UNIVERSITÀ DEGLI STUDI DI SALERNO

UNIVERSITÀ DEGLI STUDI DI SALERNO
Dipartimento di Farmacia

PhD Program
in **Drug Discovery and Development**
XXXII Cycle — Academic Year 2019/2020

PhD Thesis in

Interactome analysis of bioactive molecules: optimization of a functional proteomics platform

Candidate

Elva Morretta

Supervisor(s)

Prof. *Agostino Casapullo*

PhD Program Coordinator: Prof. Dr. *Gianluca Sbardella*

Alla mia famiglia.

Table of Contents

Abstract

The identification of natural products (NPs) target proteins is pivotal to understand their mechanism of action, in order to develop molecular probes and/or potential drugs. In the last 15 years, affinity chromatography-coupled to mass spectrometry (AP-MS) has been the top-choice technique in the Drug Target Deconvolution field, having brought brilliant results in the *targetome* profiling of a multitude of bioactive compounds.

Unfortunately, since a chemical modification of the molecule to be investigated is mandatory, AP-MS is not suitable for compounds that do not exhibit properly reactive structural features.

Furthermore, in the absence of information on the structure-activity relationships (SAR) of the analyte (e.g. in the case of natural products, often isolated in a too low amount for this investigation), the covalent modification would leave some concerns about whether or not the molecule original bioactivity is retained.

In this scenario, my PhD project was devoted to the development of an alternative Functional Proteomics Drug Target Deconvolution platform, based on two complementary *label-free* methods avoiding any chemical modification of the molecules for a proteome-wide profiling of their *targetome*.

Indeed, this platform takes advantage of both Drug Affinity Responsive Target Stability (DARTS), useful to identify the target protein(s) of a bioactive molecule, and targeted Limited Proteolysis coupled to Multiple Reaction Monitoring Mass Spectrometry (t-LiP-MRM), for the characterization of the interaction features between a protein and its ligand, enlightening the protein(s) region(s) directly or distally influenced by the NP binding.

After preliminary optimization steps, this *label-free* platform allowed the *interactome* characterization of three structurally different marine metabolites: the norcembranoid 5-epi-Sinuleptolide interacting with actins, as also validated by the AP-MS approach, the sulfated bis-steroid Crellastatin A interacting with poly [ADP ribose] polymerase 1 and the acetogenin BrACG, interacting with both the cytosolic and mitochondrial Hsp70 isoforms.

Moreover, the DARTS results were validated through Western Blotting, while the t-LiP-MRM evidences were corroborated by Molecular Docking, and i*n vitro*/*in cell* activity assays were also performed, to assess the bioactivities of the marine metabolites.

CHAPTER 1

Introduction.

1.1 Introduction

The dynamic and irreplaceable role of certain biomolecules in supporting life has been documented since the very early stages of biological research to the point that in 1838 Berzelius named them *proteins*, a word originating from the Greek *proteios* and meaning *first rank*. Indeed, proteins catalyze and control all cellular processes, forming a highly structured and complex entity known as *proteome,* which dynamically adapts to a series of external and internal (i.e. genetic) stimuli or perturbations, thereby defining the cell functional state and phenotype.

The term *proteome* was first used in 1996 by Marc Wilkins to denote the *PROTEin complement of a genOME*[1], a definition that has later on been deeply expanded to better feature the proteome complexity (Figure 1), indicating the overall protein content of a cell, with regard to protein localization, interactions, post-translational modifications (PTMs) and turnover at a particular time. Describing and understanding those properties has been, and still is, the key challenge to obtain a global and integrated view of cellular activities, disease processes and networks.

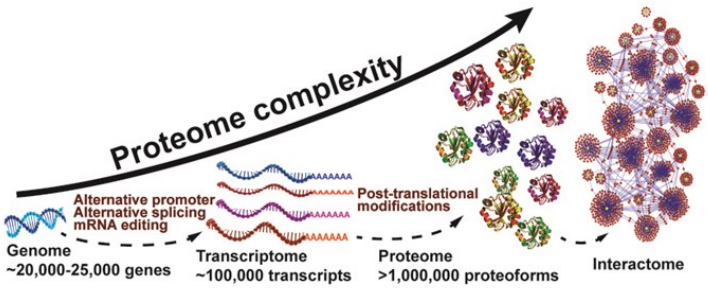


Figure 1: *Proteome complexity overview.*

In the last 20 years, mass spectrometry has emerged as a core tool for this large scale protein analysis, hustling the development and progress of a multidisciplinary science named

proteomics. Integrating synthetic organic chemistry, biochemistry, molecular biology and mass spectrometry, proteomics aims to characterize protein expression, interactions, PTMs, cellular localization and turnover as a function of time, space and cell type.

Eukaryotic cells proteomes are fairly complex and exhibit an extensive dynamic range in terms of quantity, molecular size and hydrophobicity/hydrophilicity of their constituents[2]. Thus, a rapid advance in the resolution, mass accuracy, sensitivity and scan rate of mass spectrometers and the introduction of hybrid mass analyzers (e.g. Linear Ion Trap/Orbitrap series[3,4]) have significantly improved proteomic analysis, together with the development of new strategies for peptides/proteins fractionation, quantification and data analysis.

Protein characterization can be achieved following two main paths, the *bottom-up* and *top-down* approaches[5] (Figure 2).

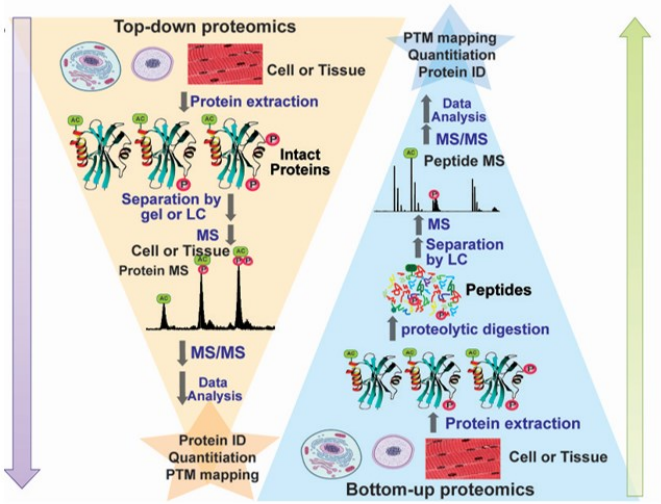


Figure 2: *Top-down and Bottom-up proteomics pipelines.*

The *top-down* strategy[6–8] allows the direct analysis of intact proteins, which are introduced into the mass spectrometer through electrospray ionization (ESI) and subsequently fragmented to obtain both their molecular mass and fragment ions. This analysis provides a complete description of the protein primary structure and modifications, thus becoming a precious tool

for the analysis of PTMs. Furthermore, improvements on both *non-denaturing* separation methods and fragmentation techniques, as well as mass spectrometers advances on transferring and detecting high mass/charge (i.e. m/z) ions, have brought native mass spectrometry and *top-down* proteomics together, to investigate *native-like* quaternary structures of up to several mega-Dalton complexes[9–11]. In fact, whereas conventional denaturing *top-down* MS can determine the mass of the proteins constituting a particular complex, native MS can be used to measure the mass of protein complexes and to address questions about their stoichiometry. Furthermore, this approach proved to be ideal to study both subunits assembly and exchange into complexes and also to monitor their activity over time[12].

Unfortunately, despite the above described potentialities and tremendous advances in instrumentation, *top-down* proteomics is not commonly employed because of issues in terms of sample handling, protein samples heterogeneity and transient or weak interactions of certain protein subunits in specific complexes[12].

For all of these reasons the *bottom-up* approach, which recovers proteins structural information from the analysis of their peptides, represents the currently most employed approach for the high-throughput identification and quantitation of proteins in a great variety of samples[13].

In a typical *bottom-up* workflow, proteins are submitted to an enzymatic digestion, typically with trypsin, and the resulting peptides analyzed through tandem mass spectrometry. Peptides identification is achieved by comparing mass spectra derived from their fragmentations with the theoretical ones generated from *in silico* digestion of a protein database. Protein inference is then accomplished by assigning peptide sequences to proteins.

Alongside *top-down* and *bottom-up*, a so-called *middle-down* approach has also become popular, since it benefits from the undeniable advantages of both the above presented strategies. *Middle-down* proteomics workflows exploit a limited digestion carried out through enzymes like Glu-C or Asp-N, leading to peptides that are significantly larger than the tryptic ones. This

makes *middle-down* suitable for studies requiring both an extensive protein sequence coverage and the localization of covalent modifications, such as PTMs and single-point mutations[14,15], which are largely preserved in this analytical strategy.

1.2 Bottom-up proteomics & tandem mass spectrometry

Bottom-up proteomics is widely employed for identifying proteins and determining details of their sequence. As briefly mentioned in the previous paragraph, in this approach proteins are typically digested with trypsin to obtain peptide mixtures that are then separated through liquid chromatography, introduced into a mass spectrometer through a soft ionization method such as ESI, fragmented, sequenced and used to identify the parent proteins via database searching. When the digestion is performed on a complex mixture of proteins without a prior fractionation step, the method is called *shotgun proteomics*[16–18], as defined by Yates *et al*., because of its analogy to shotgun genomic sequencing[19].

In contrast to mass-spectrometric strategies like peptide mass fingerprint[20] (PMF), suitable for simplified peptide mixtures and being based on the comparison of the tryptic peptide ions m/z values with a database to identify the corresponding proteins, the analysis of complex samples needs tandem mass spectrometry to achieve peptides fragmentation, necessary to disclose the amino acid sequence of a peptide[21].

The rate at which a mass spectrometer performs fragmentation is called MS/MS duty cycle, an important determinant of sampling depth and dynamic range, that is the largest challenge of the proteomics technology development.

Even if *state of the art* high duty cycle instruments can fragment hundreds of peptides in a single experiment, a typical biological sample contains a huge quantity of tryptic peptides, which largely co-elute in the chromatographic step. This phenomenon would quickly overwhelm the MS/MS acquisition speed of even the highest duty cycle instruments and lead

to a preferential fragmentation of high abundance peptides, introducing, as a consequence, an overall sequencing bias towards the highly expressed proteins, at the expense of the less abundant ones[22].

For all of these reasons, in order to save the depth of a *bottom-up* procedure, fractionation strategies at the protein or peptide level are crucial, prior to submit the sample to the mass spectrometric analysis (Figure 3).

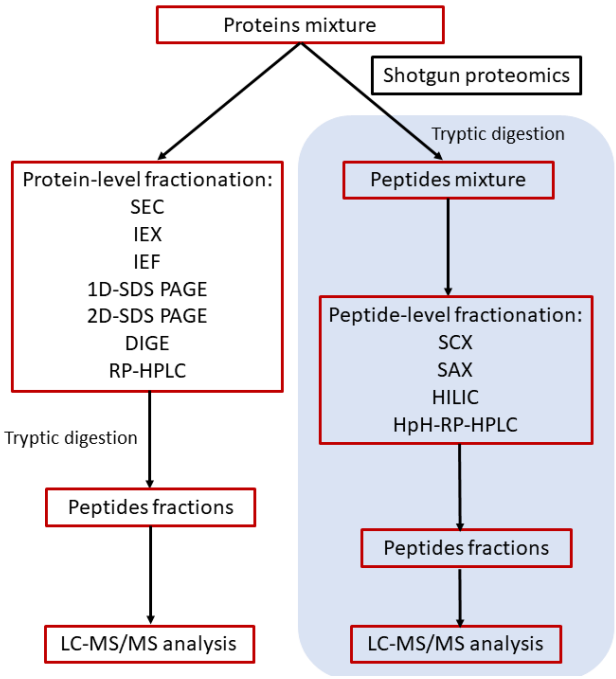


Figure 3: *Bottom-up proteomics & proteins/peptides fractionation strategies.*

A variety of separation modes have been employed to achieve protein level separation, including size exclusion chromatography[23] (SEC), ion exchange chromatography[24] (IEX), isoelectric focusing[25,26] (IEF), 1D gel separation[27,28] (1D SDS-PAGE), 2D gel separation (2D SDS-PAGE)[29,30] and reverse phase chromatography[31,32] (RP-HPLC).

Among them, *gel-based* fractionation methods are the most exploited.

In a typical 2D SDS-PAGE experiment, proteins are first separated on the basis of their isoelectric point (pI) through isoelectric focusing (IEF), and then by their molecular weight.

Briefly, the sample is loaded onto polyacrylamide gels or pH gradient strips, then an electrical field is applied, allowing proteins to migrate through the pH gradient until they reach their pI, where they stop due to the grasped neutral state of charge. After this step, proteins are treated with sodium dodecyl sulfate (SDS), which confers them a negative charge proportional to their molecular weight and allows their separation when an electric potential is applied. Protein spots can be detected by means of visible (e.g. silver staining, Coomassie Blue) or fluorescent stains and digested *in situ*[33,34]. 2D-SDS-PAGE has a good resolving power but is time consuming, requires large sample amount[30] and in case of many samples to run (e.g. comparison of proteins from different proteomes/experiments) additional problems related to the *gel-to-gel* variability can emerge. To overcome this limitation, two-dimensional differential gel electrophoresis (2D-DIGE) has been developed, in which proteins from several experimental conditions are previously tagged with different fluorescent dyes and then run on the same gel and quantified by detecting dyes intensities at different wavelengths[35–37].

Despite its undoubtable advantages over the classical 2D-PAGE approach, 2D-DIGE also has its drawbacks, in terms of the need to pre-label the protein samples and because is a time-consuming technique.

For these reasons 1D-SDS-PAGE, in which proteins are separated only according to their molecular weight, is largely preferred and commonly exploited as a protein fractionation strategy. Despite this evidence, *gel-based* fractionation methods still suffer of some drawbacks, mainly related to a loss of sample after the *in situ* tryptic digestion, during the peptides extraction from the gel slices. Therefore, efforts oriented to build-up *gel-free* methods to analyze peptide mixtures without an upstream protein-level separation brought up *shotgun* proteomics. As already mentioned, in this procedure the protein sample is directly subjected to proteolytic digestion, usually with both Lys-C and trypsin, generating a very complex mixture of peptides, which needs to be properly fractionated through two orthogonal liquid

chromatography strategies, to tackle dynamic range issues prior to the mass spectrometric analysis.

Several separation techniques have been implemented for this purpose, including strong cation exchange[38] (SCX), strong anion exchange[39] (SAX), hydrophilic interaction liquid chromatography[40,41] (HILIC) and High pH fractionation[42] (HpH-RP).

Among them, the most commonly employed technique is SCX chromatography, in which the separation is achieved exploiting the net charge of the peptides in solution. Indeed, tryptic peptides present a lysine or arginine residue at their C-terminus and one free amine group at their N-terminus: at a low pH, most of the acidic residues will be neutral and peptides will possess two positive charges. Thus, SCX stationary phase is negatively charged to capture the positively charged peptides. A typical SCX chromatogram of a tryptic digest shows a very intense peak corresponding to the largely co-eluting doubly charged peptides population, followed by peaks of triply or higher charged peptides, resulting from trypsin missed cleavages or histidine containing peptides.

Nevertheless SAX, which exploits an anion exchanger as stationary phase, and HILIC, which in turn uses a polar stationary phase to retain peptides on the basis of their hydrophilicity, have been reported as valid alternatives to SCX chromatography (Figure 4).

