL by employing an experimental strategy that relied on limited proteolysis and crosslinking reagents.

2.3.2 Experimental methods

2.3.2.1 ELISA-based interaction assay

OPN (60 nM) (Bio-techne) in PBS was used to coat Nunc MaxiSorp™ flat-bottom plates (Life Technologies, Carlsbad, California, USA) overnight at 4°C. After one wash with PBS + 0.25% Triton X-100 (Merck, Darmstadt, Germany), 3% of Bovine Serum Albumin ((BSA) Merck) in PBS + 0.05% Tween-20 (Merck) was added for 1h at 25°C. After three washes, the plate was incubated with titrated amounts (from 60 nM to 1.8 nM) of ICOSL-Fc (Bio-techne system) in PBS + 0.05% Tween-20 for 1 h at 25°C with or without 60nM of ICOS-Fc (competition assay). After washing, ICOSL-Fc binding to OPN was evaluated using HRP conjugated anti-human IgG1 antibodies (1:4000) (Dako, Santa Clara, California, USA) in PBS + 0.05% Tween-20 for 1h at 25°C, followed by washing and addition of the TMB substrate (Merck). The reaction was stopped after 2 min with H_2SO_4 2M (Merck), and absorbance was assessed at 450 nm using a Victor-X1 plate reader (PerkinElmer, Waltham, Massachusetts, USA).

For the inverted ELISA, 60 nM of ICOSL-His in PBS was used to coat the ELISA plates, and titrated amounts of OPN (from 60 nM to 1.8 nM) were incubated in the wells to assess binding. After washing, the bound OPN was revealed using polyclonal anti-OPN antibodies Biotinylated (Bio-techne) with an incubation of 2h at 25°C; after three washes Streptavidin-HRP (Bio-techne) was added for 20 min at 25°C; the signal was developed following the above protocols.

In other experiments, 60 nM of OPN-GST or home-made OPN-FL were used to compare the interaction between commercial OPN and the others, and 60 nM of OPN-N terminal and OPN-C terminal respectively were used to map the interaction site of ICOSL-OPN following the below protocol.

2.3.2.2 Limited Proteolysis experiments

Limited Proteolysis experiments were carried out on commercial recombinant His-tagged Osteopontin (OPN), produced by R&D (Minneapolis, Minnesota, US) in a mouse myeloma cell line, containing the physiological post-translational modifications but lacking 15 aa of exon 5, on ICOSL-Fc provided by R&D (Minneapolis, Minnesota, US) and on OPN/ICOSL-Fc. The three samples were treated with trypsin and chymotrypsin (Sigma-Aldrich) in phosphate Dulbecco's buffer (pH=7.4), following set-up experiments to optimize enzyme:substrate (E:S) ratios for each protein. For both proteases, the E:S ratio was 1:5000 for both isolated OPN and OPN/ICOSL-Fc complex and 1:100 for isolated ICOSL- Fc. In the complex analysis, OPN and ICOSL-Fc were pre-incubated with a 2:1 molar ratio for 1h at room temperature before starting the proteolysis experiment to allow complex formation. Enzymatic digestions were run for 30 min, after which 2% TFA was added to halt the reaction. The proteolysis mixtures were resolved on a 15% acrylamide SDS-PAGE gel, and differential bands were cut and subjected to *in situ* hydrolysis with trypsin. Tryptic peptide mixtures were analyzed both by 4800 MALDI TOF/TOF Analyzer (Sciex, Framingham,

Massachusetts, USA) and by nanoLC-MS/MS using a Proxeon-nanoEasy II-LTQ Orbitrap XL (Thermo Fisher Scientific). In this latter, peptide fractionation was performed on a C18 capillary reverse-phase column (200 mm, 75 µm, 5 µm) working at 250 nl/min flow rate, using a step gradient of eluent B (0.2% formic acid, 95% acetonitrile LC-MS Grade) from 10% to 60% over 69 min and 60% to 95% over 3 min. Mass spectrometric analyses were performed using data-dependent acquisition (DDA) mode over the 400 to 1800 m/z range, followed by acquisition in MS/MS mode of the five most abundant ions present in each MS scan. Peptides were identified using MASCOT software (Matrix Science Boston, USA) searching in a database containing only OPN and ICOSL-Fc sequences.

2.3.2.3 Cross-linking experiments

In-house purified GST-OPN (OPNa accession number NP_001035157.1) and His-tagged ICOSL (Biotechne, Minneapolis, Minnesota, US) were pre-incubated at 2:1 molar ratio for 1h at room temperature, as described above, and once the complex was formed the DTSSP cross-linker (3,3'- dithiobis (sulfosuccinimidyl propionate, Life technologies, Carlsbad, California, USA) was added in a molar excess of 1:50 and incubated for 30 min at 4°C. The reaction mixtures were loaded onto 8% non reducing SDS-PAGE gel. Bands corresponding to isolated proteins and to the complex were excised from the gel and subjected to in situ hydrolysis with trypsin, skipping the reduction and alkylation steps so as to preserve DTSSP disulfide integrity. Peptide mixtures were analyzed by MALDI-MS and by nanoLC-MS/MS, as described above. Peptide identification was carried out both manually and using software such as MASCOT and Batch-Tag, MS-Bridge, and MS-Product tools within the Protein Prospector package. An aliquot of each peptide mixture was reduced in 20mM dithiothreitol for 2h at 37°C and then alkylated with 55mM of iodoacetamide for 30 min in the dark at r.t.; the mixtures thus obtained were analyzed by MALDI-MS and nanoLC-MS/MS, as described above.

2.3.3 Results and Discussion

2.3.3.1 Conformational analysis of the OPN/ICOSL complex by limited proteolysis

Limited proteolysis and chemical cross-linking experiments were employed for conformational studies to map the interacting regions of OPN and ICOSL. Limited proteolysis experiments were run in parallel both on isolated and complexed proteins; the best experimental conditions were set up to limit the number of hydrolytic events on each protein. In these conditions, only the most exposed and flexible regions of OPN and ICOSL were hydrolyzed: by comparing proteolytic patterns of isolated proteins with those obtained for the complex, the interacting regions can be recognized as those exposed to proteolysis in the isolated proteins and hidden in the complex. In set-up experiments carried out to optimize hydrolysis conditions, OPN demonstrated to be more susceptible to hydrolysis than ICOSL-Fc; indeed, the enzyme vs. substrate ratio (E:S) need for controlling the proteolysis was 50 times lower for OPN than for ICOSL-Fc. These data demonstrated that OPN has greater flexibility and less compact structure than ICOSL-Fc, according to the intrinsically disordered nature of OPN[17]. This extremely different structural degree between the two proteins made hard the setting of common experimental conditions to probe both conformations contemporarily. Therefore, to better address and control proteases activity, the experiments were performed on the complex using the lower E:S ratio, thus giving priority to the conformational analysis of OPN rather than ICOSL-Fc, whose properly digestion would require more protease. Trypsin and chymotrypsin were chosen as proteolytic probes. Once hydrolyzed, samples of complex and of isolated proteins were fractionated by SDS-PAGE.

As shown in panel A of Fig. 2.3.1, some differences among the digestion profiles of isolated proteins and complex were highlighted in the SDS-PAGE gel. As

expected, in both tryptic and chymotryptic hydrolyses of complex, a large amount of ICOSL-Fc remained undigested. By contrast, OPN was almost completely digested. Moreover, the proteolytic profiles of isolated OPN displayed low-molecular-weight bands (indicated in the Figure with the numbers from 1 to 5) that are absent in the complex lanes, suggesting that the cleavage sites that generate them might be located in regions of OPN exposed in isolated protein but buried upon interaction with ICOSL (Fig. 2.3.1, panel A).

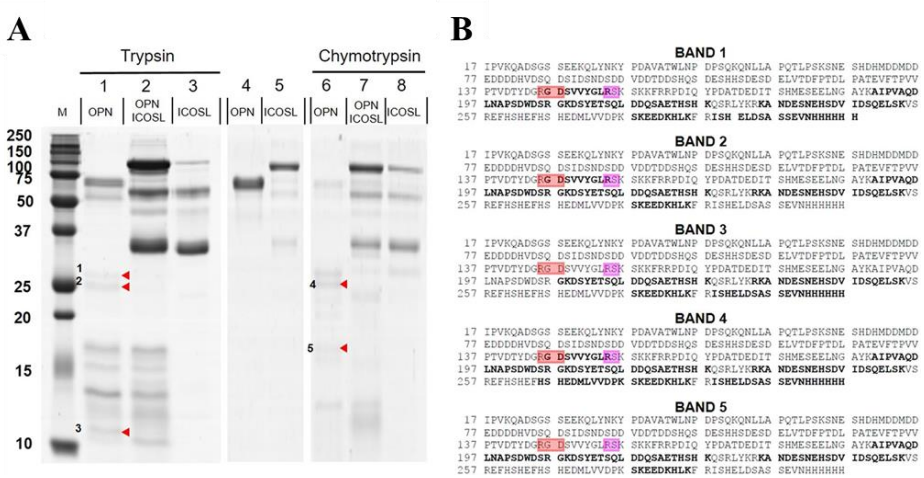

Fig. 2.3.1 Conformational analysis of the OPN/ICOSL complex by limited proteolysis. (A) SDS-PAGE of the proteolytic mixture of OPN, ICOSL-Fc, and their complex, after 30 min of proteolysis with trypsin and chymotrypsin. Lane M: molecular weight markers; lane 1: OPN (5 µg) digested with trypsin, E:S 1:5000; lane 2: OPN/ICOSL complex digested with trypsin, E:S 1:5000; lane 3: ICOSL (5 µg) digested with trypsin, E:S 1:100; lane 4: OPN (2 µg); lane 5: ICOSL (2 µg); lane 6: OPN (5 µg) digested with chymotrypsin, E:S 1:5000; lane 7: OPN–ICOSL 2:1 complex digested with chymotrypsin, E:S 1:5000; lane 8: ICOSL (5 µg) digested with chymotrypsin, E:S 1:100. (B) Sequence coverages were obtained by tryptic mass mapping experiments performed on bands 1, 2, 3, 4, and 5, which were discriminant among isolated proteins and complex. In bold the regions mapped by LC-MS/MS analyses. RGD is boxed in red, and the Thrombin cleavage site in pink.

These specific bands were analyzed using a mass mapping approach. Results showed that the bands with electrophoretic mobility of about 25-26kDa (bands 1, 2, and 4) mapped in the middle of OPN, upstream of position 130, suggesting that

residues close to this amino acid residue are exposed and accessible to proteases in the isolated protein, but become hidden in the OPN/ICOSL-Fc complex (Fig. 2.3.1, panel B). Bands 3 and 5 mapped in the same OPN C-terminal portion as bands 1, 2, and 4, but they covered a substantially shorter sequence, which suggests that they might be generated by sub-digestion events. These data suggest that the region of OPN interaction with ICOSL-Fc is located in the central part of the protein near the RGD motif highlighted in pink and the physiological thrombin cleavage site highlighted in purple. Conversely, these experiments gave no hint concerning the ICOSL-Fc counterparts, since it was not substantially digested at the E:S ratio used.

2.3.3.2 Conformational analysis of the OPN/ICOSL complex by cross-linking

In order to describe in more detail the OPN-ICOSL interacting regions, additional experiments were carried out using the bifunctional cross-linker 3,3'-dithiobis(sulfosuccinimidyl propionate) (DTSSP). DTSSP is a lysine-lysine specific reactive with a spacer arm 12 Å in length that can be cleaved under reducing conditions. Cross-linkers can also provide information on the tertiary structure since they may be considered as actual chemical labels: the presence/absence of DTSSP-mediated modification of specific lysine residue is indicative of their exposure within isolated proteins or in the complex. The experiments were carried out in parallel on both isolated proteins and their complex. Then, the samples were separated by non-reducing SDS-PAGE and stained by colloidal Coomassie (Fig. 2.1.2, panel A). In lane 2, where the complex was loaded after reaction with DTSSP, a 100 kDa band (marked with a bracket) with electrophoretic mobility in agreement with that expected for the GST-OPN/ICOSL complex was present. This band was cut out and *in situ* digested with trypsin in non-reducing conditions to preserve the integrity of the cross-linker moiety.

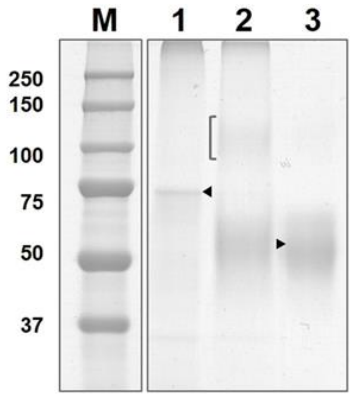

Fig.2.3.2 Non-reducing SDS-PAGE of DTSSP treated samples. Lane 1: GST-OPN; lane 2: complex GST-OPN–ICOSL; lane 3: ICOSL. The bracket indicates a band with the electrophoretic mobility expected for the complex GST-OPN–ICOSL.

LC-MS/MS and MALDI-MS analyses of the peptide mixture detected the presence of peptides derived from both OPN and ICOSL, confirming the identity of complex and the occurrence of direct interaction between the two molecules. However, the mass spectrometer data analysis, performed both manually and using suitable software for cross-linked peptide identification (i.e., Batch-Tag, MS-Bridge, and MS-Product tools within the Protein Prospector package), did not provide direct and conclusive identification of the OPN-ICOSL cross-linking sites. Detecting actual cross-linked peptides within digested mixtures has long been problematic because some cross-linked peptides are too large to be detected, their fragmentation is more complicated than usual linear peptides, and not every protein or peptide is cross-linked in the same way[18]. Therefore, we decided to analyze data considering the reactive as a conformational probe, capable of labeling only exposed and reactive lysine residues both on isolated proteins and when they are bound[19]. To this aim, an aliquot of each peptide sample produced by digestion of the bands containing the cross-linked proteins was reduced, alkylated with iodoacetamide, hydrolyzed by trypsin, and peptide mixtures analyzed by MALDI-MS and LC-MS/MS. In the presence of modification, lysine

residues are not recognized anymore by trypsin as a potential hydrolysis site, becoming missing cleavages within longer peptides.

Overall comparison of MS data from oxidized and reduced samples was informative concerning lysine exposure in OPN and ICOSL, or their complex (Tables 2.3.1 and 2.3.2, respectively).

Lysine Residue	Peptides from OPN	Peptides from (OPN+DTSSP)	Peptides from the complex (OPN-ICOSL + DTSSP)	Peptides from the complex (OPN-ICOSL + DTSSP) REDCAM
K20	/	/	/	/
K30	/	/	21-35 Exp 1682.84 Theor 1682.77	/
K35	31-51 Exp 2447.01 Theor 2447.21	31-51(+1CL$_{H2O}$) Exp 2639.19 Theor 2639.21	21-35 Exp 1682.84 Theor 1682.77	/
K51	36-51 Exp 1800.78 Theor 1800.87	36-51 Exp 1800.80 Theor 1800.87	/	/
K70	52-77 Exp 2887.33 Theor 2877.42	52-77 Exp 2887.37 Theor 2877.42	52-77 Exp 2887.31 Theor 2877.42	/
K77	52-77 Exp 2887.33 Theor 2877.42	52-77 Exp 2887.37 Theor 2877.42	52-77 Exp 2887.31 Theor 2877.42	/
K170	/	160-172 Exp 1394.60 Theor 1394.75	169-175 (+1CL$_{H2O}$+1CL$_{INTRA}$) Exp 1245.46 Theor 1245.53	169-175 (+3CL+3CAM) Exp 1314.54 Theor 1314.53
K172	/	160-172 Exp 1394.60 Theor 1394.75	169-175 (+1CL$_{H2O}$+1CL$_{INTRA}$) Exp 1246.46 Theor 1245.53	169-175 (+3CL+3CAM) Exp 1314.54 Theor 1314.53
K173	/	/	169-175 (+1CL$_{H2O}$+1CL$_{INTRA}$) Exp 1246.46	169-175 (+3CL+3CAM) Exp 1314.54

			Theor 1245.53	Theor 1314.53
K203	176-203 Exp 3223.33 Theor 3223.42	176-203 Exp 3223.38 Theor 3223.42	176-203 Exp 3223.25 Theor 3223.42	/
K222	/	221-244 (+1CL_{H2O}) Exp 2925.22 Theor 2925.22 221-244 (+1CL_{INTRA}) Exp 2907.24 Theor 2907.22	221-244 (+1CL_{H2O}) Exp 2925.13 Theor 2925.22 221-244 (+1CL_{INTRA}) Exp 2907.14 Theor 2907.22	221-244 (+1CL+1CAM) Exp 2878.16 Theor 2878.22 221-244 (+2CL+2CAM) Exp 3023.18 Theor 3023.22
K241	223-241 Exp 2176.82 Theor 2176.91	223-244 (+1CL_{H2O}) Exp 2740.08 Theor 2740.11 221-244 (+1CL_{H2O}) Exp 2925.22 Theor 2925.22	221-244 (+1CL_{H2O}) Exp 2925.13 Theor 2925.22 221-244 (+1CL_{INTRA}) Exp 2907.14 Theor 2907.22	223-244 (+1CL+1CAM) Exp 2693.04 Theor 2693.11 221-244 (+1CL+1CAM) Exp 2878.17 Theor 2878.22 221-244 (+2CL+2CAM) Exp 3023 .18 Theor 3023.22
K247	242-268 Exp 3175.62 Theor 3175.55	242-268 Exp 3175.57 Theor 3175.55 245-268 (+1CL_{H2O}) Exp 2996.10 Theor 2996.36	245-268 (+1CL_{H2O}) Exp 2996.27 Theor 2996.36	/
K249	249-268 Exp 2243.92 Theor 2244.01	242-268 Exp 3175.57 Theor 3175.55 249-268 (+1CL_{H2O}) Exp 2435.99 Theor 2436.01 245-268 (+1CL_{H2O}) Exp 2996.10 Theor 2996.36 249-271 (+1CL_{INTRA})	249-268 (+1CL_{H2O}) Exp 2435.95 Theor 2436.01 245-268 (+1CL_{H2O}) Exp 2996.27 Theor 2996.36	249-268 (+1CL+1CAM) Exp 2388.94 Theor 2389.01

		Exp 2760.20 Theor 2760.22		
K268	249-268 Exp 2243.92 Theor 2244.01	249-271 (+1CL_{H2O}) Exp 2778.19 Theor 2778.22 249-271 (+2CL_{H2O}) Exp 2970.22 Theor 2970.22 249-271 (+1CL_{INTRA}) Exp 2760.20 Theor 2760.22	249-271 (+1CL_{H2O}) Exp 2778.19 Theor 2778.22 249-271 (+2CL_{H2O}) Exp 2970.12 Theor 2970.22 249-271 (+1CL_{INTRA}) Exp 2760.12 Theor 2760.22	249-271 (+2CL+2CAM) Exp 2876.17 Theor 2876.22
K290	272-292 Exp 2534.01 Theor 2534.16	/	/	272-299 (+2CL+2CAM) Exp 3703.41 Theor 3703.61
K292	272-292 Exp 2534.01 Theor 2534.16	/	/	272-299 (+2CL+2CAM) Exp 3703.41 Theor 3703.61
K296	/	/	/	272-299 (+2CL+2CAM) Exp 3703.41 Theor 3703.61
K299	/	/	/	272-299 (+2CL+2CAM) Exp 3703.41 Theor 3703.61

Table 2.3.1 List of peptides identified by the mass mapping procedure and relative to isolated OPN, isolated OPN + DTSSP (indicated with CL); OPN/ICOSL complex +DTSSP (indicated with CL); OPN/ICOSL + DTSSP (indicated with CL) following reduction and alkylation treatment (indicated with CAM). REDCAM term indicates the procedure of reduction and thiol carbamidomethylation.

Lysine Residue	**Peptides from ICOSL**	**Peptides from (ICOSL+DTSSP)**	**Peptides from the complex (OPN-ICOSL + DTSSP)**	**Peptides from the complex (OPN-ICOSL + DTSSP) REDCAM**
N-TERM	19-23 Exp 619.24 Theor 619.28 19-26 Exp 1003.47 Theor	19-26(+1CL_{INTRA}) Exp 1177.44 Theor 1177.49 19-26(+1CL_{H2O}) Exp 1195.45 Theor 1195.49	19-26(+1CL_{INTRA}) Exp 1177.44 Theor 1177.49 19-26(+1CL_{H2O}) Exp 1195.45 Theor 1195.49	19- 26(+1CL+1CAM) Exp 1148.47 Theor 1148.49 19- 26(+2CL+2CAM) Exp 1293.47

	1003.49			Theor 1293.49
K23	19-26 Exp 1003.47 Theor 1003.49	19-26(+1CL$_{INTRA}$) Exp 1177.44 Theor 1177.49 19-26(+1CL$_{H2O}$) Exp 1195.45 Theor 1195.49	19-26(+1CL$_{INTRA}$) Exp 1177.44 Theor 1177.49 19-26(+1CL$_{H2O}$) Exp 1195.45 Theor 1195.49	19-26(+1CL+1CAM) Exp 1148.47 Theor 1148.49 19-26(+2CL+2CAM) Exp 1293.47 Theor 1293.49
K60	45-60 Exp 1992.86 Theor 1992.91	45-60 Exp 1992.86 Theor 1992.91 61-79 Exp 2158.01 Theor 2158.06	45-60 Exp 1992.83 Theor 1992.91 61-79 Exp 2158.00 Theor 2158.06	/
K110	100-110 Exp 1317.61 Theor 1317.66	/	/	/
K174	175-194 Exp 2307.13 Theor 2307.12	175-194 Exp 2307.99 Theor 2307.12	/	/
K241	240-252 Exp 1416.67 Theor 1416.71	240-252(+1CL$_{H2O}$) Exp 1608.70 Theor 1608.71	240-252 Exp 1416.64 Theor 1416.71 240-252(+1CL$_{H2O}$) Exp 1609.64 Theor 1609.71	240-252 (+1CL+1CAM) Exp 1562.66 Theor 1562.71
K252	242-264 Exp 2626.16 Theor 2626.22	242-264 (+1CL$_{H2O}$) Exp 2818.18 Theor 2818.22 253-264 Exp 1470.61 Theor 1470.64	242-264(+1CL$_{H2O}$) Exp 2818.18 Theor 2818.22 253-264 Exp 1470.61 Theor 1470.64	253-264 Exp 1470.61 Theor 1470.64

Table 2.3.2 Peptides identified by the mass mapping procedure and relative to isolated ICOSL, isolated ICOSL + DTSSP (indicated with CL); OPN/ICOSL complex +DTSSP (indicated with CL); OPN/ICOSL + DTSSP (indicated with CL) following reduction and alkylation treatment (indicated with CAM). REDCAM term indicates the procedure of reduction and thiol carbamidomethylation.

Residues modified have been reported in Fig. 2.3.3 on the sequences of both proteins.

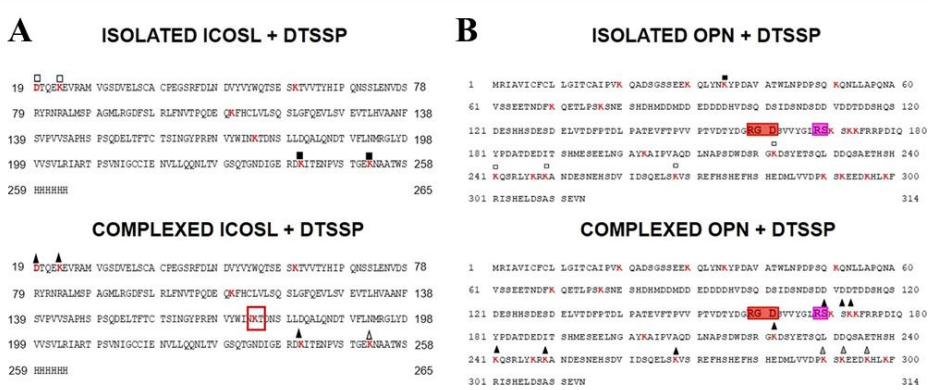

Fig. 2.3.3 ICOSL (B) and OPN (C) sequences with all lysine residues indicated in red and lysine residues modified with squares and triangles in both the isolated proteins (upper panels) and the complex (lower panels). Filled squares indicate lysine residues modified as dead-end by DTSSP; empty squares indicate those modified as dead-ends or intra-molecule cross-linking. Black triangles on OPN/ICOSL complex indicate lysine residues found modified in the complex before and after the REDCAM reaction (reduction and thiol carbamidomethylating). Gray triangles represent lysine residues found modified only in the complex after the REDCAM reaction. Empty triangles are lysine residues found modified only before the REDCAM reaction. RGD is boxed in green and the thrombin cleavage site in orange

With regard to ICOSL, all expected peptides containing a lysine residue were identified, in both isolated ICOSL and the complex, except for K_{174}. This residue was recognized by trypsin in isolated and untreated ICOSL, generating the peptide 175-194, but not in the complex. Our hypothesis was that, once modified, the residue escaped trypsin hydrolysis and occurred within an excessively large tryptic peptide that would be difficult to detect. This finding suggested that the protein region including this residue might be involved in a conformational change upon OPN binding, which made K_{194} more exposed than in the isolated

ICOSL (Fig. 2.3.3, panel A). Concerning OPN, in both the complex and the isolated GST-OPN, the lysine residues K_{222}, K_{241}, K_{249}, and K_{268} were labeled with a single dead-end molecule of DTSSP in all cases. Since these residues are equally exposed in both isolated and complexed GST-OPN, the possibility that they are located in regions involved in the interaction with ICOSL may be ruled out. Moreover, peptides 221-244 and 249-271 displayed a mass increased of 174Da (accounting for the introduction of a DTSSP moiety cross-linking two lysine residues within the linear peptides), suggesting the occurrence of intramolecular cross-links between K_{222} and K_{241} and between K_{249} and K_{268}. Therefore the two pairs of cross-linked lysine residues are spatially close, and their side chains are oriented appropriately to be cross-linked.

K_{247} was always detected in peptides including K_{249}, both in the isolated protein (242-268 and 245- 268+1CL_{H2O}) and in the complex (245-268+1CL_{H2O}); although a certain assignment of its state is not allowed, it is plausible to suppose it behaves similarly to K_{241} and K_{249} given its nearby location (Fig. 2.3.3, panel A).

Residues K_{20}, K_{30}, K_{35}, K_{51}, K_{70} K_{77}, and K_{203} were in no case labeled by DTSSP within the complex, as is indicated by the occurrence of unlabeled peptides including some of these residues (e.g., 21-35, 52-77, 176-203), nor they were generated by a cleavage event in correspondence with a non-modified lysine (e.g., 21-35, 52-77). These findings suggest that, in the GST-OPN/ICOSL complex as well as in OPN alone, none of these residues is ever exposed to be modified by the cross-linking agent. Intriguingly, K_{170}, K_{172}, and K_{173} acted differently from the other residues, in that they were modified by DTSSP only when OPN was bound to ICOSL, suggesting that the region including these three residues undergoes to a significant conformational change upon complex formation, exposing residues whose side chains are concealed or are involved in local interactions in the isolated protein. Interestingly, these three lysine residues are

located next to the thrombin cleavage site (R_{168}-S_{169}) and are thus downstream of both the RGD motif and the potential interaction region with ICOSL, as suggested by the limited proteolysis experiments. When the cross-linked complex was reduced and alkylated, the peptide 272-299 was identified, suggesting that K_{299} may not be modified by DTSSP. Moreover, this peptide, which contains three other lysine residues (K_{290}, K_{292}, and K_{296}), was found to be doubly modified: upon both reduction and alkylation reactions. This finding suggests that two of these three lysine residues are labeled within the complex. The reduced and alkylated forms of this peptide might be generated either from inter- or from intra-molecular cross-links, not ruling out the possible involvement of this C- terminal region in ICOSL binding (Fig. 2.3.3, panel B).

Collectively, the limited proteolysis and cross-linking experiments suggested that ICOSL binds both the N-term and the C-term thrombin-generated fragments of OPN. Collectively these data suggest that, on ICOSL, the OPN binding site is located in the IgC domain (Fig. 2.3.4, panel A), whereas within OPN, two different binding sites for ICOSL exist. One is upstream of the RGD domain, and the other on the C-terminal portion of the protein (Fig. 2.3.4, panel B).

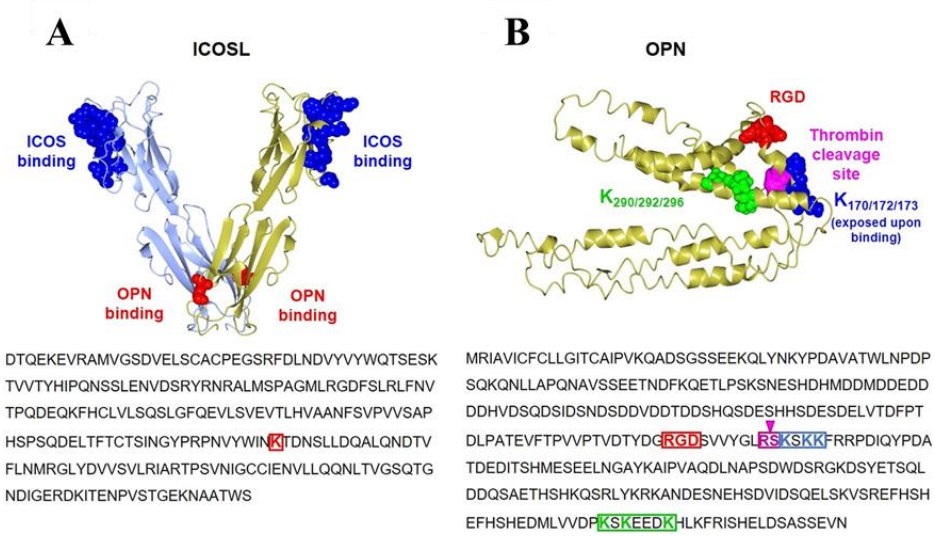

Fig. 2.3.4 Conformational analysis of the OPN/ICOSL complex. (A) Model for ICOSL created with the SWISS-MODEL server using 4f9p.1.A (CD277/ Butyrophilin-3) as a template. Residues buried at the interface with ICOS are blue-stained, K_{174} is red-stained. (B) OPN is an intrinsically disordered protein that simultaneously exhibits extended, random coil-like conformations and stable, cooperatively-folded conformations[15]. OPN model from Dianzani et al., 2017[20]. RGD in red, K_{170}, K_{172}, and K_{172} in blue, while K_{290}, K_{292}, and K_{296} are in green. The thrombin cleavage site (RS) is shown in pink.

The direct interaction between OPN and ICOSL was also validated by an ELISA-based assay using OPN as the capture protein and evaluating the binding of titrated amounts of soluble recombinant ICOSL-Fc. A commercial OPN produced in mammalian cells was used to ensure the presence of the physiological post-translational modifications. Moreover, the titration was carried out also in the presence of ICOS-Fc, thus verifying a possible competition of OPN and ICOS on the same interacting region on ICOSL. These experiments confirmed that ICOSL binds OPN in a concentration-dependent manner and that ICOSL binds OPN at different sites with respect to ICOS since they do not compete for binding (Fig. 2.3.5). This finding agrees with the results of chemical cross-linking experiments, which suggested that the OPN binding site on ICOSL involves a region around K_{174} located in the membrane-proximal IgC domain of ICOSL, whereas ICOS is

known to bind the membrane distal to the IgV domain[21].

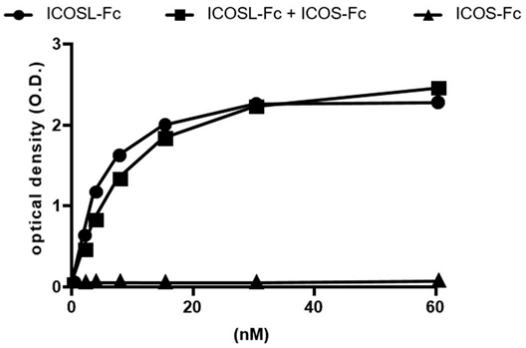

Fig. 2.3.5 ELISA-based interaction assay. Recombinant OPN was used as capture protein on 96 well plates and the binding of titrated amounts of soluble recombinant ICOSL-Fc was evaluated alone (black circle) or mixed 1:1; with ICOS-Fc (black square) or of ICOS-Fc alone (black triangle). Data are expressed as means ± standard error (n = 3 technical replicate).

The role of OPN post-translational modifications in OPN/ICOSL binding was also assessed by ELISA assay, using a recombinant OPN (OPN-GST) produced in *E.Coli*. As shown, with both preparations, the results were similar to those obtained using the commercial mammalian OPN-b and OPN-a (Fig. 2.3.6), indicating that ICOSL/OPN binding is not influenced or mediated by post-translational modifications.

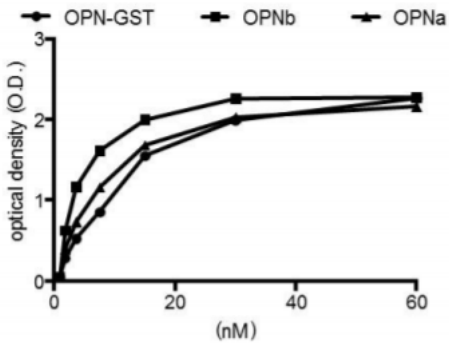

Fig. 2.3.6 ELISA-based interaction assay. The graph shows the interaction of titrated amounts of soluble ICOSL-Fc with a fixed amount of OPN-GST, OPN isoform-a, and OPN isoform-b coated on the plate. (black circle) shows OPN-GST, (black square) OPN-b, and (black triangle) OPN-a. Data are expressed as means ± standard error (n=3 technical replicate).

Since the putative interaction region, shown by the limited proteolysis experiments on OPN, is located upstream of the thrombin cleavage site ($R_{168}S_{169}$) and does not contain lysine residues, for the long stretch to $K_{77,}$ the cross-linking experiments were uninformative for that region. However, they suggested the possible existence of a second interaction site for ICOSL, comprising the OPN residues K_{290}, K_{292}, and K_{296}, located close to the OPN C-terminus. ELISA-based assays using recombinant truncated forms of OPN (Fig. 2.3.7), corresponding to N-term (aa 17-167) and C-term (aa 168-314) portions of OPN, confirmed that both retained the capability to bind ICOSL, although to lesser extents than the full-length molecule.