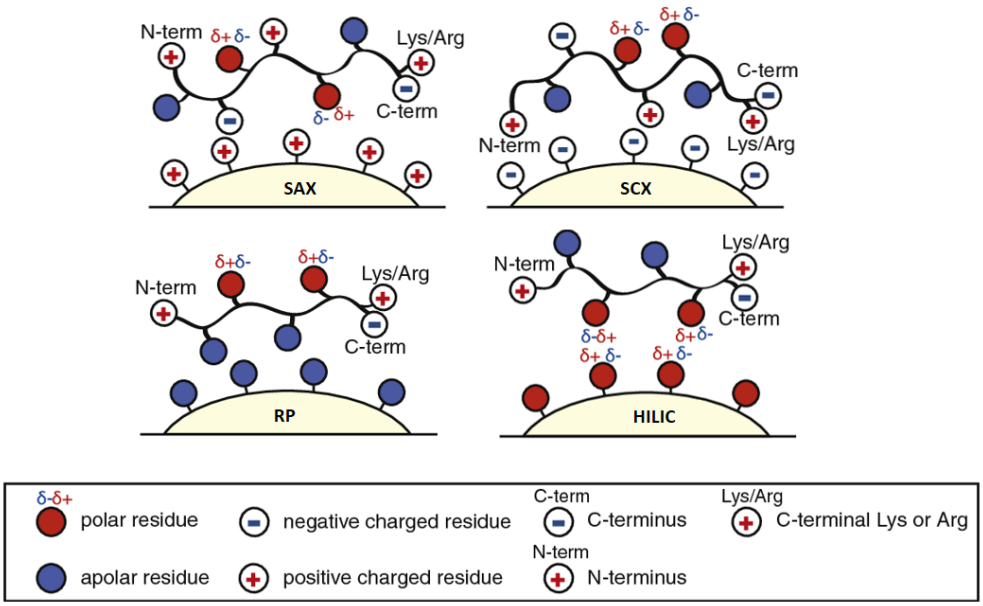


Figure 4: *Schematic representation of SAX, SCX, RP and HILIC liquid chromatography peptide separation principles.*

Furthermore, High-pH (HpH) fractionation has recently pushed through the other chromatographic separation strategies, showing great potential for the separation of complex peptide mixtures. This method is based on the same principle as RP-HPLC but exploits an alkaline mobile phase (pH=10) to improve its resolving power (Figure 5).

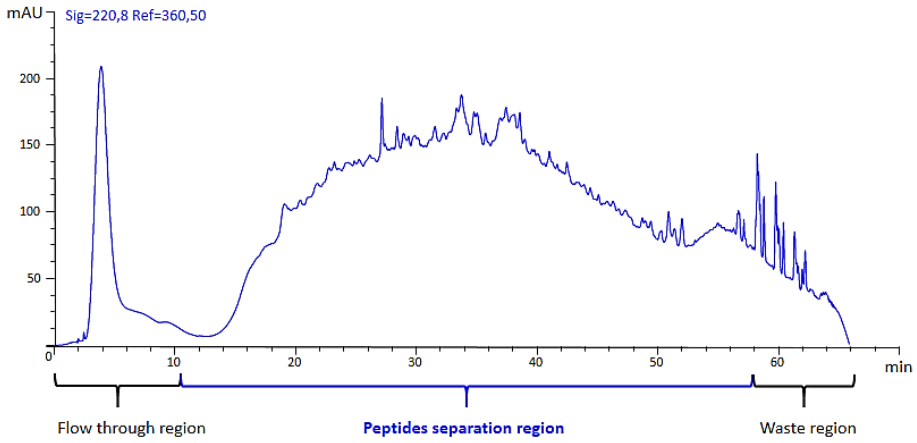


Figure 5: *HpH fractionation output chromatogram.*

The tryptic peptides mixtures obtained after digestion, whether they have been fractionated at the protein or at the peptide level, are then submitted to ultra-pressure liquid chromatography coupled to mass-spectrometry (UPLC-MS). At this stage, the liquid chromatography step is usually performed in the nano-flow scale, to improve both peptide separation and sensitivity, through columns with a particle size in the range of 1 μm interfaced with the nano-electrospray (nESI) sources of high resolution mass spectrometers[43].

Indeed, in recent years multiple hybrid instruments have been developed with different mass analyzers (e.g. Linear Ion Trap [LIT]/Orbitrap tandem mass analyzers), ion optics and fragmentation sources, improving the accuracy of peptide precursors and fragment ions masses measurements through various fragmentation methods.

The most common and robust fragmentation strategy used for peptide analysis is collision induced/activated dissociation[44] (CID/CAD), which can be performed both in the collision cell of a tandem in space analyzer and in the ion trap of a tandem in time one[45]. In this fragmentation strategy, peptides are excited by an electric field that increases their kinetic energy and activated via multiple collisions with an inert gas (i.e. helium, nitrogen, argon or xenon): the rapid vibrational redistribution of the peptides internal energy determines an event that can be compared to a sort of *internal heating*. When the internal energy reaches the fragmentation threshold, the weakest bonds are cleaved, preferentially along the single C-N amide bond[46], thus obtaining *b* (N-terminal) and *y* (C-terminal) ions, according to the Roepstorff-Fohlman nomenclature[47], which are the most observed in the resulting MS/MS spectrum (Figure 6).

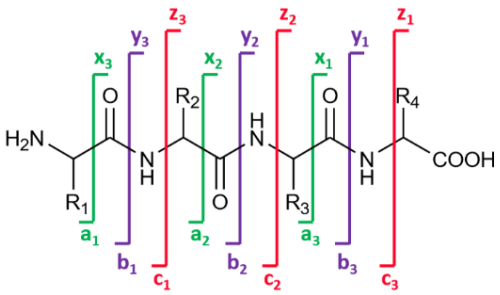


Figure 6: *Roepstorff and Fohlman peptide ion fragmentation nomenclature. Three types of fragments can be observed, leading to N-terminal (a, b and c) and C-terminal (x, y and z) ions.*

The efficiency of CID fragmentation depends on peptide length, amino acid composition and PTMs, thus it is sub-optimal in case of heavy peptides, for which the energy has to be distributed along a large number of amide bonds, and post translationally modified peptides, since internal fragmentation and loss of PTMs frequently occur[48].

For those reasons, other types of fragmentations, named electron transfer dissociation[49] (ETD) and high collision dissociation[50] (HCD), have been developed (Figure 7).

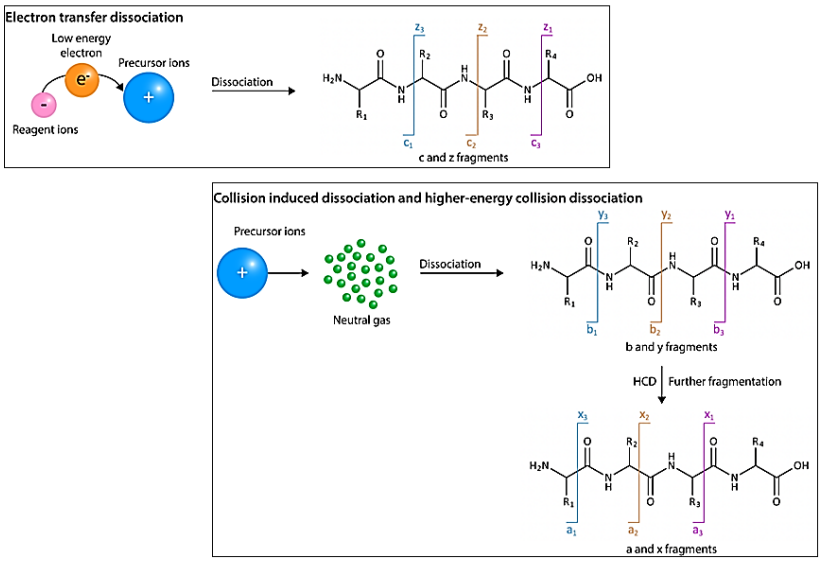


Figure 7: *Commonly exploited fragmentation strategies mechanisms.*

Once the full MS and the MS/MS spectra have been generated, the obtained fragmentation profiles are matched with the theoretical ones retrieved from web databases (i.e. SwissProt or NCBInr), through a number of different search engines and algorithms, such as Mascot[51], Sequest[52] or Andromeda[53]. Those algorithms compare the calculated masses of the precursor ions to a theoretical mass, within a defined mass window dependent on the accuracy and precision of the MS analyzer. In a second step, the MS/MS spectra of the selected precursor ions are matched against theoretical MS/MS spectra generated by the engine itself through an *in silico* digestion of all the protein sequences of the database. By using an appropriate scoring algorithm, the closest matches between theoretical and experimental spectra can be identified to accomplish protein inference by assigning the retrieved peptide sequences to the protein(s) to whom they better match.

To assess the confidence of peptide spectrum matches (PSMs), reducing the risk of false positives, several statistical tools are employed. For instance, estimation of the false discovery rate (FDR) can be done by a second search using a reversed or *decoy* database[54]. If a PSM is matching against the peptides generated in this database, it is considered as a false positive (i.e. decoy) PSM. The false discovery rate is obtained dividing the number of decoy PSMs by the number of total PSMs[55] and is usually set at 1%. However, despite the increasing number of scientific works assessing the relevance of these tools, database search algorithms and statistical reliability are still an open field in proteomics research.

1.3 Bottom-up Proteomics & Drug-Target Deconvolution

The majority of the employed drugs acts on protein targets that, being embedded in cellular pathways and networks, are physically and functionally interconnected with many other proteins and cellular components.

Among all the therapeutically relevant drugs, natural products (NPs) have long been explored as an invaluable source for drug discovery, primarily due to their chemical diversity (i.e. a large range of molecular weights) and complexity (i.e. a significant number of stereo-specific carbon centers) which offer a range of uncharted *chemotypes*[56–58], often featuring biologically relevant scaffolds and pharmacophore patterns that have naturally evolved as preferred ligand–protein binding motifs[59,60].

Historically, drugs were discovered on the basis of their ability to elicit a certain biological outcome, often screening crude natural products mixtures and following the retained bioactivity during the purification steps, until the identification of the active component(s) and its (their) mechanism of action could be accomplished[61].

The idea that a drug might specifically interact with a suitable protein target to yield a biological effect caused a change in the drug discovery field, placing the emphasis on finding suitable *druggable* proteins for a particular disease and then developing high-throughput screening assays, to discover potential hits. Unfortunately, while these pretty artificial *in vitro* assays tend to yield small molecule hits, a following *in vivo* significant biological activity is often not retrieved.

Indeed, small molecules should be expected to bind to more than one protein: the one-drug-one-target paradigm based on the *magic bullet drugs* concept has been progressively replaced by the *magic shotguns* one, in which poly-pharmacology is achieved through *selectively non-selective molecules*, capable to act in the cell at multiple levels.

In a physiological context, a molecule binding more than one protein might show desirable or undesirable consequences. Knowledge about the spectrum of proteins interacting with a small molecule could, in an early stage, inform about drug safety (i.e. by the identification of potential *off-targets* and their related toxicity effects), help in the decision making along the development process (e.g. which lead series to prioritize), and even lead to the repositioning of existing drugs (e.g. by the identification of additional targets).

In a later stage, knowing the drug multiple targets would enable a drug either to be therapeutically applied in several unrelated diseases or, as for the harmful side effects, to use adjunct therapies or dose adjustments.

Therefore, in addition to carefully determining the molecule primary target(s), it is important to identify a drug target spectrum as thoroughly as possible, to exploit its full therapeutic potential and minimize the toxicity caused by the drug *off-targets*.

The process aiming at the identification of the full spectrum of protein targets associated with a bioactive molecule and their induced cellular phenotype is called *Drug-Target Deconvolution*.

To date, because of their versatile characteristics, *bottom-up* proteomics strategies are considered pivotal techniques in this process. These approaches can be broadly categorized into two types: *global proteomics*, directed at a cell-wide characterization of proteomes or their PTMs upon different physiological/pathological conditions, internal/external stimuli or drug treatments, and *functional proteomics*, devoted to the characterization of the *interactome* of small molecules, in both activity and affinity-based fashions.

1.3.1 Global Proteomics

In view of the Drug-Target deconvolution process, global proteomics serves two different purposes: (1) the identification of proteins associated with a specific disease or pathological condition and their disrupted or altered biological pathways, that can be targeted with drugs[62], and (2) the evaluation of a cell proteome and/or its PTMs modifications upon drug treatment[63]. In both cases, a whole-cell and/or organ-wide proteome analysis can be carried out, in an unbiased nature, through *shotgun proteomics* approaches coupled with protein quantitation strategies.

Quantitative mass spectrometry strategies can be grouped into *label-free* and *label-based* approaches.

Label-free strategies have become increasingly popular in recent years because of their ease and feasibility to a wide range of proteomics workflows. They are generally based upon two distinct approaches (Figure 8): (1) measuring the MS signal intensity of peptide precursor ions associated with a given protein[64] (i.e. peptide ion intensity strategy) and (2) counting the total number of MS/MS spectra that map to peptides of a given protein[22,65] (i.e. spectral counting [SC] strategy).

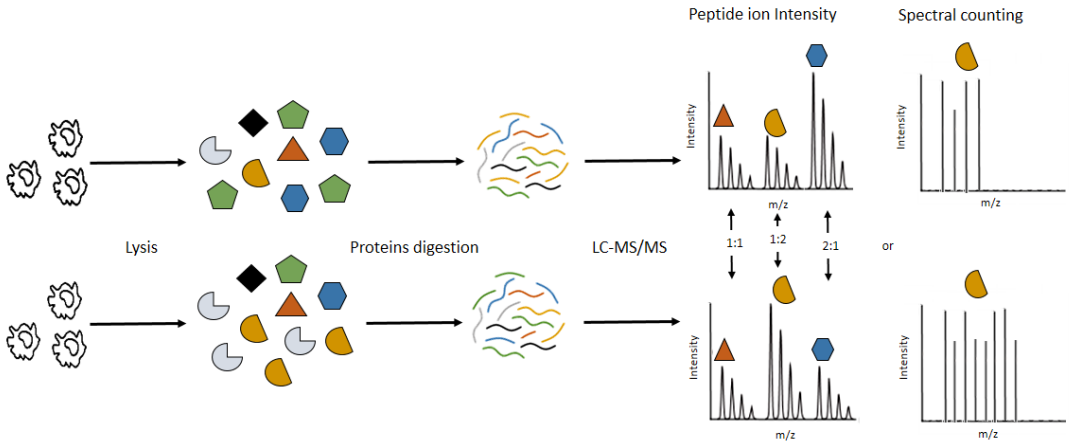


Figure 8: *Label-free proteomics approaches.*

Both of these techniques are usually performed with Data Dependent Acquisition (DDA)-based mass spectrometric methods. More in details, the mass spectra of all the ion species that co-elute at a specific point of the LC gradient (i.e. precursor ions spectra) are recorded in a survey MS-level experiment (i.e. full-scan). The mass spectrometer alternates between the acquisition of full-scan data and fragment-ion spectra, in which precursors are sequentially isolated and fragmented. In a typical *top N* experiment, with N denoting the number of MS/MS spectra to be acquired, a full scan is followed by about ten fragment-ion scans[66].

In the peptide ion intensity method, peak areas corresponding to a set of peptides coming from each protein are integrated, and their values are compared among different samples: the intensities of chromatographic peaks can be measured in the extracted ion chromatogram (XIC) and the area under the curve (AUC), correlating with protein abundances[67], can be obtained. This approach has several limitations: even though a higher number of MS/MS experiments improves the chance of peptide sequencing and consequently increases the number of identified proteins, a multiple sampling of the chromatographic peak by survey mass spectra is required for an accurate estimation of protein abundance by ion signal intensities, at the expense of the MS/MS experiments. Thus, the optimization of a mass spectrometric method for protein abundance estimation via ion signal intensities can give a lower number of identified proteins. To avoid this *quantitation versus identification* issue, data-independent acquisition (DIA) could be used instead of DDA. Indeed, in a DIA method the mass spectrometer does not cycle between MS and MS/MS modes for peptide identification, but rapidly alternates low and high levels of collision energy, simultaneously providing precursor and product ions data for all the isotopes of every charge state across the full chromatographic peak width[68–70]. Furthermore, signal-intensity calculation requires complex optimization and normalization routines to identify corresponding peptides across different experiments and to reduce background signals and/or signals from co-eluting compounds[69].

Quantitation based upon SC (spectral counting) is a more straightforward process: in its simplest form, a relative abundance estimation is obtained by counting the total number of peptides identified for a given protein. Depending on the number of MS/MS spectra corresponding with high confidence to a particular peptide, there is no *quantitation versus identification* problem linked to this method, so that its optimization also favors proteins identification. Furthermore, to allow the estimation of the absolute content of a certain protein between different samples, the so-called exponentially modified Protein Abundance Index (emPAI) can also be exploited[71]. This index is calculated on the basis of the empirical

relationship between the number of spectra identified for a given protein and the overall protein abundance in a sample, dividing the number of observed spectra for each protein out of the number of possibly observable peptides.

The *label-based* methodologies exploit stable isotope labeled molecules to distinguish the same analyte in different samples, so that they can be mixed and analyzed in a single LC-MS run.

These methods are traditionally grouped into metabolic and chemical labeling strategies.

The most exploited metabolic approach is Stable Isotope Labeling by Amino acid in Cell culture[72,73] (SILAC), in which essential (i.e. not produced by mammalian cells) auxotrophic isotope-labeled amino acids are introduced in the growth media of cultured cells, in order to be incorporated in newly synthesized proteins. In common practice, fully labeled $^{13}C_6{}^{15}N_2$-lysine and $^{13}C_6{}^{15}N_4$-arginine are used in combination so that all peptides arising from trypsin digestion, except for those at the C terminus of the protein, can be systematically quantified. Since there is no significant chemical difference between the labeled and the natural amino acids, labeled and unlabeled cells will behave exactly in the same way. In a typical SILAC workflow, cells are differentially labeled, treated or not with a drug and then proteins are extracted from both samples and mixed before the enzymatic digestion and the 2D-LC-MS/MS analysis. The quantitation takes place at the peptide mass spectrum level (Figure 9).

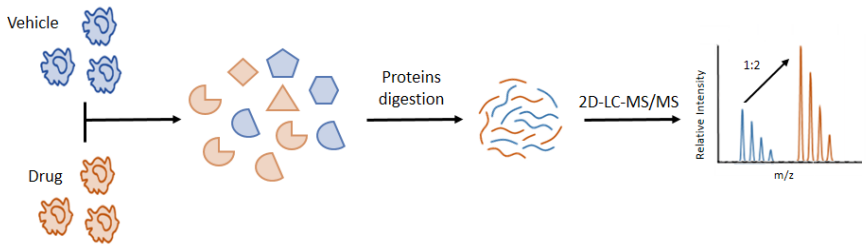


Figure 9: *SILAC workflow.*

Despite the fact that this method allows to mix the sample upstream at the protein level, thus removing the *sample-to-sample* variable handling issue, SILAC has some limitations in requiring long incubation times to fully incorporate the labels and in also being a very expensive technique.

Among the chemical labeling strategies, isotope-coded affinity tag[74] (ICAT) makes use of peculiar reagents to introduce stable isotope in a protein sample. Briefly, the ICAT reagent consists of three elements: an affinity tag (i.e. biotin) to isolate ICAT-labeled peptides, an oxyethylene linker containing stable isotope signatures to be differentiated by mass spectrometry and a reactive group with specificity towards the cysteine thiol moiety. The reagent exists in two forms: heavy and light. In the original ICAT reagent, the linker heavy form contains eight deuteriums and the light one contains only hydrogens, resulting in an 8 Da mass difference between the two tags (Figure 10).

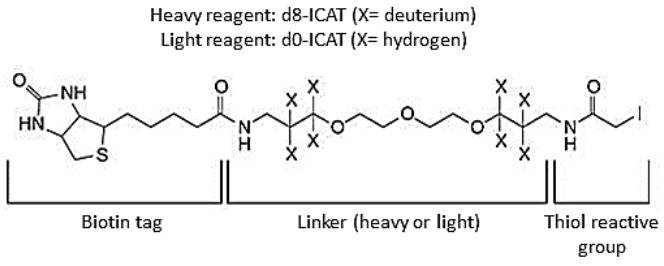


Figure 10: *Schematic representation of an ICAT tag.*

In a new version of cleavable ICAT reagents, ^{13}C is used in the linker instead of deuterium, resulting in a mass difference of 9 Da between heavy and light forms. Incorporation of ^{13}C instead of deuterium promotes co-elution of heavy and light ICAT-labeled peptides that displayed partial separation on reverse-phase chromatography as a result of an isobaric shift in the original ICAT reagents. Furthermore, since the biotin group might interfere with database

search, this cleavable ICAT reagent also presents an acid-cleavable linker connected to the biotin tag, for its removal prior to the MS step.

In the ICAT strategy, two differentially treated protein mixtures are labeled with light and heavy ICAT reagents, which covalently attach to the cysteine residues of every protein. The two labeled mixtures are then combined and proteolyzed with trypsin, so that ICAT-labeled peptides can be purified by biotin-streptavidin affinity chromatography. If cleavable ICAT reagents are used, the biotin tag is released by acid cleavage after the affinity chromatography step. The purified ICAT-labeled peptides are then analyzed through LC-MS/MS: each pair of light and heavy labeled peptides are chemically identical and co-fractionate, but present a mass difference of 8 Da for the original ICAT or 9 Da for the cleavable ICAT, which is exploited for the relative quantification step, where the ratio of peptide pairs signal intensities is determined.

Stable Isotope Dimethyl Labeling (SIDL) has been introduced in 2003 as a straightforward, fast and inexpensive strategy for high-throughput proteomics experiments[75].

In this method, all primary amines (i.e. the N-terminus and the side chain of lysine residues) in a peptide mixture are converted into dimethylamines, with the only exception being N-terminal prolines, for which monomethylamines are formed. More in details, SIDL is based on the reaction of a peptide primary amine with formaldehyde to generate a Schiff base that is rapidly reduced by the addition of cyanoborohydride to the mixture. The combination of regular formaldehyde and cyanoborohydride generates a mass increase of 28 Da per primary amine on a peptide (light label), whereas using deuterated formaldehyde a mass increase of 32 Da per primary amine is obtained (intermediate label), and combining deuterated and ^{13}C-labeled formaldehyde with cyanoborodeuteride brings to a mass increase of 36 Da (heavy label). These stable isotope dimethyl labels can be employed to differentially label up to three different samples (Figure 11).

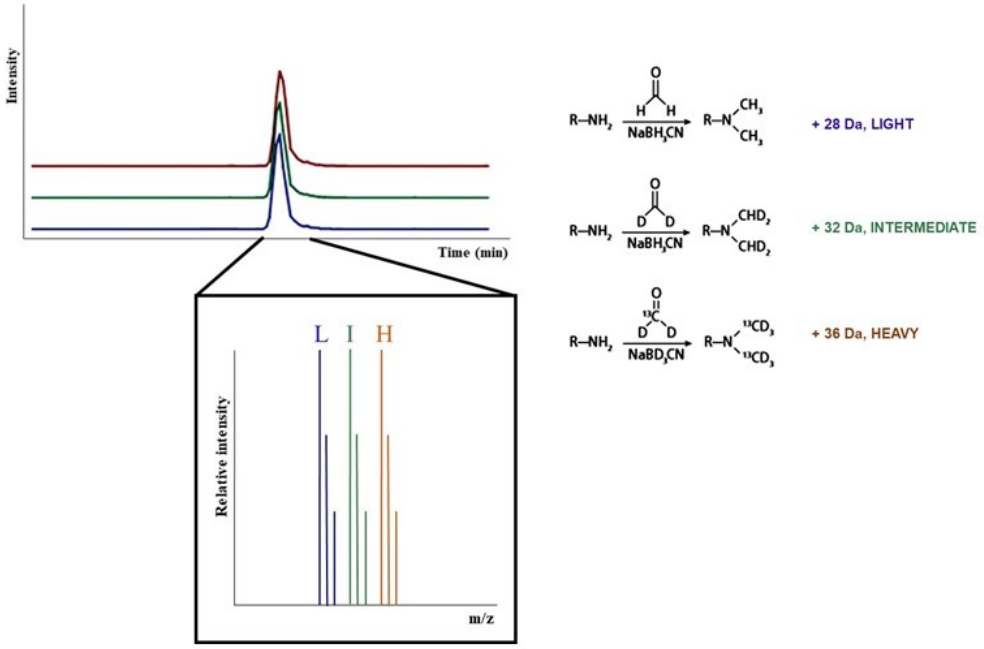


Figure 11: *Stable Isotope Dimethyl Labeling.*

In the SIDL workflow, differentially treated samples are digested and the obtained peptides are separately labeled with either of the dimethyl *isotopomers* prior to mixing the samples: as the different isotopes do not affect the labeled peptides behavior in LC-MS, they can be recognized by their known mass difference and quantification can be performed by comparing the signal intensity of the differentially labeled peptides at the MS level.

Indeed, when labeling tryptic peptides, most of them will differ in mass by either 4 Da (when cleaved after an arginine residue because only the N terminus is labeled) or 8 Da (when cleaved after a lysine residue, since both the N terminus and the lysine side chain amino group are labeled) between the light and intermediate and between the intermediate and heavy labels (Figure 12).

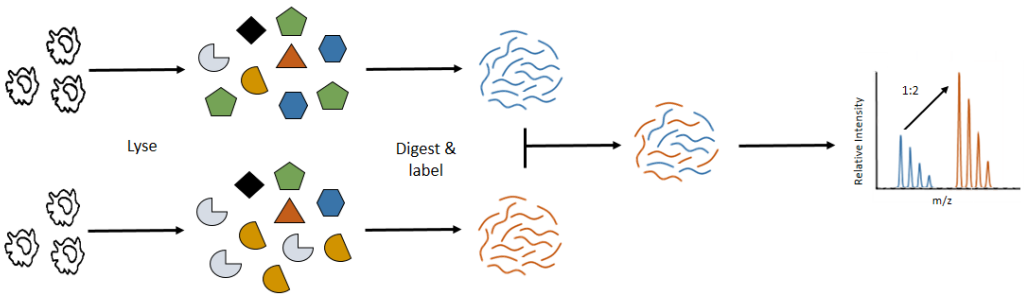


Figure 12: *SIDL workflow.*

SIDL has the huge advantage of having been optimized to be performed in three different modes[76] (i.e. in solution, on column and on line), according to the samples to be labeled. Furthermore, it is a very fast and inexpensive technique.

In the attempt to increase the number of samples that can be pooled together before the MS analysis, alternative techniques have been developed, such as isobaric Tags for Absolute and Relative Quantification[77] (iTRAQ) and Tandem Mass Tag[78] (TMT) that, in contrast to all of the above presented strategies, rely on a MS/MS level quantification step.

Both iTRAQ and TMT exploit *isotopomeric* reagents consisting of an amine-reactive group (i.e. an N-hydroxysuccinimide ester), an N-methylpiperazinic mass reporter for the quantification and a mass normalizer group called balancer.

The iTRAQ and TMT tags have been developed so that the total mass of the reporter region and the balancer is the same in all of their versions, with the individual weights of the counterparts changing. This means that same peptides from different samples are isobaric, chromatographically indistinguishable and yield a single peak in the mass spectrum. The relative abundances of the tagged peptides can be revealed during MS/MS events, when the tags are fragmented to release their reporter ions presenting different masses[79]: the different

mass reporter ions abundances reflect the amount of the same peptide in different experimental conditions (Figure 13).

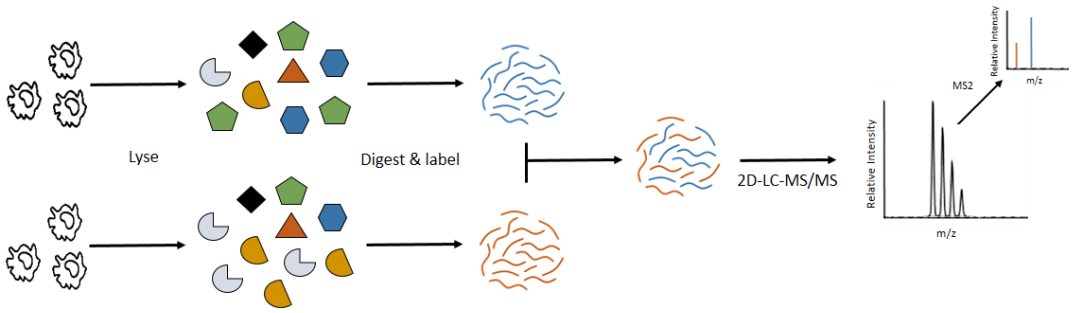


Figure 13: *iTRAQ/TMT workflow.*

To date, the commercially available TMT kits allow the multiplexing of up to 10 experimental conditions, but they are very expensive and the complex quantitative information they give requires an extensive interpretation, thus complicating the data analysis step.

1.3.2 Functional Proteomics: conventional activity and affinity-based techniques

Functional proteomics is a multifaceted discipline, exploited for the characterization of the functional profile of proteins, multi-protein complexes and signaling pathways. In a Drug Target Deconvolution optics, this strategy has become pivotal in the identification of the protein target(s) of bioactive molecules.

Functional proteomics strategies can be broadly classified into the following approaches: (1) activity-based protein profiling (ABPP), which investigates the enzymatic activity of a particular protein family using small-molecular probes, and (2) compound-centric chemical proteomics (CCCP), devoted to the *interactome* characterization of an individual bioactive molecule. CCCP strategies include affinity chromatography coupled to mass spectrometry (AP-

MS), which requires the compound to be covalently modified for being investigated, and a pretty consistent number of relatively new *label-free* methods, which take the distance from the conventional ABPP and AP-MS approaches, since they don't require any chemical modification of the examined molecule (Figure 14). Indeed, in these methods, the detection of the binding partner(s) of the *label-free compound* is carried out by evaluating proteins responses to a series of perturbations (e.g. thermal or proteolytic treatment) or chemical events. These techniques, comprising cross-linking MS (XL-MS), Drug Affinity Responsive Target Stability (DARTS), Limited Proteolysis (LiP), Pulse Proteolysis (PP), Stability of Proteins from Rates of Oxidation (SPROX), Cell Thermal Shift Assay (CETSA) and Thermal Proteome Profiling (TPP), will be discussed in the subsequent paragraphs.