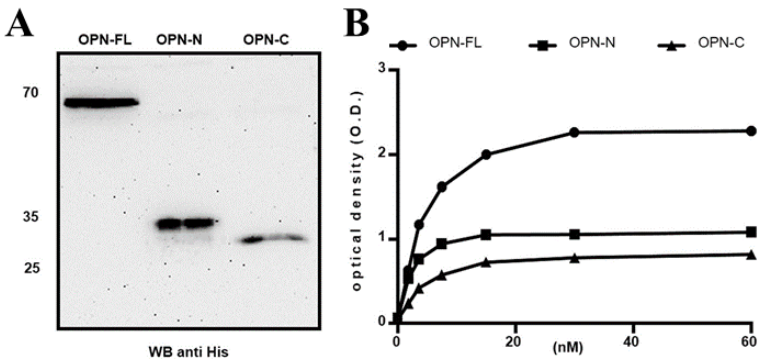

Fig. 2.3.7 ICOSL binds both the N-term and the C-term thrombin-generated fragments of OPN. (A) Western blot showing the OPN recombinant proteins (FL: full length, N and C-terminal fragments) probed with the anti-His-tag antibody. (B) ELISA-based interaction assay. The graph shows the interaction of titrated amounts of soluble ICOSL-Fc with a fixed amount of OPN-FL, OPN N- or C-fragments coated on the plate. (black circle) shows OPN full length, (black square) OPN-N, and (black triangle) OPN-C. Data are expressed as means ± standard error (n = 3 technical replicates).

In the light of these multiple interactions, it is worth noting that the cross-linking experiments suggest that, upon its binding to ICOSL, the C-terminal portion of

OPN, involving the residues K_{170}, K_{172}, and K_{173}, located close to the ICOSL binding site, undergoes a conformational change, which might influence the accessibility to integrin of the nearby RGD motif of OPN. The flexible and dynamic structure of OPN enables it to adopt different functional structures, i.e., folding upon binding, while permitting it to enter into multiple interactions with different binding partners[17]. It is, therefore, possible that OPN acts as a scaffold that can interact contemporarily with several different cell-surface receptors, and its interaction with ICOSL might modulate this activity. This would be in line with the hypothesis that integrins co-operate with other receptors (e.g., RTK, Tie2, and growth hormones receptors) on the cell surface[22]. However, the present findings indicate ICOSL to be an essential component of the signaling machinery recruited by soluble OPN, and it appears to be required for soluble OPN-induced motility even in the presence of integrins. This requirement might be due to the mechano-sensing nature of integrins, which require traction for full activation, making soluble molecules poor activators[23].

2.3.4 Conclusions

To delineate the ICOSL binding sites on OPN, we performed limited proteolysis and crosslinking experiments coupled to mass spectrometry analysis. The first approach mapped the ICOSL binding site in the central part of OPN, upstream of the RGD motif and the thrombin cleavage site. Furthermore, cross-linking experiments indicate that the region downstream the RGD domain (including K_{170}, $K_{172,}$ and K_{173}) becomes exposed upon ICOSL binding, which in turn indicates that, although OPN is known to behave as an IDP in solution, it may acquire a specific conformation in the complex with ICOSL. This conformational change might influence the accessibility of the nearby thrombin ($R_{168}S_{169}$) and ($S_{162}LAYGLR_{168}$) MMP3/7 cleavage sites. This may be biologically relevant, since the full-length OPN, and each fragment generated by its enzymatic processing in inflammatory conditions, have partly distinct activities. For

instance, full-length OPN induces both cell adhesion and migration, whereas the N-terminal and C-terminal fragments have selective effects on cell adhesion and cell migration, respectively. The OPN fragments generated by thrombin cleavage also increase hepatocellular carcinoma cell invasion[24].

The results discussed in this chapter (including figures and tables) have been published in the following article: Raineri, D.; Dianzani, C.; Cappellano, G.; Maione, F.; Baldanzi, G.; **Iacobucci, I.**; Clemente, N.; Baldone, G.; Boggio, E.; Gigliotti, C. L.; Boldorini, R.; Rojo, J. M.; Monti, M.; Birolo, L.; Dianzani, U.; Chiocchetti, A. Osteopontin Binds ICOSL Promoting Tumor Metastasis. *Commun Biol* **2020**, *3* (1), 615. https://doi.org/10.1038/s42003-020-01333-1.

2.3.5 References

(1) Buonfiglio, D.; Bragardo, M.; Bonissoni, S.; Redoglia, V.; Cauda, R.; Zupo, S.; Burgio, V. L.; Wolff, H.; Franssila, K.; Gaidano, G.; Carbone, A.; Janeway, C. A.; Dianzani, U. Characterization of a Novel Human Surface Molecule Selectively Expressed by Mature Thymocytes, Activated T Cells and Subsets of T Cell Lymphomas. *Eur J Immunol* **1999**, *29* (9), 2863–2874. https://doi.org/10.1002/(SICI)1521-4141(199909)29:09<2863::AID-IMMU2863>3.0.CO;2-W.

(2) Hutloff, A.; Dittrich, A. M.; Beier, K. C.; Eljaschewitsch, B.; Kraft, R.; Anagnostopoulos, I.; Kroczek, R. A. ICOS Is an Inducible T-Cell Co-Stimulator Structurally and Functionally Related to CD28. *Nature* **1999**, *397* (6716), 263–266. https://doi.org/10.1038/16717.

(3) Redoglia, V.; Dianzani, U.; Rojo, J. M.; Portolés, P.; Bragardo, M.; Wolff, H.; Buonfiglio, D.; Bonissoni, S.; Janeway, C. A. Characterization of H4: A Mouse T Lymphocyte Activation Molecule Functionally Associated with the CD3/T Cell Receptor. *Eur J Immunol* **1996**, *26* (11), 2781–2789. https://doi.org/10.1002/eji.1830261134.

(4) Swallow, M. M.; Wallin, J. J.; Sha, W. C. B7h, a Novel Costimulatory Homolog of B7.1 and B7.2, Is Induced by TNFalpha. *Immunity* **1999**, *11* (4), 423–432. https://doi.org/10.1016/s1074-7613(00)80117-x.

(5) Yoshinaga, S. K.; Whoriskey, J. S.; Khare, S. D.; Sarmiento, U.; Guo, J.; Horan, T.; Shih, G.; Zhang, M.; Coccia, M. A.; Kohno, T.; Tafuri-Bladt, A.; Brankow, D.; Campbell, P.; Chang, D.; Chiu, L.; Dai, T.; Duncan, G.; Elliott, G. S.; Hui, A.; McCabe, S. M.; Scully, S.; Shahinian, A.; Shaklee, C. L.; Van, G.; Mak, T. W.; Senaldi, G. T-Cell Co-Stimulation through B7RP-1 and ICOS. *Nature* **1999**, *402* (6763), 827–832. https://doi.org/10.1038/45582.

(6) Warnatz, K.; Bossaller, L.; Salzer, U.; Skrabl-Baumgartner, A.; Schwinger,

W.; van der Burg, M.; van Dongen, J. J. M.; Orlowska-Volk, M.; Knoth, R.; Durandy, A.; Draeger, R.; Schlesier, M.; Peter, H. H.; Grimbacher, B. Human ICOS Deficiency Abrogates the Germinal Center Reaction and Provides a Monogenic Model for Common Variable Immunodeficiency. *Blood* **2006**, *107* (8), 3045–3052. https://doi.org/10.1182/blood-2005-07-2955.

(7) Dianzani, C.; Minelli, R.; Gigliotti, C. L.; Occhipinti, S.; Giovarelli, M.; Conti, L.; Boggio, E.; Shivakumar, Y.; Baldanzi, G.; Malacarne, V.; Orilieri, E.; Cappellano, G.; Fantozzi, R.; Sblattero, D.; Yagi, J.; Rojo, J. M.; Chiocchetti, A.; Dianzani, U. B7h Triggering Inhibits the Migration of Tumor Cell Lines. *J.I.* **2014**, *192* (10), 4921–4931. https://doi.org/10.4049/jimmunol.1300587.

(8) Dianzani, C.; Minelli, R.; Mesturini, R.; Chiocchetti, A.; Barrera, G.; Boscolo, S.; Sarasso, C.; Gigliotti, C. L.; Sblattero, D.; Yagi, J.; Rojo, J. M.; Fantozzi, R.; Dianzani, U. B7h Triggering Inhibits Umbilical Vascular Endothelial Cell Adhesiveness to Tumor Cell Lines and Polymorphonuclear Cells. *J Immunol* **2010**, *185* (7), 3970–3979. https://doi.org/10.4049/jimmunol.0903269.

(9) Hedl, M.; Lahiri, A.; Ning, K.; Cho, J. H.; Abraham, C. Pattern Recognition Receptor Signaling in Human Dendritic Cells Is Enhanced by ICOS Ligand and Modulated by the Crohn's Disease ICOSLG Risk Allele. *Immunity* **2014**, *40* (5), 734–746. https://doi.org/10.1016/j.immuni.2014.04.011.

(10) Occhipinti, S.; Dianzani, C.; Chiocchetti, A.; Boggio, E.; Clemente, N.; Gigliotti, C. L.; Soluri, M. F.; Minelli, R.; Fantozzi, R.; Yagi, J.; Rojo, J. M.; Sblattero, D.; Giovarelli, M.; Dianzani, U. Triggering of B7h by the ICOS Modulates Maturation and Migration of Monocyte-Derived Dendritic Cells. *J Immunol* **2013**, *190* (3), 1125–1134. https://doi.org/10.4049/jimmunol.1201816.

(11) Tang, G.; Qin, Q.; Zhang, P.; Wang, G.; Liu, M.; Ding, Q.; Qin, Y.; Shen, Q. Reverse Signaling Using an Inducible Costimulator to Enhance Immunogenic Function of Dendritic Cells. *Cell Mol Life Sci* **2009**, *66* (18), 3067–3080. https://doi.org/10.1007/s00018-009-0090-7.

(12) Boggio, E.; Dianzani, C.; Gigliotti, C. L.; Soluri, M. F.; Clemente, N.; Cappellano, G.; Toth, E.; Raineri, D.; Ferrara, B.; Comi, C.; Dianzani, U.; Chiocchetti, A. Thrombin Cleavage of Osteopontin Modulates Its Activities in Human Cells In Vitro and Mouse Experimental Autoimmune Encephalomyelitis In Vivo. *J Immunol Res* **2016**, *2016*, 9345495. https://doi.org/10.1155/2016/9345495.

(13) Castello, L. M.; Raineri, D.; Salmi, L.; Clemente, N.; Vaschetto, R.; Quaglia, M.; Garzaro, M.; Gentilli, S.; Navalesi, P.; Cantaluppi, V.; Dianzani, U.; Aspesi, A.; Chiocchetti, A. Osteopontin at the Crossroads of Inflammation and Tumor Progression. *Mediators Inflamm* **2017**, *2017*, 4049098. https://doi.org/10.1155/2017/4049098.

(14) Clemente, N.; Raineri, D.; Cappellano, G.; Boggio, E.; Favero, F.; Soluri, M. F.; Dianzani, C.; Comi, C.; Dianzani, U.; Chiocchetti, A. Osteopontin Bridging Innate and Adaptive Immunity in Autoimmune Diseases. *J Immunol Res* **2016**, *2016*, 7675437. https://doi.org/10.1155/2016/7675437.

(15) Liu, Z.; Huang, Y. Advantages of Proteins Being Disordered. *Protein Sci* **2014**, *23* (5), 539–550. https://doi.org/10.1002/pro.2443.

(16) Gigliotti, C. L.; Boggio, E.; Clemente, N.; Shivakumar, Y.; Toth, E.; Sblattero, D.; D'Amelio, P.; Isaia, G. C.; Dianzani, C.; Yagi, J.; Rojo, J. M.; Chiocchetti, A.; Boldorini, R.; Bosetti, M.; Dianzani, U. ICOS-Ligand Triggering Impairs Osteoclast Differentiation and Function In Vitro and In Vivo. *J Immunol* **2016**, *197* (10), 3905–3916. https://doi.org/10.4049/jimmunol.1600424.

(17) Kurzbach, D.; Platzer, G.; Schwarz, T. C.; Henen, M. A.; Konrat, R.;

Hinderberger, D. Cooperative Unfolding of Compact Conformations of the Intrinsically Disordered Protein Osteopontin. *Biochemistry* **2013**, *52* (31), 5167–5175. https://doi.org/10.1021/bi400502c.

(18) O'Reilly, F. J.; Rappsilber, J. Cross-Linking Mass Spectrometry: Methods and Applications in Structural, Molecular and Systems Biology. *Nat Struct Mol Biol* **2018**, *25* (11), 1000–1008. https://doi.org/10.1038/s41594-018-0147-0.

(19) Scaloni, A.; Monti, M.; Acquaviva, R.; Tell, G.; Damante, G.; Formisano, S.; Pucci, P. Topology of the Thyroid Transcription Factor 1 Homeodomain-DNA Complex. *Biochemistry* **1999**, *38* (1), 64–72. https://doi.org/10.1021/bi981300k.

(20) Dianzani, C.; Bellavista, E.; Liepe, J.; Verderio, C.; Martucci, M.; Santoro, A.; Chiocchetti, A.; Gigliotti, C. L.; Boggio, E.; Ferrara, B.; Riganti, L.; Keller, C.; Janek, K.; Niewienda, A.; Fenoglio, C.; Sorosina, M.; Cantello, R.; Kloetzel, P. M.; Stumpf, M. P. H.; Paul, F.; Ruprecht, K.; Galimberti, D.; Martinelli Boneschi, F.; Comi, C.; Dianzani, U.; Mishto, M. Extracellular Proteasome-Osteopontin Circuit Regulates Cell Migration with Implications in Multiple Sclerosis. *Sci Rep* **2017**, *7*, 43718. https://doi.org/10.1038/srep43718.

(21) Levin, S. D.; Evans, L. S.; Bort, S.; Rickel, E.; Lewis, K. E.; Wu, R. P.; Hoover, J.; MacNeil, S.; La, D.; Wolfson, M. F.; Rixon, M. W.; Dillon, S. R.; Kornacker, M. G.; Swanson, R.; Peng, S. L. Novel Immunomodulatory Proteins Generated via Directed Evolution of Variant IgSF Domains. *Front Immunol* **2019**, *10*, 3086. https://doi.org/10.3389/fimmu.2019.03086.

(22) Streuli, C. H.; Akhtar, N. Signal Co-Operation between Integrins and Other Receptor Systems. *Biochem J* **2009**, *418* (3), 491–506. https://doi.org/10.1042/BJ20081948.

(23) Kechagia, J. Z.; Ivaska, J.; Roca-Cusachs, P. Integrins as Biomechanical

Sensors of the Microenvironment. *Nat Rev Mol Cell Biol* **2019**, *20* (8), 457–473. https://doi.org/10.1038/s41580-019-0134-2.

(24) Takafuji, V.; Forgues, M.; Unsworth, E.; Goldsmith, P.; Wang, X. W. An Osteopontin Fragment Is Essential for Tumor Cell Invasion in Hepatocellular Carcinoma. *Oncogene* **2007**, *26* (44), 6361–6371. https://doi.org/10.1038/sj.onc.1210463.

Chapter 3 – Investigation of molecular mechanisms in host-virus interactions

3.1 Long-chain inorganic polyphosphates bind ACE2 receptor impairing SARS-CoV-2 infection and replication

3.1.1 Introduction and preliminary results

In December 2019, a novel coronavirus was discovered because of several cases of pneumonia. It was denominated "Severe acute respiratory syndrome by Coronavirus 2" (SARS-CoV-2), or Covid 19. The emergency started in China, in Wuhan, a city in Hubei province, but it rapidly wafted globally, causing a pandemic.

SARS-CoV-2 belongs to the *Coronaviridae* family, which includes 4 genera known as α-/β-/γ-/δ-CoV. Among them, α- and β-CoV can infect mammals, while γ- and δ-CoV primarily infect birds[1]. The SARS-CoV-2 virus belongs to the β-CoV[2] and is transmitted via respiratory droplets, with possible fecal-oral transmission[3]. All coronaviruses contain a positive-sense, single-stranded RNA that encodes for many proteins: non-structural RNA-dependent RNA polymerase (RdRp), and some structural proteins, namely Nucleocapsid protein (N), Matrix protein (M), Envelope protein, and Spike- glycoprotein (S)[4] (Fig. 3.1.1). The latter is located on the spherical surface of the virus, protruding from the viral membrane and giving the peculiar "halo" appearance, and it is responsible for the entering of the virus in human cells since it binds the ACE2 receptor (Angiotensin-Converting-Enzyme 2)[5].

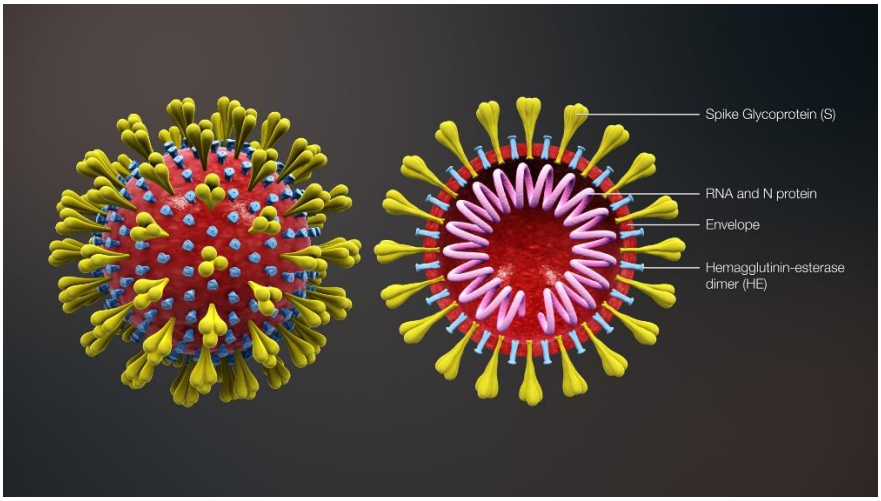

Fig. 3.1.1 External and internal structure of the new SARS-CoV-2 (credit: https://www.scientificanimations.com/wiki-images).

The S-glycoprotein is an integral membrane protein (known as Spike) consisting of an extracellular N-terminal and a C-terminal intracellular domain and transmembrane segment, showing a molecular about 200 kDa[6]. The Spike protein enables the viral entrance in the host cell, which is also possible thanks to the polysaccharide coat (common in Spike proteins) that camouflages them to avoid the immune system[7]. It consists of 1273 total amino acids, a signal peptide, the S1 domain, which binds the cell receptor, and the S2 domain consisting of heptapeptide repeat sequence 1 (HR1) and heptapeptide repeat sequence 2 (HR2) that plays an essential role in the fusion of the membranes[8] (Fig. 3.1.2, panel a). The full Spike is structured in a tri-domains protein, with a characteristic bulbous form, which gives the virus the peculiar "halo" structure. It exists as an inactive form that activates thanks to hydrolysis using specific host cell proteases, which cleave the S protein into S1 and S2 domains[9,10] (Fig. 3.1.2, panel b).

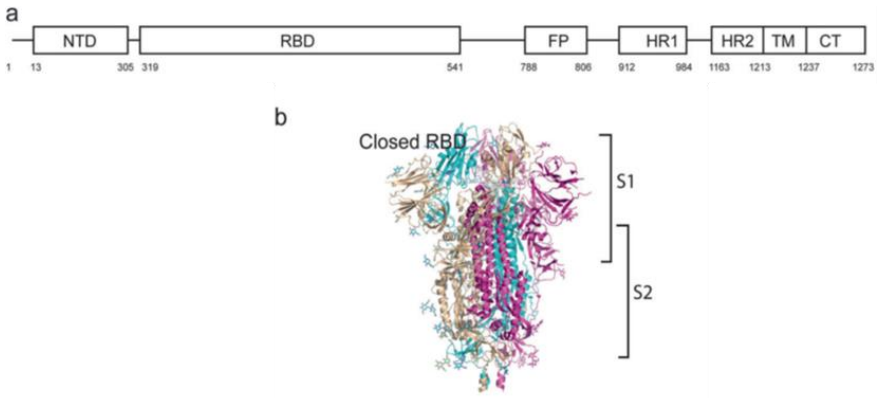

Fig. 3.1.2 a) Schematic representation of the SARS-CoV-2 spike. b-c) The S protein RBD closed and opened status[9].

Long-chain inorganic polyphosphates (PolyPs) comprise chains of a few to many hundreds of inorganic phosphates (Pi). They are ubiquitous, with prevalence in peripheral blood mononuclear cells and erythrocytes[11] and in the subcellular locations, namely nucleus, cytoplasm, plasma membranes, and mitochondria[12] PolyPs are involved in blood coagulation[13] and inhibition of the complement pathway[14]. They are also involved in the chelation of calcium for bone mineralization[15] and activation of apoptosis[11], and they act as chaperone-like[16] and neuronal excitability molecules[17,18]. For their large diffusion within human cellular and extracellular districts, PolyPs are not toxic and do not trigger immune responses. Moreover, linear PolyPs have been reported to have cytoprotective and antiviral activities against human immunodeficiency virus type 1 (HIV-1) infection *in vitro*[19]. Furthermore, and of particular relevance here, PolyPs (as Pi ×40) were recently shown to significantly inhibit the interaction of the receptor-binding domain of the SARS-CoV-2 S protein with ACE2 *in vitro*[20]. All these features make PolyPs molecules of inestimable potential.

In this project carried out in collaboration with Professor Massimo Zollo of the Department of Molecular Medicine and Medical Biotechnologies of the University of Naples “Federico II”, we show that long-chain inorganic PolyPs

have antiviral activities against SARS-CoV-2 in different cellular models (i.e., Vero E6).

Antiviral activities of linear PolyPs with average chain lengths of 15, 34, and 91 Pi residues have been shown previously for HIV-infected cells[19]. To measure the antiviral effects of PolyPs in SARS-CoV-2 infection, Prof. Zollo and collaborators used viral particles obtained from a Korean patient who was positive for COVID-19 (BetaCov/Korea/KCDC03/2020; registered ID: EPI_ISL_407193) for infecting Vero E6 cells. At 24 h from infection, the cells were treated with increasing concentrations of PolyPs (9.37-300 µM) of different chain lengths (Pi ×8, ×16, ×64, ×94, ×120). After 24 h of PolyP treatment (i.e., 48 h after infection), the viral RNA in the cell culture supernatants was quantified for RdRp expression using quantitative real-time polymerase chain reaction (qRT-PCR). These data showed that PolyP ×120 (PolyP120) significantly decreased values of RdRp at all concentrations tested. None of the other PolyPs tested here had significant effects at doses <300 µM, except PolyP94, which also significantly decreased the RdRp values at 150 µM and 300 µM (Fig. 3.1.3).

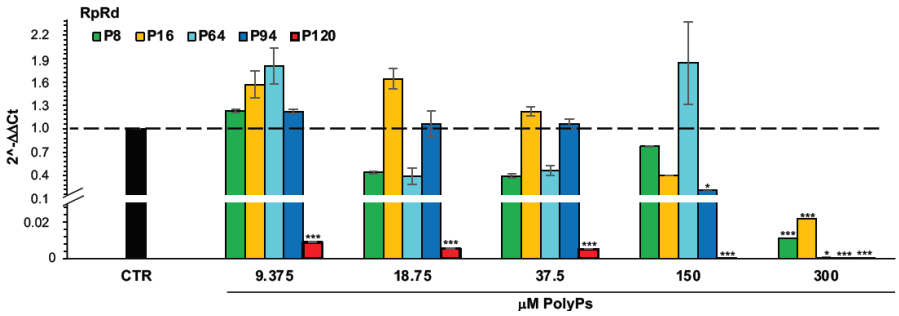

Fig 3.1.3 Viral RdRp expression from real-time RT-PCR performed on viral RNA extracted from cell-culture supernatants from Vero E6 cells (4×10^5) infected with SARS-CoV-2 for 24 h, and then treated with increasing concentrations of PolyPs (9.375, 18.75, 37.5, 150, 300 µM) of the different chain lengths (as indicated: P8, PolyP8; P16, PolyP16; P64, PolyP64; P94, PolyP94; P120, PolyP120) for additional 24 h. Of note, ΔCt was calculated as the difference between the cycle threshold (Ct) for RdRp expression in SARS-CoV-2–infected cells treated with PolyPs and the Ct for RdRp expression in SARS-

CoV-2–infected cells without PolyPs. Data are means ±standard deviation. *, P <0.05; ***P<0.001 (unpaired two-tailed Student's t-tests; RdRp vs. untreated Vero E6 cells [black column]; n = 3 independent experiments per group).

These results prompted us to study in deep the antiviral effects of P120 and its target on host cells.

In order to understand how PolyPs exert this role, multidisciplinary approaches were employed. Among all, the interaction between ACE2 and a specific PolyP (P120) was investigated through size exclusion chromatography (SEC) and limited proteolysis-mass spectrometry approaches.

3.1.2 Experimental methods

3.1.2.1 Size Exclusion Chromatography (SEC) experiments

The SEC runs were performed injecting 2.5µg of the sample onto a Superdex 200 Increase 3.2/300. The proteins were eluted with a Phosphate buffer (pH=7.4) in isocratic conditions at a flow rate of 50 µL/min by monitoring the peptide bond's absorbance wavelength (215nm). To highlight the effects of P120 on ACE2 and S1, they were individually incubated with a 13 folds molar excess of P120 overnight, and then the chromatographic experiment was done. To assess if P120 could affect the S1-ACE2 complex, the addition of the PolyP was done on the preformed protein complex. In particular, ACE2 was first incubated with S1 for 1h at room temperature, performed SEC on half of the sample, and then P120 was added to the mixture for 2h.

3.1.2.2 Limited proteolysis-mass spectrometry experiment

The limited proteolysis experiment was carried out onto a commercial human recombinant Fc tagged ACE2 expressed in a human cell line. ACE2-Fc was pre-incubated either with P120 and P8 for 2 hours at room temperature in a 10 times molar excess of polyPs. The limited proteolysis experiment was executed onto the three samples constituted by ACE2-Fc, ACE2-Fc/P120, and ACE2-Fc/P8 using

chymotrypsin with an enzyme:substrate (E:S) ratio of 1:50 w/w. The digestion reactions were run for 30 minutes, 1 hour, and 2h. The digestion products were resolved onto a 10% SDS-PAGE gel, as well as an aliquot of 4µg of intact ACE2-Fc used for protein apparent molecular weight reference. The differential bands displayed in the three samples were cut and *in situ* hydrolyzed with trypsin. The peptide mixtures were analyzed by nanoLC-MS/MS using a Proxeon-nanoEasy II-LTQ Orbitrap XL. The peptide fractionation was performed onto a C18 capillary reverse-phase column (200 mm, 75 µm, 5 µm) working at 250 nl/min flow rate, using a step gradient of eluent B (0.2% formic acid, 95% acetonitrile LC-MS Grade) from 10 to 60% over 69 min and 60 to 95% over 3 min. The mass spectrometric data were collected in data-dependent acquisition (DDA) mode by fragmenting the most 10 intense ions with a dynamic exclusion of 40 seconds. The peptide identification was performed using MASCOT software searching in a database containing only the sequence of ACE2-Fc using the search parameters: 10 ppm as peptides mass tolerance for MS and 0.6 Da for MS/MS search; carbamidomethyl (Cys) as fixed modification and Gln->pyro-Glu (N-term Gln), Oxidation (Met), Pyro-carbamidomethyl (N-term Cys) as variable modifications.

3.1.3 Results and Discussion

3.1.3.1 Interactions of PolyP120 with ACE2 by size-exclusion chromatography and limited proteolysis

To further evaluate the binding of PolyP120, hereafter P120, and a His-tagged form of ACE2, we estimated the retention times shift of the isolated and PolyP incubated protein by a gel filtration experiment. The PolyP was incubated in a 13 folds molar excess overnight at room temperature. The SEC experiment exhibited an earlier retention time (RT) of the ACE2 protein incubated with P120 (Fig. 3.1.4, panel B) compared with the isolated protein (Fig. 3.1.4, panel A).

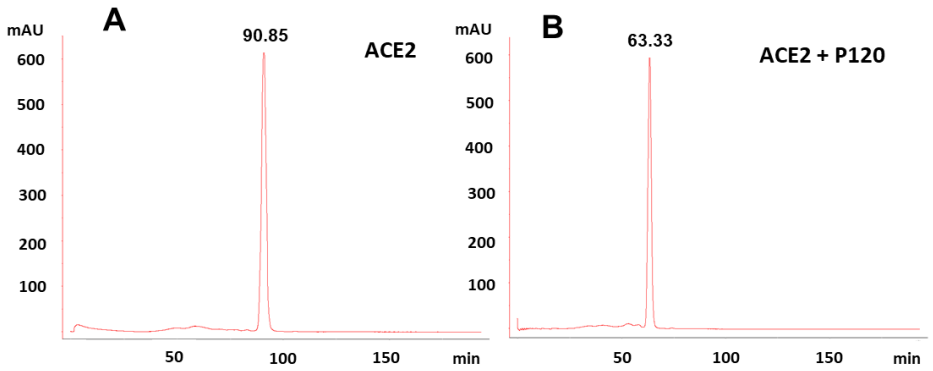

Fig. 3.1.4 Size exclusion chromatogram of (A) ACE2 and (B) ACE2 with P120. The absorbance wavelength was set at 215 nm.

The interaction of PolyPs and ACE2 was also confirmed by our collaborators by enzyme-linked immunosorbent assays (ELISA). In this assay, PolyPs of different chain lengths, as PolyP8 (P8), PolyP34 (P34), PolyP120 (P120), PolyP137 (P137), and PolyP189 (P189) (all at 416 nM coating concentration). The interaction with ACE2increases with PolyPs molecular weight (Fig. 3.1.5). In particular, PolyP120 showed the greatest binding here ($p = 9.4E^{-05}$; unpaired two-tailed t-test, adjusted with the Bonferroni method).

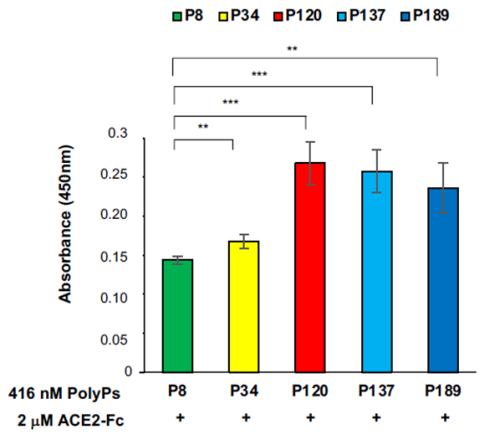

Fig. 3.1.5 ELISA assays using 2 μM human ACE2-Fc chimera protein and 416 nM PolyPs

at different chain lengths (as indicated: P8, P34, P120, P137, P189) coated on 96-well plates. Absorbance at 450 nm was measured after 1 h at 25 °C. Comparison among the different PolyPs for the binding to ACE2-Fc was evaluated comparing them to PolyP8 (i.e., 0.1 µM). Data are means ±standard deviation. **, P <0.01; ***, P <0.001 (unpaired two-tailed Student's t-tests; n = 3 independent experiments per group).

The stoichiometry of the binding is not readily calculable because of the chemical nature of the PolyP, which could alter the hydrodynamic properties of the protein upon binding.

Then, the interaction between viral spike S1 and P120 was also tested in the same experimental conditions used for ACE2. Two chromatographic peaks at comparable retention times (Fig. 3.1.6, panel A and B) were recorded for the isolated and P120-incubated protein, thus suggesting the absence of direct interaction between the viral Spike S1 protein and P120.

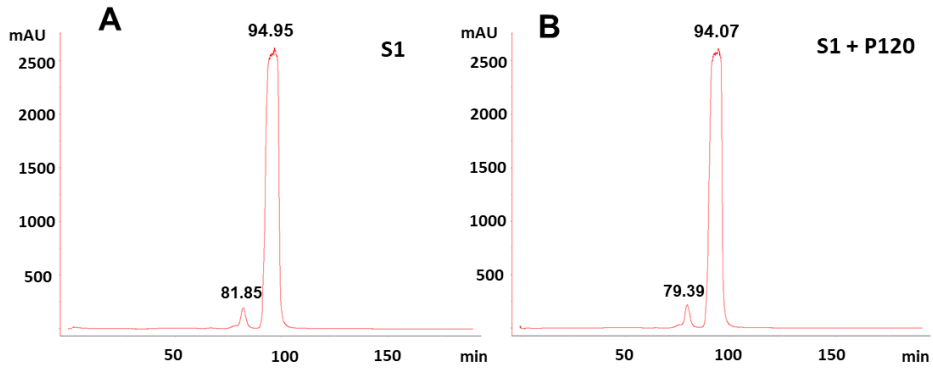

Fig. 3.1.6 Size exclusion chromatogram of (A) Spike S1 and (B) Spike S1 with P120. The absorbance wavelength was set at 215 nm.

P120 was then added to the ACE2-S1 complex, which was previously analyzed and showed an RT of 79.63min (Fig. 3.1.7, panel A).

The retention time shift at earlier minutes (Fig. 3.1.7, panel B) showed by the unique chromatographic peak suggests that P120 can still bind ACE2 even if it is involved in the complex with S1, probably occupying a different binding site respect S1 on the receptor.

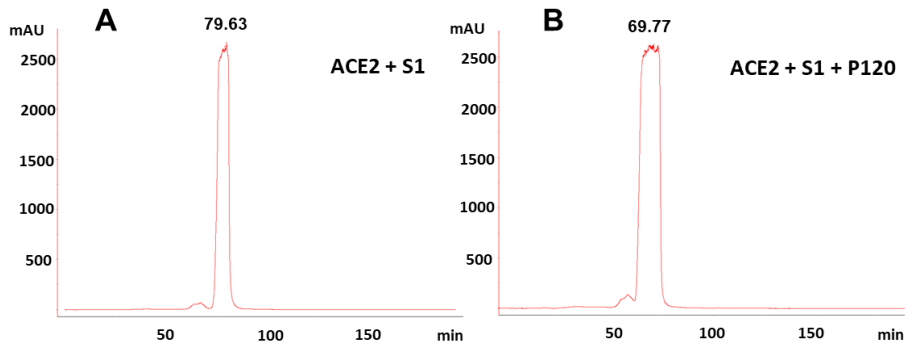

Fig. 3.1.7 Size exclusion chromatogram of (A) Spike S1-ACE2 and (B) Spike S1-ACE2 with P120. The absorbance wavelength was set at 215 nm.