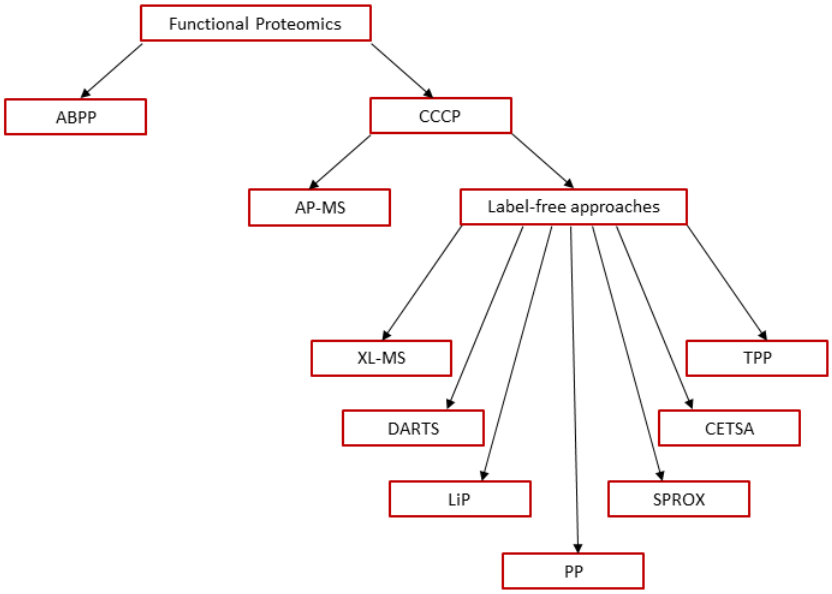


Figure 14: *Functional Proteomics strategies.*

Activity-based protein profiling (ABPP) is devoted to directly capturing, visualizing, identifying and quantifying active enzymes in complex biological systems[80,81], taking advantage of small active site-directed chemical probes (i.e. activity-based probes [ABPs])

designed to interact with active-site residues of target enzymes forming stable covalent bonds[82,83].

An ABP comprises three fundamental building blocks: (1) a warhead (or reactive group) that covalently binds to the active-site residues of a given class of enzymes, (2) a flexible linker that serves as the recognition element for the enzyme-binding pocket(s) and (3) a reporter tag for the detection and enrichment of the probe-labeled enzymes (Figure 15).

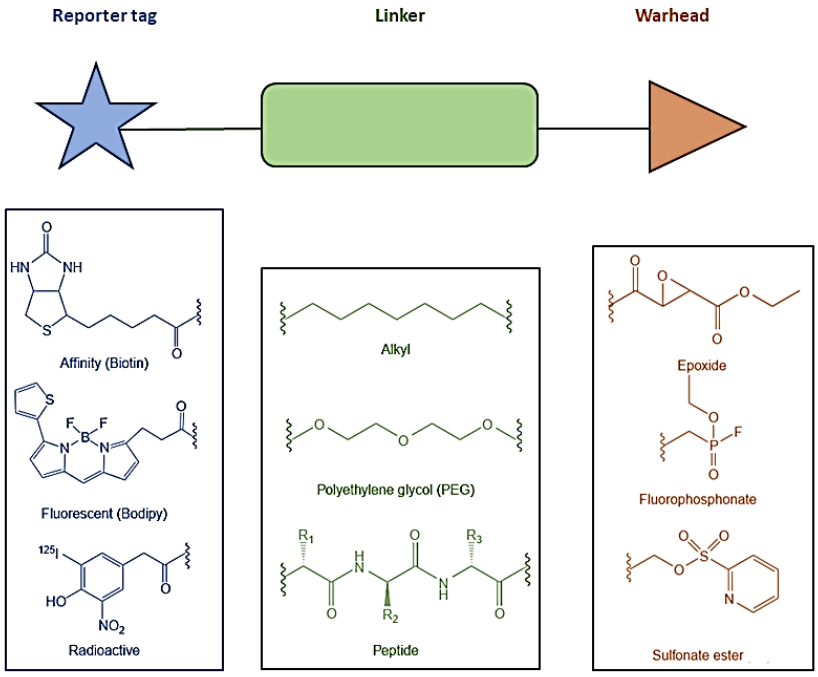


Figure 15: *Schematic representation of ABPs key features.*

The chemical warhead is usually designed as an electrophilic trap that irreversibly reacts in the target enzyme active site to form a new irreversible covalent bond. In order to label proteases that do not use a side-chain nucleophile to process a protein substrate, affinity-based probes containing a photoreactive group have also been designed to form a non-covalent bond with a specific site on a protein, followed by a nonspecific covalent bond forming upon irradiation with UV light[84,85].

The most commonly exploited linkers are based on polyethylene glycol, alkyl or peptide skeletons. Their primary role is to drive the specificity of the ABP and provide a space between the reactive group and the tag, to prevent steric hindrance that may block the enzyme access at the warhead, or the accessibility of the tag, for identification and purification purposes.

An ABP must also possess a reporter tag to facilitate detection. Common examples of reporter tags include fluorophores, biotin, hemagglutinin and alkynes or azides, which can be modified by click chemistry methods (e.g. copper-catalyzed Huisgen's azide–alkyne cycloaddition) to visualize protein targets postlabeling[86–88].

Multiple ABPP based methods have emerged in the past few years, with the most common being *comparative ABPP* for target discovery, *competitive ABPP* for inhibitors discovery and *in vivo imaging*.

Comparative ABPP exploits the comparison of two or more proteomes from differentially treated systems (e.g. exposed or not to a drug) to detect enzymes whose activity is regulated or altered. The second major application of ABPP, in its competitive mode, is the discovery of enzyme inhibitors: small molecules can compete with the ABPs for the binding to the enzyme active sites, thereby slowing the rate of probe labeling. Furthermore, when an ABP presents a fluorogenic tag, it can be employed for imaging applications, through *in vivo* enzyme labeling and microscopy localization (Figure 16).

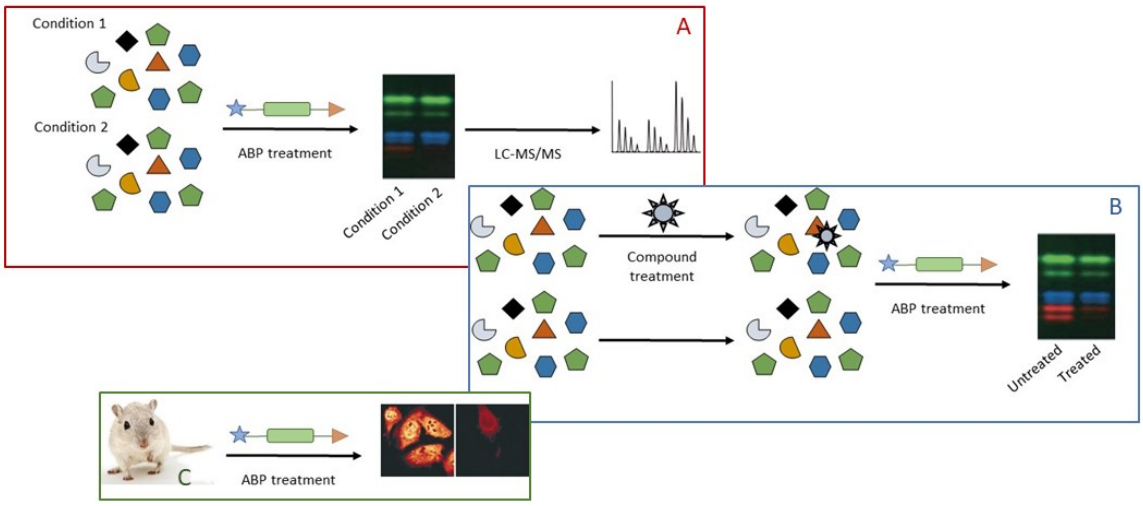


Figure 16: *ABPP strategies: (A) comparative ABPP; (B) competitive ABPP; (C) in vivo ABPP.*

As already stated in this paragraph, Compound Centric Chemical Proteomics (CCCP) or Affinity-Based Chemical Proteomics (ABCP) is devoted to the characterization of the molecular mechanism of action of a given bioactive compound through the identification of its protein target(s).

Among the CCCP strategies, the affinity purification mass spectrometry-based technique (AP-MS), also known as pull-down or *Fishing for Partners*, is a consolidated and powerful strategy in the drug discovery field, applied during the last 15 years for the *targetome* discovery of a large number of bioactive compounds[89–95].

This strategy can be carried out in either a heterogeneous or a homogeneous fashion.

The heterogeneous approach consists of three main steps: (1) the chemical immobilization of the small molecule to be studied onto a commercially available solid matrix (e.g. agarose, sepharose or magnetic beads); (2) the affinity enrichment of the compound targets from a complex protein mixture (e.g. a cell lysate) and (3) the elution and identification of the captured proteins by means of tandem mass spectrometry coupled to bioinformatics (Figure 17).

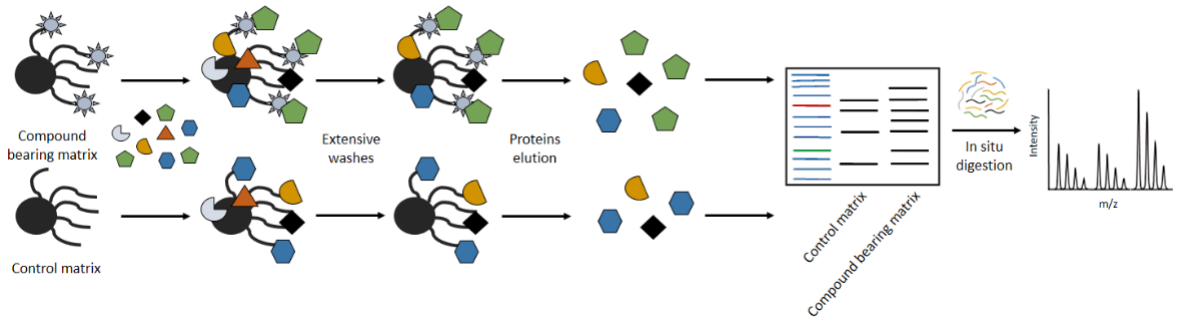


Figure 17: *Heterogeneous AP-MS workflow.*

The homogeneous strategy, also known as *in solution pull-down*, takes advantage of the strong non-covalent interaction between biotin and streptavidin. Indeed the compound is first modified with a biotin tag and then incubated with the protein mixture, to allow the interaction with its protein partner(s). In a last step, streptavidin bearing beads are added to bind the molecule and, consequently, to fish out the linked proteins, which are eluted and analyzed, as for the previous strategy.

In both approaches the small molecule needs to contain reactive chemical groups (i.e. sulfhydryl, amino, hydroxyl, carbonyl or carboxyl groups), in order to be immobilized onto the resin or functionalized with the tag, respectively.

Furthermore, the attachment of a spacer arm between the small molecule and the solid matrix is mandatory to avoid matrix steric interference in the interaction with the proteins, thus assuring this step is carried out under *solution-like* conditions.

Different classes of linkers have been designed (e.g. alkyl or PEG linkers, peptide-like linkers and cleavable linkers), whose length and chemical features have a pivotal role in a pull-down experiment. Indeed, more then 15-atoms-long spacers are not ideal, since they might auto-aggregate in aqueous solutions and, especially if they are hydrophobic alkyl linkers, favor the non-specific binding of the so called *sticky proteins* (i.e. proteins interacting with hydrophobic

and/or charged surfaces in a non-specific fashion). Thus, a hydrophilic polyethylene glycol (PEG) spacer is usually employed[96] (Figure 18).

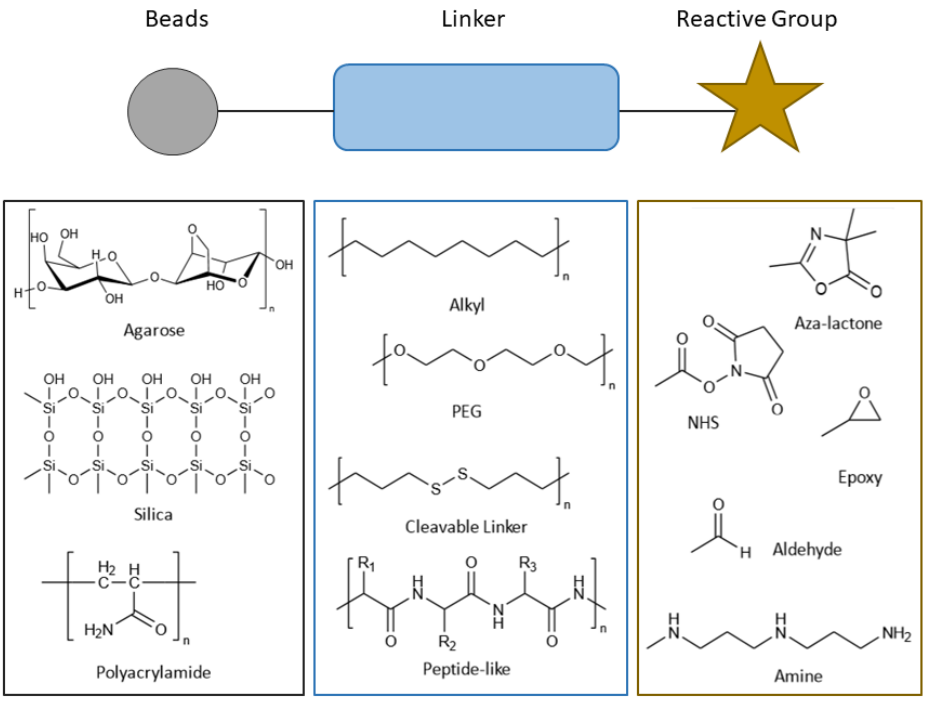


Figure 18: *Schematic representation of commonly exploited matrix, linkers and reactive groups.*

For the incubation step, a cell extract is prepared, either from cells or tissues, under mild conditions to preserve a pseudo-cellular environment in which proteins retain their native state and features. Subsequently, the lysate is incubated with the affinity matrix which, after a settling step, is extensively washed to reduce the amount of proteins un-specifically interacting with the linker or the matrix. Depending on the experimental strategy, washes of different strenghts can be used (i.e. detergents, salts or pseudo-physiologic buffers washes).

In the last step, the small molecule interacting proteins are eluted from the matrix system (e.g. through high-salts containing buffer, change of pH, denaturing agents or an excess of free compound), separated and visualized, usually through 1D-SDS-PAGE, and finally digested

with trypsin. The resulting peptides mixtures are then analyzed by nano-ESI LC-MS/MS and submitted to bioinformatics search using protein databases (e.g. SwissProt) and search engines (e.g. Mascot) to give protein identification.

The ability of providing a high-throughput and unbiased investigation of a small molecule *interactome*, carried out through a quite simple procedure and in pseudo-physiological conditions, made AP-MS one of the top proteomics technique in the drug discovery field.

Indeed, this approach identifiies a small molecule interacting proteins in their natural state and environment, reflecting their endogenous abundance levels, PTMs, splice variants and the presence of their natural binding partners, such as other proteins, nucleic acids or endogenous small molecules. Furthermore, AP-MS could potentially be performed with any cell type, tissue or species, to investigate possible disease-relevant drugs in virtually any clinically meaningful sample.

However, some undeniable caveats have to be taken into account.

The preparation of the cell lysate for the incubation step has the strict requirement to be performed with a gentle procedure to preserve pseudo-physiological conditions: the lysis method will not capture all the proteins, resulting in a bias especially versus membrane proteins that require peculiar solubilization conditions. Aiming to overcome this issue, *in vivo* AP-MS has been optimized, taking advantage of bio-orthogonal chemistry, based on the design of warheads with functional groups unable to react with biomolecules but highly reactive with an orthogonally-functionalized reporter tag. As an example, the copper (I)-catalayzed variant of the Huisgen 1,3-dipolar cycloaddition takes advantage of the reactivity between azides and terminal alkynes to give 1,2,3-triazoles[97,98]. Indeed, the azide or alkyne functionalized version of the compound to be studied can be incubated with living cells and, after proteins extraction, alkyne or azide reporter tags can be introduced into the system to give click chemistry reactions, prior to pull-down to identify the compound *interactors*.

Another AP-MS shortcoming could be the elution step: when the molecule-linked beads are treated with denaturant agents, salts, pH change or an excess of the free compound, only proteins interacting with the molecule in a non-covalent fashion will be released, and this will prevent the identification of potential covalent interactors. To overcome this limitation, the small molecule could be functionalized with a cleavable linker (e.g. a disulfide bond containing linker or an acid cleavable one) prior to be anchored onto the beads: during the elution step, when treated with proteases[99], acids[100] or reducing agents[101], the linker should release the compound covalently and non-covalently interacting proteins.

Another AP-MS issue is represented by the false positive target(s) that could be identified through the pull-down step. Cell lysates are indeed packed with proteins, so that a compound could be involved in a plethora of interactions, among which the high-affinity interactions, often with low abundance proteins, are most likely to be the truly physiologically relevant. During the affinity purification step, highly abundant proteins with low affinity for the immobilized compound will often co-elute, as well as *sticky* proteins, prone to interact with either hydrophobic or charged surfaces of the matrix/linker system.

It is therefore mandatory to perform suitable negative control experiments, to unequivocally distinguish between proteins and non-specific interactors. Many approaches can be employed to address this issue. Proteins that bind to the matrix itself can be identified with blocked control beads samples, performing a parallel pull-down experiment with the same naked matrix or with the linker-bearing matrix. To tackle low-affinity interactions with the drug itself or with specific drug-binding proteins, an inactive compound closely resembling the molecule to be investigated in structure, size, polarity and charge can be immobilized onto the control beads: both the compound and the analog bearing beads will pull-down similar background proteins, differing for the specific targets[102,103]. However, finding a suitable inactive structural analog of the sample molecule is a difficult task, and sometimes could give false negative results.

Additional strategies to discriminate between true and false interactor(s) of the bioactive molecule use *competitive and serial affinity purification*.

In competitive AP-MS (Figure 19), an excess of the bioactive compound is added to the protein mixture prior to the pull-down step: on the basis of competition for the binding pocket of the target protein(s), the free compound is able to prevent or reduce the binding of the beads-immobilized species. Thus, looking at the 1D SDS-PAGE profiles, only the proteins contended by the free ligand are considered as real targets[104].

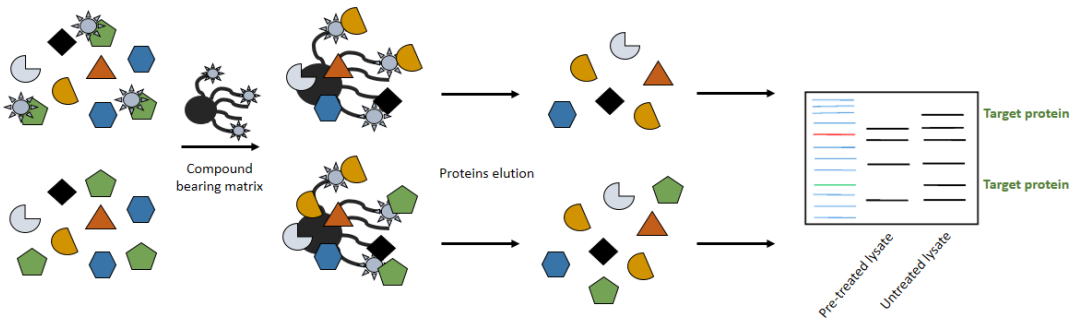


Figure 19: *Competitive AP-MS workflow.*

When the compound is poorly soluble in the aqueous buffer, a serial AP-MS approach can be carried out (Figure 20), in which a sample of the compound-bearing matrix is incubated with the protein lysate, then, after beads settling, the remaining proteins are added to a fresh sample of the compound-bearing matrix: analysis of the 1D SDS-PAGE profiles will show a preferential enrichment of the target candidates on the first matrix sample, whereas the level of the non-specific binders is similar for both [62,105,106].

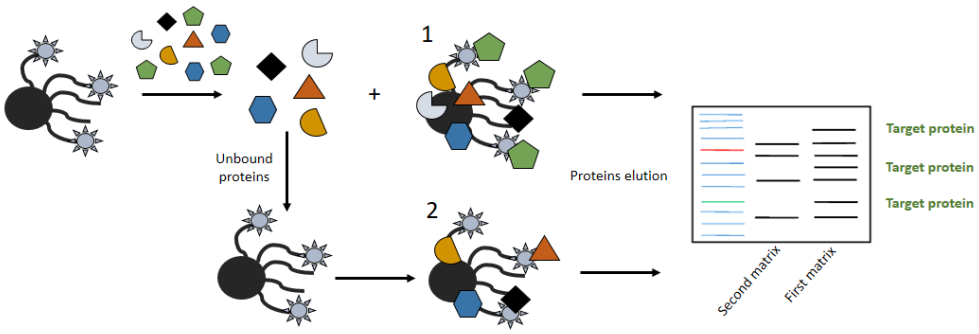


Figure 20: *Serial AP-MS workflow.*

Even though properly performed, a chemical proteomics experiment does not give any clue on the functional implications or biological relevance of the interaction between the molecule and its target(s). Thus, further investigations are usually required to deepen the nature of these interactions (e.g. generating quantitative affinity parameters through surface plasmon resonance), and *in vitro* or *in vivo* biological assays are needed to investigate the capability of the small molecule to modulate its partner(s) activity in a biological context.

1.3.3 Functional proteomics: label-free CCCP approaches

Even if in the last 15 years AP-MS has been the top-choice technique in the CCCP-driven Drug Target Deconvolution field, a major disadvantage of this technique is the need of a chemical modification of the molecule to be characterized. For this reason, AP-MS is not suitable for compounds that do not exhibit properly reactive structural features (e.g. sulfhydryl, amino, hydroxyl, carbonyl or carboxyl groups). Furthermore, even when the covalent modification with a resin or a linker is possible, the reaction of a molecule functional group would leave some concerns about whether or not the compound is able to retain its original bioactivity.

To overcome these AP-MS major drawbacks, a new toolbox of mass-spectrometry based techniques has been established over the past decade for the Drug Target Deconvolution purpose, avoiding any chemical modification of the small molecules for a proteome-wide profiling of their *targetome*. Furthermore, compared to the pull-down based techniques, these new approaches work on *in solution* interactions and consequently need smaller amounts of the sample compound, which is very important in the analysis of natural products. Moreover, even the identification of possible covalent *interactors* of the sample molecule to be studied is feasible.

This new series of proteomics techniques enables the study of protein-ligand binding by evaluating the ligand-induced protein conformational changes, which either translate into a major stability of the target to perturbation events like heating, denaturing agents or proteolytic enzymes, or in structural fluctuations that can be spotted through chemical cross-linking.

As briefly mentioned in paragraph 1.3.2, these methods encompass cross-linking MS (XL-MS), Drug Affinity Responsive Target Stability (DARTS), Limited Proteolysis (LiP), Pulse Proteolysis (PP), Stability of Proteins from Rates of Oxidation (SPROX), Cell Thermal Shift Assay (CETSA) and Thermal Proteome Profiling (TPP), all exploiting common *bottom-up* proteomics procedure for protein targets identification and/or quantitation.

These techniques can be broadly classified in two groups, based on the readout nature of the LC−MS/MS approaches they rely on. In particular, DARTS, PP, CETSA and TPP are *protein-centered* strategies, meaning that the *bottom-up* proteomics data produced in the LC−MS/MS analyses are used to generate quali/quantitative information about the interacting proteins of the sample compound, whereas XL-MS, LiP and SPROX are *peptide-centered* methods, because the differential behavior of the proteins exposed or not to a molecule are directly ascertained from the peptides identified and/or quantified in the *bottom-up* proteomics analysis.

Among these approaches, DARTS and targeted LiP strategies have been the focus of this PhD project and will be comprehensively discussed in the following paragraphs.

As for the peptide-centric strategies, cross-linking MS (XL-MS) has emerged as a useful tool for probing protein conformations in physiologically relevant conditions, providing valuable spatial information for the elucidation of protein structural changes and dynamics associated with their functions[107,108] or due to interactions with a ligand[109,110]. In this approach, spatial proximities between amino acid residues in protein chains are engaged into covalent bonds via cross-linkers, which thereby define the maximal distance restraints between the residues they pair. In a typical XL-MS experiment, differentially treated protein samples (i.e. exposed or not to a bioactive compound) are processed with a chemical cross-linking reagent containing two protein-reactive functional groups (e.g. thiol-reactive maleimides, carboxyl group reactive diazoacetate-esters or amine reactive N-hydroxysuccinimide esters) that are connected by a linker region. The samples are then digested with trypsin and the cross-linked peptides, enriched through SCX prior to the LC-MS/MS analysis, are subsequently identified from their fragmentation patterns via database search (Figure 21).

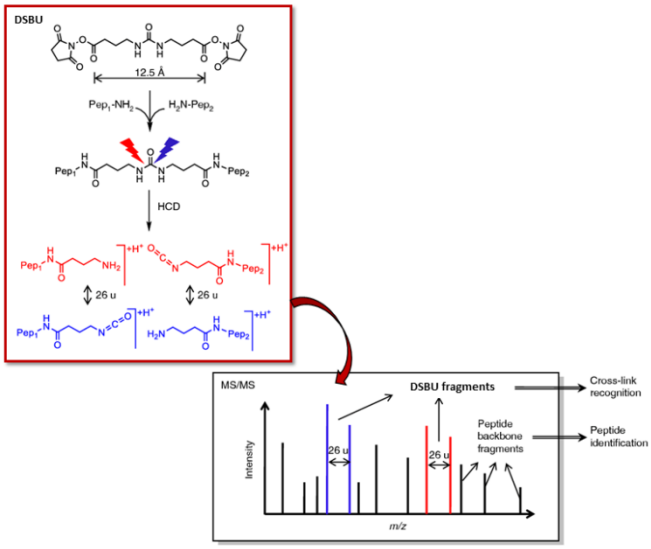


Figure 21: *Disuccinimidyl Dibutyric Urea (DSBU) cross-linked peptides & their identification through LC-MS/MS.*

Cross-linked peptides can also be quantified through stable-isotope labeling strategies or *label-free* approaches. Isotope-labeling can be carried out to pairwise compare ligand-induced protein conformational changes, either by labeling the already cross-linked peptides or by exploiting stable-isotope labeled cross-linkers[111]. A 1:1 mixture of the two labeled and cross-linked peptide samples is then analyzed by LC-MS/MS: for each cross-linked peptide pair, the signals derived from the two protein samples are distinguished by their different masses and intensities.

In *label-free* quantitation workflows, differentially treated protein samples are separately submitted to LC-MS/MS: for a given cross-linked peptide pair, abundance is compared based on the signal intensities across individual MS acquisitions on the different samples.

Despite its tremendous potential, applications of XL-MS has been restricted due to the objective difficulties in the data interpretation and quantitation step.

Stability of Proteins from Rates of Oxidation (SPROX) is a covalent protein labeling mass spectrometry-based method for evaluating the *in solution* thermodynamic properties of protein-ligand complexes, through the dependence of the hydrogen peroxide-mediated oxidation of methionine side chains on a certain denaturant amount. In doing so, this method enables to measure both the folding free energy of proteins and the K_D values of protein-ligand complexes.

SPROX was originally coupled with an *intact protein* mass-spectrometry readout, but it has recently been optimized and interfaced with *bottom-up* proteomics to facilitate the detection and quantitation of methionine containing-oxidized peptides, through *label-free* or isotope-based quantitation approaches[112,113].

More in details, after the incubation with a bioactive compound or just the vehicle, the SPROX protocol involves the equilibration of the protein samples in a series of buffers with increasing concentrations of a chemical denaturant (e.g. urea or guanidinium hydrochloride). The samples are then exposed to hydrogen peroxide to selectively oxidize the thioether group in the methionine side chain. Oxidation at other susceptible amino acid side chains (e.g., cysteine and tryptophan) is generally not observed, since their oxidation rates are much slower than the methionine one.

When a protein forms a complex with a ligand, the bound conformation is stabilized according to the dissociation free energy of the complex, at a given ligand concentration. This stabilization results in changes in the energetic properties of the target protein, such as the increase in its thermodynamic stability and the decrease in its unfolding rates. This makes the equilibrium shift towards the ligand-bound form of a particular protein subtracting its unbound form, which would in turn be more prone to the chemical-induced denaturation. Since methionines are oxidized when are solvent-exposed, and this condition is achieved during the protein unfolding in the denaturant-containing buffers, the interaction with a ligand should hamper their oxidation (Figure 22). The interacting proteins can thus be identified because they

need a significantly higher denaturant concentration for their methionine to be oxidized (e.g. > 0.5 M Guanidinium hydrochloride or > 1.0 M urea)[112].

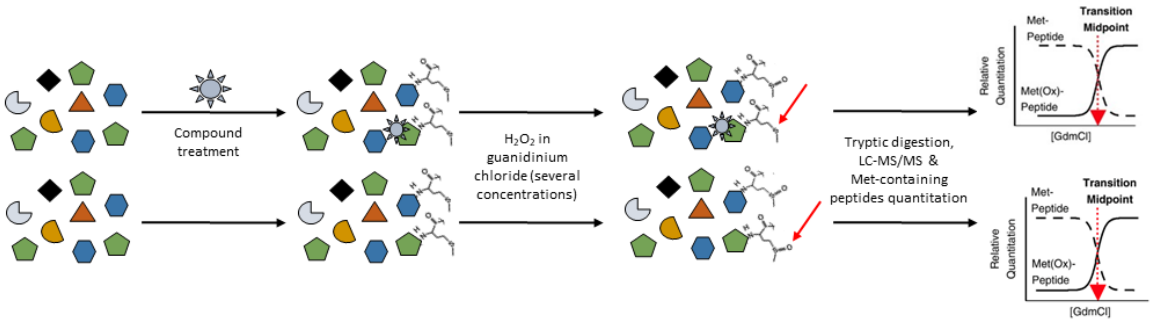


Figure 22: *Schematic representation of SPROX workflow.*

In several cases, as for the other *label-free* CCCP approaches, a bioactive compound could preferentially bind to the locally unfolded state of a protein, which is in equilibrium with the folded one. This would bring to a consequent destabilization of the protein partner(s), thus leading in the SPROX procedure to an easier oxidation of its (their) methionine residues.

Despite the huge advantage of providing thermodynamic parameters associated with a small molecule/protein interaction, SPROX has a limitation when the sample compounds are unstable in presence of hydrogen peroxide and, furthermore, only methionine containing peptides are suitable for the readout, so that they often have to be enriched prior to the mass spectrometry analysis.

Pulse Proteolysis (PP) is a protein-centered limited proteolysis-based approach, whose basilar concept is not so far from the SPROX one. Indeed, this technique exploits the chemical denaturant dependence of an unspecific proteolytic digestion to evaluate the thermodynamic properties of protein-ligand complexes[114].

An effective proteolytic digestion requires to unfold the protein substrate, either globally or locally[115]: folded and unfolded proteins have very different proteolytic susceptibilities.

When a protease is added to a mixture presenting folded and unfolded proteins at the equilibrium, unfolded proteins are rapidly digested, whereas proteolysis of the folded ones is much slower, unless they have intrinsically unstructured regions. Therefore, a short incubation time (i.e. pulse proteolysis) will digest only unfolded proteins, leaving the folded intact, for them to be determined after the pulse step by *bottom-up* proteomics.