To map more in detail the interaction regions of ACE2 and P120, limited proteolysis coupled to mass spectrometry experiment was performed. The proteolytic probe used was chymotrypsin, which preferentially cleaves peptide bonds at the C-term of aromatic amino acid residues. The isolated recombinant ACE2-Fc was digested in parallel with protein incubated with PolyP120 or PolyP8 (P8). The latter was used as a control because it does not bind ACE2 (Fig. 3.1.5). The reactions were monitored over time for 30, 60, and 120 minutes. As shown in Fig. 3.1.8, the proteolytic profiles of isolated ACE2 and its complex with PolyP8 were identical, confirming that P8 does not influence ACE2 conformation.

In contrast, these proteolytic profiles were different from those for the complex ACE2-Fc–P120, indicating such an interaction, with the main difference being a band with electrophoretic mobility of about 50 kDa (Fig. 3.1.8, panel A, highlighted with a red arrow). The band was excised, *in situ* hydrolyzed, and the peptide mixture was analyzed by nano liquid chromatography-tandem mass spectrometry. The mass spectrometry data allowed the identification of peptides belonging to the C-term moiety of the ACE2 amino acidic sequence (Fig. 3.1.8, panel B and Table 3.1.1). The mass mapping data and the band's apparent molecular weight analyzed suggested that, both in the isolated and P8 treated ACE2, the chymotrypsin cleavage site is located upstream of R514, within

fragment 510-518, which contains several aromatic amino-acid residues. The band disappears in the P120 pre-treated sample, indicating that the cleavage site is hidden in the complex of ACE2-Fc–PolyP120. The SEC and limited proteolysis findings are in accordance with Lan and colleagues[21], which structurally characterized the Spike S1-ACE2 binding region mainly at the N-term of ACE2, while our work suggested an ACE2 pocket located at the C-term of the protein able to bind P120. However, further experiments have to be carried out with other enzymatic probes to define the ACE2-P120 interacting region/s better.

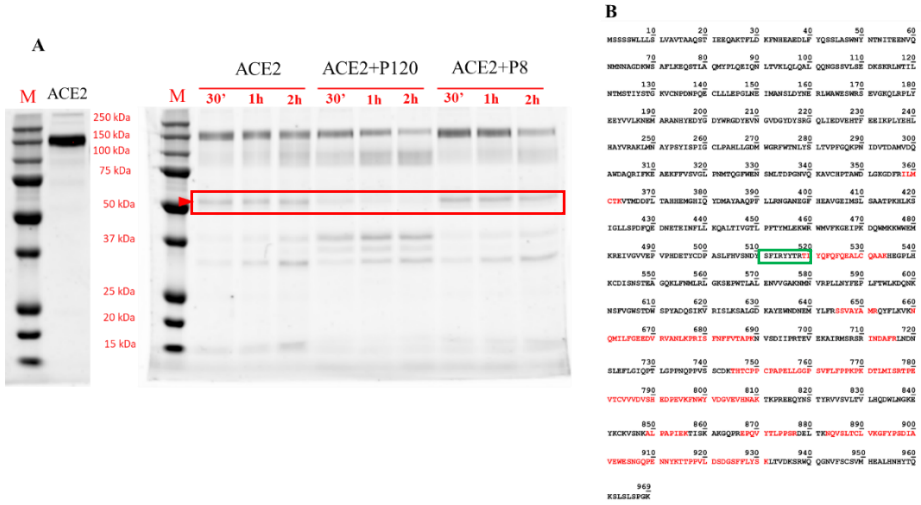

Fig. 3.1.8 (A) SDS-PAGE of intact ACE2-Fc and chymotrypsin limited proteolysis profiles of ACE2-Fc, ACE2-Fc pre-incubated with either P120 and P8. The proteolytic reaction mixtures after 30 minutes, 1 hour, and 2 hours are displayed. In the red box are indicated the discriminant bands that were analyzed by mass mapping (B) Sequence coverage obtained by tryptic mass mapping of the ACE2 and ACE2+P8 slices indicated within the red box. Peptides identified by LC-MS/MS are highlighted in red. The green box indicates the putative position of the chymotrypsin cleavage site.

Peptide	Observed m/z	Theoretical MW	Calculated MW	Mass error (ppm)	Ion score	Aminoacidic peptide sequence
358-363	383.2061	764.3976	764.3925	6.79	28	ILMCTK
519-534	973.4823	1944.95	1944.9404	4.98	37	TLYQFQFQEALCQAAK
645-652	442.7219	883.4292	883.4222	8.02	71	SSVAYAMR
660-671	725.8596	1449.705	1449.6922	8.59	69	NQMILFGEEDVR
672-678	399.2562	796.4978	796.4919	7.47	38	VANLKPR
679-689	635.8442	1269.674	1269.6758	-1.51	23	ISFNFFVTAPK
711-716	368.1963	734.378	734.3711	9.43	45	INDAFR
745-770	711.8764	2843.477	2843.4503	9.22	39	THTCPPCPAPELLGGPSVFLFPPKPK
771-777	418.2242	834.4338	834.4269	8.3	44	DTLMISR
778-796	1070.028	2138.041	2138.0202	9.94	78	TPEVTCVVVDVSHEDPEVK
797-810	839.4105	1676.806	1676.7947	7	57	FNWYVDGVEVHNAK
849-856	419.7588	837.503	837.496	8.45	26	ALPAPIEK
867-877	643.8464	1285.678	1285.6666	9.03	50	EPQVYTLPPSR
883-892	581.3228	1160.631	1160.6223	7.5	54	NQVSLTCLVK
893-914	1272.5791	2543.144	2543.1241	7.7	51	GFYPSDIAVEWESNGQPENNYK
915-931	937.4736	1872.933	1872.9146	9.65	84	TTPPVLDSDGSFFLYSK

Table 3.1.1 List of peptides identified by a bottom-up proteomic approach based on LC-MS/MS analysis of the 50kDa gel band released from ACE2-Fc following 2 hours of chymotrypsin limited proteolysis experiment. For each peptide, the detected m/z, the theoretical and calculated molecular weight, the mass error, the MASCOT ion score, and the amino acid sequence are reported.

The binding region suggested by limited proteomic results was very close to the R514 residue, which is one of the residues included in the binding pocket described by the docking calculations. The latter, indeed, described a potential binding pocket for shorter PolyPs (e.g., PolyP20) in which the binding was mainly mediated by four ACE2 amino-acid residues: Arg514, His401, His378, Arg393 (Fig. 3.1.9).

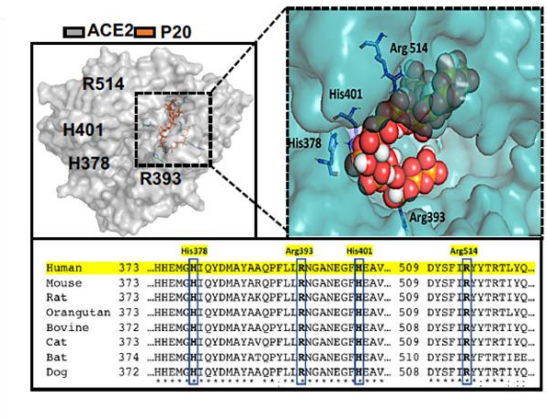

Fig. 3.1.9 Molecular docking of PolyP20 on the SARS-CoV-2 ACE2 domain (PDB structure: 6M0J chain A). Left: ACE2 is depicted transparent molecular surface colored according to electrostatic potential. PolyP20: orange sticks. Right: Expanded view of the ACE2 receptor binding domain as a cyan transparent surface, to indicate the binding interface. Bottom: Alignment analysis of ACE2 protein regions with the potential binding sites for PolyP20. The amino-acid residues mainly responsible for the interactions between ACE2 and PolyP20 are shown as blue boxes (His378, Arg393, His401, Arg514).

3.1.4 Conclusions

In this work, we have shown that long-chain PolyPs (i.e., PolyP120, with 120 Pi residues) can inhibit the expression of SARS-CoV-2 viral genes (genomic and sub-genomic transcripts), and thus impair the replication of the virus in the host cells targeting ACE2. The long-chain PolyP120 has been demonstrated to directly bind ACE2 receptor imposing a conformational change differently to the short polyP8 with SEC and limited proteolysis experiments. Furthermore, the SEC data suggests that P120 and Spike S1 have different binding sites on ACE2. Although these need to be verified *in cells*, the data presented here highlight the potential importance of using PolyPs *in vivo* to limit virus entrance into human host cells. Of note, in directing any treatment at ACE2 as the receptor for SARS-CoV-2 entry into cells, the use of ACE inhibitors has been discouraged for patients with COVID-19. This is because of the consequent increased levels of ACE2 mRNA,

which might facilitate engagement and entry of SARS-CoV-2 into cells[22]. However, the future use of PolyPs might overcome this side effect of using ACE inhibitors. Thus, the combination of ACE inhibitors with PolyPs can be postulated to treat patients with COVID-19. These activities might be important for preventing infections and progression of SARS-CoV-2 and other viruses, underlining the importance of PolyPs due to their powerful antiviral functions.

The results discussed in this chapter (including figures and tables) have been reported in the following article under revision: Ferrucci V., Kong D.Y, Asadzadeh F., Marrone L., Boccia A., Siciliano R., Anastasio C., Quarantelli F., Comegna M., Pisano I., Passariello M, **Iacobucci I**., Della Monica R., Izzo B., Cerino P., Fusco G., Viscardi M., Brandi S., Pierri B.M, Tiberio C, Atripaldi L., Bianchi M., Paolella G., Capoluongo E., Castaldo G., Chiariotti L., Monti M., De Lorenzo C, Yun K.S., Pascarella S., Cheong J.H, Kim H.Y and Zollo M. Long-chain polyphosphates impair SARS-CoV-2 infection and replication. *Science Signaling*. ***Under Revision***

3.1.5 References

(1) Cirrincione, L.; Plescia, F.; Ledda, C.; Rapisarda, V.; Martorana, D.; Moldovan, R. E.; Theodoridou, K.; Cannizzaro, E. COVID-19 Pandemic: Prevention and Protection Measures to Be Adopted at the Workplace. *Sustainability* **2020**, *12* (9), 3603. https://doi.org/10.3390/su12093603.

(2) Andersen, K. G.; Rambaut, A.; Lipkin, W. I.; Holmes, E. C.; Garry, R. F. The Proximal Origin of SARS-CoV-2. *Nat Med* **2020**, *26* (4), 450–452. https://doi.org/10.1038/s41591-020-0820-9.

(3) Guan, W.; Ni, Z.; Hu, Y.; Liang, W.; Ou, C.; He, J.; Liu, L.; Shan, H.; Lei, C.; Hui, D. S. C.; Du, B.; Li, L.; Zeng, G.; Yuen, K.-Y.; Chen, R.; Tang, C.; Wang, T.; Chen, P.; Xiang, J.; Li, S.; Wang, J.; Liang, Z.; Peng, Y.; Wei, L.; Liu, Y.; Hu, Y.; Peng, P.; Wang, J.; Liu, J.; Chen, Z.; Li, G.; Zheng, Z.; Qiu, S.; Luo, J.; Ye, C.; Zhu, S.; Zhong, N. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* **2020**, *382* (18), 1708–1720. https://doi.org/10.1056/NEJMoa2002032.

(4) Kim, D.; Lee, J.-Y.; Yang, J.-S.; Kim, J. W.; Kim, V. N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181* (4), 914-921.e10. https://doi.org/10.1016/j.cell.2020.04.011.

(5) Ludwig, S.; Zarbock, A. Coronaviruses and SARS-CoV-2: A Brief Overview. *Anesth Analg* **2020**, *131* (1), 93–96. https://doi.org/10.1213/ANE.0000000000004845.

(6) Bosch, B. J.; van der Zee, R.; de Haan, C. A. M.; Rottier, P. J. M. The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. *J Virol* **2003**, *77* (16), 8801–8811. https://doi.org/10.1128/jvi.77.16.8801-8811.2003.

(7) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* **2020**, *369* (6501), 330–333. https://doi.org/10.1126/science.abb9983.

(8) Xia, S.; Zhu, Y.; Liu, M.; Lan, Q.; Xu, W.; Wu, Y.; Ying, T.; Liu, S.; Shi, Z.; Jiang, S.; Lu, L. Fusion Mechanism of 2019-NCoV and Fusion Inhibitors Targeting HR1 Domain in Spike Protein. *Cell Mol Immunol* **2020**, *17* (7), 765–767. https://doi.org/10.1038/s41423-020-0374-2.

(9) Huang, Y.; Yang, C.; Xu, X.-F.; Xu, W.; Liu, S.-W. Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacol Sin* **2020**, *41* (9), 1141–1149. https://doi.org/10.1038/s41401-020-0485-4.

(10) Hörnich, B. F.; Großkopf, A. K.; Schlagowski, S.; Tenbusch, M.; Kleine-Weber, H.; Neipel, F.; Stahl-Hennig, C.; Hahn, A. S. SARS-CoV-2 and SARS-CoV Spike-Mediated Cell-Cell Fusion Differ in the Requirements for Receptor Expression and Proteolytic Activation. *J Virol* **2021**. https://doi.org/10.1128/JVI.00002-21.

(11) Abramov, A. Y.; Fraley, C.; Diao, C. T.; Winkfein, R.; Colicos, M. A.; Duchen, M. R.; French, R. J.; Pavlov, E. Targeted Polyphosphatase Expression Alters Mitochondrial Metabolism and Inhibits Calcium-Dependent Cell Death. *Proceedings of the National Academy of Sciences* **2007**, *104* (46), 18091–18096. https://doi.org/10.1073/pnas.0708959104.

(12) Angelova, P. R.; Agrawalla, B. K.; Elustondo, P. A.; Gordon, J.; Shiba, T.; Abramov, A. Y.; Chang, Y.-T.; Pavlov, E. V. In Situ Investigation of Mammalian Inorganic Polyphosphate Localization Using Novel Selective Fluorescent Probes JC-D7 and JC-D8. *ACS Chem Biol* **2014**, *9* (9), 2101–2110. https://doi.org/10.1021/cb5000696.

(13) Müller, F.; Mutch, N. J.; Schenk, W. A.; Smith, S. A.; Esterl, L.; Spronk, H. M.; Schmidbauer, S.; Gahl, W. A.; Morrissey, J. H.; Renné, T. Platelet Polyphosphates Are Proinflammatory and Procoagulant Mediators in Vivo. *Cell* **2009**, *139* (6), 1143–1156. https://doi.org/10.1016/j.cell.2009.11.001.

(14) Wat, J. M.; Foley, J. H.; Krisinger, M. J.; Ocariza, L. M.; Lei, V.; Wasney,

G. A.; Lameignere, E.; Strynadka, N. C.; Smith, S. A.; Morrissey, J. H.; Conway, E. M. Polyphosphate Suppresses Complement via the Terminal Pathway. *Blood* **2014**, *123* (5), 768–776. https://doi.org/10.1182/blood-2013-07-515726.

(15) Hoac, B.; Kiffer-Moreira, T.; Millán, J. L.; McKee, M. D. Polyphosphates Inhibit Extracellular Matrix Mineralization in MC3T3-E1 Osteoblast Cultures. *Bone* **2013**, *53* (2), 478–486. https://doi.org/10.1016/j.bone.2013.01.020.

(16) Xie, L.; Jakob, U. Inorganic Polyphosphate, a Multifunctional Polyanionic Protein Scaffold. *J Biol Chem* **2019**, *294* (6), 2180–2190. https://doi.org/10.1074/jbc.REV118.002808.

(17) Stotz, S. C.; Scott, L. O.; Drummond-Main, C.; Avchalumov, Y.; Girotto, F.; Davidsen, J.; Gómez-Gárcia, M. R.; Rho, J. M.; Pavlov, E. V.; Colicos, M. A. Inorganic Polyphosphate Regulates Neuronal Excitability through Modulation of Voltage-Gated Channels. *Mol Brain* **2014**, *7* (1), 42. https://doi.org/10.1186/1756-6606-7-42.

(18) Holmström, K. M.; Marina, N.; Baev, A. Y.; Wood, N. W.; Gourine, A. V.; Abramov, A. Y. Signalling Properties of Inorganic Polyphosphate in the Mammalian Brain. *Nat Commun* **2013**, *4* (1), 1362. https://doi.org/10.1038/ncomms2364.

(19) Lorenz, B.; Leuck, J.; Köhl, D.; Muller, W. E.; Schröder, H. C. Anti-HIV-1 Activity of Inorganic Polyphosphates. *J Acquir Immune Defic Syndr Hum Retrovirol* **1997**, *14* (2), 110–118. https://doi.org/10.1097/00042560-199702010-00003.

(20) Neufurth, M.; Wang, X.; Tolba, E.; Lieberwirth, I.; Wang, S.; Schröder, H. C.; Müller, W. E. G. The Inorganic Polymer, Polyphosphate, Blocks Binding of SARS-CoV-2 Spike Protein to ACE2 Receptor at Physiological Concentrations. *Biochem Pharmacol* **2020**, *182*, 114215.

https://doi.org/10.1016/j.bcp.2020.114215.

(21) Lan, J.; Ge, J.; Yu, J.; Shan, S.; Zhou, H.; Fan, S.; Zhang, Q.; Shi, X.; Wang, Q.; Zhang, L.; Wang, X. Structure of the SARS-CoV-2 Spike Receptor-Binding Domain Bound to the ACE2 Receptor. *Nature* **2020**, *581* (7807), 215–220. https://doi.org/10.1038/s41586-020-2180-5.

(22) Kai, H.; Kai, M. Interactions of Coronaviruses with ACE2, Angiotensin II, and RAS Inhibitors—Lessons from Available Evidence and Insights into COVID-19. *Hypertens Res* **2020**, *43* (7), 648–654. https://doi.org/10.1038/s41440-020-0455-8.

3.2 Study of the Spike S1 “receptorome” in human cells

3.2.1 Introduction

The pandemic of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that emerged in the last part of 2019 in China is still posing significant threats worldwide. Although vaccines and drugs are in development, effective therapeutic and preventive approaches are not still available. Furthermore, in recent times many SARS-CoV-2 variants[1] have been isolated that complicate the scenario. The most significant part of the mutations is borne by the viral glycoprotein spike (S). Indeed, it is the crucial viral protein responsible for the virus attachment and fusion to the host cell receptor angiotensin-converting enzyme 2 (ACE2)[2]. In particular, the first step of recognition has been reported to be played by the S1 subunit of the glycoprotein, while the S2 and S2’ take place in the membrane fusion and viral entry. The subunits are formed and activated by human proteases (e.g., furin)[3]. Since ACE2 shows a low expression level in the lung and the upper respiratory tract[4,5], alternative receptors could be responsible for the viral recognition on the host cells. Furthermore, Yang and colleagues investigated the binding of the S1 subunit to ACE2 on living cells by force-distance (FD) curve-based atomic force microscopy (FD-curve-based AFM), founding that S1 was able to interact even on control cells with a frequency of ≈10%, although the expression levels of ACE2 were low[6].

S1 residue contains the receptor-binding domain (RBD), which binds the ACE2 receptor on the host cell through the opening of the RBD domain (Fig. 3.2.1), and the C-terminal and N-terminal domains (CTD and NTD).

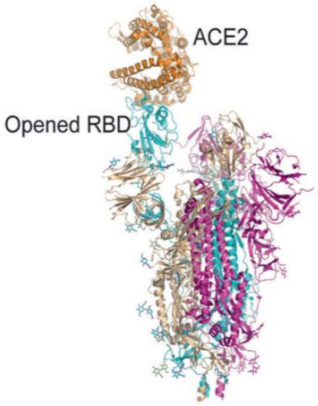

Fig 3.2.1 The S protein binds to ACE2 with opened RBD in the S1 subunit[7].

S2 subunit is responsible for membrane fusion and viral entry, and it is made of a fusion peptide (FP), a heptapeptide repeating sequence 1 and 2 (HR1, HR2), TM domain, and cytoplasmic domain fusion (CT). FP portion is made of 15-20 hydrophobic amino acids (Glycine, Alanine) and plays an important role in the membrane disruption and fusion with the membrane of the host cell[8]. HR1 and HR2 are composed of a repetitive heptapeptide and form a six-helical bundle, essential for the membrane fusion and entry function of the S2 subunit[9]. The first is located on the C-terminal of the FP portion, while the second is attached to the TM domain, which is essential to anchor the S protein to the viral membrane, where the S2 subunit ends in a CT tail[10].

Although ACE2 is the primary Spike S1 host receptor currently studied, it shows a low expression level in the lung and the upper respiratory tract[4]. This result indicates that SARS-CoV-2 may have alternative receptors on the host cells that allow virus entry. Above the lung, ACE2 is reported to be mainly expressed in other tissues such as the enterocytes of colon and kidneys[11]. To date, alternative SARS-CoV-2 binding receptors have been proposed mainly based on computational approaches or meta-proteomics analyses. The most extensive parts of the information on SARS-CoV-2 have been retrieved by comparing the

previous SARS-CoV, other Coronaviruses, and/or other viruses. ACE2 receptor itself was initially investigated because it was known to be the receptor responsible for SARS-CoV entry[12]. Bioinformatic and computational analyses have also been carried out to predict alternative spike S binding proteins based on known human receptors for other viruses[13] or coronavirus host-virus interaction[14]. Very few or preliminary experimental data are available on these topics. In particular, the investigation of alternative receptors can better elucidate the pathogenicity of SARS-CoV-2, also in other tissues than lungs.

In this project, we studied the Spike S1 protein-protein interactions with membrane protein extracts derived from kidney (HK-2 cell line) and colon (NCM460D cell line) human cells employing a pulldown assay on the membrane proteins. The finding of other S1 interacting proteins may be responsible for SARS-CoV-2 infectivity and might suggest different targets in COVID-19 treatment.

3.2.2 Experimental methods

3.2.2.1 Cell lysis and protein quantification

To obtain the membrane protein extracts, the two cell lines were treated with Mem-PER™ Plus Membrane Protein Extraction kit, according to the manufacturer protocol. The cytosol and membrane proteins were quantified by the Bradford assay as described in section 2.1.2.1.

3.2.2.2 Western Blot assay for the verification of fractionated lysis

We performed a Western Blot assay to verify the membrane proteins enrichment for the sample used in the pulldown assay. In particular, 20µg of cytosolic and membrane extracts were loaded onto a 10% SDS-PAGE gel. The samples were prepared with Laemli Buffer (100mM Tris HCl pH=6.8, 4% SDS, 20% glycerol, bromophenol blue, 100mM DTT) and boiled at 99°C for 10 minutes. The

electrophoretic run started at 100V. Separated proteins were transferred onto a nitrocellulose membrane using Trans-blot Turbo Transfer System (standard protocol: 1.5 A constant; up to 25 V; 10 minutes). Membranes were blocked with 5% nonfat milk in PBS for 1 hour and were incubated overnight at 4°C with rabbit polyclonal anti-Caveolin 1 (Abcam) (1:1000 in PBS, 0.2% Tween 20), as membrane marker, and 90 minutes at room temperature with rabbit polyclonal anti-GAPDH (Santa Cruz, California, US) (1:200 in PBS, 0.2% Tween20), as a cytosolic marker. Membranes were incubated with an anti-rabbit HRP secondary antibody (Invitrogen, 1:10000 in PBS, 0.2% Tween20) for 45 minutes at room temperature. Immunoreactive bands were developed by ECL Western Blotting Substrate, and the proteins were visualized by exposing the membrane to autoradiography films.

3.2.2.3 Isolation of Spike S1 interacting proteins

The spike S1 protein used for the pulldown experiment was purchased by Abcam (Cambridge, UK) as a chimera protein fused with a sheep Fc. Spike S1 was immobilized on Dynabeads protein G (Thermo Fisher Scientific), adding 2.5µg of protein each 40µl of slurry resin and incubating 2h at 4°C. For the pre-clearing step, the membrane extract of HK-2 (1.1 mg) and NCM460D (0.65 mg) cell lines were incubated with Dynabeads Protein G (Thermo Fischer Scientific) for 2h at 4°C considering 40µl of slurry resin for 1 mg of extract. The supernatants were incubated with the resin derivatized with the S1 protein overnight at 4°C. The supernatants were removed, and the resin washed five times with membrane extraction buffer provided by the Mem-PER kit. The resin immobilized with S1, containing its interactors, and the pre-clearing resin, on which the S1 non-specific proteins were absorbed, were eluted with 5% SDS. The pre-clearing eluates were used as control.

3.2.2.4 Protein digestion with S-Trap cartridges and mass spectrometry analysis

S1 pulled-down and control eluates from the three cell lines and were digested with the S-Trap cartridges (Protifi, New York, US) according to the manufacturer protocol. The latter led to the obtainment of a non-separated peptide mixture suitable for a shotgun proteomics approach. In particular, the mixture peptides were analyzed by nano LC-MS/MS, using the Easy-nLC II chromatographic system coupled with an LTQ Orbitrap XL mass spectrometer. The peptide fractionation was performed onto a C18 capillary reverse-phase column (200 mm, 75 μm, 5 μm) working at 250 nl/min flow rate, using a step gradient of eluent B (0.2% formic acid, 95% acetonitrile LC-MS Grade) was employed and a non-linear 5% to 50% gradient of eluent B (95% acetonitrile, 0.2% formic acid) over 260min was applied. Mass analyses were performed in Data Dependent Acquisition (DDA) mode by fragmenting the 10 most intense ions in the collision-induced dissociation (CID) modality. Protein identification was carried out by MaxQuant software (v.1.5.2.8), using UniProt Homo Sapiens as the database. The peptide tolerance was set as 10 ppm and 1% FDR cutoff for each peptide and protein identification. The fixed modification was the carbamidomethyl (Cys), while the variable modifications were Gln->pyro-Glu (N-term Gln) and Oxidation (Met).

3.2.2.5 Bioinformatic analysis

The datasets derived from the pulldown of HK-2 and NCM460D were analyzed with the platform STRING[15]. The latter allowed the visualization of the protein-protein interaction networks based both on functional and physical interaction. The interaction's confidence was set as high (score: 0.7) for both cell lines. Then, the over-representation analysis was done considering statistically significantly enriched categories carrying a p-value<0.05.

3.2.3 Results and Discussion

To identify new potential receptors for the Spike S1 domain, HK-2 and NCM460D were subjected to a lysis protocol suitable to separate membrane proteins from the remaining cell protein content. To verify the fractionated lysis procedure, we performed a Western Blot assay. The presence of Caveolin-1, a marker for membrane proteins, and GAPDH, a marker for soluble cytoplasmic proteins, was verified in cytoplasmic and membrane protein extracts. As clearly shown in Fig. 3.2.2, enrichment in the membranous protein content in all the cell lines was reached.

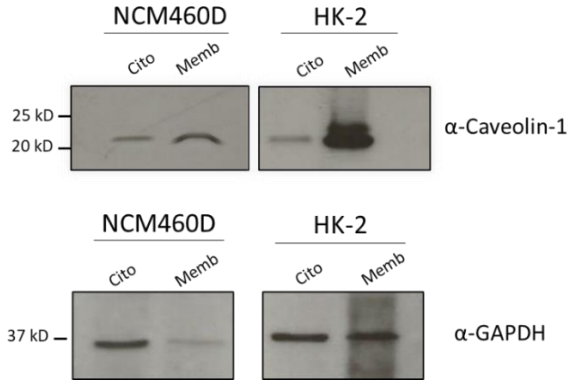

Fig 3.2.2 Western Blot of membrane and cytoplasmic protein extract for NCM460D and HK-2 cell lines developed using anti-Caveolin-1 (α-Caveolin-1) and anti-GAPDH (α-GAPDH) antibodies.

The membrane protein extracts of HK-2 and NCM460D were incubated with a Protein G-derivatized solid support on which S1-Fc recombinant protein was immobilized. The control was represented by the protein extract incubated with the Protein G solid support alone (pre-clearing step). The proteins eluted both from S1-derivated and pre-clearing resins were digested with the S-Trap cartridges, and the peptide mixtures were analyzed with a "shotgun" proteomics approach using LC-MS/MS methodologies. The proteins identified both in the pulled-down and control samples were discarded, obtaining a list of specific

putative spike S1 binding partners. This approach aimed to identify other receptors/ancillary proteins responsible for the viral recognition in several human tissues. The presence of other mechanisms of entry of SARS-CoV-2 above the ACE2 receptor could explain the observed infectivity of SARS-CoV-2 at an even low expression level of ACE2[6,13].

3.2.3.1 Results on HK-2 cell line

The pulldown experiment associated with shotgun proteomics analysis allowed the identification of 50 putative S1-interacting proteins from the HK-2 cell membrane. The identified proteins are reported in the table below (Table 3.2.1).

Protein name	Gene name	Uniprot code	Peptides	Sequence coverage (%)
Peroxiredoxin-1	PRDX1	Q06830	10	64.8
Protein S100-A9	S100A9	P06702	6	63.2
Prohibitin-2	PHB2	Q99623	15	59.9
Prohibitin	PHB	P35232	12	52.9
Cytoskeleton-associated protein 4	CKAP4	Q07065	25	49.8
ATP synthase subunit O, mitochondrial	ATP5PO	P48047	7	45.5
40S ribosomal protein S19	RPS19	P39019	8	44.1
Voltage-dependent anion-selective channel protein 1	VDAC1	P21796	10	42.4
Calreticulin	CALR	P27797	10	40.8
Phosphoglycerate kinase 1	PGK1	P00558	12	40.8
Fructose-bisphosphate aldolase A	ALDOA	P04075	11	39.3
Stomatin-like protein 2, mitochondrial	STOML2	Q9UJZ1	7	35.7
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 1	RPN1	P04843	16	35.1
Ras-related protein Rab-7a	RAB7A	P51149	6	34.8
ATP synthase subunit alpha, mitochondrial	ATP5F1A	P25705	14	34.4
Cytochrome c oxidase subunit 4	COX4I1	P13073	5	32

isoform 1, mitochondrial				
Ras-related protein Rab-1A	RAB1A	P62820	6	31.2
Ubiquitin carboxyl-terminal hydrolase isozyme L1	UCHL1	P09936	4	29.1
Dolichyl-diphosphooligosaccharide--protein glycosyltransferase subunit 2	RPN2	P04844	11	27.9
Endoplasmic reticulum resident protein 29	ERP29	P30040	5	27.2
Calnexin	CANX	P27824	12	24.8
Calumenin	CALU	O43852	5	24.1
Elongation factor 2	EEF2	P13639	14	24.1
L-lactate dehydrogenase A chain	LDHA	P00338	6	23.5
Ras-related protein Rab-5C	RAB5C	P51148	4	22.2
Cytochrome b-c1 complex subunit 2, mitochondrial	UQCRC2	P22695	6	21
Clusterin	CLU	P10909	6	20.9
Aldo-keto reductase family 1 member B1	AKR1B1	P15121	4	20.6
ATP synthase subunit gamma, mitochondrial	ATP5F1C	P36542	4	20.5
Transcription intermediary factor 1-beta	TRIM28	Q13263	10	20.1
OCIA domain-containing protein 1	OCIAD1	Q9NX40	4	17.6
Serine hydroxymethyltransferase, mitochondrial	SHMT2	P34897	7	17.3
Polyadenylate-binding protein 1	PABPC1	P11940	9	17.1
Moesin	MSN	P26038	9	16.6
Sorting and assembly machinery component 50 homolog	SAMM50	Q9Y512	5	16.6
Aspartate aminotransferase, mitochondrial	GOT2	P00505	6	14.9
Integrin beta-1	ITGB1	P05556	8	13.2
Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	P31930	4	13.1
4F2 cell-surface antigen heavy chain	SLC3A2	P08195	7	13
Signal transducer and activator of transcription 1-alpha/beta	STAT1	P42224	7	12.8

Protein ERGIC-53	LMAN1	P49257	6	12.7
Inhibitor of nuclear factor kappa-B kinase-interacting protein	IKBIP	Q70UQ0	4	12
Prolyl 4-hydroxylase subunit alpha-1	P4HA1	P13674	5	11.8
Fascin	FSCN1	Q16658	5	11.6
Neuroblast differentiation-associated protein AHNAK	AHNAK	Q09666	18	11.4
T-complex protein 1 subunit delta	CCT4	P50991	5	11.3
Mitochondrial antiviral-signaling protein	MAVS	Q7Z434	4	10.7
Leucine-rich PPR motif-containing protein, mitochondrial	LRPPRC	P42704	9	9.9
Laminin subunit gamma-1	LAMC1	P11047	4	4.4
DNA-dependent protein kinase catalytic subunit	PRKDC	P78527	4	1.4

Table 3.2.1 List of putative S1 interactors. The table reports the protein name and gene, the Uniprot code, the identified peptides, and protein sequence coverage.

The proteins in Table 3.2.1 were used as input for the STRING platform to detect the protein-protein network (Fig. 3.2.3) and perform an over-representation analysis.

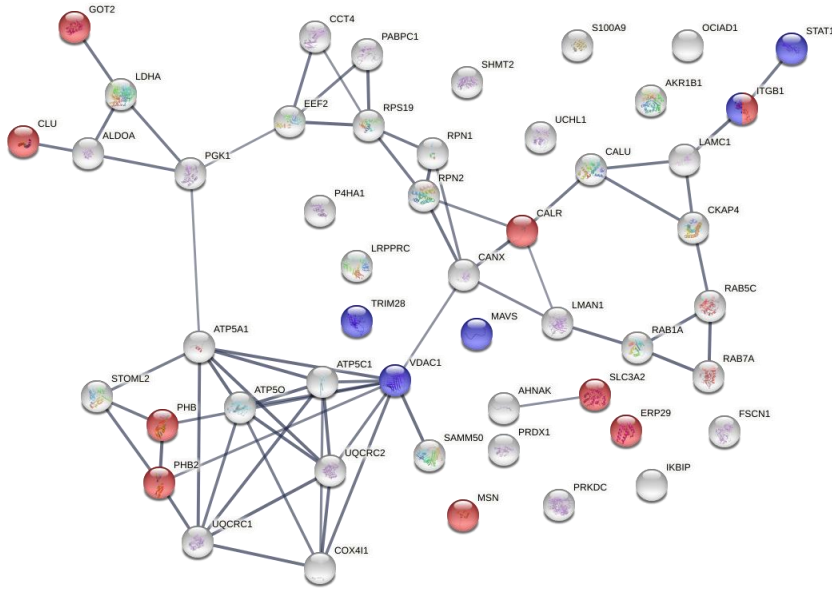

Fig. 3.2.3 STRING protein-protein network of potential S1 binding proteins in HK-2 cell line. In red are highlighted proteins belonging to the cell surface (p-value=5.7E-4) category; in blue, those are associated with virus-host interaction (p-value=0.037).