The PP protocol involves equilibrating the different protein samples (e.g. pre-incubated or not with a drug) in a series of buffers containing increasing concentrations of urea as a denaturing agent. The mixtures are then submitted to *pulse proteolysis* with an unspecific enzyme (e.g. thermolysin), to selectively digest the unfolded protein population in each urea-containing buffer.

As already mentioned for SPROX, ligands commonly stabilize proteins during the interaction[116], thus a target protein will be less prone to be denatured by urea and, in turn, less susceptible to the protease action in the pulse phase.

After the proteolysis, the undigested proteins are separated by the pulse products in order to be identified and quantified through several mass-spectrometry or *gel-based* readouts[117–120] (Figure 23).

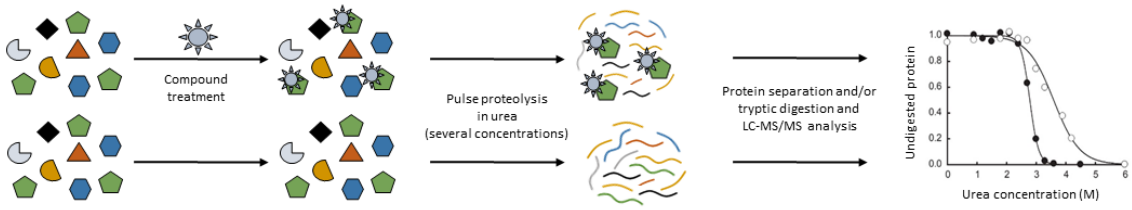


Figure 23: *Schematic and simplified representation of pulse proteolysis workflow. White dots in the graph represent the compound-interacting protein behaviour, whereas black dots the untreated protein one.*

To obtain reliable results, the pulse proteolysis length must be chosen properly, to ensure complete digestion of the unfolded proteins and a minimal digestion of the folded ones.

While the lower limit of the pulse can be estimated by the intrinsic proteolytic kinetics of the enzyme, the upper limit depends on the protein targets, even if 1 min pulse is a reasonable choice.

As SPROX, PP has the tremendous advantage of being capable of identifying the protein target(s) of a bioactive compound in a simple and quantitative fashion, and of detecting the thermodynamic parameters associated with the interaction.

If, anyway, a protein tends to unfold quite quickly in its transition zone (i.e. when both the folded and unfolded species are detectable in the denaturant containing solution) or has an unstructured region in its native conformation, PP is not a reliable choice to study protein-ligand interactions, because of the extensive proteolysis of the folded protein that would occur within the pulse.

Melting temperature (or thermal) shift assays (TSAs) have been of great value in both drug discovery and academic research fields, for the deconvolution of ligand-protein(s) interactions[121–124]. In Cell Thermal Shift Assay (CETSA) the ligand-induced protein stabilization is investigated in cell lysates, intact cells and tissues, to preserve proteins subcellular localization, PTMs and interactions. CETSA relies on the same principle as TSA, which is the well-known ability of a protein interacting with a ligand to become more resistant to heat-induced unfolding. As such, CETSA relies on the irreversible aggregation and precipitation of proteins that follows their thermal denaturation.

Thus, in a typical CETSA experiment, vehicle and ligand-treated cell lysates or cells are heated at different temperatures and the target protein(s), which do not aggregate and thus remain soluble, are detected after separation from the aggregated ones by centrifugation or filtration[125].

A typical CETSA output gives a comparison between apparent melting curves (i.e. temperature-induced aggregation curves) in which the protein, in presence and absence of its ligand, is subjected to a panel of temperatures to assess a potential compound-induced thermal stabilization through the shift in the apparent melting point (i.e. the temperature at which a protein unfolds, aggregates and precipitates) of the protein itself (Figure 24).

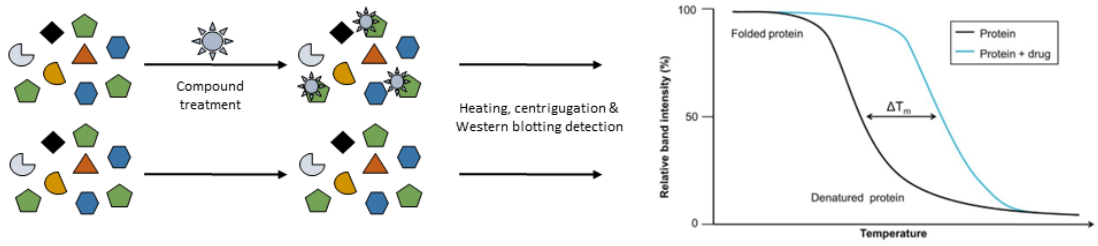


Figure 24: *Ligand-induced protein stabilization evaluation through CETSA.*

Alternatively, if the melting point of a given protein is known, an isothermal dose-response curve (ITDR) can be obtained, in which the stabilization of a protein can be followed as a function of increasing ligand concentrations, to evaluate its potency as the molecule concentration at which 50% of the total stabilizing effect can be observed (EC_{50}).

The initial readout from CETSA and ITDR experiments was restricted to Western blotting, which surely enabled to verify the engagement of a compound interaction with its target and the potency of the molecule itself, but lacked the potential for detecting unknown targets. When combined with *bottom-up* proteomics for the latter purpose, the mandatory optimization of the thermal gradient, requiring a large number of different experimental settings to evaluate the compound induced thermo-stabilization of its *interactors*, posed some drawbacks on the CETSA approach.

To overcome these limitations, CETSA was recently coupled with multiplexed quantitative mass spectrometry in the so-called thermal proteome profiling (TPP) approach, which enables to monitor changes in proteins thermal stability across the whole proteome and to identify direct and indirect drugs targets in an unbiased manner[126]. TPP key feature is exploiting isobaric tandem mass tag 10-plex reagents[126,127] to label digested protein samples corresponding to each soluble fraction after the heating step, which allows multiplexing of up to ten temperature conditions in a single experiment.

In a TPP experiment, CETSA workflow is followed until centrifugation is performed after the heating step. Then, the relative soluble amounts of proteins across the differentially heated and treated (i.e. compound exposed or not) samples are submitted to tryptic digestion to obtain peptides mixtures that are individually labeled with different isobaric tags and combined, so that each temperature series is further processed as a single sample (Figure 25).

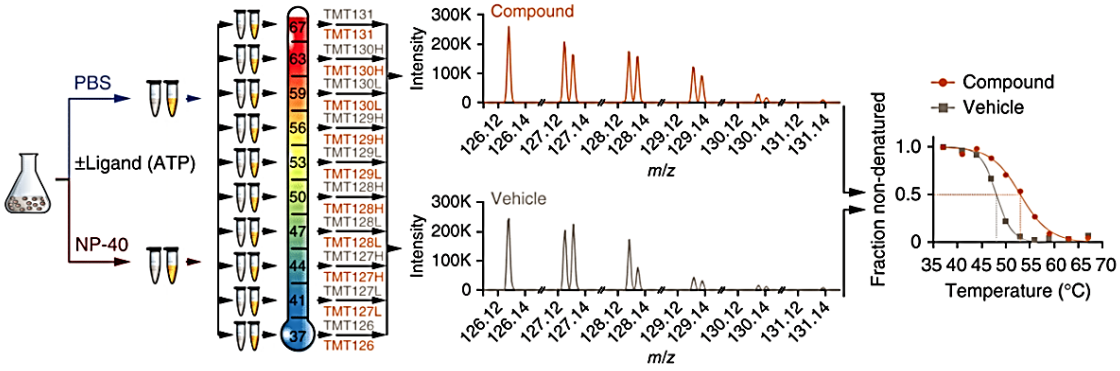


Figure 25: *TPP schematic representation. Adapted from Franken et al.[126].*

Consequently, one TPP experiment requires two labeled samples to be provided, one for the compound-exposed proteins and one for the vehicle-treated ones. For the sake of proteome coverage, samples are fractionated offline prior to LC-MS/MS, usually with the HpH strategy (paragraph 1.2). The resulting fractions are then separated by liquid chromatography and

analyzed with a Q Exactive mass spectrometer, to finally perform protein identification and quantification and assess the potential compound-induced thermal stabilization of its target protein(s).

1.3.3.1 Drug Affinity Responsive Target Stability (DARTS)

Pioneered by Lomenick and co-workers in 2009[128], DARTS is a straightforward and universally applicable protein-centered Drug Target Deconvolution approach, sought to overcome the limitations of the top choice proteomics method for the identification of protein partners of bioactive compounds, AP-MS.

As the other *label-free* CCCP approaches presented in the previous paragraph, DARTS exploits the thermodynamic stabilization conferred by a molecule to its protein target(s).

Indeed, whereas under physiological conditions a protein is in a dynamic equilibrium with multiple alternative conformations also exhibiting local and reversible unfolding[129,130], the interaction with a specific ligand will shift the equilibrium to highly favor the molecule-bound conformation. This phenomenon results from a change in the negative free energy due to hydrophobic, hydrogen bonding and/or electrostatic interactions that are formed between the protein and the ligand itself. This leads to a thermodynamically more stable state in which the target protein(s) conformational fluctuations (i.e. *breathing*) and unfolding rates are dramatically decreased, thus markedly increasing resistance to limited proteolysis, which is the perturbation event exploited by DARTS.

In a typical DARTS workflow, cell lysates are incubated with the molecule to be studied at several concentrations, then the protein samples are submitted to limited proteolysis with an unspecific protease (e.g. thermolysin, subtilisin or pronase). Because of the reduction in *breathing*, proteins interacting with the molecule will become thermodynamically more stable and thus less exposed to the proteolytic action, in respect to the other proteins in the mixture

that do not engage in any interaction with the compound. The protease is then quenched and the proteic mixtures separated through 1D-SDS–PAGE to give evidence, after staining the gel (e.g. with Coomassie Blue or silver stain), of the proteins with altered proteolytic patterns in presence of the binding molecule. For this purpose, an *on eye* inspection of the gel is carried out, in which the lanes corresponding to the differentially treated protein samples (with or without the small molecule) are compared, looking for bands that are more intense in the treated versus the untreated sample, evidence of a major abundance of the corresponding protein(s) because of the compound-exerted stabilization. Each of these bands is then cut out and submitted to *in situ* tryptic digestion for the subsequent nano-ESI-LC-MS/MS analysis (Figure 26).

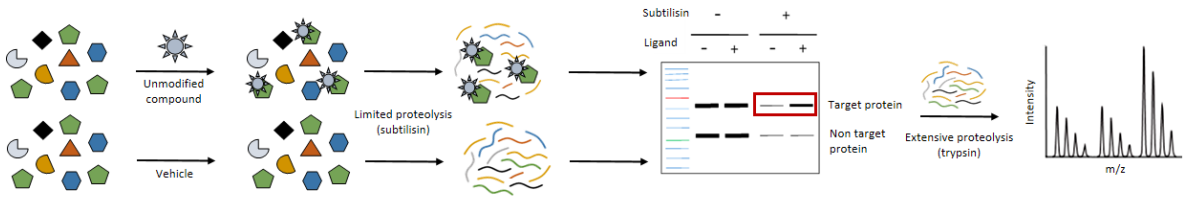


Figure 26: *DARTS workflow.*

The MS results are then searched against a database for the protein identification step and a *label-free* quantitation is carried out to evaluate their abundance level. In particular, a semi-quantitative analysis can be performed either by measuring the MS signal intensity of peptide precursor ions associated with a given protein (i.e. peptide ion intensity) or by counting the total number of MS/MS spectra that map for the peptides of a given protein (i.e. spectral counting). Both of these quantitation methods have already been discussed in paragraph 1.3.1.

The initial breakthrough of DARTS in respect to the most commonly exploited AP-MS procedure was likely due to the broader applicability of this method over the pull-down based experiments. Indeed, as already mentioned, the matrix-based strategies require three basic conditions to be satisfied for their general application: (1) the presence of a reactive group on the small molecule structural core, needed for the derivatization step on the matrix, (2) the compound bioactivity and binding specificity should be unaffected by its chemical derivatization and (3) the matrix should not hinder the interaction between the target protein(s) and the small molecule. The latter two criteria are impossible to be predicted *a priori,* while the first one strongly depends on the structural features of each molecule, thus making AP-MS not universally applicable.

On the other hand, DARTS is based on the *in-solution* interactions between the unmodified compound and the proteins of a cell lysate, thus eliminating the limitations due to its chemical features, as far as the molecule is soluble in the incubation buffer with possible addition of only little amounts of organic solvents, to preserve proteins in their native state.

As straightforward as it might seem, DARTS presents some tricky aspects that need to be established prior to perform the experimental procedure. Given the huge dynamic range in terms of protein sequences, structures, conformational dynamics and interactions, it is not so difficult to picture that also proteins sensitivity to proteolysis has a large dynamic range. This issue is particularly important since different proteases target specific protein residues and also show preferences for folding or location of the recognition site within a polypeptide. Those aspects make the proteome-wide susceptibility to proteolysis strongly dependent on both proteins and proteases.

In the very first DARTS experiments, the unspecific metallo-endopeptidase thermolysin, very efficient only on unfolded proteins, has been used. In non-denaturing conditions, this protease

can only digest unfolded proteins or proteins presenting unstructured regions, when used for short time and in low amounts, thus showing a low dynamic range in its substrates pool.
In these experiments, the protection levels from thermolysin action was remarkably robust since the small molecule binding prevented the unfolding of the target proteins, but at the same time many background proteins were also refractory to digestion, complicating the proteomics analysis. Thus, thermolysin was rapidly replaced with other proteases, such as subtilisin, able to give a broad response to the digestion of differentially folded proteins for a less biased drug target identification[131].

Another important parameter to be optimized is the protease to proteins ratio. Indeed, in order to point out the different intensities of the gel bands, a high digestion level of the non-target proteins beside a low digestion level of the protected target proteins should be maximized. This ideal situation is however unlikely to be reached, and a fine tuning of the protease amount to be used must be performed, depending on both the chosen enzyme and the molecule behavior.

In spite of being a general, straightforward and fast technique, in its original version DARTS presented some drawbacks related to the poor dynamic range of its 1D-SDS-PAGE readout.

In order to be spotted on gel *by eye*, the target proteins need to be highly abundant and not extremely sensitive or resistant to the exploited protease. Indeed, if a protein is not sufficiently abundant or is highly sensitive to proteolysis, it could be hardly visible after gel staining. On the other hand, if a protein is not sensitive to the enzyme, DARTS technique cannot obviously be applied.

Furthermore, protein enrichment could be masked in case of co-migration between proteins of the same or similar molecular weight: a highly abundant protein will hide the the target one, while the presence of several low or moderately abundant proteins could just mask the differential abundance of the targets. This is common for binders whose molecular weight is

below 30 kDa: the low masses region on a 1D-SDS-PAGE will be largely enriched in limited proteolysis products (i.e. partially digested proteins of various lengths), thus the identification of low-molecular weight targets is particularly daunting by an *on eye* visualization of the gel bands. Even if this issue could be partially circumvented by exploiting high poly-acrylamide percentage gels, for a better resolution of the low molecular weight regions, the peptide mixtures coming from digestion of a gel slice would anyway be highly complex, partially hampering the mass spectrometric analysis.

To overcome those limitations, several alternative DARTS approaches have been proposed in the last few years, going towards higher dynamic range *gel-based* strategies, coupling DARTS with 2D-SDS-PAGE or DIGE[132], or pursuing *gel-free* DARTS[133]. In the latter case, differences in proteins abundances due to a compound-exerted protection can be accomplished after separation of the limited proteolysis products from the remaining intact proteins by dialysis or filtration[131] (e.g. through molecular weight filters), prior to submit the samples to *in-solution* digestion and 2D-LC-MS/MS.