The STRING network analysis showed 61 edges and a protein-protein interactions (PPIs) p-value of 2.33E-15, indicating that the nodes (proteins) are not random and the number of edges is significant.

Furthermore, the over-representation analysis of the cell component category displayed proteins belonging to the cell surface (p-value=5.7E-4, red nodes in Fig. 3.2.3) and proteins involved in the host-virus interaction process (p-value=0.037, blue nodes in Fig. 3.2.3).

These indications could be very attractive to depict SARS-CoV-2 still unknown entry molecular mechanisms relying on the interaction with new targets on the host cells. Among these, Integrin beta-1 stands out. As reported above, integrins are emerging as novel receptors for SARS-CoV-2 due to virus integrin-binding RGD motif in spike protein[16]. Furthermore, ITGB1 is known to interact with

SARS-CoV-2 protein ORF3a[17]. Other proteins listed in the cell membrane compartment are found responsible for other viruses' viral processes but not yet in the *Coronaviridae* family. The amino acid transporter SLC3A2 is specifically required for the entry step but not for other stages of the Hepatitis C virus life cycle[18]; the chaperone Clusterin (CLU) has been identified as related to herpes simplex virus cellular entry and intracellular transport[19]; the actin cytoskeleton-to-plasma membrane crosslink Moesin (MSN) promotes HIV binding and entry through the polymerization of actin that is required to assemble high concentrations of the co-receptors at the plasma membrane of the target cell[20].

An additional comment should be made for PHB/PHB2/STOML2 complex, MAVS, and Calnexin/Calreticulin chaperones. Prohibitin-1 (PHB), -2 (PHB2), and Stomatin-like 2 (STOML2) proteins are involved in mitochondria homeostasis and localize in the inner mitochondrial membrane[21]. Cornillez-Ty et al. previously highlighted SARS-CoV non-structural protein 2 (nsP2) interactions with prohibitins PHB and PHB2[22], and STOML2 was also present in their dataset. Interestingly, Gordon et al. also identified an interaction of ORF3B from SARS-CoV-2 with STOML2[23]. These cellular proteins regulate mitochondrial homeostasis and are involved in processes such as mitophagy and mitochondrial fusion. By targeting these proteins, coronavirus proteins may also impact essential mitochondrial functions such as respiration, lipid homeostasis, and innate immunity. As prohibitins also regulate mitochondrial fusion and fission, for which the impact on antiviral signaling has been well documented[24], this could have indirect consequences on the antiviral response. Recent studies have established physical interactions between components of the prohibitin complex and Mitochondrial antiviral-signaling protein (MAVS), a pivotal adaptor in viral RNA sensing and interferon induction[25]. An interaction between MAVS and SARS-CoV-2 M membrane glycoprotein has also been identified, leading to attenuation of the innate antiviral response.

Calnexin (CANX) and calreticulin (CALR) are homologous lectin chaperones that assist the maturation of cellular and viral glycoproteins in the mammalian endoplasmic reticulum. Fukushi et al. have found that SARS-CoV spike glycoprotein (S protein), a key molecule for viral entry, binds to calnexin, a molecular chaperone in the endoplasmic reticulum (ER), but not to calreticulin, a homolog of calnexin[26]. However, this interaction might occur following virus entrance during the synthesis, folding, and assembly process of viral proteins.

3.2.3.2 Results on NCM460D cell line

Proteins identified in NCM460D cells as putative S1-binding proteins were 115 and are listed in Table 3.2.2.

Protein name	Gene name	Uniprot code	Peptides	Sequence coverage %)
Peptidyl-prolyl cis-trans isomerase A	PPIA	P62937	8	64.2
Protein S100-A7	S100A7	P31151	5	53.5
Ras-related protein Rab-7a	RAB7A	P51149	8	44.9
Protein S100-A9	S100A9	P06702	4	43.9
Endoplasmic reticulum resident protein 29	ERP29	P30040	9	42.5
Ras-related protein Rab-11A	RAB11A	P62491	7	41.2
Hypoxia up-regulated protein 1	HYOU1	Q9Y4L1	28	40
Cathepsin D	CTSD	P07339	12	39.8
Protein S100-A8	S100A8	P05109	4	39.8
Electron transfer flavoprotein subunit alpha, mitochondrial	ETFA	P13804	9	38.7
Alpha-actinin-4	ACTN4	O43707	29	38.6
Protein disulfide-isomerase A3	PDIA3	P30101	16	37.4
Ras-related protein Rab-1B	RAB1B	Q9H0U4	6	37.3
14-3-3 protein beta/alpha	YWHAB	P31946	4	36.6
14-3-3 protein theta	YWHAQ	P27348	8	35.9
14-3-3 protein zeta/delta	YWHAZ	P63104	4	35.5

Protein disulfide-isomerase	P4HB	P07237	15	33.7
Pyruvate dehydrogenase E1 component subunit beta, mitochondrial	PDHB	P11177	9	32.9
Heat shock protein 75 kDa, mitochondrial	TRAP1	Q12931	18	32.5
Protein disulfide-isomerase A4	PDIA4	P13667	16	32.1
Adipocyte plasma membrane-associated protein	APMAP	Q9HDC9	10	32
Lipocalin-1	LCN1	P31025	4	31.8
Enoyl-CoA hydratase, mitochondrial	ECHS1	P30084	6	31.7
Delta(3,5)-Delta(2,4)-dienoyl-CoA isomerase, mitochondrial	ECH1	Q13011	8	31.4
Peroxiredoxin-5, mitochondrial	PRDX5	P30044	5	31.3
Galectin-3	LGALS3	P17931	7	31.2
Chloride intracellular channel protein 1	CLIC1	O00299	4	30.7
Membrane-associated progesterone receptor component 1	PGRMC1	O00264	5	30.3
Myosin light polypeptide 6	MYL6	P60660	4	29.8
14-3-3 protein epsilon	YWHAE	P62258	5	29.4
Chromobox protein homolog 5	CBX5	P45973	5	29.3
Thioredoxin-dependent peroxide reductase, mitochondrial	PRDX3	P30048	5	28.9
Caspase-14	CASP14	P31944	6	28.1
Calreticulin	CALR	P27797	9	27.8
Tubulin beta-4B chain	TUBB4B	P68371	10	27
Peroxiredoxin-2	PRDX2	P32119	4	26.8
Ras-related protein Rab-18	RAB18	Q9NP72	4	26.7
Transferrin receptor protein 1	TFRC	P02786	16	26.7
Succinate--CoA ligase [ADP-forming] subunit beta, mitochondrial	SUCLA2	Q9P2R7	12	26.3
Transcription intermediary factor 1-beta	TRIM28	Q13263	15	26.1
Protein ABHD11	ABHD11	Q8NFV4	7	25.7

Malectin	MLEC	Q14165	6	25.7
Phosphoglycerate kinase 1	PGK1	P00558	8	24.9
GrpE protein homolog 1, mitochondrial	GRPEL1	Q9HAV7	4	24.4
Nuclease-sensitive element-binding protein 1	YBX1	P67809	5	24.4
Persulfide dioxygenase ETHE1, mitochondrial	ETHE1	O95571	5	24
Ras-related protein Rab-10	RAB10	P61026	4	24
Alpha-enolase	ENO1	P06733	7	23.7
60S acidic ribosomal protein P0	RPLP0	P05388	4	23.7
Fumarate hydratase, mitochondrial	FH	P07954	7	23.3
Endoplasmic reticulum resident protein 44	ERP44	Q9BS26	7	22.4
Electron transfer flavoprotein subunit beta	ETFB	P38117	6	22.4
Transforming protein RhoA	RHOA	P61586	4	22.3
NADPH--cytochrome P450 reductase	POR	P16435	12	22
Carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	P40199	4	21.5
Methylcrotonoyl-CoA carboxylase beta chain, mitochondrial	MCCC2	Q9HCC0	8	21
L-lactate dehydrogenase B chain	LDHB	P07195	5	20.4
Chromobox protein homolog 3	CBX3	Q13185	4	20.2
Aldehyde dehydrogenase, mitochondrial	ALDH2	P05091	8	19.7
Glutamate dehydrogenase 1, mitochondrial	GLUD1	P00367	9	19.4
Stomatin-like protein 2, mitochondrial	STOML2	Q9UJZ1	5	19.4
Cytochrome b-c1 complex subunit 1, mitochondrial	UQCRC1	P31930	7	19.2
Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex,	DLST	P36957	6	19

mitochondrial				
Semenogelin-1	SEMG1	P04279	7	18.4
Peptidyl-prolyl cis-trans isomerase FKBP4	FKBP4	Q02790	7	18.1
2-oxoglutarate dehydrogenase, mitochondrial	OGDH	Q02218	13	17.8
Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	ACADM	P11310	5	17.6
Glutaminase kidney isoform, mitochondrial	GLS	O94925	8	17.6
HLA class I histocompatibility antigen, A-68 alpha chain	HLA-A	P01891	5	17.5
RuvB-like 2	RUVBL2	Q9Y230	7	17.5
Transitional endoplasmic reticulum ATPase	VCP	P55072	9	17.4
Heat shock protein HSP 90-alpha	HSP90AA1	P07900	5	17.2
Serine hydroxymethyltransferase, mitochondrial	SHMT2	P34897	7	17.1
Progranulin	GRN	P28799	9	17
Aconitate hydratase, mitochondrial	ACO2	Q99798	10	16.9
Clathrin heavy chain 1	CLTC	Q00610	20	16.7
L-lactate dehydrogenase A chain	LDHA	P00338	5	16.6
Aldehyde dehydrogenase X, mitochondrial	ALDH1B1	P30837	6	16.4
Calumenin	CALU	O43852	4	16.2
Lactotransferrin	LTF	P02788	10	16.2
Pyruvate dehydrogenase E1 component subunit alpha, somatic form, mitochondrial	PDHA1	P08559	6	16.2
Neutral alpha-glucosidase AB	GANAB	Q14697	13	15.7
NAD kinase 2, mitochondrial	NADK2	Q4G0N4	5	14.3
Synaptic vesicle membrane protein VAT-1 homolog	VAT1	Q99536	4	13.5
Glucose-6-phosphate isomerase	GPI	P06744	6	13.1
Polyadenylate-binding protein 1	PABPC1	P11940	8	12.9

Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	DLAT	P10515	5	12.7
Dipeptidase 1	DPEP1	P16444	5	12.2
Importin subunit beta-1	KPNB1	Q14974	7	11.8
Neutral amino acid transporter B(0)	SLC1A5	Q15758	5	11.8
Nucleobindin-1	NUCB1	Q02818	5	11.5
Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	ATP2A2	P16615	9	11.2
ERO1-like protein alpha	ERO1A	Q96HE7	4	11.1
Mitochondrial import inner membrane translocase subunit TIM44	TIMM44	O43615	4	11.1
Lon protease homolog, mitochondrial	LONP1	P36776	7	10.7
Thioredoxin domain-containing protein 5	TXNDC5	Q8NBS9	4	10.6
Elongation factor G, mitochondrial	GFM1	Q96RP9	7	10.4
Succinate dehydrogenase [ubiquinone] flavoprotein subunit, mitochondrial	SDHA	P31040	5	10.1
Plastin-1	PLS1	Q14651	4	10
Heat shock protein 105 kDa	HSPH1	Q92598	7	9.7
Prosaposin	PSAP	P07602	4	8.8
Transketolase	TKT	P29401	4	8.8
Vesicle-fusing ATPase	NSF	P46459	6	8.7
Isoleucine--tRNA ligase, mitochondrial	IARS2	Q9NSE4	8	8.4
Lysosome-associated membrane glycoprotein 1	LAMP1	P11279	4	8.4
Nuclear pore complex protein Nup93	NUP93	Q8N1F7	5	8.3
Mitochondrial intermediate peptidase	MIPEP	Q99797	5	8.1
Integrin alpha-2	ITGA2	P17301	6	7.3
Transmembrane 9 superfamily member 2	TM9SF2	Q99805	4	7.2
Reticulon-4	RTN4	Q9NQC3	5	7

Procollagen galactosyltransferase 1	COLGALT 1	Q8NBJ5	4	6.9
Cell cycle and apoptosis regulator protein 2	CCAR2	Q8N163	4	6.1
Non-specific lipid-transfer protein	SCP2	P22307	4	5.7
Nuclear pore complex protein Nup155	NUP155	O75694	4	4.3
UDP-glucose:glycoprotein glucosyltransferase 1	UGGT1	Q9NYU2	5	3.4

Table 3.2.2 List of putative S1 interactors. The table reports the protein name and gene, the Uniprot code, the identified peptides, and protein sequence coverage.

The proteins listed in the Table 3.2.2 were used as input for STRING obtaining the network in Fig. 3.2.4. STRING displayed the overall presence of 230 edges in the network with a total PPIs interaction p-value <1.0E-16.

The biological processes enrichment analysis showed the presence of proteins related to the viral entry into the host cell (p-value=3E-3, blue in Fig. 3.2.4) and the host cell receptor for virus entry (p-value=7E-3, red in Fig. 3.2.4).

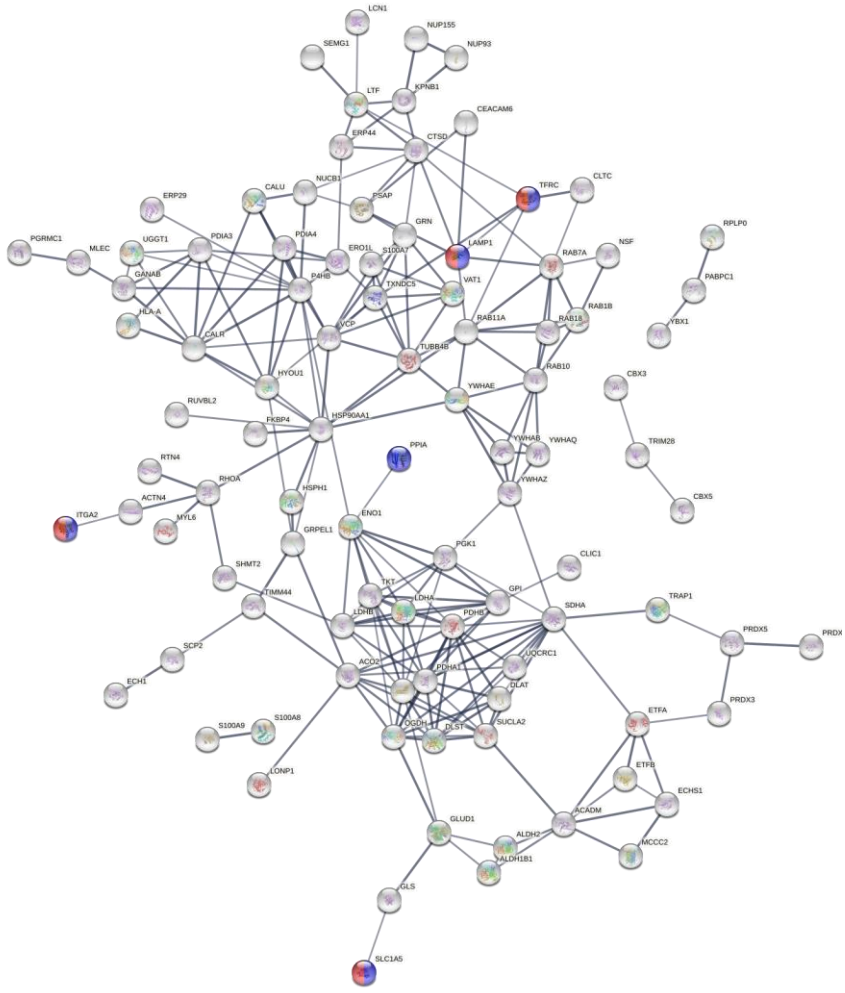

Fig. 3.2.4 STRING protein-protein network of potential S1 binding proteins in NCM460D cell line. In blue are highlighted proteins belonging to the "viral entry into host cell" category (p-value=3E-3); in red, those annotate as the "host cell receptor for virus entry" category (p-value=7E-3).

STRING over-representation analysis of the colon dataset displayed 4 proteins as potential receptors responsible for virus entry. Among these, Lysosome associated membrane glycoprotein 1 (LAMP1), a known lysosome marker, was reported to mediate the entrance of human Lassa virus (LASV), belonging to the *Arenaviridae* family, through the interaction with a subunit of the trimeric class 1 viral

glycoprotein complex (the spike complex) in a pH-dependent manner[27,28]. Of interest, the transferrin receptor protein 1 (TFRC) was also identified, which was reported as a target of several mammalian retroviruses[29,30]. A recent work proposed TFRC as an alternative SARS-CoV-2 receptor for viral entry[31]; it was identified as closely interacting with ACE2 in infected cardiomyocytes[32] and up-regulated in the severe pathogenesis of COVID-19 disease.

The remaining component of the "host cell receptor for virus entry" category are Integrin alpha-2 (ITGA2) and Neutral amino acid transporter B(0) (SLC1A5), which are receptors for human and/or animal viruses[33,34,35]. Integrins are of particular interest and are supposed to be the main alternative as S protein receptors in alveolar epithelial cells through its RGD consensus sequence to accelerate the infection process[36].

Three other proteins have caught our attention as known to participate in viral and/or SARS-CoV-2 processes: Carcinoembryonic antigen-related cell adhesion molecule 6 (CEACAM6), Cyclophilin A (PPIA), and Galectin-3 (LGALS3).

CEACAM6 is a cell surface glycoprotein that plays a role in cell adhesion and tumor progression, but this family of proteins was reported to bind S1 in other coronaviruses and trigger membrane fusion reactions mediated by integral membrane S2 fragments[37]. Cyclophilin A is a member of a family of highly conserved and ubiquitous proteins with peptidyl-prolyl cis/trans isomerase activity, crucial for protein folding[38]. In addition to its canonical roles, the PPIase activity of PPIA has recently been demonstrated to have other effects, including intracellular trafficking, signal transduction, transcription regulation, cell cycle regulation, and stress response[39]. In recent years many studies showed that PPIA was involved in the pathogenesis of viral infection. PPIA plays a critical role in the successful replication of viruses such as human immunodeficiency virus (HIV), hepatitis C virus (HCV), hepatitis B virus (HBV), etc.[40] PPIA was also

reported to regulate SARS-CoV replication by binding to the nucleocapsid protein and incorporating it into particles[41,42]. The inhibition of cyclophilins with drugs has been reported to block the replication of CoVs of all genera, including SARS-CoV, human CoV-229E and -NL-63, feline CoV[43], and recently it was also proposed as a therapeutic target for SARS-CoV-2[44]. Galectin-3, a member of the galectins family, is a carbohydrate-binding protein that exhibits pleiotropic effects and has been implicated in the disease process of various inflammatory conditions[45]. LGALS3 is highly expressed in the lung, followed by the gastrointestinal tract (stomach, duodenum, small intestine, colon, and rectum) and brain. Recent discoveries have begun to shed light on its role in viral infections [46]. For example, in HIV and HTLV, Gal-3 serves as an attachment factor that facilitates viral entry into T-cells[46]. A role also in SARS-CoV-2 infection and/or entry has been suggested through the triggering of the "cytokine storm" typical in COVID-19 patients[47] and the interaction with Spike S1 Receptor Binding Domain (RBD), respectively[48,49].

3.2.4 Conclusions

In conclusion, we identified some putative Spike S1 targets on renal and colon cells and which could be associated with viral entry through a proteomics approach. High-throughput experiments have not yet been done to elucidate the S1 interactome. The combined results obtained by NCM460D and HK-2 cell lines are displayed in the Cytoscape network in Fig. 3.2.5, finding twelve common proteins. Above the previously discussed proteins, putative S1 targets not strictly related to the cell surface were detected. These proteins may be further investigated to detect potential host-virus interactions downstream involved in the viral internalization processes.

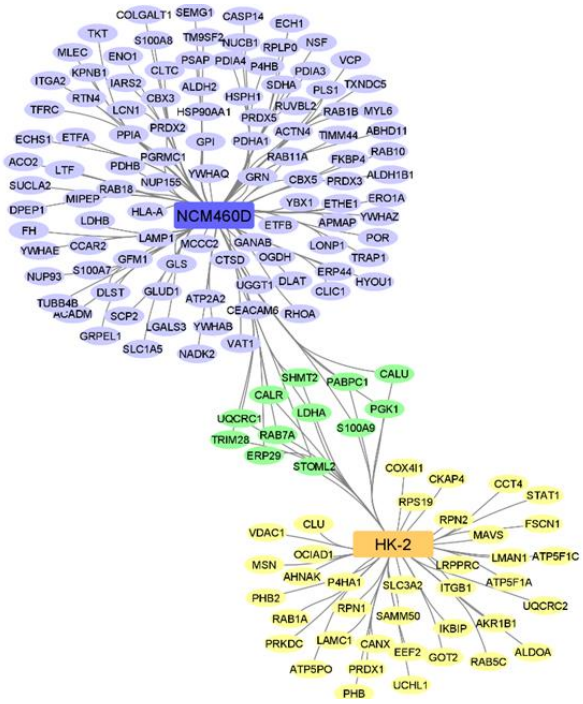

Fig 3.2.5 Cytoscape networks for NCM460D (blue nodes), HK-2 (yellow nodes) cell lines. In green, the proteins present in both datasets are displayed.

Overall, the interactions detected in our results have to be confirmed and functionally investigated to go deeper in their meaning. Experiments on full-length Spike (S1+S2) and/or on SARS-CoV-2 infected cells are needed to better unravel the complete molecular mechanisms of these alternative ways used by the virus to enter host cells.

3.2.5 References

(1) Durmaz, B.; Abdulmajed, O.; Durmaz, R. Mutations Observed in the SARS-CoV-2 Spike Glycoprotein and Their Effects in the Interaction of Virus with ACE-2 Receptor. *Medeni Med J* **2020**, *35* (3), 253–260. https://doi.org/10.5222/MMJ.2020.98048.

(2) Shang, J.; Wan, Y.; Luo, C.; Ye, G.; Geng, Q.; Auerbach, A.; Li, F. Cell Entry Mechanisms of SARS-CoV-2. *Proc Natl Acad Sci U S A* **2020**, *117* (21), 11727–11734. https://doi.org/10.1073/pnas.2003138117.

(3) Chambers, J. P.; Yu, J.; Valdes, J. J.; Arulanandam, B. P. SARS-CoV-2, Early Entry Events. *J Pathog* **2020**, *2020*, 9238696. https://doi.org/10.1155/2020/9238696.

(4) Qi, F.; Qian, S.; Zhang, S.; Zhang, Z. Single Cell RNA Sequencing of 13 Human Tissues Identify Cell Types and Receptors of Human Coronaviruses. *Biochem Biophys Res Commun* **2020**, *526* (1), 135–140. https://doi.org/10.1016/j.bbrc.2020.03.044.

(5) Zhang, H.; Kang, Z.; Gong, H.; Xu, D.; Wang, J.; Li, Z.; Li, Z.; Cui, X.; Xiao, J.; Zhan, J.; Meng, T.; Zhou, W.; Liu, J.; Xu, H. Digestive System Is a Potential Route of COVID-19: An Analysis of Single-Cell Coexpression Pattern of Key Proteins in Viral Entry Process. *Gut* **2020**, *69* (6), 1010–1018. https://doi.org/10.1136/gutjnl-2020-320953.

(6) Yang, J.; Petitjean, S. J. L.; Koehler, M.; Zhang, Q.; Dumitru, A. C.; Chen, W.; Derclaye, S.; Vincent, S. P.; Soumillion, P.; Alsteens, D. Molecular Interaction and Inhibition of SARS-CoV-2 Binding to the ACE2 Receptor. *Nat Commun* **2020**, *11* (1), 4541. https://doi.org/10.1038/s41467-020-18319-6.

(7) Huang, Y.; Yang, C.; Xu, X.-F.; Xu, W.; Liu, S.-W. Structural and Functional Properties of SARS-CoV-2 Spike Protein: Potential Antivirus Drug Development for COVID-19. *Acta Pharmacol Sin* **2020**, *41* (9), 1141–

1149. https://doi.org/10.1038/s41401-020-0485-4.

(8) Millet, J. K.; Whittaker, G. R. Physiological and Molecular Triggers for SARS-CoV Membrane Fusion and Entry into Host Cells. *Virology* **2018**, *517*, 3–8. https://doi.org/10.1016/j.virol.2017.12.015.

(9) Xia, S.; Zhu, Y.; Liu, M.; Lan, Q.; Xu, W.; Wu, Y.; Ying, T.; Liu, S.; Shi, Z.; Jiang, S.; Lu, L. Fusion Mechanism of 2019-NCoV and Fusion Inhibitors Targeting HR1 Domain in Spike Protein. *Cell Mol Immunol* **2020**, *17* (7), 765–767. https://doi.org/10.1038/s41423-020-0374-2.

(10) Robson, B. Computers and Viral Diseases. Preliminary Bioinformatics Studies on the Design of a Synthetic Vaccine and a Preventative Peptidomimetic Antagonist against the SARS-CoV-2 (2019-NCoV, COVID-19) Coronavirus. *Comput Biol Med* **2020**, *119*, 103670. https://doi.org/10.1016/j.compbiomed.2020.103670.

(11) Chen, Q.-L.; Li, J.-Q.; Xiang, Z.-D.; Lang, Y.; Guo, G.-J.; Liu, Z.-H. Localization of Cell Receptor-Related Genes of SARS-CoV-2 in the Kidney through Single-Cell Transcriptome Analysis. *Kidney Dis (Basel)* **2020**, *6* (4), 258–270. https://doi.org/10.1159/000508162.

(12) Turner, A. J.; Hiscox, J. A.; Hooper, N. M. ACE2: From Vasopeptidase to SARS Virus Receptor. *Trends Pharmacol Sci* **2004**, *25* (6), 291–294. https://doi.org/10.1016/j.tips.2004.04.001.

(13) Zhang, Z.; Ye, S.; Wu, A.; Jiang, T.; Peng, Y. Prediction of the Receptorome for the Human-Infecting Virome. *Virol. Sin.* **2020**. https://doi.org/10.1007/s12250-020-00259-6.

(14) Perrin-Cocon, L.; Diaz, O.; Jacquemin, C.; Barthel, V.; Ogire, E.; Ramière, C.; André, P.; Lotteau, V.; Vidalain, P.-O. The Current Landscape of Coronavirus-Host Protein–Protein Interactions. *J Transl Med* **2020**, *18* (1), 319. https://doi.org/10.1186/s12967-020-02480-z.

(15) Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas,

J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; Jensen, L. J.; Mering, C. von. STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res* **2019**, *47* (D1), D607–D613. https://doi.org/10.1093/nar/gky1131.

(16) Sigrist, C. J.; Bridge, A.; Le Mercier, P. A Potential Role for Integrins in Host Cell Entry by SARS-CoV-2. *Antiviral Res* **2020**, *177*, 104759. https://doi.org/10.1016/j.antiviral.2020.104759.

(17) Sardar, R.; Satish, D.; Gupta, D. Identification of Novel SARS-CoV-2 Drug Targets by Host MicroRNAs and Transcription Factors Co-Regulatory Interaction Network Analysis. *Front Genet* **2020**, *11*, 571274. https://doi.org/10.3389/fgene.2020.571274.

(18) Nguyen, N. N. T.; Lim, Y.-S.; Nguyen, L. P.; Tran, S. C.; Luong, T. T. D.; Nguyen, T. T. T.; Pham, H. T.; Mai, H. N.; Choi, J.-W.; Han, S.-S.; Hwang, S. B. Hepatitis C Virus Modulates Solute Carrier Family 3 Member 2 for Viral Propagation. *Sci Rep* **2018**, *8* (1), 15486. https://doi.org/10.1038/s41598-018-33861-6.

(19) Carter, C. J. APP, APOE, Complement Receptor 1, Clusterin and PICALM and Their Involvement in the Herpes Simplex Life Cycle. *Neurosci Lett* **2010**, *483* (2), 96–100. https://doi.org/10.1016/j.neulet.2010.07.066.

(20) Maniti, O.; Carvalho, K.; Picart, C. Model Membranes to Shed Light on the Biochemical and Physical Properties of Ezrin/Radixin/Moesin. *Biochimie* **2013**, *95* (1), 3–11. https://doi.org/10.1016/j.biochi.2012.09.033.

(21) Hernando-Rodríguez, B.; Artal-Sanz, M. Mitochondrial Quality Control Mechanisms and the PHB (Prohibitin) Complex. *Cells* **2018**, *7* (12). https://doi.org/10.3390/cells7120238.

(22) Cornillez-Ty, C. T.; Liao, L.; Yates, J. R.; Kuhn, P.; Buchmeier, M. J. Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein 2 Interacts

with a Host Protein Complex Involved in Mitochondrial Biogenesis and Intracellular Signaling. *J Virol* **2009**, *83* (19), 10314–10318. https://doi.org/10.1128/JVI.00842-09.

(23) Gordon, D. E.; Jang, G. M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K. M.; O'Meara, M. J.; Rezelj, V. V.; Guo, J. Z.; Swaney, D. L.; Tummino, T. A.; Hüttenhain, R.; Kaake, R. M.; Richards, A. L.; Tutuncuoglu, B.; Foussard, H.; Batra, J.; Haas, K.; Modak, M.; Kim, M.; Haas, P.; Polacco, B. J.; Braberg, H.; Fabius, J. M.; Eckhardt, M.; Soucheray, M.; Bennett, M. J.; Cakir, M.; McGregor, M. J.; Li, Q.; Meyer, B.; Roesch, F.; Vallet, T.; Mac Kain, A.; Miorin, L.; Moreno, E.; Naing, Z. Z. C.; Zhou, Y.; Peng, S.; Shi, Y.; Zhang, Z.; Shen, W.; Kirby, I. T.; Melnyk, J. E.; Chorba, J. S.; Lou, K.; Dai, S. A.; Barrio-Hernandez, I.; Memon, D.; Hernandez-Armenta, C.; Lyu, J.; Mathy, C. J. P.; Perica, T.; Pilla, K. B.; Ganesan, S. J.; Saltzberg, D. J.; Rakesh, R.; Liu, X.; Rosenthal, S. B.; Calviello, L.; Venkataramanan, S.; Liboy-Lugo, J.; Lin, Y.; Huang, X.-P.; Liu, Y.; Wankowicz, S. A.; Bohn, M.; Safari, M.; Ugur, F. S.; Koh, C.; Savar, N. S.; Tran, Q. D.; Shengjuler, D.; Fletcher, S. J.; O'Neal, M. C.; Cai, Y.; Chang, J. C. J.; Broadhurst, D. J.; Klippsten, S.; Sharp, P. P.; Wenzell, N. A.; Kuzuoglu-Ozturk, D.; Wang, H.-Y.; Trenker, R.; Young, J. M.; Cavero, D. A.; Hiatt, J.; Roth, T. L.; Rathore, U.; Subramanian, A.; Noack, J.; Hubert, M.; Stroud, R. M.; Frankel, A. D.; Rosenberg, O. S.; Verba, K. A.; Agard, D. A.; Ott, M.; Emerman, M.; Jura, N.; von Zastrow, M.; Verdin, E.; Ashworth, A.; Schwartz, O.; d'Enfert, C.; Mukherjee, S.; Jacobson, M.; Malik, H. S.; Fujimori, D. G.; Ideker, T.; Craik, C. S.; Floor, S. N.; Fraser, J. S.; Gross, J. D.; Sali, A.; Roth, B. L.; Ruggero, D.; Taunton, J.; Kortemme, T.; Beltrao, P.; Vignuzzi, M.; García-Sastre, A.; Shokat, K. M.; Shoichet, B. K.; Krogan, N. J. A SARS-CoV-2 Protein Interaction Map Reveals Targets for Drug Repurposing. *Nature* **2020**, *583* (7816), 459–468.

https://doi.org/10.1038/s41586-020-2286-9.

(24) Pourcelot, M.; Arnoult, D. Mitochondrial Dynamics and the Innate Antiviral Immune Response. *FEBS J* **2014**, *281* (17), 3791–3802. https://doi.org/10.1111/febs.12940.

(25) Yasukawa, K.; Kinoshita, D.; Yaku, K.; Nakagawa, T.; Koshiba, T. The MicroRNAs MiR-302b and MiR-372 Regulate Mitochondrial Metabolism via the SLC25A12 Transporter, Which Controls MAVS-Mediated Antiviral Innate Immunity. *J Biol Chem* **2020**, *295* (2), 444–457. https://doi.org/10.1074/jbc.RA119.010511.

(26) Fukushi, M.; Yoshinaka, Y.; Matsuoka, Y.; Hatakeyama, S.; Ishizaka, Y.; Kirikae, T.; Sasazuki, T.; Miyoshi-Akiyama, T. Monitoring of S Protein Maturation in the Endoplasmic Reticulum by Calnexin Is Important for the Infectivity of Severe Acute Respiratory Syndrome Coronavirus. *J Virol* **2012**, *86* (21), 11745–11753. https://doi.org/10.1128/JVI.01250-12.

(27) Jae, L. T.; Raaben, M.; Herbert, A. S.; Kuehne, A. I.; Wirchnianski, A. S.; Soh, T. K.; Stubbs, S. H.; Janssen, H.; Damme, M.; Saftig, P.; Whelan, S. P.; Dye, J. M.; Brummelkamp, T. R. Lassa Virus Entry Requires a Trigger-Induced Receptor Switch. *Science* **2014**, *344* (6191), 1506–1510. https://doi.org/10.1126/science.1252480.