Lastly, it should be noted that DARTS was meant to be performed on cell lysates and not on living cells, in which proteins degradation is carried out by supra-molecular machines (e.g. proteasomes) and controlled by PTMs (e.g. ubiquitinylation), making the *in vivo* protein stability a complex and generally unpredictable matter. Indeed, drug binding has been shown to both positively and engatively modulate the proteolytic susceptibility of the target protein *in vivo*[134–137]. Furthermore, an *in vivo* experiment could be complicated by possible drug-induced up or down regulation of the expression of certain proteins, which could be misinterpreted as the molecule binders.

1.3.3.2 Targeted Limited Proteolysis coupled to Multiple Reaction Monitoring Mass Spectrometry (t-LiP-MRM)

T-LiP-MRM is a recently developed method[138] that enables to probe proteins conformational changes in complex biological environments by coupling limited proteolysis (LiP) with targeted proteomics tools.

T-LiP-MRM is based on the same phenomenon exploited by DARTS, which is the thermodynamic stabilization of a target protein interacting with a small molecule when it comes to limited proteolysis. Neverthless, DARTS readout is protein-centered while the t-LiP-MRM one is peptide-centered. Indeed, the molecule-induced proteolytic resistance of a protein is perceived through changes in abundance of its tryptic peptides.

T-LiP shares a part of its experimental procedure with DARTS, that is (1) the incubation, in native conditions, of cell lysate samples with the molecule of interest (or just with the vehicle for the negative control) and (2) the limited proteolysis of the protein mixtures with a broad-specificity protease (e.g., proteinase K, subtilisin, papain, chymotrypsin or elastase) at a low enzyme:substrate ratio and for a short time, so that the initial cleavage sites are dependent on the proteins structural features and generate large protein fragments[139]. After quenching the protease, differently by DARTS, samples are shifted to denaturing conditions and submitted to an extensive *in-solution* tryptic digestion, to give peptide mixtures suitable for a *bottom-up* MS analysis. As a positive control, an aliquot of the cell lysate is only subjected to trypsin proteolysis in denaturing conditions. After a desalting step (e.g. through C18 packed cartridges) peptides are analyzed by Multiple Reaction Monitoring Mass Spectrometry (MRM MS)[140], taking advantage of its high sensitivity and background filtering capabilities to probe the molecule target protein(s) altered proteolytic patterns between differentially treated samples, in complex matrices (Figure 27).

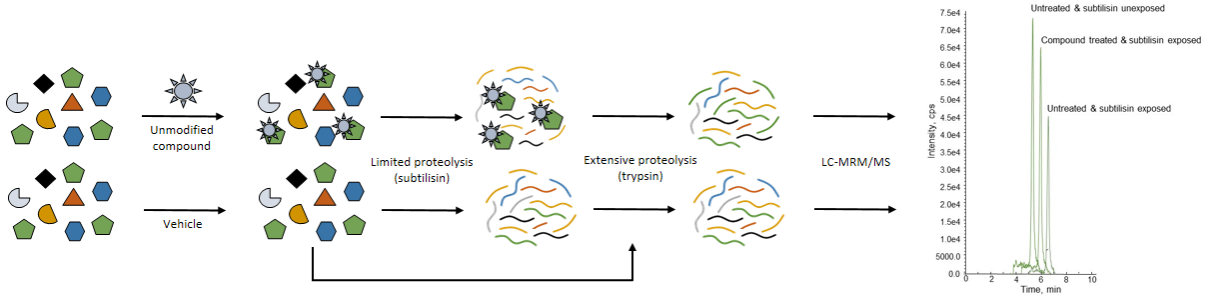


Figure 27: *T-LiP-MRM workflow.*

MRM (also referred to as *Selected Reaction Monitoring* [SRM]) is a mass spectrometry technique that takes advantage of the unique capability of triple quadrupole (QQQ and Q-TRAP) mass spectrometers to act as mass filters, for the selective monitoring of specific parent ions and one or several of their daughter ions[141], so that multiple species with known fragmentation properties can be detected and quantified even in complex systems. In this technique, the precursor/fragment ion pairs are termed transitions.

When applied to t-LiP, MRM analyzes the peptides produced by tryptic digestion of differentially treated protein samples: peptide ions of a designated protein are selected in the first mass analyzer (Q1), fragmented by CID in Q2 and then one or several fragment ions (uniquely derived from the Q1 selected peptides) are analyzed by the Q3 quadrupole (for QQQ instruments) or by the LIT (for Q-TRAP instruments).

Integration of the chromatographic peaks for each transition allows the relative quantification of the targeted peptides and, by inference, of the corresponding proteins of interest.

The development of a robust MRM assay needs the selection of peptides unique to the target protein(s) (i.e. proteotypic peptides[142,143]) and easily detectable by mass spectrometry.

In order to choose the correct peptides to analyze, information from previous experiments stored in repositories like *PeptideAtlas*[144] can be exploited. Indeed, *PeptideAtlas* is an on-line collection of observed peptides and their related annotations, retrieved by a large number of data sets that have been deposited and reprocessed through the trans-proteomic pipeline[145]. More in details, raw files from MS/MS experiments are collected and converted into mzXMl files for the sequence searching step, performed with either Sequest[52] or X!tandem[146] (MS/MS data analysis programs). The top hits are then validated and the results combined to retrieve protein-level probability matches. The spectral library-building tool SpectraSt is then used to create a consensus spectrum library comprising all of the observed high-scoring peptide ions. All the raw data are then subjected to a second round of searching and validation, and all of the identified high-scoring peptides are finally assigned to a single reference Ensembl build[147] and mapped to the genome. All these information are then loaded into the PeptideAtlas database for browsing or downloading purposes.

For a given protein, tryptic peptides having a high-probability to be produced and detected by MS are presented and ranked on the basis of different criteria, and proteotypic peptides can thus be picked as the best choice for a targeted proteomics experiment.

Once chosen the peptides, a selection of the proper transitions should be made: for each precursor-ion charge state, fragment ions providing the highest intensity and lowest level of interfering signals should be preferred. Such selection can be guided by *SRMAtlas*, a public available resource of SRM assays generated for entire proteomes by the analysis of known synthetic peptides (and to lesser extent natural peptides) on different mass spectrometers, mainly having a quadrupole as detector. *SRMAtlas* allows the user to retrieve high probability fragmentations for a given peptide, searching into one of the available SRM builds (i.e. *Complete Human SRMAtlas*, *M. tuberculosis SRMAtlas*, *Yeast SRMAtlas*). More in details, it is possible to either query the peptides transitions of a given protein or to retrieve them by searching the proteotypic peptides of interest, previously selected through *PeptideAtlas*.

Furthermore, it is possible to specify the number of fragment ions to be reported from each peptide (i.e. Number of highest Intensity Fragment Ions to Keep), to select the mass spectrometer to be used to run the retrieved MRM assays (i.e. Target Instrument), to select the MS instrument that should be considered to build the MRM assays (i.e. Transition Source), to search for labeled peptides if necessary, to select the fragment ions types included in the assays (i.e. Allowed ions types, usually b and y ions), and to select the peptide modifications to return the query results (i.e. Allowed peptide modifications).

Once specified all of these search parameters, the interface will return a query result table in which *m/z* ratios of both precursors and their fragments are reported, as well as their charge state, ion type and predicted retention time.

With the appropriate MRM assays in hand the samples, treated (or not) with a small molecule and submitted (or not) to the double-digestion procedure, are analyzed by LC-MRM-MS, to give quantification of the maximum number of fully tryptic peptides for a given protein (i.e. targeted-LiP-MRM: the molecule target protein needs to be already disclosed from previous experiments). Indeed, when trypsin is added to partially digested samples, its cleavages can either occur on the non-specific subtilisin products or in undigested protein regions: in the former case semi-tryptic peptides can be produced, whereas in the latter circumstance fully-tryptic peptides are generated.

Since the interaction with a small molecule stabilizes a protein, inducing structural changes that might reduce subtilisin action onto the regions directly or distally involved in the interaction event, a larger amount of fully tryptic peptides will be observed in respect of the untreated sample, where the limited proteolysis trend is only dictated by the conformational features of the protein itself. On the other hand, no significant change should be detectable between a molecule-exposed and unexposed sample for fully tryptic peptides corresponding to protein regions not involved in the interaction and thus not undergoing structural fluctuations.

Despite being experimentally quite similar to DARTS, t-LiP-MRM offers the undeniable advantage of both confirming the interaction a compound engages with its target protein(s) and, most importantly, of pinpointing the specific target(s) region(s) locally or distally affected by the compound binding, such as direct binding sites or allosteric paths.

With the development and continuous updating of *PeptideAtlas* and *SRMAtlas* repositories, what could in principle represent LiP major challenge (i.e. the development of a robust MRM assay) is now becoming a not so difficult to accomplish task for proteins whose data are actually deposited in these on-line resources.

T-LiP is a fast and straightforward technique that can be exploited to detect several protein structural changes, such as subtle alterations in secondary structure, larger-scale allostery movements and more pronounced transitions (i.e. the swap between folded and unfolded states, multimerization events and protein(s) aggregation)[138].

Furthermore, t-LiP-MRM offers the advantages of being a simple and high throughput technique (i.e. the time required to perform such an experiment is comparable with that required for a classic quantitative proteomic experiment, with multiple samples being handled in parallel) directly applicable to complex biological matrices.

Nevertheless, t-LiP also presents several caveats.

As an example, some proteins might not be susceptible to limited proteolysis in any of their conformations, thus being refractory to this kind of approach. Furthermore, t-LiP requires extraction and handling of a proteome under non-denaturing conditions, which restricts its use to soluble proteins and hampers the characterization of membrane proteins, requiring peculiar extraction and solubilization strategies. Moreover, being a targeted approach, t-LiP requires *a priori* known data about the protein to be analyzed, thus the target protein of a bioactive compound must have already been discovered through other CCCP techniques, like DARTS.

To overcome this limitation, a discovery driven LiP application, called LiP-small molecule mapping (LiP-SMap), has very recently been presented to enable the systematic and unbiased analysis of protein/small molecules interactions in the biologically relevant context of a cell lysate[148]. Thus, LiP-SMap aims to detect ligand-induced structural alterations on a proteome-wide scale, investigating proteins that become differentially susceptible to protease cleavage upon binding of a small molecule added to a proteome extract.

The LiP-SMap method shares the same experimental workflow with t-LiP-MRM, but the peptide mixtures are analyzed in an un-targeted manner, through quantitative and unbiased *shotgun* proteomics, comparing LiP patterns of proteomes exposed or not to the molecule and using a *label-free* quantitative MS approach.

Although the small molecule of interest can be applied to cultured cells, addition of the compound to a proteome extract (e.g. after cell or tissue lysis) is preferable, to minimize the detection of indirect effects such as protein structural changes in pathways modulated by the compound.

Despite offering the undeniable advantage of probing protein/small molecule interaction in a complex matrix and through a simple and fast experimental procedure, LiP-SMap necessitates of high-performance and high-accuracy mass spectrometers with fast scanning rates (e.g. Q-exactive plus mass spectrometers) and of an extensive and time-consuming data analysis to identify peptides whose intensity significantly changes in presence of a compound. This step typically involves searching the MS/MS raw data on a dedicated software (e.g. MaxQuant[149,150] or similar) to analyze LiP samples by treating every single peptide as an independent entity to obtain peptide-level abundance changes. The retrieved data should then be submitted to statistical testing to identify significantly changed peptides and to calculate median abundance changes and associated p values corrected for multiple testing (i.e. q values). A filtering step should then be performed, selecting protein whose $\log_2$ abundance change is higher than two-fold and whose q values are lesser than 0.01. The obtained results could then be plotted in the

form of a volcano plot, representing peptide abundance changes versus the associated q values: peptides that significantly change abundance in the differentially treated LiP samples identify the specific protein(s) region(s) undergoing structural changes in presence of the compound, and the proteins they map to, are recognized as the molecule interacting partners.

1.4 Aim of the PhD project

Natural products (NPs) have long been explored as invaluable sources for drug discovery. In this scenario, the functional proteomics AP-MS approach has been the last 15 years top-choice technique, having brought brilliant results in the *targetome* profiling of bioactive compounds. Briefly, this strategy relies on a pull-down experiment in which a chemical functionalized NP is used as a bait to *fish out* its interacting partners from a complex mixture.

Nevertheless, despite being a well consolidated, robust and straightforward approach having brought undeniable results, this strategy is not universally applicable, being limited by the chemical features of the molecule to be characterized for the mandatory *on-beads* immobilization step. This strictly requirement makes AP-MS a non-universally applicable strategy: compounds that do not exhibit properly reactive functional groups cannot be taken into analysis, as well as compounds showing a promiscuous and difficult-to-handle reactivity. Furthermore, if no information on the analyte SAR is available, as it's often the case for NPs, even when the covalent modification is possible blocking or modifying a functional group would leave some concerns about the compound retaining its original bioactivity. Furthermore, AP-MS is not capable of providing information about NPs covalent interactors and of shedding light on NP/target protein(s) interaction features.

Thus, to overcome AP-MS major drawbacks and to provide a universally applicable and more comprehensive Drug-Target Deconvolution strategy, my PhD project has been devoted

to the development and optimization of a functional proteomics platform based on two complementary *label-free* strategies, DARTS and t-LiP-MRM. This platform has the undeniable advantage of being universally applicable to any molecule, regardless of its chemical features, not requiring any chemical modification of the compound to be studied for both the identification of its interacting protein partner(s) and the characterization of their interaction features. Indeed, at first DARTS is exploited to identify NPs most reliable cellular partners, then t-LiP-MRM is carried out to investigate the molecules/target proteins interaction features.

Briefly, Chapter 2 presents the preliminary analyses carried out to optimize DARTS and t-LiP-MRM, as well as the subsequent *proof of concept* achieved through the characterization of the well-known radicicol/Heat shock 90kDa proteins system through this functional proteomics platform.

Then, Chapter 3, 4 and 5 report the *interactome* analysis of three marine-derived NPs presenting distinctive scaffolds and reactive moieties (i.e. a norcembranoid, a sulfated-bis-steroid and an acetogenin), to evaluate the DARTS/t-LiP-MRM platform feasibility on compounds showing different chemical features. When needed, AP-MS was also performed in parallel to DARTS to achieve a clearer picture of complex NPs target profiles.

Moreover, the proteomics results were validated by Western Blotting, to confirm NPs interaction with their DARTS-identified target, and by *in silico* molecular docking, to corroborate t-LiP-MRM information about the target region(s) involved in the NP binding. The direct effects of the disclosed interactions were also assessed through proper *in vitro* and/or *in cell* biological assays.

RESULTS AND DISCUSSION

CHAPTER 2

A *label-free* functional proteomics platform: optimization and coupling of Drug Affinity Responsive Target Stability (DARTS) and targeted Limited Proteolysis-Multiple Reaction Monitoring Mass Spectrometry (t-LiP-MRM).