(28) Cohen-Dvashi, H.; Israeli, H.; Shani, O.; Katz, A.; Diskin, R. Role of LAMP1 Binding and PH Sensing by the Spike Complex of Lassa Virus. *J. Virol.* **2016**, *90* (22), 10329–10338. https://doi.org/10.1128/JVI.01624-16.

(29) Coffin, J. M. Virions at the Gates: Receptors and the Host-Virus Arms Race. *PLoS Biol* **2013**, *11* (5), e1001574. https://doi.org/10.1371/journal.pbio.1001574.

(30) Demogines, A.; Abraham, J.; Choe, H.; Farzan, M.; Sawyer, S. L. Dual Host-Virus Arms Races Shape an Essential Housekeeping Protein. *PLoS Biol* **2013**, *11* (5), e1001571. https://doi.org/10.1371/journal.pbio.1001571.

(31) Lai, R.; Tang, X.; Yang, M.; Duan, Z.; Liao, Z.; Liu, L.; Cheng, R.; Fang, M.; Wang, G.; Liu, H.; Xu, J.; Kamau, P.; Zhang, Z.; Yang, L.; Zhao, X.; Peng, X. *Transferrin Receptor Is Another Receptor for SARS-CoV-2 Entry*; preprint; In Review, 2020. https://doi.org/10.21203/rs.3.rs-96962/v1.

(32) Wicik, Z.; Eyileten, C.; Jakubik, D.; Simões, S. N.; Martins, D. C.; Pavão, R.; Siller-Matula, J. M.; Postula, M. ACE2 Interaction Networks in COVID-19: A Physiological Framework for Prediction of Outcome in Patients with Cardiovascular Risk Factors. *J Clin Med* **2020**, *9* (11). https://doi.org/10.3390/jcm9113743.

(33) Graham, K. L.; Halasz, P.; Tan, Y.; Hewish, M. J.; Takada, Y.; Mackow, E. R.; Robinson, M. K.; Coulson, B. S. Integrin-Using Rotaviruses Bind Alpha2beta1 Integrin Alpha2 I Domain via VP4 DGE Sequence and Recognize AlphaXbeta2 and AlphaVbeta3 by Using VP7 during Cell Entry. *J Virol* **2003**, *77* (18), 9969–9978. https://doi.org/10.1128/jvi.77.18.9969-9978.2003.

(34) Rasko, J. E.; Battini, J. L.; Gottschalk, R. J.; Mazo, I.; Miller, A. D. The RD114/Simian Type D Retrovirus Receptor Is a Neutral Amino Acid Transporter. *Proc Natl Acad Sci U S A* **1999**, *96* (5), 2129–2134. https://doi.org/10.1073/pnas.96.5.2129.

(35) Scalise, M.; Pochini, L.; Console, L.; Losso, M. A.; Indiveri, C. The Human SLC1A5 (ASCT2) Amino Acid Transporter: From Function to Structure and Role in Cell Biology. *Front. Cell Dev. Biol.* **2018**, *6*, 96. https://doi.org/10.3389/fcell.2018.00096.

(36) Sigrist, C. J.; Bridge, A.; Le Mercier, P. A Potential Role for Integrins in Host Cell Entry by SARS-CoV-2. *Antiviral Research* **2020**, *177*, 104759. https://doi.org/10.1016/j.antiviral.2020.104759.

(37) Lewicki, D. N.; Gallagher, T. M. Quaternary Structure of Coronavirus Spikes in Complex with Carcinoembryonic Antigen-Related Cell Adhesion

Molecule Cellular Receptors. *Journal of Biological Chemistry* **2002**, *277* (22), 19727–19734. https://doi.org/10.1074/jbc.M201837200.

(38) Fischer, G.; Wittmann-Liebold, B.; Lang, K.; Kiefhaber, T.; Schmid, F. X. Cyclophilin and Peptidyl-Prolyl Cis-Trans Isomerase Are Probably Identical Proteins. *Nature* **1989**, *337* (6206), 476–478. https://doi.org/10.1038/337476a0.

(39) Lammers, M.; Neumann, H.; Chin, J. W.; James, L. C. Acetylation Regulates Cyclophilin A Catalysis, Immunosuppression and HIV Isomerization. *Nat Chem Biol* **2010**, *6* (5), 331–337. https://doi.org/10.1038/nchembio.342.

(40) Watashi, K.; Shimotohno, K. Cyclophilin and Viruses: Cyclophilin as a Cofactor for Viral Infection and Possible Anti-Viral Target. *Drug Target Insights* **2007**, *2*, 9–18.

(41) Luo, C.; Luo, H.; Zheng, S.; Gui, C.; Yue, L.; Yu, C.; Sun, T.; He, P.; Chen, J.; Shen, J.; Luo, X.; Li, Y.; Liu, H.; Bai, D.; Shen, J.; Yang, Y.; Li, F.; Zuo, J.; Hilgenfeld, R.; Pei, G.; Chen, K.; Shen, X.; Jiang, H. Nucleocapsid Protein of SARS Coronavirus Tightly Binds to Human Cyclophilin A. *Biochem Biophys Res Commun* **2004**, *321* (3), 557–565. https://doi.org/10.1016/j.bbrc.2004.07.003.

(42) Chen, Z.; Mi, L.; Xu, J.; Yu, J.; Wang, X.; Jiang, J.; Xing, J.; Shang, P.; Qian, A.; Li, Y.; Shaw, P. X.; Wang, J.; Duan, S.; Ding, J.; Fan, C.; Zhang, Y.; Yang, Y.; Yu, X.; Feng, Q.; Li, B.; Yao, X.; Zhang, Z.; Li, L.; Xue, X.; Zhu, P. Function of HAb18G/CD147 in Invasion of Host Cells by Severe Acute Respiratory Syndrome Coronavirus. *J Infect Dis* **2005**, *191* (5), 755–760. https://doi.org/10.1086/427811.

(43) Pfefferle, S.; Schöpf, J.; Kögl, M.; Friedel, C. C.; Müller, M. A.; Carbajo-Lozoya, J.; Stellberger, T.; von Dall'Armi, E.; Herzog, P.; Kallies, S.; Niemeyer, D.; Ditt, V.; Kuri, T.; Züst, R.; Pumpor, K.; Hilgenfeld, R.;

Schwarz, F.; Zimmer, R.; Steffen, I.; Weber, F.; Thiel, V.; Herrler, G.; Thiel, H.-J.; Schwegmann-Weßels, C.; Pöhlmann, S.; Haas, J.; Drosten, C.; von Brunn, A. The SARS-Coronavirus-Host Interactome: Identification of Cyclophilins as Target for Pan-Coronavirus Inhibitors. *PLoS Pathog* **2011**, *7* (10), e1002331. https://doi.org/10.1371/journal.ppat.1002331.

(44) Softic, L.; Brillet, R.; Berry, F.; Ahnou, N.; Nevers, Q.; Morin-Dewaele, M.; Hamadat, S.; Bruscella, P.; Fourati, S.; Pawlotsky, J.-M.; Ahmed-Belkacem, A. Inhibition of SARS-CoV-2 Infection by the Cyclophilin Inhibitor Alisporivir (Debio 025). *Antimicrob Agents Chemother* **2020**, *64* (7). https://doi.org/10.1128/AAC.00876-20.

(45) Uchino, Y.; Woodward, A. M.; Mauris, J.; Peterson, K.; Verma, P.; Nilsson, U. J.; Rajaiya, J.; Argüeso, P. Galectin-3 Is an Amplifier of the Interleukin-1 *β* -Mediated Inflammatory Response in Corneal Keratinocytes. *Immunology* **2018**, *154* (3), 490–499. https://doi.org/10.1111/imm.12899.

(46) Wang, W.-H.; Lin, C.-Y.; Chang, M. R.; Urbina, A. N.; Assavalapsakul, W.; Thitithanyanont, A.; Chen, Y.-H.; Liu, F.-T.; Wang, S.-F. The Role of Galectins in Virus Infection - A Systemic Literature Review. *Journal of Microbiology, Immunology and Infection* **2020**, *53* (6), 925–935. https://doi.org/10.1016/j.jmii.2019.09.005.

(47) Caniglia, J. L.; Guda, M. R.; Asuthkar, S.; Tsung, A. J.; Velpula, K. K. A Potential Role for Galectin-3 Inhibitors in the Treatment of COVID-19. *PeerJ* **2020**, *8*, e9392. https://doi.org/10.7717/peerj.9392.

(48) Pirone, L.; Del Gatto, A.; Di Gaetano, S.; Saviano, M.; Capasso, D.; Zaccaro, L.; Pedone, E. A Multi-Targeting Approach to Fight SARS-CoV-2 Attachment. *Front. Mol. Biosci.* **2020**, *7*, 186. https://doi.org/10.3389/fmolb.2020.00186.

(49) Lenza, M. P.; Oyenarte, I.; Diercks, T.; Quintana, J. I.; Gimeno, A.; Coelho, H.; Diniz, A.; Peccati, F.; Delgado, S.; Bosch, A.; Valle, M.; Millet, O.;

Abrescia, N. G. A.; Palazón, A.; Marcelo, F.; Jiménez-Osés, G.; Jiménez-Barbero, J.; Ardá, A.; Ereño-Orbea, J. Structural Characterization of N-Linked Glycans in the Receptor Binding Domain of the SARS-CoV-2 Spike Protein and Their Interactions with Human Lectins. *Angew. Chem. Int. Ed.* **2020**, *59* (52), 23763–23771. https://doi.org/10.1002/anie.202011015.

3.3 The interaction between the F55 virus-encoded transcription regulator and the RadA host recombinase reveals a common strategy in Archaea and Bacteria to sense the UV-induced damage to the host DNA

3.3.1 Introduction

Sulfolobus spindle-shaped virus 1 (SSV1) is an Archaeal virus, well known as extremophiles, that can thrive in some of the most extreme and inhospitable environments. SSV1 was initially isolated from its natural host (Saccharolobus shibatae B12) in a (>70°C) hot spring in Beppu, Japan[1], but it has also been shown to infect the archaeon Saccharolobus solfataricus isolated from a solfataric field near Naples, Italy[2]. To date, SSV1 is the only UV-inducible member of the *Fuselloviridae* family, comprising 9 members (Table 3.3.1)[3,4,5].

Virus name NCBI number	Sampling site	Genome size (bp)	NCBI number	Reference
SSV1	Japan	15,465	NC_001338	Palm et al. 1991
SSV2	Iceland	14,796	NC_005265	Stedman et al. 2003
SSV4	Iceland	15,135	EU030938	Redder et al. 2009
SSV5	Iceland	15,330	EU030939	Redder et al. 2009
SSV6	Iceland	15,684	NC_013587	Redder et al. 2009
SSV7	Iceland	17,602	NC_013588	Redder et al. 2009
SSV8	USA	16,473	NC_005360	Wiedenheft et al. 2004

SSV9	Kamchatka	17,385	NC_005361	Wiedenheft et al. 2004
ASV1	USA	24,186	NC_013585	Redder et al. 2009

Table 3.3.1 Features of all known *Fuselloviridae.*

The complete sequence of its genome was determined in the '90s[6], after which SSV1 promptly became a model to study the genetics of archaeal viruses[7,8,9,10]. For instance, transcription analyses of fuselloviruses and other episomic genetic elements have laid the basis for the understanding of how gene expression is regulated in archaeal microorganisms[11,12]. However, the lack of sequence homology between the SSV1-encoded proteins and those already characterized in other viruses, has hindered the *in silico* identification of the SSV1 transcription regulators and a reliable prediction of their function[13,14]. Therefore, biochemical studies[15,16,17] are essential to reveal the physiological role of these viral proteins as well as some unexpected[18,19] or cryptic[20,21,22] properties.

Upon the infection of the host cell, one copy of the SSV1 genome (Fig. 3.3.1, panel B) is site-specifically integrated into the host chromosome (provirus)[23], whereas about 5-6 additional copies are maintained as episomes[24]. Once the infection is established, the virus expresses a minimal set of genes that are required for the replication and packaging of its genome into virus particles. These genes include those encoding for: i) the structural proteins (VP1, VP2, and VP3) which form the spindle-shaped viral particles (Fig. 3.3.1, panel A), ii) the integrase (D335), iii) the transcription repressor F55 as well as iv) proteins of unknown functions (A291 and C124)[25].

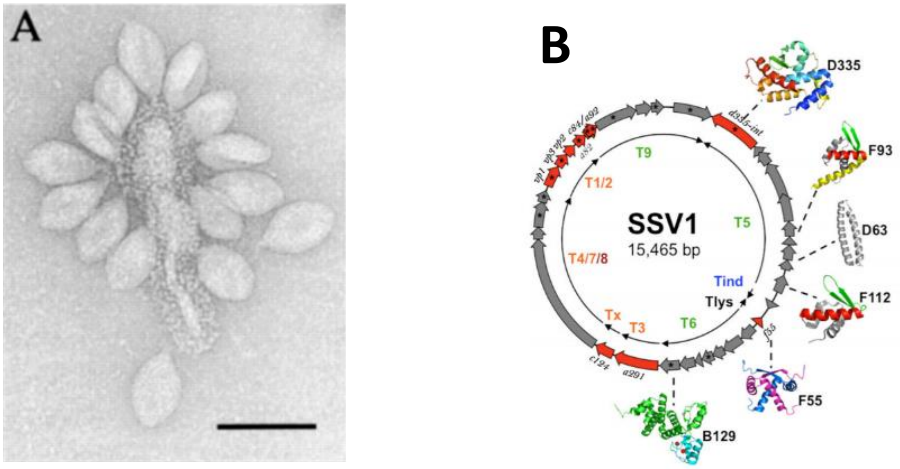

Fig 3.3.1 Panel A: electron microscopy images of *Fuselloviridae* virions displaying the typical lemon-/spindle-shaped morphology[26]. Panel B: genomic map of SSV1 and the three-dimensional structures of SSV1 proteins (determined for D63, F93, F112, B129, and D335-int C-terminal domain and predicted for F55) are shown[3].

Unlike λ lysogeny, in which the phage DNA is present in the host cells only as a provirus and the only viral gene expressed is cI, in the SSV1 lysogeny (better defined as carrier state), the SSV1 provirus coexists with some episomal copies, and a constitutive extrusion of the viral particles occurs without causing cell lysis[27]. The F55 regulator binds, in a concentration-dependent manner, to tandem repeat sequences that overlap the transcription start sites and/or the B recognition element (BRE) within the promoter regions of: i) the early transcripts T_5 and T_6, ii) the UV-inducible T_{ind} as well as iii) the F55 transcript (named T_{lys}) (Fig. 3.3.2).

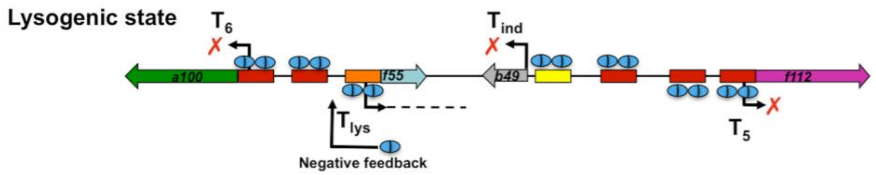

Fig. 3.3.2 In the lysogenic state, F55 (cyan ovals) binds as dimers to the target sequences in the promoters of T_5, T_6, T_{ind} (red and yellow boxes) as well as to its promoter (orange box). As shown by red crosses, transcription of T_5, T_6, and T_{ind} is locked, while the expression of its gene is progressively turned off following a negative feedback control[3].

Sequence analysis of these repeats revealed a 22-bp consensus sequence (5-ATAGATAGAGTATAGATAGAGT-3). A dimer of F55 interacts with a minimal binding site of 11-bp (5-ATAGATAGAGT-3) to form a stable dsDNA-protein complex[24]. The transcription of $T_{1/2}$ (VP1, VP2, and VP3), T_3 (A291), and T_x (C124) during the carrier state is in accordance with the absence of F55 binding sites in the promoter region of these transcripts[25]. Moreover, once F55 has reached a high intracellular concentration, it is also able to bind to target sites located in the promoter of the transcript T_{lys}, thus repressing its expression through a negative feedback mechanism (Fig. 3.3.2)[27]. Accordingly, T_{lys} is strongly downregulated during the carrier state and reaches its lowest expression level in the late-stationary phase of the host cell growth[24]. Therefore, F55 acts as a key regulator involved in the maintenance of the SSV1 carrier state, likewise the CI repressor protein of the λ bacteriophage. However, in the case of λ, CI is the only virus protein expressed in *E. coli* lysogens[28]. This repressor binds to the operators within the promoter regions P_L and P_R; thus, repressing the expression of lytic functions, i.e., cro, cII, and N[29].

As for F55, the CI repressor auto regulates its synthesis by repressing transcription at the promoter P_{RM} when it reaches a high intracellular concentration. This strategy might have been evolved to keep a functional concentration of the lysogenic regulator (CI or F55) and avoid its counterproductive overproduction[27].

Upon exposure to UV light, a temporally coordinated pattern of gene expression is triggered in SSV1 infected cells. This involves the rapid expression of a UV-inducible transcript (T_{ind}), followed by the transcription of the early (T_5, T_6, and T_9), late ($T_{1/2}$, T_3, T_x, and $T_{4/7}$), and late-extended ($T_{4/7/8}$) RNAs[7]. In accordance, F55 has been shown to dissociate first from the UV-inducible T_{ind} promoter (2 h post-UV-exposure) and subsequently from those of the early T_5 and T_6 transcripts (4 h post-UV-exposure) as shown in Fig. 3.3.3[27,30].

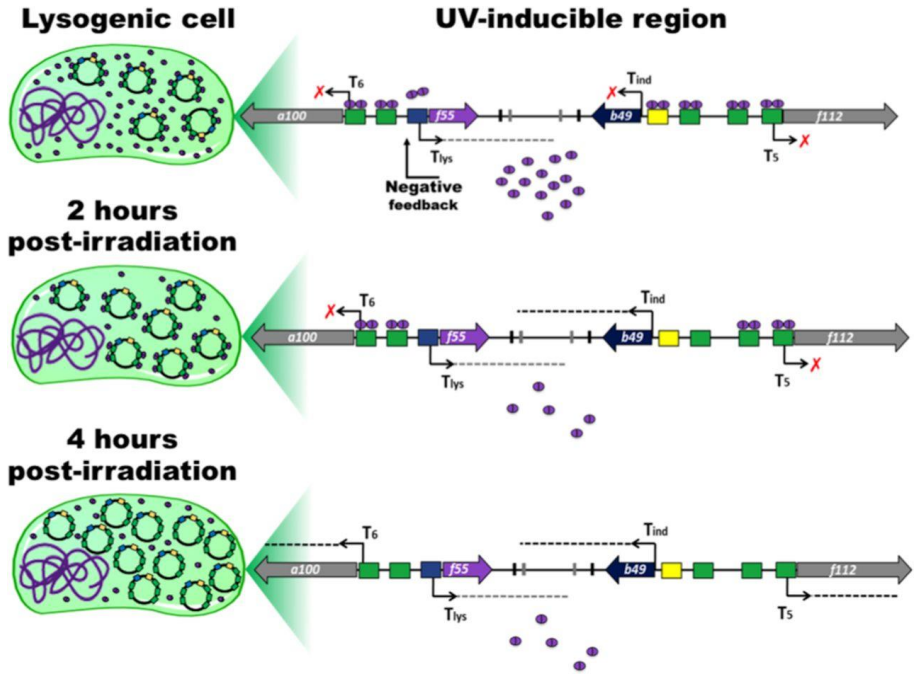

Fig.3.3.3 Schematic representations of the infected cell and the UV-inducible region of the SSV1 genome are presented. The operators recognized by F55 are in green, yellow, and blue. Bent arrows indicate the transcription start sites, and dashed lines represent transcripts. Dimers of F55 are represented by purple ovals. In the lysogenic cell, the amount of F55 is suitable to saturate most of its binding sites and to keep SSV1 in a steady carrier state. At 2 h post-irradiation, a decrease of the F55 concentration and a concurrent increase of the viral copy number led to the dissociation from the lower-affinity operators in the promoter of T_{ind} and T_{lys}. Later, at 4 h post-irradiation, the dilution effect is enhanced by a further accumulation of the viral DNA, which results in the release of the early promoters (i.e., those of T_5 and T_6), thus allowing transcription derepression[27].

This progressive release of the binding sites by F55 is a consequence of two conditions: i) the differential affinity of F55 for the target sequences in the promoter regions of these transcripts (Kd: $T_5 \simeq T_6 < T_{ind}$) and ii) the reduction of the F55:binding sites ratio, upon UV-induced copy number increase of SSV1[24]. This cascade of events is then followed by the onset of the induction of SSV1 genome replication, packaging, and release of virus particles.

Although SSV1 has been extensively characterized over the last three decades, the

regulatory mechanisms governing the switch from the carrier to the induced state are still not completely unraveled. This project, carried out in collaboration with the research group of the Professor Patrizia Contrusi of the Department of Biology of the University "Federico II" of Naples, focused on the investigation of the protein-protein interaction network responsible for the crosstalk between the host *Saccharolobus solfataricus* and the virus SSV1.

3.3.2 Experimental methods

3.3.2.1 Protein identification by mass spectrometry analysis

The molecular mechanisms influenced by F55 were investigated through the identification of host *S. solfataricus* proteins interacting with F55 onto T_6-promoter probe by using an original EMSA-MS approach. Protein crude extract of *S. solfataricus* supplemented with 2.2 µM F55 was incubated in the presence of a T_6-promoter probe. The crude extract incubated only with the oligonucleotide probe was used as control. Sample and control were loaded onto a 10% polyacrylamide native gel, and the band showing an electrophoretic mobility shift, detected by the UV lamp, was excised from the sample lane. A gel band was cut in the control lane at the same migration point.

Gel bands were *in situ* hydrolyzed by trypsin[31], and peptide mixtures were analyzed by nano LC-MS/MS using LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). Peptides were fractionated onto a C18 reverse-phase capillary column (5µm biosphere, 75µm internal diameter, 200mm length) working at 250nL/min flow rate and adopting a step gradient from 10% to 60% of eluent B (0.2% formic acid, 95% acetonitrile LC-MS Grade) over 69 minutes and 60% to 95% over 3 minutes. Mass spectrometric analyses were carried out in data depending acquisition (DDA): from each MS scan, spanning from 400 to 1800 m/z, the five most abundant ions were selected and fragmented. Output data were further processed into mgf file extension to perform protein identification by using

Mascot licensed software (Matrix Science Boston, USA) and a protein in-house database containing *S. solfataricus* and SSV1 protein sequences in FASTA format. Proteins identification was carried out by using 10 ppm as peptides mass tolerance for MS and 0.6 Da for MS/MS search; carbamidomethyl (C) as fixed modification and Gln->pyro-Glu (N-term Q), Oxidation (M), Pyro-carbamidomethyl (N-term C) as variable modifications.

3.3.3 Results and Discussion

3.3.3.1 Isolation and identification of F55 protein partners

To find F55 interactors in the crude extracts of *S. solfataricus* cells, we have set up an original EMSA-MS assay: the recombinant F55 was added to the crude extract of *S. solfataricus* cells, and the T_6-promoter was used as a "bait" for the dsDNA-F55-interactors complexes that were subsequently separated on a native electrophoretic gel. Surprisingly, the addition of an increasing amount of F55 in the crude extract led to the appearance of a fast-migrating (FM) specific complex at the expense of a slow-migrating (SM) unspecific complex (Fig. 3.3.7 panel A) already present at 0 concentration of F55. To be sure of the presence and position of F55 along the gel, a WEMSA using total IgG against F55 was performed. By comparing the western blot profile containing the F55 positive band with the preparative EMSA gel, we could unambiguously recognize in the FM band the solely containing F55 (Fig. 3.3.7 panel B).

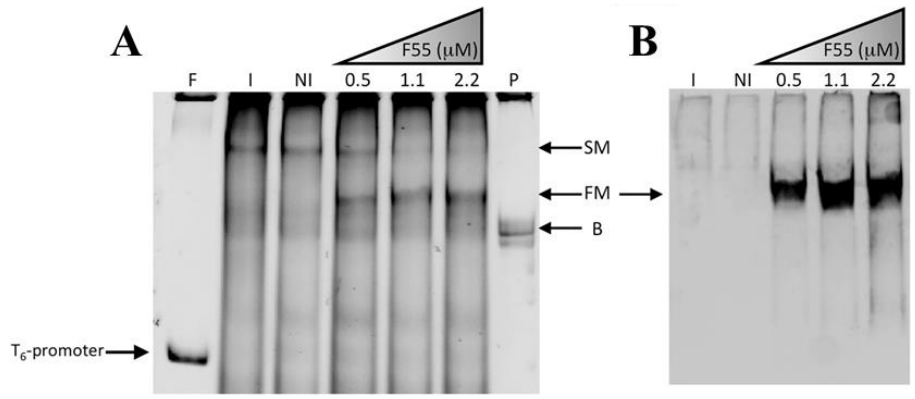

Fig 3.3.4 Fishing of molecular interactors of F55 through WEMSA. Panel A: EMSA assay using the T_6-promoter probe and *S. Solfataricus* protein extracts from SSV1-infected (I) and not infected (NI) cells. Increasing amounts of F55 (0.5–2.2 µM) were added to the crude extract. The shifted signal B is the F55/dsDNA complex formed when the recombinant F55 is mixed with the probe T_6-promoter (positive control, P). The shifted signal FM corresponds to a specific protein/ DNA complex formed when F55 is supplemented to the protein extract from *S. Solfataricus* cells, whereas SM is detectable when crude extracts from both infected and not infected *S. Solfataricus* cells are mixed with the probe T_6-promoter. Panel B: WEMSA assay using an anti-F55 antibody. The only detectable complex containing the F55/interactors/dsDNA was (FM), which was identified using chemiluminescence imaging. A re- presentative result of three independent experiments is shown.

This band was then excised from the preparative EMSA gel (Fig. 3.3.5, black frame lane 2), *in situ* hydrolyzed by trypsin and the proteins identified by nanoLC-MS/MS strategy. To discriminate between the proteins belonging to the dsDNA-F55 complex(es) from those randomly migrating in the same region, a gel band with the same electrophoretic mobility was excised from the control lane 3 (Fig. 3.3.5, black frame lane 3) containing the oligonucleotide probe and *S. solfataricus* protein extract but not F55.

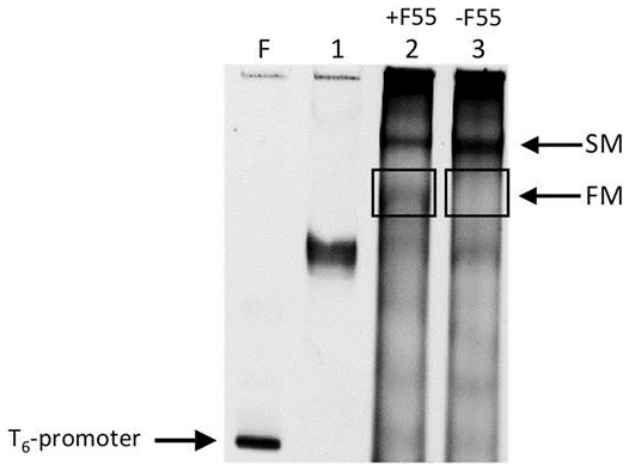

Fig 3.3.5 Preparative EMSA gel for the identification of the F55 interactors. The band labeled as FM (black framed) in lane 2 (containing F55), and the control band with the same electrophoretic mobility from lane 3 were cut and processed for protein identification.

In the FM band, besides F55, 23 proteins, absent in the control, were identified, suggesting that they bind T_6-promoter probe only in the presence of F55 (Table 3.3.2).

Protein name	Gene name	Uniprot code	Peptides	MW (kDa)	% Sequence coverage
RHH transcription regulator (F55)	LysR	R7RTW8	5	6	74
3-Ketoacyl-ACP reductase	fabG	A0A0E3K740	9	27	44
Glutamate dehydrogenase	gdhA-4	Q97WS2	9	46	30
Thermosome subunit	SULA_1299	A0A0E3K8S3	18	60	28
Universal stress protein A (UspA)	SSOP1_3274	A0A0E3K798	2	14	26
Succinate dehydrogenase	sdhA	A0A0E3MCA7	12	62	24
Glutamine synthetase	SSOP1_0348	A0A0E3K7R3	10	53	24

50S ribosomal protein L14e	rpl14e	A0A0E3K7T1	2	11	22
Acyl-CoA dehydrogenase	SSOP1_2886	A0A0E3MF90	5	44	21
Adenylate kinase	adkA	A0A0E3MJD3	3	21	20
Cupin	SSOP1_1694	A0A0E3K920	2	15	16
50S ribosomal protein L4	rpl4lp	A0A0E3GWP6	4	29	15
Isocitrate dehydrogenase, probable (Idh)	idh	Q97WN0	3	46	13
Uncharacterized protein	SULA_2252	A0A0E3JVF8	2	20	12
Allantoate amidohydrolase	SSOP1_2839	A0A0E3GW22	4	45	11
Tyrosine—tRNA ligase	tyrS	A0A0E3GWC2	3	41	11
DNA repair and recombination protein (RadA)	radA	A0A0E3MIV8	2	36	11
Ornithine carbamoyltransferase	SSOP1_0912	A0A0E3MGK7	3	34	10
Cytidyltransferase	SSOP1_0984	A0A0E3MD44	2	26	9
Transcriptional activator, TenA family	Ssol_2802	D0KPY0	2	24	8
UPF0173 metal-dependent hydrolase SSOP1_0092	SSOP1_0092	A0A0E3GT97	2	25	7
S-adenosylmethionine synthase	mat	A0A0E3MIU3	3	45	6
Phosphoglycerate kinase	pgk	A0A0E3K5Z0	2	45	6
FAD-linked oxidase	SSOP1_3253	A0A0E3GUQ6	2	51	5

Table 3.3.2 *S. Solfataricus* proteins binding the T_6 promoter in the presence of F55. For each protein the protein and gene name, the Uniprot code, the peptide identified by mass spectrometry, the molecular weight, and the percentage of sequence coverage are reported.

Among the F55 protein partners, the DNA repair and recombination protein (RadA; UniProt code: A0A0E3MIV8) caught our attention. RadA is the archaeal homolog of the bacterial RecA, and it is a recombinase enzyme essential for genome stability[32]. In particular, it promotes the repair of double-stranded DNA breaks and the rescue of collapsed DNA replication forks[33].

The DNA-binding activity of RadA was tested by band-shift assays using the double-stranded T_6-promoter as a probe. As shown (Fig. 3.3.6, panel A), unlike F55, the protein RadA alone does not cause a band shift. Therefore, the interaction observed through mass spectrometry between RadA and T_6-promoter was not direct and might be mediated by F55. To test this hypothesis, a band shift assays with both proteins was performed, showing a slower band in the lane containing both proteins in respect to the control, containing the probe incubated with F55 (Fig. 3.3.6, panel B). This finding confirmed that F55 acts as a molecular bridge for the interaction between RadA and the dsDNA. As above-mentioned, upon UV exposure, RadA mediates the repair of double-stranded DNA breaks and the rescue of collapsed DNA replication forks by interacting with ssDNA regions and ATP. Therefore, we have added ssDNA as a competitor to the mixture containing F55, RadA, and the T_6-promoter probe. The increasing amounts of ssDNA cause the appearance of the T_6-F55 complex at the expense of the T_6-F55-RadA band as a consequence of the recruitment of RadA onto the ssDNA to form the activated RadA* (i.e., RadA bound to ATP and ssDNA). The formation of the fast-migrating complex (Fig. 3.3.6 panel C, FM) clearly indicates that F55 does not interact with the activated RadA*. Noteworthy, unlike one would expect by analogy with the bacteriophage λ, the presence of RadA* does not apparently induce the proteolysis of F55 *in vitro*. Altogether, these evidences indicate that although the archaeal molecular partners of response to DNA damage (RadA-F55) are functionally homologs to the bacterial counterparts (RecA-CI), the host-viral mechanisms underlying this process are different.

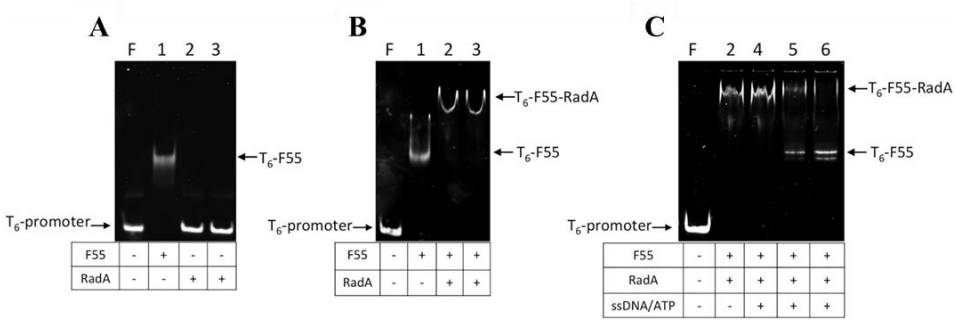

Fig. 3.3.6 Analysis of the interaction between F55 and RadA. The band-shifts were performed using the T_6-promoter dsDNA probe (22 nM) and a fixed concentration of F55 (2.2 μM). Panel A: two different RadA concentrations (2.2 μM lane 2 and 4.4 μM lane 3) were tested. Panel B: molar ratios of F55:RadA of 1:1 and 1:2 were employed in the lane 2 and 3, respectively. Panel C: RadA (2.2 μM) is displaced from the complex T6-F55-RadA by adding dsDNA:ssDNA at a ratio of 1:1 (lane 4), 1:10 (lane 5) and 1:100 (lane 6). The F55-DNA complex, as well as the free probe (T_6-promoter), are indicated by arrows.

Therefore, these results highlight a strategic crosstalk between a host-encoded protein, which is involved in the surveillance of the genome integrity, and a virus protein, which regulates the transition from the carrier to the induced state of the SSV1 life cycle. Although our data do not demonstrate unequivocally that the interaction of F55 with its operator sites is destabilized by the detachment of RadA as a consequence of its recruitment at ssDNA regions, we propose that the equilibrium of association/dissociation of RadA from F55 is the molecular sensor evolved by SSV1 to detect the compromised viability of the host cell (Fig. 3.3.7). In our model, the carrier state the concentration of F55 is optimal to saturate all the target sites in the UV-induced region of the SSV1 genome[24,27]; thus, repressing the expression of the early transcripts. RadA interacts with F55 bound to the dsDNA. Upon irradiation, UV-induced host DNA damage causes the accumulation of stalled replication forks as a consequence of DNA unrepaired lesions or lesions undergoing repair[34]. The increase of ssDNA regions fosters the recruitment of RadA that forms nucleoprotein filaments on the exposed ssDNA. This event results in the progressive release of RadA from F55 and, later on, in

the dissociation of F55 from its target sites with the consequent activation of the temporary coordinated transcription pattern of SSV1.

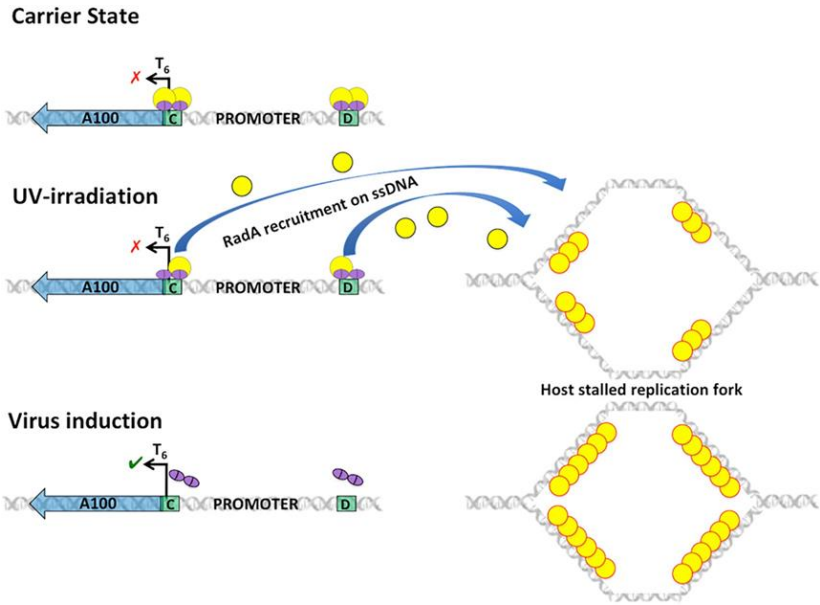

Fig. 3.3.7 Model that describes how SSV1 senses the host DNA damage.

This kind of molecular crosstalk between a key virus transcription regulator and one of the main players of the host DNA repair mechanisms has already been described for the phage λ that infects *E. coli*. In this system, the CI transcription regulator inhibits the expression of all λ genes in *E. coli* lysogens and this transcriptional block is released upon CI autoproteolysis that, in turn, triggers the switch from the lysogenic to the lytic cycle. The CI proteolytic cleavage is induced by the direct interaction with the activated RecA* (i.e., RecA bound to ATP and ssDNA), which is formed as a consequence of UV-induced damage of the host DNA. In the SSV1-Saccharolobus system, F55 was found to interact with the not-activated RadA, and the formation of the RadA* does not apparently induce F55 proteolysis. Therefore, although the λ and SSV1 responses to the host DNA damage seems to be mechanistically different, both viruses have evolved a

transcriptional repressor able to interact with a host regulator of the DNA repair pathways.

In the case of the bacteriophage λ, the transcriptional derepression of the lytic genes is an irreversible process because the cleavage of CI leads to its complete inactivation. Besides the control at the protein level, the switch from lysogeny to the lytic cycle is amplified by a strict transcription regulation of CI (repression)[35] and RecA (induction)[36]. Conversely, as it concerns SSV1, we did not observe such a tightly regulated response neither at protein nor at transcription level for both F55[27] and RadA[37]. Instead, the switch from the carrier to the induction of the viral life cycle is characterized by a fluctuation in the levels of the bound F55 to DNA and its free form[27]. From an evolutionary point of view, this kind of host-virus "communication" that is not based on an all-or-none response mirrors the non-lytic nature of this fusellovirus. Indeed, unlike the case of λ, the SSV1 response to UV irradiation does not impose a metabolic burden on the host physiology that evolves into cell lysis but consists in the amplification of the viral particles production, possibly through the dynamic variation of the intracellular concentration of the dsDNA-F55-RadA and dsDNA-F55 complexes. Furthermore, the fairly constant expression levels of F55 and RadA transcripts, even upon UV exposure, indicate that the system is self-consistent in restoring the carrier state once the host has recovered from the stress, as one could expect in the case of the harmonic host-virus coexistence between SSV1 and Saccharolobus. Noteworthy, the absence of CRISPR spacers for SSV1 in the Saccharolobus genome supports the hypothesis of a mutual beneficial interaction in this host/virus system.

3.3.4 Conclusions

Using an original EMSA-MS approach, we unraveled the biological function of the SSV1 protein F55 by identifying its interactors. Among all the interactions

with RadA was investigated. Functional experiments allowed us to propose a model explaining the effect of the interaction of the F55-RadA complex on the T_6 promoter. Of course, we cannot exclude that the switch from the carrier to the induction phase of SSV1 relies on an even more complex interactome involving other factors that complement the primary signal of RadA detachment. However, the proteomic approach turned out to be powerful in straightforward identification of the main F55 interacting factors and, therefore, in unraveling the critical components of the virus/host cross-talk.

The results discussed (including figures and tables) have been published in the following article: Fusco, S.; Aulitto, M.; **Iacobucci, I.**; Crocamo, G.; Pucci, P.; Bartolucci, S.; Monti, M.; Contursi, P. The Interaction between the F55 Virus-Encoded Transcription Regulator and the RadA Host Recombinase Reveals a Common Strategy in Archaea and Bacteria to Sense the UV-Induced Damage to the Host DNA. *Biochim Biophys Acta Gene Regul Mech* **2020**, *1863* (5), 194493. https://doi.org/10.1016/j.bbagrm.2020.194493.

3.3.5 References

(1) Martin, A.; Yeats, S.; Janekovic, D.; Reiter, W. D.; Aicher, W.; Zillig, W. SAV 1, a Temperate u.v.-Inducible DNA Virus-like Particle from the Archaebacterium Sulfolobus Acidocaldarius Isolate B12. *EMBO J* **1984**, *3* (9), 2165–2168.

(2) Contursi, P.; Pisani, F. M.; Grigoriev, A.; Cannio, R.; Bartolucci, S.; Rossi, M. Identification and Autonomous Replication Capability of a Chromosomal Replication Origin from the Archaeon Sulfolobus Solfataricus. *Extremophiles* **2004**, *8* (5), 385–391. https://doi.org/10.1007/s00792-004-0399-y.

(3) Contursi, P.; Fusco, S.; Cannio, R.; She, Q. Molecular Biology of Fuselloviruses and Their Satellites. *Extremophiles* **2014**, *18* (3), 473–489. https://doi.org/10.1007/s00792-014-0634-0.

(4) Prangishvili, D.; Bamford, D. H.; Forterre, P.; Iranzo, J.; Koonin, E. V.; Krupovic, M. The Enigmatic Archaeal Virosphere. *Nat Rev Microbiol* **2017**, *15* (12), 724–739. https://doi.org/10.1038/nrmicro.2017.125.

(5) Krupovic, M.; Cvirkaite-Krupovic, V.; Iranzo, J.; Prangishvili, D.; Koonin, E. V. Viruses of Archaea: Structural, Functional, Environmental and Evolutionary Genomics. *Virus Res* **2018**, *244*, 181–193. https://doi.org/10.1016/j.virusres.2017.11.025.

(6) Palm, P.; Schleper, C.; Grampp, B.; Yeats, S.; McWilliam, P.; Reiter, W. D.; Zillig, W. Complete Nucleotide Sequence of the Virus SSV1 of the Archaebacterium Sulfolobus Shibatae. *Virology* **1991**, *185* (1), 242–250. https://doi.org/10.1016/0042-6822(91)90771-3.

(7) Fröls, S.; Gordon, P. M. K.; Panlilio, M. A.; Schleper, C.; Sensen, C. W. Elucidating the Transcription Cycle of the UV-Inducible Hyperthermophilic Archaeal Virus SSV1 by DNA Microarrays. *Virology* **2007**, *365* (1), 48–59. https://doi.org/10.1016/j.virol.2007.03.033.

(8) Iverson, E. A.; Goodman, D. A.; Gorchels, M. E.; Stedman, K. M. Genetic Analysis of the Major Capsid Protein of the Archaeal Fusellovirus SSV1: Mutational Flexibility and Conformational Change. *Genes (Basel)* **2017**, *8* (12). https://doi.org/10.3390/genes8120373.

(9) Stedman, K. M.; Schleper, C.; Rumpf, E.; Zillig, W. Genetic Requirements for the Function of the Archaeal Virus SSV1 in Sulfolobus

Solfataricus: Construction and Testing of Viral Shuttle Vectors. *Genetics* **1999**, *152* (4), 1397–1405.

(10) Iverson, E. A.; Goodman, D. A.; Gorchels, M. E.; Stedman, K. M. Extreme Mutation Tolerance: Nearly Half of the Archaeal Fusellovirus Sulfolobus Spindle-Shaped Virus 1 Genes Are Not Required for Virus Function, Including the Minor Capsid Protein Gene Vp3. *J Virol* **2017**, *91* (10). https://doi.org/10.1128/JVI.02406-16.

(11) Reiter, W. D.; Palm, P.; Yeats, S.; Zillig, W. Gene Expression in Archaebacteria: Physical Mapping of Constitutive and UV-Inducible Transcripts from the Sulfolobus Virus-like Particle SSV1. *Mol Gen Genet* **1987**, *209* (2), 270–275. https://doi.org/10.1007/BF00329653.

(12) Prato, S.; Vitale, R. M.; Contursi, P.; Lipps, G.; Saviano, M.; Rossi, M.; Bartolucci, S. Molecular Modeling and Functional Characterization of the Monomeric Primase-Polymerase Domain from the Sulfolobus Solfataricus Plasmid PIT3. *FEBS J* **2008**, *275* (17), 4389–4402. https://doi.org/10.1111/j.1742-4658.2008.06585.x.

(13) Contursi, P.; Fusco, S.; Limauro, D.; Fiorentino, G. Host and Viral Transcriptional Regulators in Sulfolobus: An Overview. *Extremophiles* **2013**, *17* (6), 881–895. https://doi.org/10.1007/s00792-013-0586-9.

(14) Contursi, P.; D'Ambrosio, K.; Pirone, L.; Pedone, E.; Aucelli, T.; She, Q.; De Simone, G.; Bartolucci, S. C68 from the Sulfolobus Islandicus Plasmid-Virus PSSVx Is a Novel Member of the AbrB-like Transcription Factor Family. *Biochem J* **2011**, *435* (1), 157–166. https://doi.org/10.1042/BJ20101334.

(15) Contursi, P.; Farina, B.; Pirone, L.; Fusco, S.; Russo, L.; Bartolucci, S.; Fattorusso, R.; Pedone, E. Structural and Functional Studies of Stf76 from the Sulfolobus Islandicus Plasmid-Virus PSSVx: A Novel Peculiar Member of the Winged Helix-Turn-Helix Transcription Factor Family. *Nucleic Acids Res* **2014**, *42* (9), 5993–6011. https://doi.org/10.1093/nar/gku215.

(16) Kraft, P.; Kümmel, D.; Oeckinghaus, A.; Gauss, G. H.; Wiedenheft, B.; Young, M.; Lawrence, C. M. Structure of D-63 from Sulfolobus Spindle-Shaped Virus 1: Surface Properties of the Dimeric Four-Helix Bundle Suggest an Adaptor Protein Function. *J Virol* **2004**, *78* (14), 7438–7442. https://doi.org/10.1128/JVI.78.14.7438-7442.2004.

(17) Kraft, P.; Oeckinghaus, A.; Kümmel, D.; Gauss, G. H.; Gilmore, J.; Wiedenheft, B.; Young, M.; Lawrence, C. M. Crystal Structure of F-93 from Sulfolobus Spindle-Shaped Virus 1, a Winged-Helix DNA Binding Protein. *J Virol* **2004**, *78* (21), 11544–11550. https://doi.org/10.1128/JVI.78.21.11544-11550.2004.

(18) Quax, T. E. F.; Daum, B. Structure and Assembly Mechanism of Virus-Associated Pyramids. *Biophys Rev* **2018**, *10* (2), 551–557. https://doi.org/10.1007/s12551-017-0357-4.

(19) Quax, T. E. F.; Lucas, S.; Reimann, J.; Pehau-Arnaudet, G.; Prevost, M.-C.; Forterre, P.; Albers, S.-V.; Prangishvili, D. Simple and Elegant Design of a Virion Egress Structure in Archaea. *Proc Natl Acad Sci U S A* **2011**, *108* (8), 3354–3359. https://doi.org/10.1073/pnas.1018052108.

(20) Notomista, E.; Falanga, A.; Fusco, S.; Pirone, L.; Zanfardino, A.; Galdiero, S.; Varcamonti, M.; Pedone, E.; Contursi, P. The Identification of a Novel Sulfolobus Islandicus CAMP-like Peptide Points to Archaeal Microorganisms as Cell Factories for the Production of Antimicrobial Molecules. *Microb Cell Fact* **2015**, *14*, 126. https://doi.org/10.1186/s12934-015-0302-9.

(21) Gaglione, R.; Pirone, L.; Farina, B.; Fusco, S.; Smaldone, G.; Aulitto, M.; Dell'Olmo, E.; Roscetto, E.; Del Gatto, A.; Fattorusso, R.; Notomista, E.; Zaccaro, L.; Arciello, A.; Pedone, E.; Contursi, P. Insights into the Anticancer Properties of the First Antimicrobial Peptide from Archaea. *Biochim Biophys Acta Gen Subj* **2017**, *1861* (9), 2155–2164. https://doi.org/10.1016/j.bbagen.2017.06.009.

(22) Roscetto, E.; Contursi, P.; Vollaro, A.; Fusco, S.; Notomista, E.; Catania, M. R. Antifungal and Anti-Biofilm Activity of the First Cryptic Antimicrobial Peptide from an Archaeal Protein against Candida Spp. Clinical Isolates. *Sci Rep* **2018**, *8* (1), 17570. https://doi.org/10.1038/s41598-018-35530-0.

(23) Schleper, C.; Kubo, K.; Zillig, W. The Particle SSV1 from the Extremely Thermophilic Archaeon Sulfolobus Is a Virus: Demonstration of Infectivity and of Transfection with Viral DNA. *Proc Natl Acad Sci U S A* **1992**, *89* (16), 7645–7649. https://doi.org/10.1073/pnas.89.16.7645.

(24) Fusco, S.; She, Q.; Bartolucci, S.; Contursi, P. Tlys, a Newly Identified Sulfolobus Spindle-Shaped Virus 1 Transcript Expressed in the Lysogenic State, Encodes a DNA-Binding Protein Interacting at the Promoters of the Early Genes.

Journal of Virology **2013**, *87* (10), 5926–5936. https://doi.org/10.1128/JVI.00458-13.

(25) Fusco, S.; Liguori, R.; Limauro, D.; Bartolucci, S.; She, Q.; Contursi, P. Transcriptome Analysis of Sulfolobus Solfataricus Infected with Two Related Fuselloviruses Reveals Novel Insights into the Regulation of CRISPR-Cas System. *Biochimie* **2015**, *118*, 322–332. https://doi.org/10.1016/j.biochi.2015.04.006.

(26) Stedman, K. M.; Schleper, C.; Rumpf, E.; Zillig, W. Genetic Requirements for the Function of the Archaeal Virus SSV1 in Sulfolobus Solfataricus</Em>: Construction and Testing of Viral Shuttle Vectors. *Genetics* **1999**, *152* (4), 1397.

(27) Fusco, S.; She, Q.; Fiorentino, G.; Bartolucci, S.; Contursi, P. Unravelling the Role of the F55 Regulator in the Transition from Lysogeny to UV Induction of Sulfolobus Spindle-Shaped Virus 1. *J Virol* **2015**, *89* (12), 6453–6461. https://doi.org/10.1128/JVI.00363-15.

(28) Lewis, D.; Le, P.; Zurla, C.; Finzi, L.; Adhya, S. Multilevel Autoregulation of Repressor Protein CI by DNA Looping in Vitro. *Proceedings of the National Academy of Sciences* **2011**, *108* (36), 14807–14812. https://doi.org/10.1073/pnas.1111221108.

(29) Schubert, R. A.; Dodd, I. B.; Egan, J. B.; Shearwin, K. E. Cro's Role in the CI Cro Bistable Switch Is Critical for 's Transition from Lysogeny to Lytic Development. *Genes & Development* **2007**, *21* (19), 2461–2472. https://doi.org/10.1101/gad.1584907.

(30) Fusco, S.; Aulitto, M.; Bartolucci, S.; Contursi, P. A Standardized Protocol for the UV Induction of Sulfolobus Spindle-Shaped Virus 1. *Extremophiles* **2015**, *19* (2), 539–546. https://doi.org/10.1007/s00792-014-0717-y.

(31) Zanca, C.; Cozzolino, F.; Quintavalle, C.; Di Costanzo, S.; Ricci-Vitiani, L.; Santoriello, M.; Monti, M.; Pucci, P.; Condorelli, G. PED Interacts with Rac1 and Regulates Cell Migration/Invasion Processes in Human Non-Small Cell Lung Cancer Cells. *J Cell Physiol* **2010**, *225* (1), 63–72. https://doi.org/10.1002/jcp.22197.

(32) Komori, K.; Miyata, T.; Daiyasu, H.; Toh, H.; Shinagawa, H.; Ishino, a Y. Domain Analysis of an Archaeal RadA Protein for the Strand Exchange

Activity. *J Biol Chem* **2000**, *275* (43), 33791–33797. https://doi.org/10.1074/jbc.M004556200.

(33) Hogrel, G.; Lu, Y.; Alexandre, N.; Bossé, A.; Dulermo, R.; Ishino, S.; Ishino, Y.; Flament, D. Role of RadA and DNA Polymerases in Recombination-Associated DNA Synthesis in Hyperthermophilic Archaea. *Biomolecules* **2020**, *10* (7). https://doi.org/10.3390/biom10071045.

(34) *DNA Repair and Mutagenesis*, 2nd ed.; Friedberg, E. C., Friedberg, E. C., Eds.; ASM Press: Washington, D.C, 2006.

(35) Patel, M.; Jiang, Q.; Woodgate, R.; Cox, M. M.; Goodman, M. F. A New Model for SOS-Induced Mutagenesis: How RecA Protein Activates DNA Polymerase V. *Crit Rev Biochem Mol Biol* **2010**, *45* (3), 171–184. https://doi.org/10.3109/10409238.2010.480968.

(36) Markham, B. E.; Harper, J. E.; Mount, D. W. Physiology of the SOS Response: Kinetics of LexA and RecA Transcriptional Activity Following Induction. *Mol Gen Genet* **1985**, *198* (2), 207–212. https://doi.org/10.1007/BF00382997.

(37) Sandler, S. J.; Satin, L. H.; Samra, H. S.; Clark, A. J. RecA-like Genes from Three Archaean Species with Putative Protein Products Similar to Rad51 and Dmc1 Proteins of the Yeast Saccharomyces Cerevisiae. *Nucleic Acids Res* **1996**, *24* (11), 2125–2132. https://doi.org/10.1093/nar/24.11.2125.

Chapter 4- Study of metal-based compounds as potential pharmacological drugs

4.1 A structural study of β-lactoglobulin binding to cisplatin and oxaliplatin

4.1.1 Introduction

β-Lactoglobulin is a whey carrier protein of 18.4 kDa[1]. Thanks to its biochemical and biophysical features, including high solubility, abundance, and stability against pepsin,[2] it can be used to prepare micro- or nanoparticles for pharmaceutical and food industries.[3] β-Lactoglobulin is a good system in the fabrication of delivering vehicles with controlled release properties for orally administered bioactive molecules.[4] It has been reported that β-lactoglobulin interacts with metallodrugs.[4,5]

The discovery of cisplatin (CDDP, Fig. 4.1.1) as an anti-tumor agent by Barnett Rosenberg in 1971 and the following Food and Drug Administration (FDA) approval for the treatment of some advanced cancers has opened the doors to the investigation of metal compounds as potential drugs. However, CDDP suffers from some toxic effects, and this prompted the researchers to develop CDDP derivates able to ameliorate its undesired side effects. Several second-generation compounds containing dicarboxylate leaving groups in place of the more labile chloride ions of the parent compound cisplatin were investigated in the 1980s based on the hypothesis that platinum(II) diamine compounds containing more stable leaving groups would retain the desired anticancer properties while imparting lower toxicity and more predictable pharmacokinetics[6,7]. This hypothesis turned out to be correct for carboplatin (CBDCA, Fig. 4.1.1), which was granted FDA approval in 1989, and is now widely used in clinics primarily

in the treatment of ovarian cancer. A prominent third-generation drug as an anticancer agent is represented by oxaliplatin (L-OHP, Fig. 4.1.1), which was demonstrated to perform better than cisplatin while showing reduced side effects in early murine leukemia[8]. The first studies regarding the mechanism of action of CDDP and its derivatives indicated to carry out their antineoplastic effects by forming adducts with DNA.

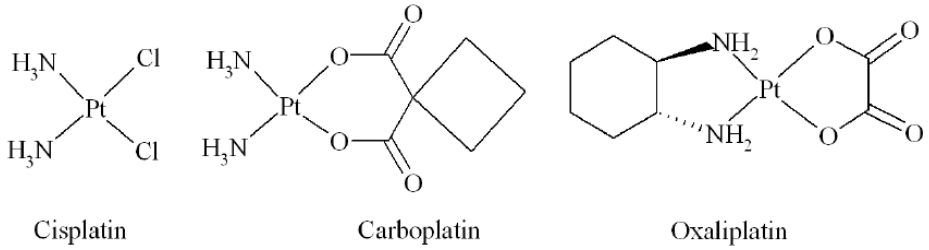

Fig 4.1.1 Chemical structures of Cisplatin (CDDP), Carboplatin (CBDCA), and oxaliplatin (L-OHP).

Beyond their mechanistically relevant DNA interactions, cisplatin, carboplatin, and oxaliplatin have been implicated in the extensive formation of Pt–protein adducts. Though the involvement of these adducts in the mode of action and toxicity of Pt drugs is still unclear, more and more attention is being placed on protein–Pt drug interactions concerning their overall pharmacological and toxicological impact[9,10].

Above cisplatin, β-lactoglobulin is also able to bind transplatin and oxaliplatin[11,12] while it is unable to bind carboplatin[12]; interestingly, in the complex with oxaliplatin, non-covalent interactions (electrostatic interactions, hydrogen bonds, and hydrophobic interactions) between the metallodrug and the protein are supposed to be formed[12]. In this work, carried out in collaboration with Professor Antonello Merlino of the Department of Chemical Sciences at the University of Naples "Federico II", the interactions between β-lactoglobulin and either cisplatin (CDDP) and oxaliplatin have been investigated[7] to shed light on complexes

stoichiometry, Pt-conjugated binding sites on the protein. In particular, we characterized the interactions of CDDP and L-OHP with β-lactoglobulin by using native electrospray ionization mass spectrometry, collecting data over time.

4.1.2 Experimental Methods

4.1.2.1 ESI-MS analysis

β-lactoglobulin reaction with CDDP and L-OHP was carried out in water by mixing them in a 1:5 (protein:CDDP) and 1:3 (protein:L-OHP) molar ratios at 25 °C. A time course was performed acquiring ESI-MS spectra upon 0 h (free protein), 3 h, 9 h, 18 h, 33/36 h, and 72 h of incubation of β-lactoglobulin with metal compounds in order to delineate the time-dependent reaction development. Following dilution in 10 mM ammonium acetate pH 6.8, each mixture was analyzed using a Q-ToF Premier (Waters, Milford, MA, USA) mass spectrometer by direct injection mode at a 10 µl min^{-1} flow rate. The source parameters were set at 3 kV for capillary voltage, 42 kV for cone voltage, and 80 °C for the temperature. The m/z acquisition range spanned from 900 to 3500, and the raw data were processed by the MassLynx 4.1 software (Waters, Milford, MA, USA).

4.1.3 Results and discussion

4.1.3.1 ESI-MS analysis of β-lactoglobulin/CDDP adducts

In order to evaluate the nature of Pt-adducts with the protein and if they increase over time, we have collected electrospray ionization mass spectra of the CDDP/β-lactoglobulin as a function of time. As expected, ESI-MS spectra of metal-free β-lactoglobulin reveal the presence of two species with molecular weights of 18362.89 ± 0.79 Da and 18276.39 ± 0.47 Da, for β-lactoglobulin variants A and B, respectively[13] (Fig. 4.1.2, panel A). Both variants are able to bind CDDP, as demonstrated by the presence in the spectra of the peaks attributable to the formation of adducts for both proteins at all incubation times (Table 4.1.1). At the

shortest reaction time (3 h), an analysis of the ESI-MS spectrum reveals the presence of metal-free protein and the formation of an adduct in which β-lactoglobulin variant A binds a [$Pt(NH_3)_2Cl^+$] fragment. This indicates that after 3h, CDDP binds the protein in a monodentate mode after releasing one Cl^- ion (Fig. 4.1.2, panel B). After 9 h, for both β-lactoglobulin variants, the spectra show the presence of a mixture of this latter form, of metal-free protein, of an adduct with a [$Pt(NH_3)_2OH^{2+}$] fragment bound to the protein and of a bidentate adduct, in which both Cl^- ions are released by CDDP.

Following 18 h of incubation, the relative amount of adducts increased compared to the metal-free proteins (for molecular weights, see Table 4.1.1). Mass spectra acquired after 36 h and 72 h of incubation display the presence of the adducts of β-lactoglobulin variants with fragments from one and two molecules of CDDP, while only at the longest incubation time adducts formed by the protein coupled to fragments from three CDDP molecules were present (Table 4.1.1).

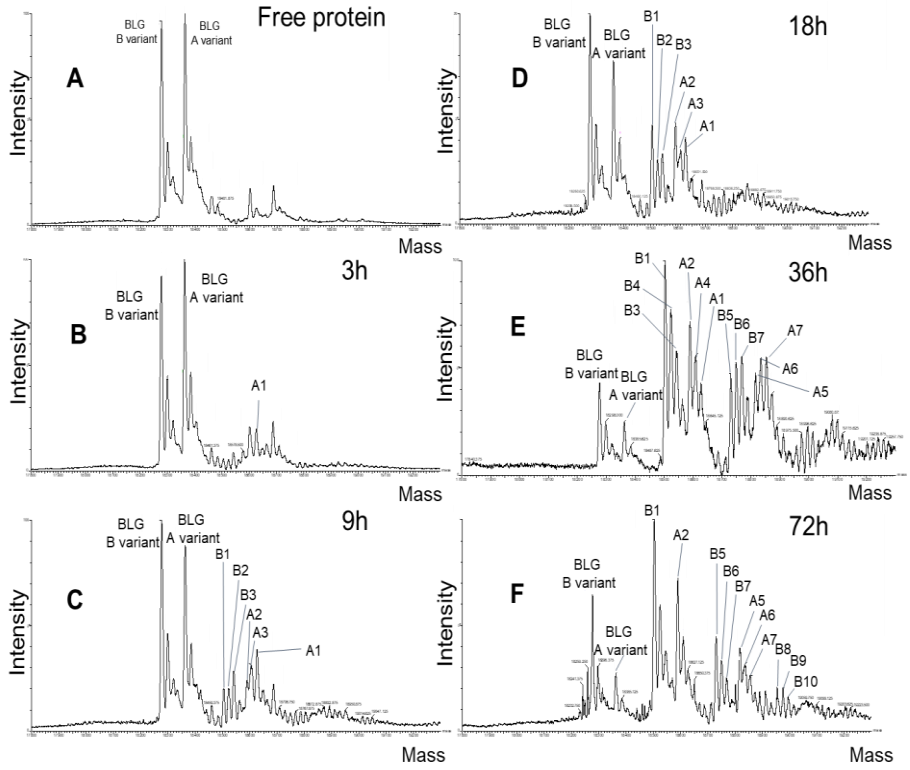

Fig. 4.1.2 Transformed ESI-MS spectra of free β-lactoglobulin (t = 0, panel A) and following 3 h (B), 9 h (C), 18 h (D), 36 h (E) and 72 h (F) of incubation with CDDP. The label on each peak represents the respective species interpretation as reported in Table 4.1.1.

Time	Experimental MW (Da)	Theoretical MW (Da)	Species	Components
0h	18276.39±0.47	18277.2	BLG (B variant)	/
	18362.89±0.79	18363.3	BLG (A variant)	/
3h	18278.03±0.93	18277.2	BLG (B variant)	/
	18363.54±0.58	18363.3	BLG (A variant)	/
	18626.36±1.09	18626.86	BLG (A)+ $[Pt(NH_3)_2Cl^+]$	A1
9h	18277.15±0.92	18277.2	BLG (B variant)	/
	18363.16±0.99	18363.3	BLG (A variant)	/
	18504.38±0.79	18504.31	BLG (B)+ $[Pt(NH_3)_2^{2+}]$	B1
	18522.97 ± 0.30	18522.53	BLG (B)+ $[Pt(NH_3)_2OH_2^{2+}]$	B2
	18540.92±0.83	18540.76	BLG (B)+ $[Pt(NH_3)_2Cl^+]$	B3
	18590.36±0.44	18590.41	BLG (A)+ $[Pt(NH_3)_2^{2+}]$	A2
	18605.33 ± 0.95	18608.61	BLG (A)+ $[Pt(NH_3)_2OH_2^{2+}]$	A3
	18626.42±1.18	18626.86	BLG (A)+ $[Pt(NH_3)_2Cl^+]$	A1
18h	18277.53±0.40	18277.2	BLG (B variant)	/

	18363.45±0.52	18363.3	BLG (A variant)	/
	18505.82±1.60	18504.31	BLG (B)+ $[Pt(NH_3)_2^{2+}]$	B1
	18524.93 ± 0.28	18522.53	BLG (B)+ $[Pt(NH_3)_2OH_2^{2+}]$	B2
	18541.58±0.06	18540.76	BLG (B)+ $[Pt(NH_3)_2Cl^{+}]$	B3
	18590.26±2.26	18590.41	BLG (A)+ $[Pt(NH_3)_2^{2+}]$	A2
	18609.53 ± 0.72	18608.61	BLG (A)+ $[Pt(NH_3)_2OH_2^{2+}]$	A3
	18627.29±1.97	18626.86	BLG (A)+ $[Pt(NH_3)_2Cl^{+}]$	A1
36h	18275.56±2.57	18277.2	BLG(B variant)	/
	18362.04±0.57	18363.3	BLG (A variant)	/
	18503.97±0.50	18504.31	BLG (B)+ $[Pt(NH_3)^{2+}]$	B1
	18522.38 ± 0.71	18522.33	BLG (B)+$[Pt(NH_3)_2OH_2^{2+}]$	B4
	18541.22±0.80	18540.76	BLG (B) +$[Pt(NH_3)_2Cl^{+}]$	B3
	18588.43±0.98	18590.41	BLG (A)+ $[Pt(NH_3)_2^{2+}]$	A2
	18608.39±0.78	18608.62	BLG (A)+ $[Pt(NH_3)_2OH_2^{2+}]$	A4
	18626.94±0.98	18626.86	BLG (A)+ $[Pt(NH_3)_2Cl^{+}]$	A1
	18730.84±0.59	18731.42	BLG (B)+ $2[Pt(NH_3)^{2+}]$	B5
	18749.26±0.52	18749.44	BLG (B)+ $[Pt(NH_3)_2^{2+}]$+ $[Pt(NH_3)_2OH_2^{2+}]$	B6
	18768.45±0.53	18767.46	BLG (B)+ $2[Pt(NH_3)_2OH_2^{2+}]$	B7
	18815.45±0.74	18817.52	BLG (A)+ $2[Pt(NH_3)_2^{2+}]$	A5
	18834.52 ± 0.78	18835.54	BLG (A)+ $[Pt(NH_3)_2^{2+}]$ + $[Pt(NH_3)_2OH_2^{2+}]$	A6
	18853.48±0.78	18853.56	BLG (A)+$2[Pt(NH_3)_2OH_2^{2+}]$	A7
72h	18276.90±0.60	18277.2	BLG (B variant)	/
	18362.31±0.22	18363.3	BLG (A variant)	/
	18502.82±0.77	18504.31	BLG (B)+ $[Pt(NH_3)_2^{2+}]$	B1
	18589.62±0.25	18590.41	BLG (A)+ $[Pt(NH_3)_2^{2+}]$	A2
	18730.97±0.41	18731.42	BLG (B)+ $2[Pt(NH_3)_2^{2+}]$	B5
	18750.96±0.79	18749.44	BLG (B)+ $[Pt(NH_3)_2^{2+}]$+ $[Pt(NH_3)_2OH_2^{2+}]$	B6
	18770.71±0.37	18767.46	BLG (B)+ $2[Pt(NH_3)_2OH_2^{2+}]$	B7
	18817.00±0.09	18817.52	BLG (A)+ $2[Pt(NH_3)_2^{2+}]$	A5
	18835.25±0.80	18835.54	BLG (A)+ $[Pt(NH_3)_2^{2+}]$+ $[Pt(NH_3)_2OH_2^{2+}]$	A6
	18854.51±0.92	18853.56	BLG (A)+ $2[Pt(NH_3)_2OH_2^{2+}]$	A7
	18957.83±0.55	18958.53	BLG (B)+ $3[Pt(NH_3)_2^{2+}]$	B8
	18976.03±0.50	18976.55	BLG (B)+ $2[Pt(NH_3)_2^{2+}]$+ $[Pt(NH_3)_2OH_2^{2+}]$	B9
	18994.26±0.18	18994.57	BLG (B)+ $[Pt(NH_3)_2^{2+}]$+ $2[Pt(NH_3)_2OH_2^{2+}]$	B10

Table 4.1.1 Results of ESI-MS analysis. Incubation times of β-lactoglobulin and CDDP, experimental and theoretical molecular weights, the chemical formula, and the acronym used in Fig.4.1.2 of the corresponding species detected in the ESI-MS spectra are reported. (A) β-lactoglobulin A variant, (B) β-lactoglobulin B variant.

The mass spectrometry time-course experiment results are summarized in the bar

graph in Fig. 4.1.3, where the relative abundance of all detected species was reported for all time of analysis.

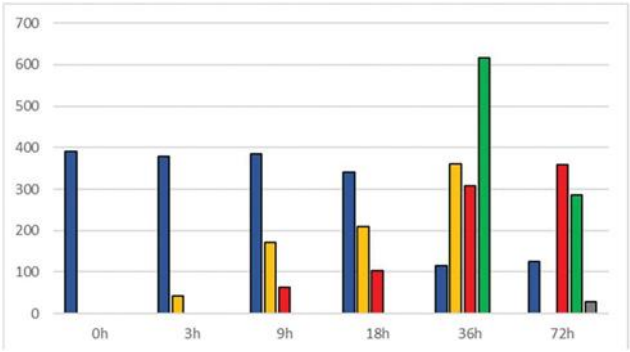

Fig. 4.1.3 Bar graph of the relative abundances of β-lactoglobulin and its adducts with CDDP as a function of time as revealed by ESI-MS. Species intensity, obtained by summing values for the A and B variants for each reaction time, are reported. Metal-free β-lactoglobulin is reported in blue, the monodentate adduct with the $[Pt(NH_3)^2Cl^+]$ or $[Pt(NH_3)_2OH^{2+}]$ fragment bound to the protein is in yellow, the adduct with the $[Pt(NH_3)^{2+}]$ fragment bound to the protein is in red, the adduct with two CDDP fragments bound to the protein is in green, while the adduct with three CDDP fragments bound to the protein is in grey.

4.1.3.2 ESI-MS analysis of β-lactoglobulin/L-OHP adducts

To investigate the β-lactoglobulin/oxaliplatin adducts we performed ESI-MS analysis. Furthermore, the time-dependent evolution of the adducts has been evaluated by a time-course experiment at 0h, 3h, 9h, 18h, 33h, 72h of β-lactoglobulin and oxaliplatin incubation. The spectrum of the free protein (0h) showed the presence of two species at a molecular weight of 18276.67±0.40 Da and 18362.46±0.52 Da corresponding to the B and A β-lactoglobulin variants, respectively (Fig. 4.1.4, panel A). Both the variants displayed the capability to bind oxaliplatin as demonstrated by the presence of signals attributable to the non-covalent interaction of the two variants with the intact metallodrug at all longer incubation times (Table 4.1.2). Nevertheless, some differences have been detected among the adducts over time.

At the shortest reaction time (3h), ESI-MS analysis revealed that both β-lactoglobulin forms bound intact metallodrug, as suggested by the detection of species of 18674.96±0.67 Da and 18759.92±0.86 Da molecular weight for the B and A variant, respectively (Fig. 4.1.4, panel B). In addition to the latter, at 9h of incubation, a small amount of both isoforms bound two molecules of oxaliplatin (19069.13±1.20 Da for B variant, and 19157.68±0.89 Da for A variant). These species did not accumulate over time while starting from 9h of incubation, a species showing a molecular weight of 18583.11±1.64 Da appeared. The latter was assigned to the B variant bound to the fragment $Pt(DACH)^{2+}$ (DACH=$(NH_2)_2C_6H_5$, Fig. 4.1.4, panel C), generated from the adduct with oxaliplatin and increasing in terms of relative abundance with respect to the free protein over 18h, 33h, and 72h (Fig. 4.1.4, panels D-F). The correspondent species involving isoform A was not immediately evident since its expected molecular weight (18672.56 Da) differs for only about 2Da from the one showed by the BLG (B variant)+1L-OHP (18674.48 Da). However, at the longest time, the decreasing of the signal of the A variant bound to 1 L-OHP (18760.27±0.76Da) and the concomitant increasing of a species showing an experimental molecular weight (18672.47±1.14Da) strictly in accordance with the expected for the A variant bound to $Pt(DACH)^{2+}$ confirmed the accumulation of this species over the time, analogously to the B isoform. This finding suggests the occurrence of a progressive releasing of oxalate moiety from the bound oxaliplatin that at longest incubation times (18h, 33h, 72h) also involved the adducts of both β-lactoglobulin isoforms carrying two ligands, in which one of the two intact oxaliplatin molecules released the oxalate fragment (for molecular weights see Table 4.1.2) generating heterogeneous adducts.

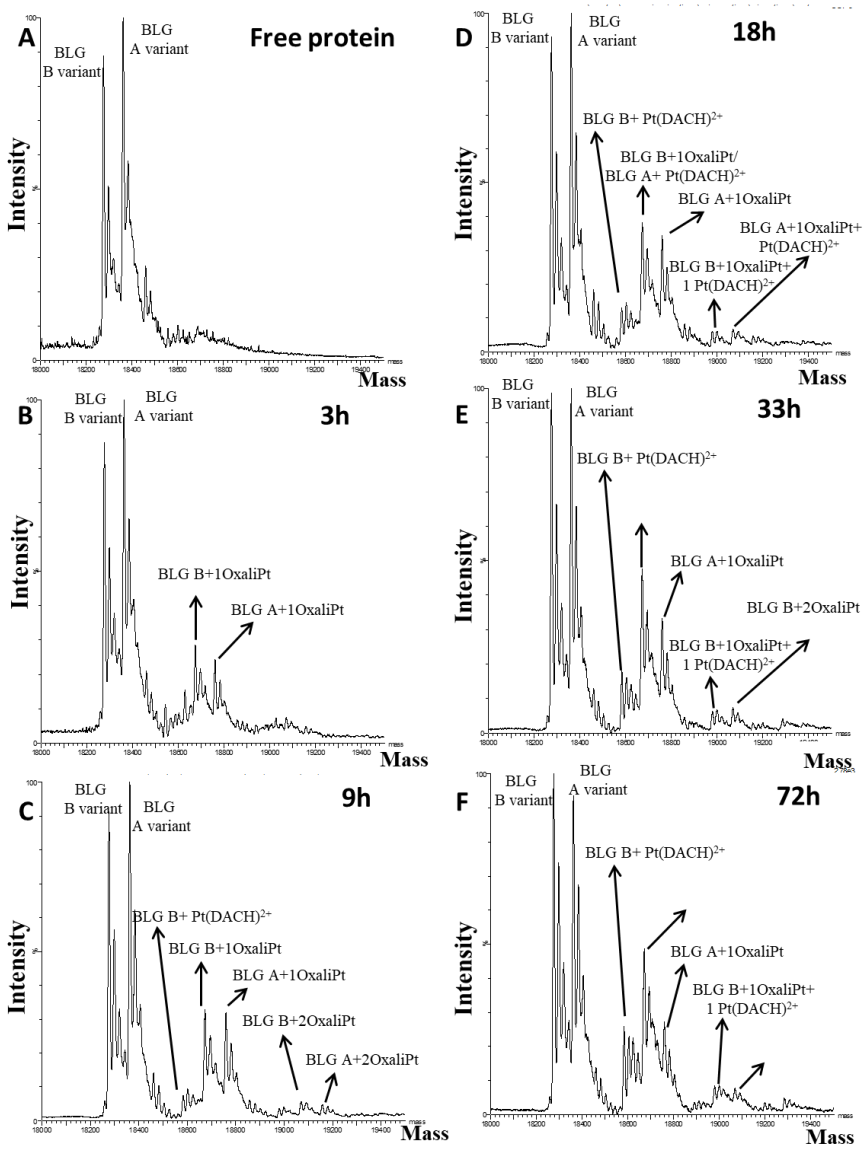

Fig. 4.1.4 Transformed ESI-MS spectra of free β-lactoglobulin (t = 0, panel A) and following 3 h (B), 9 h (C), 18 h (D), 33 h (E), and 72 h (F) of incubation with L-OHP.

Time	Experimental MW (Da)	Theoretical MW (Da)	Species
0h	18276.67±0.40	18277.2	BLG(B variant)
	18362.46±0.52	18363.3	BLG(A variant)
3h	18277.16±0.15	18277.2	BLG(B variant)

	18363.11±0.33	18363.30	BLG(A variant)
	18674.96±0.67	18674.48	BLG(B variant)+1OxaliPt
	18759.92±0.86	18760.58	BLG(A variant)+1OxaliPt
9h	18276.24±0.85	18277.2	BLG(B variant)
	18362.74±0.20	18363.3	BLG(A variant)
	18583.11±1.64	18586.46	BLG(B variant) + 1 $Pt(DACH)^{2+}$
	18673.84±1.01	18674.48	BLG(B variant)+1OxaliPt
	18760.18±0.14	18760.58	BLG(A variant)+1OxaliPt
	19069.13±1.20	19071.76	BLG (B variant)+2OxaliPt
	19157.68±0.89	19157.86	BLG(A variant)+2OxaliPt
18h	18276.78±0.16	18277.2	BLG (B variant)
	18362.59±0.56	18363.3	BLG(A variant)
	18584.21±0.29	18586.46	BLG(B variant) + 1 $Pt(DACH)^{2+}$
	18673.30±0.76	18674.48 18672.56	BLG (B variant)+1OxaliPt BLG (A variant)+ 1$Pt(DACH)^{2+}$
	18759.98±0.41	18760.58	BLG(A variant)+1OxaliPt
	18978.73±0.86	18983.74	BLG(B variant) + 1 $Pt(DACH)^{2+}$+1OxaliPt
	19069.70±0.80	19069.84	BLG(A variant) + 1 $Pt(DACH)^{2+}$+1OxaliPt
33h	18276.21±0.29	18277.2	BLG (B variant)
	18362.59±0.36	18363.3	BLG(A variant)
	18584.21±0.29	18586.46	BLG(B variant) + 1 $Pt(DACH)^{2+}$
	18673.33±0.80	18674.48 18672.56	BLG (B variant)+1OxaliPt BLG (A variant)+ 1$Pt(DACH)^{2+}$
	18759.95±0.70	18760.58	BLG(A variant)+1OxaliPt
	18977.96±1.67	18983.74	BLG(B variant) + 1 $Pt(DACH)^{2+}$+1OxaliPt
	19069.76±1.09	19069.84	BLG(A variant) + 1 $Pt(DACH)^{2+}$+1OxaliPt
72h	18277.50±1.04	18277.2	BLG (B variant)
	18363.80±0.97	18363.3	BLG(A variant)
	18584.56±0.91	18586.46	BLG(B variant) + 1 $Pt(DACH)^{2+}$
	18672.47±1.14	18674.48 18672.56	BLG (B variant)+1OxaliPt BLG (A variant)+ 1$Pt(DACH)^{2+}$
	18760.27±0.76	18760.58	BLG(A variant)+1OxaliPt
	18980.31±1.52	18983.74	BLG(B variant) + 1 $Pt(DACH)^{2+}$+1OxaliPt
	19069.38±0.96	19069.84	BLG(A variant) + 1 $Pt(DACH)^{2+}$+1OxaliPt

Table 4.1.2. Results of ESI-MS experiments. For each time of incubation the calculated, the theoretical molecular weights, and the chemical formula of the detected adducts are reported.

4.1.4 Conclusions

In the present study, we investigated the interaction of β-lactoglobulin with anticancer metallodrugs cisplatin (CDDP) and oxaliplatin (L-OHP). The ESI-MS

analysis of CDDP incubated with β-lactoglobulin evidenced the formation of adducts with more than one Pt-containing fragment coordinating protein residue side chains in a time-dependent manner. The CDDP/protein adducts amount increased over time as determined by the relative abundance of each species. Finally, the monodentate and bidentate mode of binding are both possible, but the first precedes the latter. The ESI-MS results of the L-OHP incubated with β-lactoglobulin confirmed the non-covalent and covalent binding of the metallodrug. L-OHP approaches to the protein in a non-covalent mode. β-lactoglobulin showed the capability to bind up to two oxaliplatin molecules. Together with *in vivo* and other *in vitro* evidences, these findings may suggest how to develop other cisplatin-derived drugs that can be better delivered in β-lactoglobulin nanocages[14].

The results discussed in this chapter (including figures and tables) have been published in the following article: Balasco, N.; Ferraro, G.; Loreto, D.; **Iacobucci, I.**; Monti, M.; Merlino, A. Cisplatin Binding to β-Lactoglobulin: A Structural Study. *Dalton Trans* **2020**, *49* (35), 12450–12457. https://doi.org/10.1039/d0dt02582h.

4.1.5 References

(1) Kontopidis, G.; Holt, C.; Sawyer, L. Invited Review: Beta-Lactoglobulin: Binding Properties, Structure, and Function. *J Dairy Sci* **2004**, *87* (4), 785–796. https://doi.org/10.3168/jds.S0022-0302(04)73222-1.

(2) Ko, S.; Gunasekaran, S. Preparation of Sub-100-Nm Beta-Lactoglobulin (BLG) Nanoparticles. *J Microencapsul* **2006**, *23* (8), 887–898. https://doi.org/10.1080/02652040601035143.

(3) Teng, Z.; Luo, Y.; Li, Y.; Wang, Q. Cationic Beta-Lactoglobulin Nanoparticles as a Bioavailability Enhancer: Effect of Surface Properties and Size on the Transport and Delivery in Vitro. *Food Chem* **2016**, *204*, 391–399. https://doi.org/10.1016/j.foodchem.2016.02.139.

(4) Shafaei, Z.; Ghalandari, B.; Vaseghi, A.; Divsalar, A.; Haertlé, T.; Saboury, A. A.; Sawyer, L. β-Lactoglobulin: An Efficient Nanocarrier for Advanced Delivery Systems. *Nanomedicine* **2017**, *13* (5), 1685–1692. https://doi.org/10.1016/j.nano.2017.03.007.

(5) Leilabadi-Asl, A.; Divsalar, A.; Saboury, A. A.; Parivar, K. Probing the Interaction of Two Chemotherapeutic Drugs of Oxali-Palladium and 5-Fluorouracil Simultaneously with Milk Carrier Protein of β-Lactoglobulin. *Int J Biol Macromol* **2018**, *112*, 422–432. https://doi.org/10.1016/j.ijbiomac.2018.01.067.

(6) Muggia, F. M.; Bonetti, A.; Hoeschele, J. D.; Rozencweig, M.; Howell, S. B. Platinum Antitumor Complexes: 50 Years Since Barnett Rosenberg's Discovery. *J Clin Oncol* **2015**, *33* (35), 4219–4226. https://doi.org/10.1200/JCO.2015.60.7481.

(7) Kelland, L. The Resurgence of Platinum-Based Cancer Chemotherapy. *Nat Rev Cancer* **2007**, *7* (8), 573–584. https://doi.org/10.1038/nrc2167.

(8) Lebwohl, D.; Canetta, R. Clinical Development of Platinum Complexes in Cancer Therapy: An Historical Perspective and an Update. *Eur J Cancer* **1998**, *34* (10), 1522–1534. https://doi.org/10.1016/s0959-8049(98)00224-x.

(9) Pinato, O.; Musetti, C.; Sissi, C. Pt-Based Drugs: The Spotlight Will Be on Proteins. *Metallomics* **2014**, *6* (3), 380–395. https://doi.org/10.1039/C3MT00357D.

(10) Arnesano, F.; Natile, G. "Platinum on the Road": Interactions of Antitumoral Cisplatin with Proteins. *Pure and Applied Chemistry* **2008**, *80* (12), 2715–2725. https://doi.org/10.1351/pac200880122715.

(11) Chanphai, P.; Bariyanga, J.; Bérubé, G.; Tajmir-Riahi, H. A. Complexation of Cis-Pt and Trans-Pt(NH_3)$_2$$Cl_2$ with Serum Proteins: A Potential Application for Drug Delivery. *Journal of Biomolecular Structure and Dynamics* **2020**, *38* (9), 2777–2783. https://doi.org/10.1080/07391102.2019.1654408.

(12) Ghalandari, B.; Divsalar, A.; Eslami-Moghadam, M.; Saboury, A. A.; Haertlé, T.; Amanlou, M.; Parivar, K. Probing of the Interaction between β-Lactoglobulin and the Anticancer Drug Oxaliplatin. *Appl Biochem Biotechnol* **2015**, *175* (2), 974–987. https://doi.org/10.1007/s12010-014-1341-0.
(13) Qin, B. Y.; Bewley, M. C.; Creamer, L. K.; Baker, E. N.; Jameson, G. B. Functional Implications of Structural Differences between Variants A and B of Bovine Beta-Lactoglobulin. *Protein Sci* **1999**, *8* (1), 75–83. https://doi.org/10.1110/ps.8.1.75.
(14) Zang, J.; Zheng, B.; Zhang, X.; Arosio, P.; Zhao, G. Design and Site-Directed Compartmentalization of Gold Nanoclusters within the Intrasubunit Interfaces of Ferritin Nanocage. *J Nanobiotechnol* **2019**, *17* (1), 79. https://doi.org/10.1186/s12951-019-0512-0.

Chapter 4- Study of metal-based compounds as potential pharmacological drugs

4.2 Investigation of metallodrugs potential inhibitory activity towards the aggregation of $A\beta_{21-40}$ amyloidogenic peptide

4.1.1 Introduction

The Aβ peptide is the proteolytic degradation product of a transmembrane precursor protein called APP (Amyloid Precursor Protein). Contrary to popular belief, the Aβ peptide is a regular product also present in healthy people and not only in patients affected by neurodegenerative diseases such as Alzheimer's[1]. The APP degradation process involves three different enzymes: α-secretase, β-secretase, γ-secretase and depending on how these enzymes operate. Degradation may encounter a non-amyloidogenic pathway and an amyloidogenic pathway, as shown in Fig. 4.2.1.

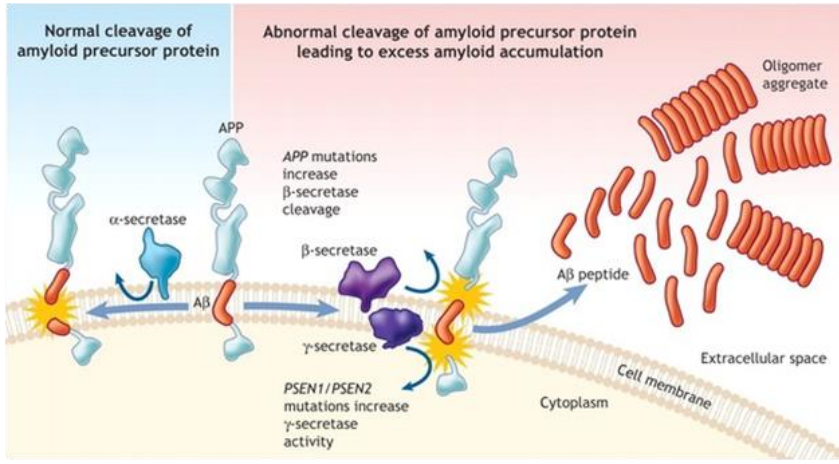

Fig. 4.2.1 APP degradation process in amyloidogenic and non-amyloidogenic pathways[2].

In the amyloidogenic pathway, the Aβ-amyloid peptide (1-40 and 1-42) forms oligo- and multimeric aggregates that evolve into fibrils recovered in the brains of patients affected by Alzheimer's disease (AD)[1]. The research field involving the use of metal-based drugs as inhibitors of amyloid fibril formation and toxicity, targeting neurodegenerative diseases like Alzheimer's (AD) and Parkinson's Disease (PD), is experiencing a great flowering[3,4]. In general, transition metal complexes have tunable properties, depending on the oxidation and spin states of the metal center as well as the coordination geometry. These features could influence the reactivity of these compounds with amyloidogenic species and, consequently, modulate their aggregation pathways[5,6].

The amyloid inhibitory activity of the metal-based drugs can be exploited through different mechanisms[7]: i) coordination chemistry, ii) oxidative, iii) proteolytic reactions[8] for peptide modifications. Metal complexes of relatively stable kinetically inert metal ions (such as Pt(II) square-planar complexes) can form stable coordinate bonds with amyloidogenic peptides in their monomeric state[9], thus, can prevent toxic effects of Aβ oligomerization and serve as potential neuro drugs. In general, Pt(II) and Ru(III)[10] complexes are more stable and redox inert with respect to Zn-, Cu-, or Fe-based compounds; thus, they can have major chances of success to prevent toxic effects of Aβ oligomerization in a coordinative mechanism and to serve as potential neurodrugs.

Starting from cisplatin as pioneering antineoplastic metal-based compounds[11], many analogs of 2nd and 3rd generation[12] compounds were investigated for their interaction with DNA, and several examples of square-planar Pt(II)-complexes bearing phenanthroline, bipyridine, and terpyridine as intercalating ligands were reported[13,14,15]. Several studies demonstrated the capability of Pt(II) complexes to interfere with amyloid-aggregation, many of these involving hydrophobic phenanthroline(phen)-based bidentate ligands along with two monodentate

ligands (e.g., chlorides)[16]. π-π interactions between the phen moiety and aromatic Phe4, Tyr10 and Phe19 aid the formation of the [Pt(II)(phen)- Aβ]$^{2+}$ adduct that suppresses aggregation and limit the neurotoxicity of Aβ in mouse hippocampal tissue[16]. Direct coordination of the Pt center to amyloid peptides/proteins is not the only strategy that has been identified to inhibit the formation of amyloid fibers. It has been shown that metal-driven oxidation of crucial residues could be also exploited[6]. Indeed, the oxidation of Met35 affects the aggregation of $A\beta_{1-40}$[17]. The formation of the hydrophilic Met35 sulfoxide influences the electronic properties of Aβ peptides causing an enhancement of peptide's polarity, which hampers hydrophobic interactions that are crucial for the initiation and progression of Aβ aggregation. Thus, it is possible to trigger the oxidation of Met in amyloid peptides utilizing metal compounds that can act as reducing agents or produce reactive oxygen species (ROS), which can be responsible for the oxidation. Beyond the Pt(II) complexes, other metal complexes have been used as amyloid fiber formation inhibitors. Due to the ligand field similarities between Pd(II), Au(II), and Pt(II) compounds, the ability of Pd(II) and Au(III) compounds to inhibit the fiber formation has been investigated[18,19,20,21]. Pd(II) compounds can exchange ligands 10^5 times faster than their Pt(II) analogs, enhancing their capability to interact with the cellular components such as sulfur-donor biomolecules.

The current project has been carried out in collaboration with Professor Daniela Marasco and Antonello Merlino research groups of the Department of Pharmacy and Chemical Sciences, respectively, of the University of Naples "Federico II". We focused our attention on the study of the 21-40 C-terminal fragment of the Aβ peptide (Table 4.2.1), for which few works are reported in the recent literature, although its propensity to form amyloid aggregations.

We have investigated the anti-aggregation effects of three cisplatin analogs bearing 1,10-phenanthroline (Pt(II)-phen) or 2,2'-bipyridine (Pt(II)-bipy), or

2,2':6',2''-terpyridine ligands (Pt(II)-terpy) (Fig. 4.2.2). These metal compounds have been demonstrated to be able to bind the entire beta-amyloid polypeptide.

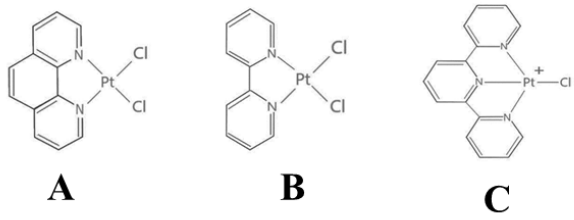

Fig. 4.2.2 Chemical structure of the Pt(II) complexes analyzed in this study. A) Dichloro(1,10-phenanthroline)platinum(II); B) (2,2'-bipyridine) dichloroplatinum(II); C) Dichloro(2,2′:6′,2″-terpyridine)platinum(II).

Peptide	Sequence	pI	Net charge at pH7
$A\beta_{21-40}$	AEDVGSNKGAIIGLMVGGVV	3.93	-1

Table 4.2.1. The aminoacidic sequence, the isoelectric point (pI), and the charge at pH=7 is reported for $A\beta_{21-40}$.

Furthermore, to investigate the incidence of the metal ion on the aggregation process of $A\beta_{21-40}$, we studied the activity of square-planar complexes of Pt(II), Pd(II), Au(III) carrying 2-(2'-pyridyl)benzimidazole and chlorides as ligands (Fig. 4.2.3) on the $A\beta_{21-40}$ peptide.

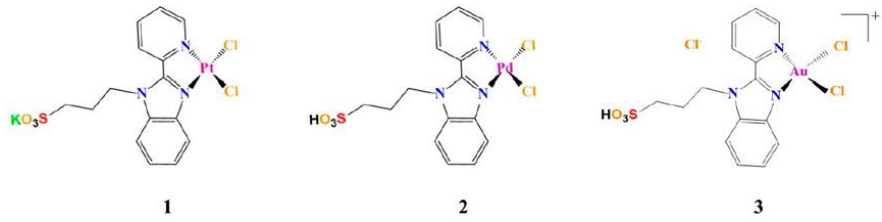

Fig. 4.2.3 Chemical structure of the 2-(2'-pyridyl)benzimidazole 1) Pt(II), 2) Pd(II), 3) Au(III) complexes.

The native ESI-MS approach, combined with other physicochemical techniques, allowed us to investigate if the metal complexes could bind the $A\beta_{21-40}$. Once detected the adducts, the binding mode and the stoichiometry were determined to

understand the different mechanisms of action of the potential neuro drugs.

4.2.2 Experimental methods

4.2.2.1 ESI-MS analysis

Solutions of 50 μM $A\beta_{21-40}$ in 15 mM AMAC (Ammonium acetate) buffer pH=6.8, at 1:1 ratios with complexes were incubated for 0 and 24h for Pt-(phen), Pt-(bipy), and Pt-(terpy), and 0h, 3h and 17h for **1**, **2**, and **3**. The samples were analyzed by Q-ToF Premier (Waters, Milliford, MA, US) mass spectrometer. The analysis were done by direct injection at 10 μL/min and the source parameters were set as 3 kV for capillary voltage and 42 kV for cone voltage. The acquisition range was spanned from 700 to 2500 m/z, and the raw data were processed with MassLynx 4.1 software (Waters, Milliford, MA, US).

4.2.3 Results and Discussion

4.2.3.1 Electrospray-mass spectrometry analysis

The ESI-MS spectra of $A\beta_{21–40}$ in the absence of metal complexes highlight a weakness of the peptide bonds under the experimental conditions used for this analysis. During spectra acquisition using standard source settings, the peptide undergoes to a fragmentation process, as evidenced by the presence in the spectra of several b-series signals (Fig. 4.2.4, panel A and B).

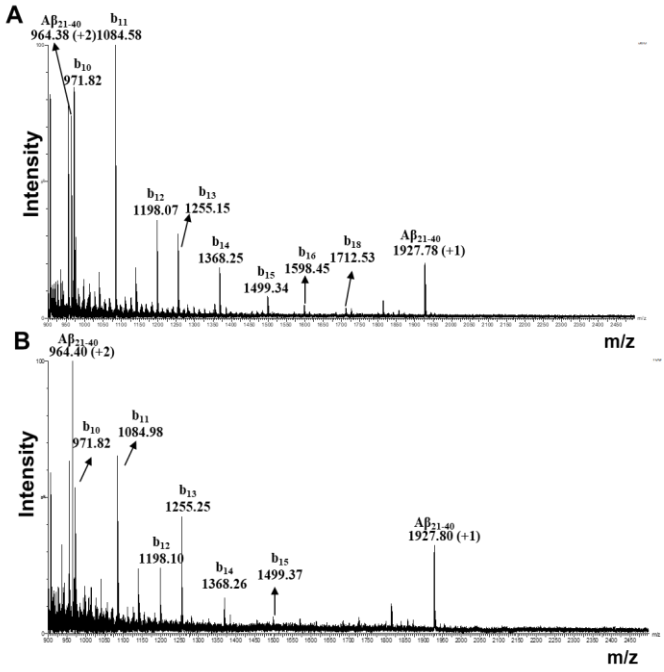

Fig. 4.2.4 ESI-MS spectrum of $A\beta_{21-40}$ peptide alone at 0h (A) and 24h (B). $A\beta_{21-40}$ fragments belonging to b-series and spontaneously in source-generated are also present.

Then, $A\beta_{21-40}$ incubated with Pt-(phen), Pt-(bipy), Pt-(terpy) compounds at 1:1 ratios, were analyzed at two different times (0 and 24h) by electrospray ionization mass spectrometry (ESI-MS)[37]. As reported in Figs. 4.2.5-4.2.7 and Tables 4.2.2-4.2.4, ESI-MS spectra showed that all the three Pt complexes are able to bind one chain of the $A\beta_{21-40}$ at longer as well as shorter incubation times. By a detailed inspection of MS spectra of $A\beta_{21-40}$ complexes with bi-dentate adducts (Pt-phen and Pt-bipy), two double-charged ions were present: one at higher m/z values (Pt-phen m/z=1168.52; Pt-bipy m/z=1156.50) generated by the loss of one chloride ion, and a second peak, at lower m/z values (Pt-phen m/z=1150.52; Pt-bipy m/z=1138.02), deriving by the loss of both chloride ions. These latter peaks increased over time. In addition, the species missing two chloride ions are flanked by the presence of an adduct deriving from the substitution of a chloride with an

acetate ion, deriving from the buffer (Pt-phen m/z=1180.04 Da; Pt-bipy m/z=1168.05 Da). $A\beta_{21-40}$ formed an adduct with one molecule of Pt-terpy losing the only chloride present, as demonstrated by the presence of the double charged ion at m/z 1176.57. Moreover, in the ESI-MS acquiring conditions, the isolated $A\beta_{21-40}$ peptide provided an in-source fragmentation phenomena, generating b series fragments. The fragmentation of free $A\beta_{21-40}$ peptide also persisted in the presence of Pt(II)-phen (Fig. 4.2.5) and Pt(II)-terpy (Fig. 4.2.7), since a large amount of peptide was unbound. Conversely, in the presence of Pt(II)-bipy, no free peptide fragmentation was detectable (Fig. 4.2.6), being the largest part of peptide bound to the Pt-compound. However, differently from other complexes, $A\beta_{21-40}$/ Pt-bipy is the unique adduct to give fragmentation, as confirmed by mono and double-charged species of b_{19} C-terminal fragment evidenced in the ESI-MS spectrum.

MS data demonstrated that all Pt-compounds bind $A\beta_{21-40}$ with 1:1 stoichiometry and that the adduct with bipy resulted more abundant and stable in respect to the others, as demonstrated by the dominant species in the spectrum associated with the complex and its b_{19} fragment.

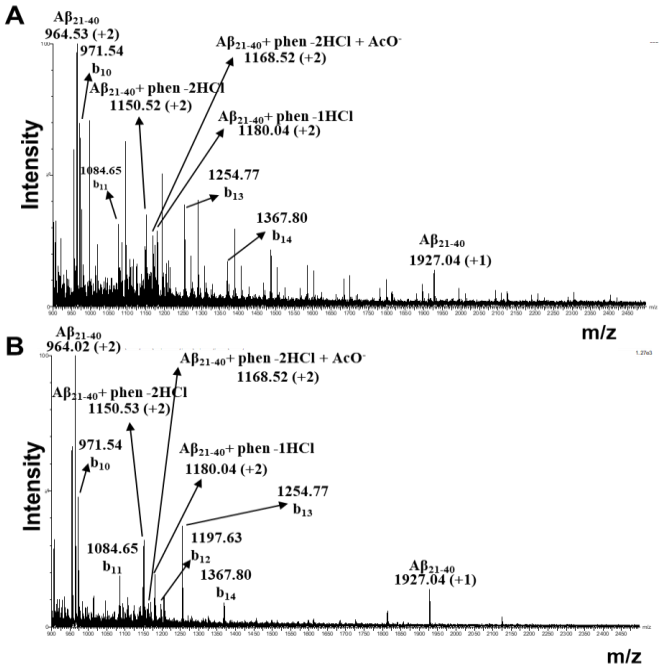

Fig. 4.2.5 ESI-MS spectra of $A\beta_{21-40}$ peptide at 0h (A) and 24h (B) of incubation with Pt-phen. Fragmentation ions are also indicated.

Time	Experimental m/z, (charge state)	Experimental Monoisotopic Mass (Da)	Theoretical Monoisotopic Mass (Da)	Pt(II)- Peptide Complexes
0h	1150.52 (+2)	2299.06	2299.29	$A\beta_{21-40}$+1 Pt(II)-phen -2HCl
	1180.04 (+2)	2358.07	2358.29	$A\beta_{21-40}$+1 Pt(II)-phen -2HCl+1 AcO^-
	1168.52 (+2)	2334.18	2335.74	$A\beta_{21-40}$+1 Pt(II)-phen -1HCl
24h	1150.53 (+2)	2299.06	2299.29	$A\beta_{21-40}$+1 Pt(II)-phen -2HCl
	1180.04 (+2)	2358.07	2358.29	$A\beta_{21-40}$+1 Pt(II)-phen -2HCl+1 AcO^-
	1168.50 (+2)	2335.00	2335.74	$A\beta_{21-40}$+1 Pt(II)-phen -1HCl

Table 4.2.2. Summary of main ion species occurring in the ESI-MS spectra of $A\beta_{21-40}$ with the Pt-phen, at 0h and 24h of incubation. Each detected adduct is reported with the corresponding experimental m/z values, the charge state, the calculated and theoretical monoisotopic molecular weight. Fragmentation ions are not reported.

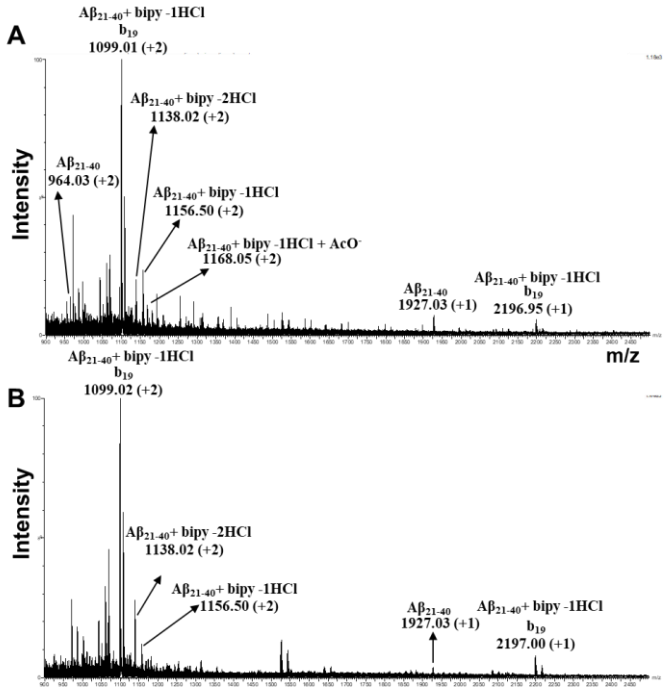

Fig. 4.2.6 ESI-MS spectra of Aβ$_{21-40}$ peptide at 0h (A) and 24h (B) of incubation with Pt-bipy. Fragmentation ions are also indicated.

Time	Experimental m/z, (charge state)	Experimental Monoisotopic Mass (Da)	Theoretical Monoisotopic Mass (Da)	Pt(II)- Peptide Complexes
0h	1138.02 (+2)	2274.04	2275.27	Aβ$_{21-40}$+1 Pt(II)-bipy-2HCl
	1168.05 (+2)	2234.10	2334.27	Aβ$_{21-40}$+1 Pt(II)-bipy -2HCl+AcO$^-$
	1156.50 (+2)	2311.00	2311.72	Aβ$_{21-40}$+1 Pt(II)-bipy -1HCl
24h	1138.02 (+2)	2274.04	2275.27	Aβ$_{21-40}$+1 Pt(II)-bipy -2HCl
	1156.50 (+2)	2310.00	2311.72	Aβ$_{21-40}$+1 Pt(II)-bipy -1HCl

Table 4.2.3. Summary of main ion species occurring in the ESI-MS spectra of Aβ$_{21-40}$ with the Pt-bipy, at 0h and 24h of incubation. Each detected adduct is reported with the corresponding experimental m/z values, the charge state, the calculated and theoretical monoisotopic molecular weight. Fragmentation ions are not reported.

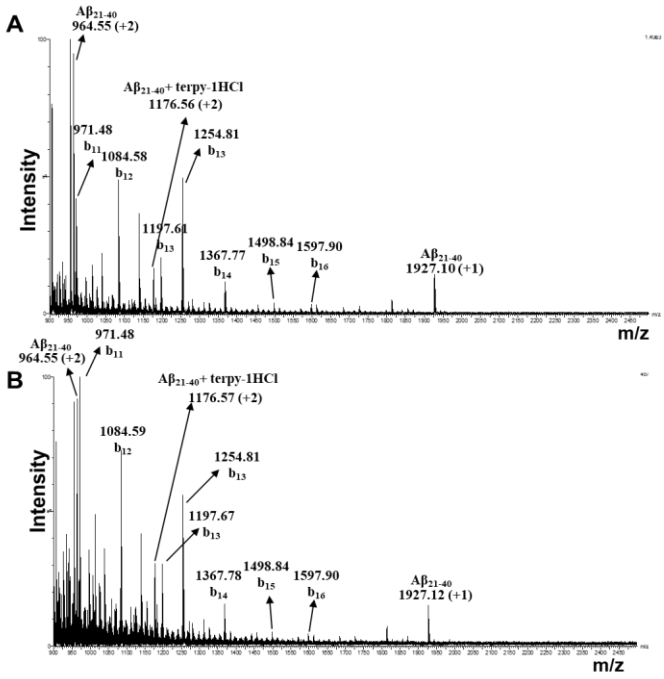

Fig. 4.2.7 ESI-MS spectra of $A\beta_{21-40}$ peptide at 0h (A) and 24h (B) of incubation with Pt-terpy. Fragmentation ions are also indicated.

Time	Experimental m/z, (charge state)	Experimental Monoisotopic Mass (Da)	Theoretical Monoisotopic Mass (Da)	Pt(II)- Peptide Complexes
0h	1176.56 (+2)	2351.14	2353.35	$A\beta_{21-40}$+1 Pt(II)-terpy -1HCl
24h	1176.57 (+2)	2351.14	2353.35	$A\beta_{21-40}$+1 Pt(II)-terpy -1HCl

Table 4.2.4. Summary of main ion species occurring in the ESI-MS spectra of $A\beta_{21-40}$ with the Pt-terpy, at 0h and 24h of incubation. Each detected adduct is reported with the corresponding experimental m/z values, the charge state, the calculated and theoretical monoisotopic molecular weight. Fragmentation ions are not reported.

The same binding mode of the Pt-based complexes described above was also found probing the interaction between $A\beta_{21–40}$ and Pt(II) carrying 2-(2'-pyridyl)benzimidazole ligand. Therefore, just after adding **1** to $A\beta_{21–40}$ (Fig. 4.2.8, panel A and Table 4.2.5), one and two chloride ligands are substituted by the peptide, as indicated by the presence of the species at 2471.54 Da and 2435.59 Da. The latter increases over time, as observed in the other Pt(II) complexes.

On the contrary, when $A\beta_{21-40}$ is incubated with **2** (Fig. 4.2.9 and Table 4.2.6), it immediately binds one molecule of the Pd(II) compound that has released two Cl^- ligands, as inferred from the species of 2346.03 Da molecular weight. Moreover, the presence of an adduct of 2029.00 ± 0.02 Da suggests that $A\beta_{21-40}$ also binds a naked Pd(II) ion. In addition to the $A\beta_{21-40}$ peptide fragmentation signals and contrarily to the other investigated compounds, the adducts formed by $A\beta_{21-40}$ with **1** and **2** showed signals belonging to b-series. This finding suggests that the $A\beta_{21-40}$ N-terminal tail is directly involved in recognizing these metal compounds.

Unlike the other metal complexes, **3** does not form adducts with $A\beta_{21-40}$ (Fig. 4.2.10, Table 4.2.7). However, the presence of the metal complex affects the oxidation state of the peptide. In particular, it promotes the oxidation of Met35, as demonstrated by a slight increase of $A\beta_{21-40}$ oxidized form over time (1941.68 Da).

These findings suggest that both the direct binding of metal complexes to the peptide and the oxidation of Met35 can be used as valuable strategies to inhibit the $A\beta_{21-40}$ aggregation.

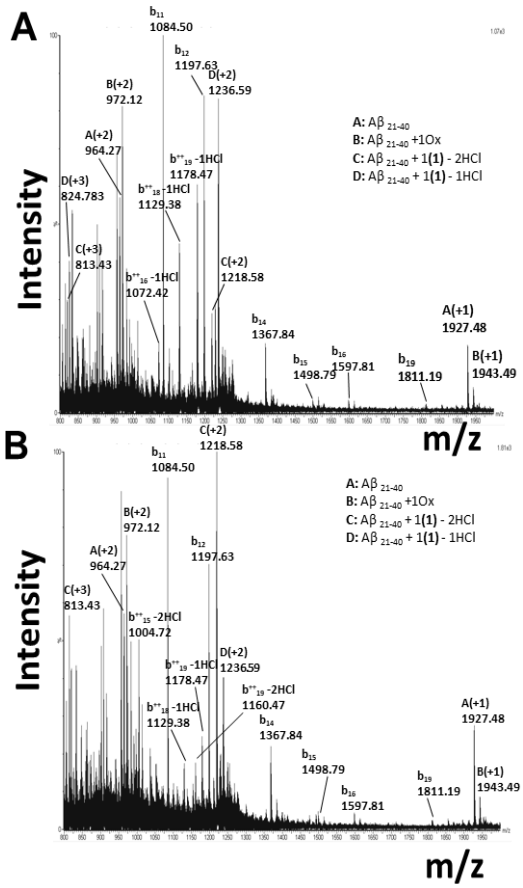

Fig. 4.2.8 ESI-MS spectra of Aβ$_{21–40}$ incubated with **1** at t=0h (A) and t=17 h (B). Signals belonging to the fragmentation b series of Aβ$_{21–40}$ are indicated.

Metal ion	Signal (m/z)	Charge	Experimental MW (Da)	Theoretical MW (Da)	Species
Pt	1927.48 964.27	A +1 A +2	1926.50±0.02	1926.00	Aβ $_{21-40}$
	1943.49 972.12	B +1 B +2	1942.35±0.13	1942.00	Aβ $_{21}$-40 + 1Ox
	1218.583 813.429	C +2 C +3	2435.59±0.56	2436.45	Aβ $_{21-40}$ + 1**(1)** - 2HCl
	1236.599 824.641	D +2 D +3	2471.54±0.36	2472.9	Aβ $_{21-40}$ + 1**(1)** - 1HCl

Table 4.2.5 Results of the ESI-MS analysis of the adducts formed upon 0h and 17h of incubation of Aβ$_{21–40}$ with **1**. The experimental m/z values, the ion charge status, the experimental and theoretical monoisotopic mass values, and the corresponding ion species are reported.

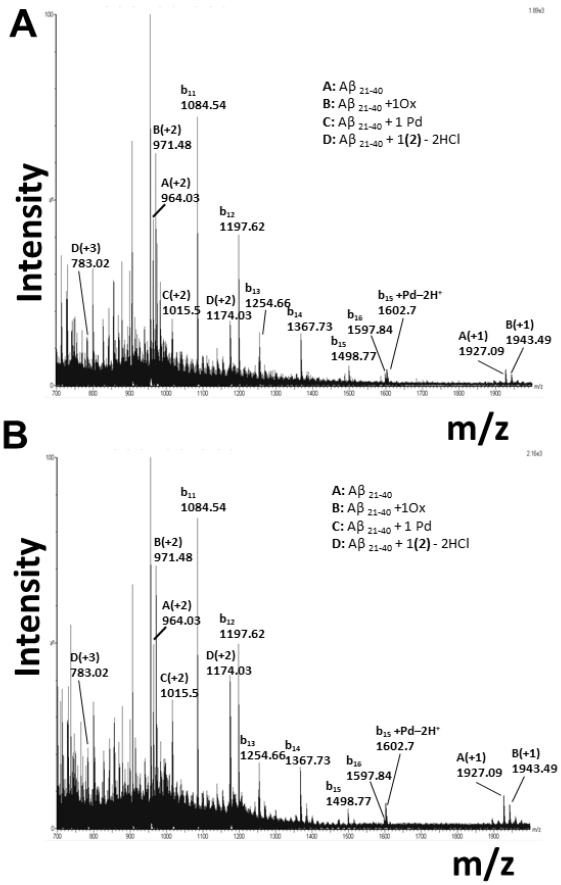

Fig. 4.2.9 ESI-MS spectra of Aβ$_{21–40}$ incubated with **2** at t=0h (A) and t=17 h (B). Signals belonging to the fragmentation b series of Aβ$_{21–40}$ are indicated.

Metal ion	Signal (m/z)	Charge	Experimental MW (Da)	Theoretical MW (Da)	Species
Pd	1927.09 964.03	A +1 A +2	1926.07±0.02	1926.00	Aβ $_{21-40}$
	1943.07 971.48	B +1 B +2	1941.50±0.56	1942.00	Aβ $_{21-40}$ + 1Ox

	2030.029 1015.500	C +1 C +2	2029.00±0.02	2028.42	Aβ $_{21-40}$ + 1 Pd
	1174.035 783.022	D +2 D +3	2346.03±0.02	2347.79	Aβ $_{21-40}$ + 1**(2)** - 2HCl

Table 4.2.6 Results of the ESI-MS analysis of the adducts formed upon 0h and 17h of incubation of Aβ$_{21–40}$ with **2**. The experimental m/z values, the ion charge status, the experimental and theoretical monoisotopic mass values, and the corresponding ion species are reported.

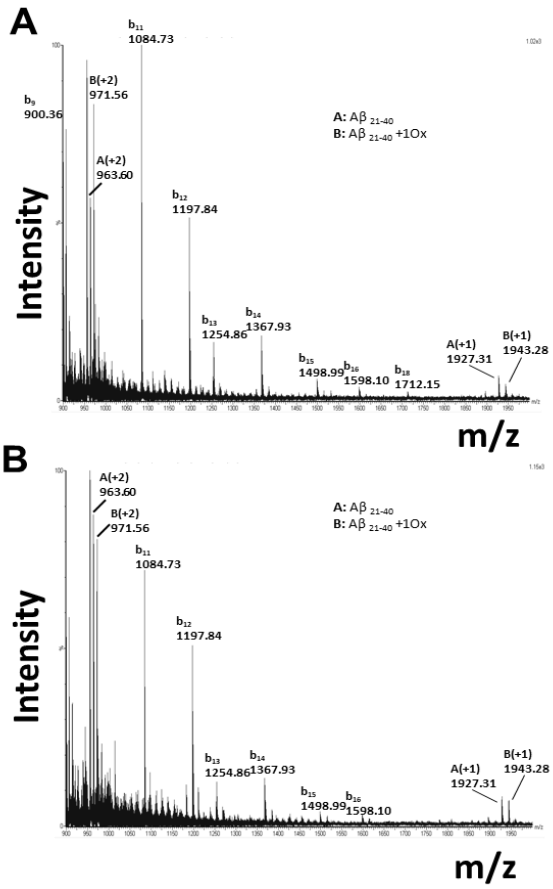

Fig. 4.2.10 ESI-MS spectra of Aβ$_{21–40}$ incubated with **3** at t=0h (A) and t=17 h (B). Signals belonging to the fragmentation b series of Aβ$_{21–40}$ are indicated.

Metal ion	Signal (m/z)	Charge	Experimental MW (Da)	Theoretical MW (Da)	Species
Au	1927.31 963.60	A +1 A +2	1925.74±0.56	1926.00	Aβ $_{21-40}$
	1943.28 971.56	B +1 B +2	1941.68±0.58	1942.00	Aβ $_{21-40}$ +1Ox

Table 4.2.7 Results of the ESI-MS analysis of the adducts formed upon 0h and 17h of incubation of $A\beta_{21–40}$ with **3**. The experimental m/z values, the ion charge status, the experimental and theoretical monoisotopic mass values, and the corresponding ion species are reported.

The use of the transition metals as potential drugs for neurodegenerative diseases treatment has been inspired from the observation that in senile plaques found in the Alzheimer Disease (AD)-affected brain, mainly composed of Aβ aggregates, highly concentrated metals, such as Cu, Zn, and Fe[22,23,24] have been found coordinated to the amino acid residues, such as His6, His13, and His14, of Aβ peptide[25,26,27].

Therefore, transition metal complexes have been investigated as a class of chemical modulators against Aβ aggregation, considering their tunable characteristics, including the oxidation state and coordination geometry around the metal center. The adducts between the transition metal complexes and Aβ can affect intramolecular and intermolecular interactions essential for the folding and aggregation of Aβ peptides (Fig. 4.2.11)[16,28,29].

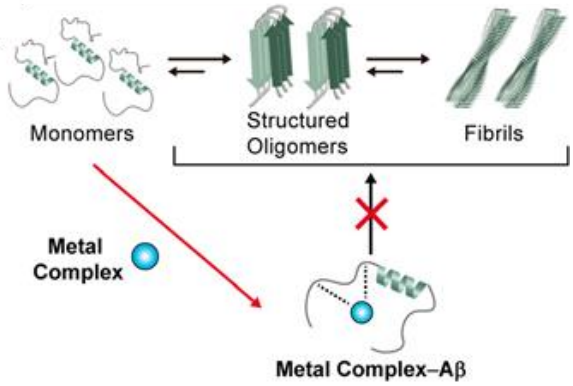

Fig. 4.2.11 Schematic drawing of modulating Aβ aggregation upon coordination of transition metal complexes to Aβ[6].

However, other mechanisms have been reported to inhibit amyloid aggregation. Among them, the Aβ oxidation is reported[30,31]. The main amino acid residues that can be oxidized are His and Met[17,31]. The higher polar generated species vary the electronic properties of Aβ peptides and cause an enhancement of peptide's polarity, which hampers hydrophobic interactions that are crucial for the initiation and progression of Aβ aggregation.

In particular, Pt-based complexes with 1,10-phenanthroline (Pt(II)-phen) or 2,2'-bipyridine (Pt(II)-bipy), or 2,2':6',2''-terpyridine ligands (Pt(II)-terpy) ligands as well as Pt(II)-, Pd(II)-, Au(III) based complexes carrying 2-(2'-pyridyl)benzimidazole were investigated (labeled as **1**, **2**, and **3**, respectively).

In this project, the physicochemical data collected by our collaborators have revealed the ability of the investigated metal complexes to modulate the aggregation of the 21-40 C-term moiety of Aβ amyloid peptide by time-course Thioflavin T (ThT) fluorescence emission. ThT is a fluorescent dye that, when bound to beta-sheet rich structures, displays an enhanced fluorescence emission at about 480 nm upon excitation at 440 nm. All the tested compounds showed to be able to suppress the aggregation process (Fig. 4.2.12, panels A and B). As a

control of non-aggregating peptide, $A\beta_{21\text{-}40}$ bearing the point mutation G37D, which was reported to suppress aggregation of the entire $A\beta_{1\text{-}42}$, was subjected to the ThT assay. The no-zero starting value of fluorescence is ascribable to an immediate partial aggregation during sample preparation as already observed[32].

Pt-(phen), Pt-(bipy), and Pt-(terpy) was detected able to inhibit amyloid aggregation of $A\beta_{21\text{-}40}$, as displayed in panel A of Fig. 4.2.12. Moreover, two ratios for peptide:metal complexes were possible only for Pt-terpy, i. e., 1:1 and 1:5, while Pt-phen and Pt-(bipy) resulted insoluble for 1:5 ratio (500 µM). In addition, $A\beta_{21\text{-}40}$ alone reached a saturated signal after 6h of stirring while mut $A\beta_{21\text{-}40}$ cannot aggregate for 20 h, as expected. The anti-aggregation ability is more evident for terpy and bipy complexes while almost negligible for Pt-(phen). A complete inhibition of the aggregation is displayed by 1:5 ratio of Pt-(terpy) complex that provides a signal similar to mut $A\beta_{21\text{-}40}$ for the entire analysis duration. The **1**, **2**, and **3** also displayed an effect on the aggregation process (Fig.4.2.12, panel B). **1** provided a slight ThT signal decrease, while the presence of **2** and **3** seems to have a faster effect in amyloid formation inhibition.

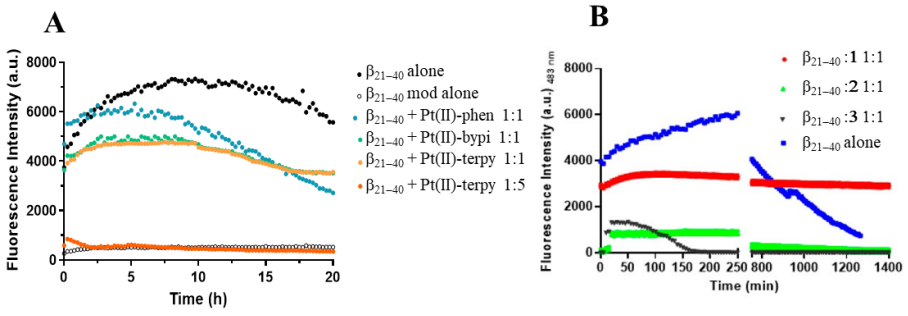

Fig. 4.2.12 Time course of ThT fluorescence emission intensity of $A\beta_{21\text{-}40}$ peptides alone and upon the incubation with (A) Pt(II)-phen, -bipy,-terpy complexes and (B) with **1**, **2**, **3** at indicated molar ratios.

The ESI-MS experiments revealed a 1:1 stoichiometry of the Aβ adducts formed with the metal complexes. In particular, all the Pt(II)- based compounds have

shown to exhibit an early recognition towards the Aβ peptide by substituting the monodentate chloride ligand. Over time, the peptide binding develops in a bidentate mode inducing the release of the chloride ion ligand.

Compared to the analog Pt(II) compound, the Pd(II)-based has shown the ability to bind the $A\beta_{21-40}$ directly in the bidentate mode. It can be explained because Pd(II) exchange ligands 10^5 times faster than their Pt(II) analogs[33].

Otherwise, the Au(III) compound was found not to bind the $A\beta_{21-40}$ peptide, as expected by the UV/Vis experiments in which the ligand field of this metal complex has not been influenced by the presence of the peptide, differently from the other compounds under investigation. However, an increase in the oxidized form of the peptide has been observed when incubated with the Au(III)-based complex. These data indicate that the compounds can use different mechanisms of action in the peptides' aggregation process. The modulation of the aggregation process of amyloid peptide by metal compounds depends on many factors that include, but are not limited to, oxidation state of the metal, redox potential, stability, strength of metal-bond, total charge of the complex.

4.2.4 Conclusions

Several metal-based compounds have been investigated to bind the $A\beta_{21-40}$ amyloidogenic peptide. In particular, the binding features were probed fixing ligands or metal ion centre. All the tested complexes were able to bind the peptide, except the Au(III)-based compound, in a 1:1 stoichiometry. The coordination of side chains of amino acids of the peptide to the Pt(II) ion occurs upon the release of the two or one coordinated Cl^- ions. On the other hand, the Au(III)-based complex may mediate the amyloid inhibition through the Met35 oxidation, suggesting a different mechanism of action.

The investigated compounds have demonstrated to modulate the overall

aggregation process of the peptide *in vitro*, strongly suggesting an inhibitory action, even though only future morphological data deriving from electronic microscopy investigations can definitively assess the inhibition of the formation of amyloid fibers. Thus, future studies carried out on the different full-length amyloid proteins could confirm the anti-aggregation properties of this class of complexes and its potential therapeutic application in neurodegenerative diseases. On the other hand, our data provide exciting information on the peptide/metal compound recognition process.

The results discussed in this chapter (including figures and tables) have been published in the following article:

(1) Florio, D.; **Iacobucci, I.*****; Ferraro, G.; Mansour, A. M.; Morelli, G.; Monti, M.; Merlino, A.; Marasco, D. Role of the Metal Center in the Modulation of the Aggregation Process of Amyloid Model Systems by Square Planar Complexes Bearing 2-(2'-Pyridyl)Benzimidazole Ligands. *Pharmaceuticals (Basel)* **2019**, *12* (4). https://doi.org/10.3390/ph12040154. ***Co-first author**

(2) L Manna, S. L.; Florio, D.; **Iacobucci, I.**; Napolitano, F.; Benedictis, I. D.; Malfitano, A. M.; Monti, M.; Ravera, M.; Gabano, E.; Marasco, D. A Comparative Study of the Effects of Platinum (II) Complexes on β-Amyloid Aggregation: Potential Neurodrug Applications. *Int J Mol Sci* **2021**, *22* (6). https://doi.org/10.3390/ijms22063015.

4.2.5 References

(1) Yoshikai, S.; Sasaki, H.; Doh-ura, K.; Furuya, H.; Sakaki, Y. Genomic Organization of the Human Amyloid Beta-Protein Precursor Gene. *Gene* **1990**, *87* (2), 257–263. https://doi.org/10.1016/0378-1119(90)90310-n.

(2) McNaull, B. B. A.; Todd, S.; McGuinness, B.; Passmore, A. P. Inflammation and Anti-Inflammatory Strategies for Alzheimer's Disease – A Mini-Review. *Gerontology* **2010**, *56* (1), 3–14. https://doi.org/10.1159/000237873.

(3) Hayne, D. J.; Lim, S.; Donnelly, P. S. Metal Complexes Designed to Bind to Amyloid-β for the Diagnosis and Treatment of Alzheimer's Disease. *Chem Soc Rev* **2014**, *43* (19), 6701–6715. https://doi.org/10.1039/c4cs00026a.

(4) Mjos, K. D.; Orvig, C. Metallodrugs in Medicinal Inorganic Chemistry. *Chem Rev* **2014**, *114* (8), 4540–4563. https://doi.org/10.1021/cr400460s.

(5) Son, G.; Lee, B. I.; Chung, Y. J.; Park, C. B. Light-Triggered Dissociation of Self-Assembled β-Amyloid Aggregates into Small, Nontoxic Fragments by Ruthenium (II) Complex. *Acta Biomater* **2018**, *67*, 147–155. https://doi.org/10.1016/j.actbio.2017.11.048.

(6) Suh, J.-M.; Kim, G.; Kang, J.; Lim, M. H. Strategies Employing Transition Metal Complexes To Modulate Amyloid-β Aggregation. *Inorg Chem* **2019**, *58* (1), 8–17. https://doi.org/10.1021/acs.inorgchem.8b02813.

(7) Mjos, K. D.; Orvig, C. Metallodrugs in Medicinal Inorganic Chemistry. *Chem. Rev.* **2014**, *114* (8), 4540–4563. https://doi.org/10.1021/cr400460s.

(8) Derrick, J. S.; Lee, J.; Lee, S. J. C.; Kim, Y.; Nam, E.; Tak, H.; Kang, J.; Lee, M.; Kim, S. H.; Park, K.; Cho, J.; Lim, M. H. Mechanistic Insights into

Tunable Metal-Mediated Hydrolysis of Amyloid-β Peptides. *J Am Chem Soc* **2017**, *139* (6), 2234–2244. https://doi.org/10.1021/jacs.6b09681.

(9) Florio, D.; Malfitano, A. M.; Di Somma, S.; Mügge, C.; Weigand, W.; Ferraro, G.; Iacobucci, I.; Monti, M.; Morelli, G.; Merlino, A.; Marasco, D. Platinum(II) O,S Complexes Inhibit the Aggregation of Amyloid Model Systems. *Int J Mol Sci* **2019**, *20* (4). https://doi.org/10.3390/ijms20040829.

(10) Gomes, L. M. F.; Bataglioli, J. C.; Jussila, A. J.; Smith, J. R.; Walsby, C. J.; Storr, T. Modification of Aβ Peptide Aggregation via Covalent Binding of a Series of Ru(III) Complexes. *Front Chem* **2019**, *7*, 838. https://doi.org/10.3389/fchem.2019.00838.

(11) Dasari, S.; Tchounwou, P. B. Cisplatin in Cancer Therapy: Molecular Mechanisms of Action. *Eur J Pharmacol* **2014**, *740*, 364–378. https://doi.org/10.1016/j.ejphar.2014.07.025.

(12) Weiss, R. B.; Christian, M. C. New Cisplatin Analogues in Development. A Review. *Drugs* **1993**, *46* (3), 360–377. https://doi.org/10.2165/00003495-199346030-00003.

(13) Pages, B. J.; Sakoff, J.; Gilbert, J.; Zhang, Y.; Kelly, S. M.; Hoeschele, J. D.; Aldrich-Wright, J. R. Combining the Platinum(II) Drug Candidate Kiteplatin with 1,10-Phenanthroline Analogues. *Dalton Trans.* **2018**, *47* (7), 2156–2163. https://doi.org/10.1039/C7DT04108J.

(14) Rillema, D. P.; Stoyanov, S. R.; Cruz, A. J.; Nguyen, H.; Moore, C.; Huang, W.; Siam, K.; Jehan, A.; KomReddy, V. HOMO–LUMO Energy Gap Control in Platinum(II) Biphenyl Complexes Containing 2,2′-Bipyridine Ligands. *Dalton Trans.* **2015**, *44* (39), 17075–17090. https://doi.org/10.1039/C5DT01891A.

(15) Nakamura, Y.; Taruno, Y.; Sugimoto, M.; Kitamura, Y.; Seng, H. L.; Kong, S. M.; Ng, C. H.; Chikira, M. The DNA Binding Site Specificity and Antiproliferative Property of Ternary Pt(Ii) and Zn(Ii) Complexes of Phenanthroline and N,N′-Ethylenediaminediacetic Acid. *Dalton Trans.* **2013**, *42* (10), 3337. https://doi.org/10.1039/c2dt32709k.

(16) Barnham, K. J.; Kenche, V. B.; Ciccotosto, G. D.; Smith, D. P.; Tew, D. J.; Liu, X.; Perez, K.; Cranston, G. A.; Johanssen, T. J.; Volitakis, I.; Bush, A. I.; Masters, C. L.; White, A. R.; Smith, J. P.; Cherny, R. A.; Cappai, R. Platinum-Based Inhibitors of Amyloid-Beta as Therapeutic Agents for Alzheimer's Disease. *Proc Natl Acad Sci U S A* **2008**, *105* (19), 6813–6818. https://doi.org/10.1073/pnas.0800712105.

(17) Cheignon, C.; Tomas, M.; Bonnefont-Rousselot, D.; Faller, P.; Hureau, C.; Collin, F. Oxidative Stress and the Amyloid Beta Peptide in Alzheimer's Disease. *Redox Biology* **2018**, *14*, 450–464. https://doi.org/10.1016/j.redox.2017.10.014.

(18) Wang, Y.; Xu, J.; Wang, L.; Zhang, B.; Du, W. Interaction of the Human Prion Protein PrP106-126 with Metal Complexes: Potential Therapeutic Agents against Prion Disease. *Chemistry* **2010**, *16* (45), 13339–13342. https://doi.org/10.1002/chem.201002207.

(19) He, L.; Zhu, D.; Zhao, C.; Jia, X.; Wang, X.; Du, W. Effects of Gold Complexes on the Assembly Behavior of Human Islet Amyloid Polypeptide. *J Inorg Biochem* **2015**, *152*, 114–122. https://doi.org/10.1016/j.jinorgbio.2015.08.020.

(20) Wang, Y.; Feng, L.; Zhang, B.; Wang, X.; Huang, C.; Li, Y.; Du, W. Palladium Complexes Affect the Aggregation of Human Prion Protein PrP106-126. *Inorg Chem* **2011**, *50* (10), 4340–4348. https://doi.org/10.1021/ic102331x.

(21) Wang, W.; Zhao, C.; Zhu, D.; Gong, G.; Du, W. Inhibition of Amyloid Peptide Fibril Formation by Gold-Sulfur Complexes. *J Inorg Biochem* **2017**, *171*, 1–9. https://doi.org/10.1016/j.jinorgbio.2017.02.021.

(22) Savelieff, M. G.; Nam, G.; Kang, J.; Lee, H. J.; Lee, M.; Lim, M. H. Development of Multifunctional Molecules as Potential Therapeutic Candidates for Alzheimer's Disease, Parkinson's Disease, and Amyotrophic Lateral Sclerosis in the Last Decade. *Chem. Rev.* **2019**, *119* (2), 1221–1322. https://doi.org/10.1021/acs.chemrev.8b00138.

(23) Pithadia, A. S.; Lim, M. H. Metal-Associated Amyloid-β Species in Alzheimer's Disease. *Curr Opin Chem Biol* **2012**, *16* (1–2), 67–73. https://doi.org/10.1016/j.cbpa.2012.01.016.

(24) Atrián-Blasco, E.; Gonzalez, P.; Santoro, A.; Alies, B.; Faller, P.; Hureau, C. Cu and Zn Coordination to Amyloid Peptides: From Fascinating Chemistry to Debated Pathological Relevance. *Coord Chem Rev* **2018**, *375*, 38–55. https://doi.org/10.1016/j.ccr.2018.04.007.

(25) Faller, P.; Hureau, C. Bioinorganic Chemistry of Copper and Zinc Ions Coordinated to Amyloid-β Peptide. *Dalton Trans.* **2009**, No. 7, 1080–1094. https://doi.org/10.1039/B813398K.

(26) Faller, P.; Hureau, C.; Berthoumieu, O. Role of Metal Ions in the Self-Assembly of the Alzheimer's Amyloid-β Peptide. *Inorg Chem* **2013**, *52* (21), 12193–12206. https://doi.org/10.1021/ic4003059.

(27) Alies, B.; Conte-Daban, A.; Sayen, S.; Collin, F.; Kieffer, I.; Guillon, E.; Faller, P.; Hureau, C. Zinc(II) Binding Site to the Amyloid-β Peptide: Insights from Spectroscopic Studies with a Wide Series of Modified Peptides. *Inorg. Chem.* **2016**, *55* (20), 10499–10509. https://doi.org/10.1021/acs.inorgchem.6b01733.

(28) Valensin, D.; Gabbiani, C.; Messori, L. Metal Compounds as Inhibitors of β-Amyloid Aggregation. Perspectives for an Innovative Metallotherapeutics on Alzheimer's Disease. *Coordination Chemistry Reviews* **2012**, *256* (19–20), 2357–2366. https://doi.org/10.1016/j.ccr.2012.04.010.

(29) Ma, G.; Huang, F.; Pu, X.; Jia, L.; Jiang, T.; Li, L.; Liu, Y. Identification of [PtCl2(Phen)] Binding Modes in Amyloid-β Peptide and the Mechanism of Aggregation Inhibition. *Chemistry* **2011**, *17* (41), 11657–11666. https://doi.org/10.1002/chem.201101859.

(30) Hou, L.; Kang, I.; Marchant, R. E.; Zagorski, M. G. Methionine 35 Oxidation Reduces Fibril Assembly of the Amyloid Abeta-(1-42) Peptide of Alzheimer's Disease. *J Biol Chem* **2002**, *277* (43), 40173–40176. https://doi.org/10.1074/jbc.C200338200.

(31) Bitan, G.; Tarus, B.; Vollers, S. S.; Lashuel, H. A.; Condron, M. M.; Straub, J. E.; Teplow, D. B. A Molecular Switch in Amyloid Assembly: Met35 and Amyloid Beta-Protein Oligomerization. *J Am Chem Soc* **2003**, *125* (50), 15359–15365. https://doi.org/10.1021/ja0349296.

(32) Di Natale, C.; Scognamiglio, P. L.; Cascella, R.; Cecchi, C.; Russo, A.; Leone, M.; Penco, A.; Relini, A.; Federici, L.; Di Matteo, A.; Chiti, F.; Vitagliano, L.; Marasco, D. Nucleophosmin Contains Amyloidogenic Regions That Are Able to Form Toxic Aggregates under Physiological Conditions. *The FASEB Journal* **2015**, *29* (9), 3689–3701. https://doi.org/10.1096/fj.14-269522.

(33) Kuznetsov, M. L.; Kukushkin, V. Y.; Pombeiro, A. J. L. Reactivity of Pt- and Pd-Bound Nitriles towards Nitrile Oxides and Nitrones: Substitution vs. Cycloaddition. *Dalton Trans* **2008**, No. 10, 1312–1322. https://doi.org/10.1039/b713425h.