2.1 Background

In the last 15 years, our research group successfully disclosed the *interactome* of several natural bioactive compounds[95,151–158], defining their mechanisms of action and revealing their biological profiles through AP-MS (paragraph 1.3.2). Briefly, this strategy relies on a *pull-down* step, performed in either a homogenous or a heterogeneous phase, in which the chemical functionalized version of the molecule under analysis is used as a bait to *fish out* its interacting partners from a complex mixture. As already mentioned, despite being a well consolidated, robust and straightforward approach, AP-MS suffers from the necessary covalent modification of the bioactive compound previous to the affinity chromatography step: on this basis, molecules that do not exhibit properly reactive structural features cannot be submitted to this strategy. Furthermore, even when the covalent modification is possible, the functionalization of a reactive group leaves some concerns about whether or not the compound original bioactivity is still retained, in the absence of SAR studies.

In order to overcome these limitations with a valid alternative Drug-Target Deconvolution strategy, my PhD project has been devoted to the development and optimization of a functional proteomics platform based on two complementary *label-free* strategies, for the identification of the interacting protein partner(s) of bioactive small molecules and the characterization of their interaction features, avoiding any chemical modification of the analyzed compounds.

This platform is essentially based on the coupling of the protein-centered Drug Affinity Responsive Target Stability strategy (DARTS, paragraph 1.3.3.1) with the peptide-centered targeted-Limited Proteolysis coupled to Multiple Reaction Monitoring approach (t-LiP-MRM, paragraph 1.3.3.2). Indeed, in a first step DARTS is exploited to characterize the natural compounds *interactome*, then t-LiP-MRM is carried out to investigate the molecule/target protein(s) interaction features.

Although DARTS and t-LiP-MRM are mostly superimposable up to the completion of the limited proteolysis step, they become very different in the following phases, from the handling of the partially digested samples to the bioinformatics tools required to build the mass spectrometric methods and to perform data analysis.

In this scenario, several attempts have been made towards the optimization of both DARTS and t-LiP-MRM procedures, which will be extensively explained in the following pharagraphs and brought to the experimental outline summarized in Figure 28.

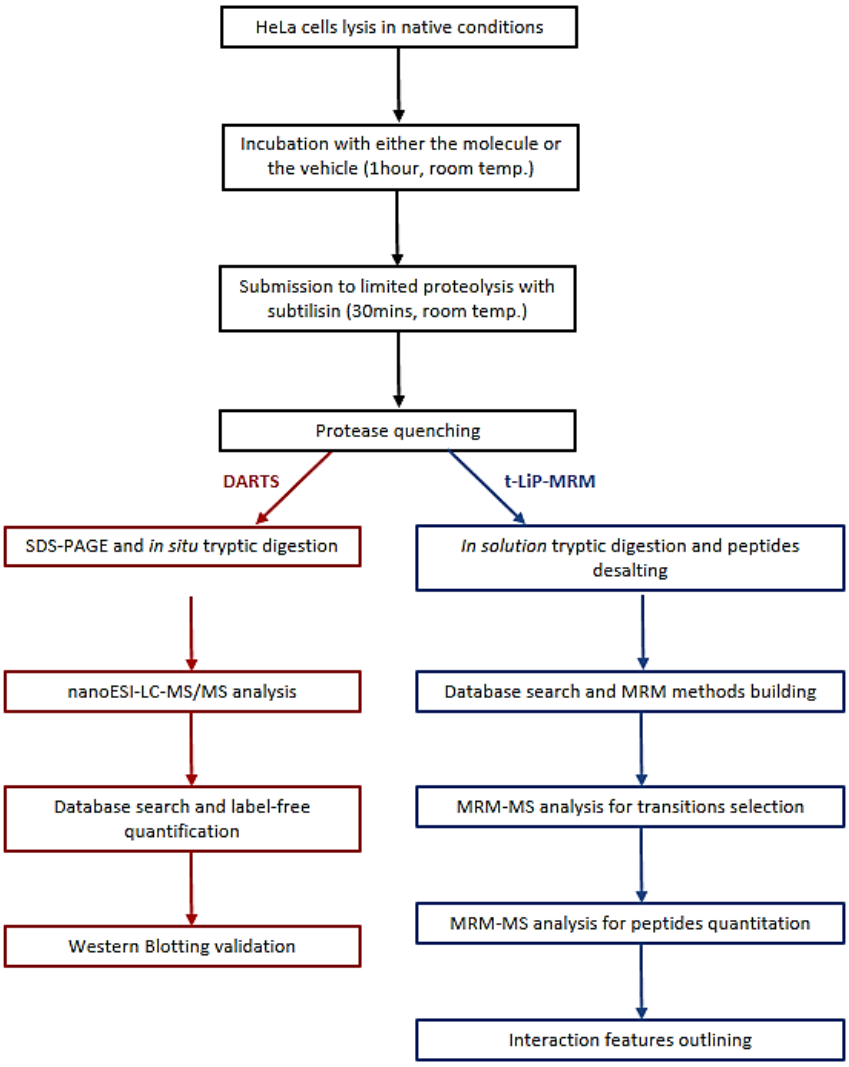


***Figure 28*:** *DARTS and t-LiP-MRM experimental outlines.*

2.2 DARTS strategy optimization

As already explained in the previous chapter, DARTS exploits the thermodynamic stabilization conferred by a molecule to its target protein(s) that develop, as a result of the interaction with the compound, resistance to limited proteolysis.

In a typical DARTS experiment, cells are lysed under non-denaturing conditions and the obtained lysates are incubated with the sample molecule (at several concentrations) and then submitted to a limited proteolysis step with a non-specific protease. The protease is then quenched and the protein mixture separated through 1D-SDS–PAGE: all the proteins with an altered proteolytic pattern, due to the interaction with the sample molecule, are spotted by an *on eye* inspection of the stained gel, looking at bands that result more intense in the treated versus the untreated sample. Once excised, these bands are submitted to *in situ* tryptic digestion and nano-ESI-LC-MS/MS analysis. The output results are searched against a database for protein identification and *label-free* quantitation is carried out to define which proteins are more abundant in the compound-treated sample, as a result of their thermodynamic stabilization.

Due to its peculiar procedure, DARTS is a very tricky technique, and several optimization steps had to be done to reach the right workflow for Drug-Target Deconvolution purposes.

2.2.1 Limited Proteolysis step optimization

The limited proteolysis step is clearly the most crucial part of a DARTS experiment, and needs to be finely tuned to find the most suitable protease and its best handling conditions. It is indeed essential to reach an appropriate equilibrium between the need to perturb the system and to appreciate the level of protection exerted by the molecule interaction with a given protein, which are dependent on the selected protease, its concentration and the incubation time.

As a best result, the *on eye* comparison of the SDS-PAGE runs, corresponding to the lysate samples treated or not with the small molecule, should detect bands showing increasing intensity accordingly with the compound concentration, due to their stabilization towards proteolisys.

Among the proteases that could be used to perform DARTS experiments, thermolysin and subtilisin are the most employed[131,159]. As already described (paragraph 1.3.3.1), thermolysin is able to efficiently digest only unfolded proteins: since the fraction of proteins exhibiting a significant unfolding rate under the pseudo-physiological conditions required to perform DARTS is just a small portion of the proteome, thermolysin has a limited range of application.

Conversely, subtilisin is an unspecific serine protease, mostly cleaving at C-terminus of large uncharged amino acids, able to digest both folded and unfolded proteins. Thus, since our purpose was to exploit DARTS as a tool for target identification, subtilisin has been preferred over thermolysin due to its higher substrate dynamic range, which allowed us to examine a higher proteome fraction carrying out a less biased target identification procedure.

Once chosen the protease, the enzyme to proteins ratio and the most appropriate digestion time were then under analysis. The former parameter is strictly dependent on the molecule protecting effect and thus might need to be tuned according to its behavior and/or to particular experimental requirements. Nevertheless, to avoid this case-to case variability, we sought to perform preliminary titration experiments with different subtilisin quantities to define a general procedure circumventing issues related to target proteins different enzymatic susceptibility.

Thus a HeLa cell lysate, obtained in non-denaturing conditions, was submitted to limited proteolysis with four different subtilisin amounts (enzyme to lysate proteins ratios of 1:2500, 1:1500, 1:1000 and 1:500 w/w respectively) for 30 minutes at room temperature and under continuous shaking. An undigested sample was kept to be used as a reference. A SDS-PAGE was then run and, after Coomassie staining, each gel lane was submitted to a densitometric analysis, instead of relying on a mere *on eye* visualization. More in details, this analysis was

performed taking into account the molecular weight ranges showed through the dashed red lines in Figure 29A and exploiting the pixel intensities of the undigested sample as normalizing factors. A graph was then plotted to show the digestion trend for each molecular weight range at each subtilisin amount (Figure 29B).

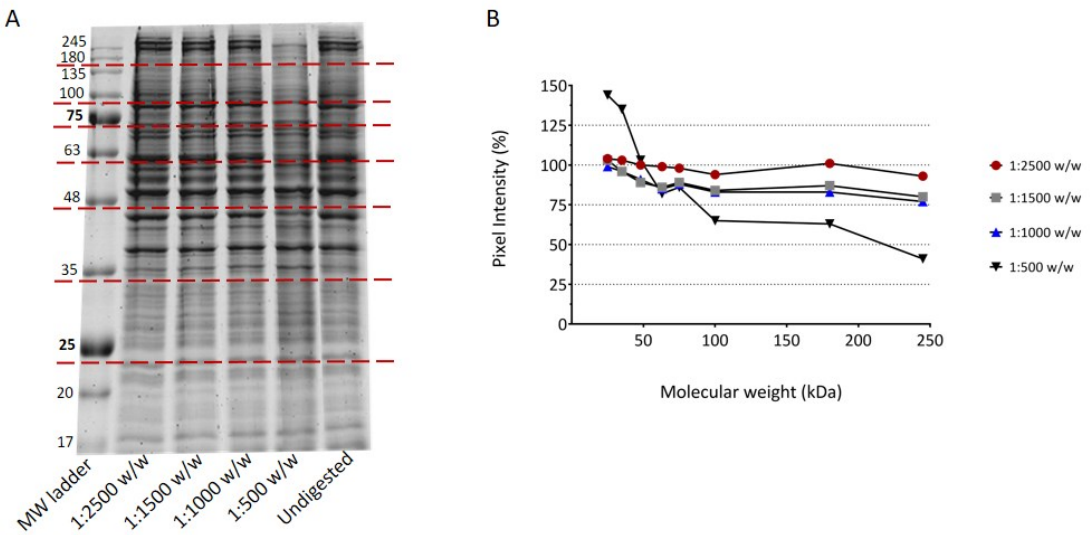


Figure 29: *Subtilisin titration experiment. (A) SDS-PAGE Coomassie stained run. Red dashed lines define the molecular weight ranges exploited for the densitometric analysis. (B) Densitometric analysis output. Undigested sample is not depicted in the graph, since its pixel intensities have been set at 100%.*

As can be observed in the figure, the subtilisin to proteins ratio of 1:2500 w/w led to an insufficient digestion, compared with the enzyme-untreated sample, for all of the molecular weight ranges. Furthermore, no substantial differences could be detected between the 1:1500 and 1:1000 w/w ratios, which both showed a quite good digestion extent. The 1:500 w/w ratio resulted as the best subtilisin amount, since it led to a substantial difference in pixel intensities as compared to the undigested sample in a quite large range of molecular weights. More in

details, the highest digestion rate could be observed above 200 kDa, where only half of the reference sample total pixel intensity was still detectable. Between 200 kDa and 100 kDa the digestion rate was still satisfactory, since only 60% of the total intensity could be detected. For molecular weights comprised between 100 kDa and 50 kDa this percentage raised to 80%, as a result of both a less efficient digestion and the co-migration on gel of the subtilisin partially digested protein fragments. The latter phenomenon could be largely appreciated at molecular weights lower than 50 kDa, where the pixel intensity became very strong, due to a huge accumulation of the limited proteolysis products.

On the basis of the digestion trends, we decided to discharge the 1:2500 and 1:1500 w/w ratios and perform another titration experiment in which the 1:1000, 1:500 and the newly added 1:100 w/w subtilisin to proteins ratios were tested at the digestion times of 30 minutes and 1 hour (Figure 30).

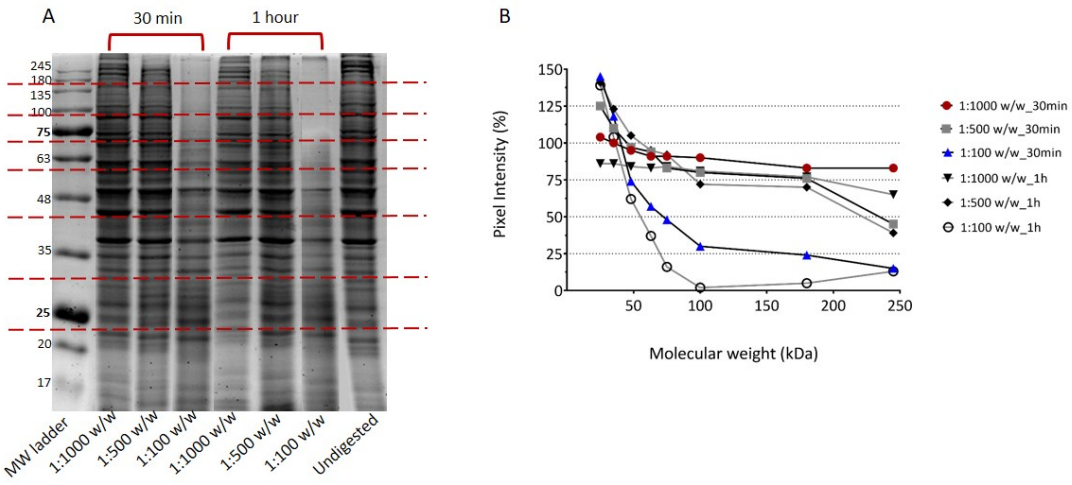


Figure 30: *Subtilisin titration experiments carried out at 30 minutes and 1 hour digestion time. (A) SDS-PAGE Coomassie stained run. Red dashed lines define the molecular weight ranges exploited for the densitometric analysis. (B) Densitometric analysis output. Undigested sample is not depicted in the graph, since its pixel intensities have been set at 100%.*

As shown in Figure 30B, no substantial difference between 30 minutes and 1 hour digestion time could be observed for the 1:1000 and 1:500 w/w subtilisin to proteins ratios, whereas the digestion rate for the 1:100 w/w ratio was markedly more pronounced after 1 hour. Despite this evidence, as can be especially noticed from Figure 30A, the last condition led to a drastic depletion of the proteins with molecular weight above 50 kDa, which could be an issue when it comes to actually appreciate the protection exerted by a molecule on its target protein(s).

Thus, to ensure a correct balance between acceptable digestion rates in the widest range of molecular weights without excessively perturbing the proteins system, the subtilisin to proteins ratios of 1:1000 and 1:500 w/w and 30 minutes digestion time were selected as the best compromise before proceeding with the following DARTS optimization steps.

2.2.2 Subtilisin quenching

A DARTS experiment requires a strict comparison between cell lysate samples treated or not with a small molecule, at a given subtilisin to proteins ratio, to visualize and identify the proteins interacting with the compound. It is thus mandatory that all of the samples are digested for the same time, at the end of which subtilisin proteolytic action needs to be promptly blocked.

For this purpose, several quenching strategies were tested to find the most suitable and reliable one.

Thus, subtilisin was added to HeLa cell lysate samples (1:500 w/w), which were immediately treated with either Laemmli buffer, 8M Urea or 8M Urea supplemented with a protease inhibitors cocktail, and leaved shaking for 30 minutes at 25°C prior to the SDS-PAGE analysis. Furthermore, one subtilisin-treated aliquot was kept as a positive control of the enzymatic digestion and an undigested sample was also reserved as a reference (Figure 31).

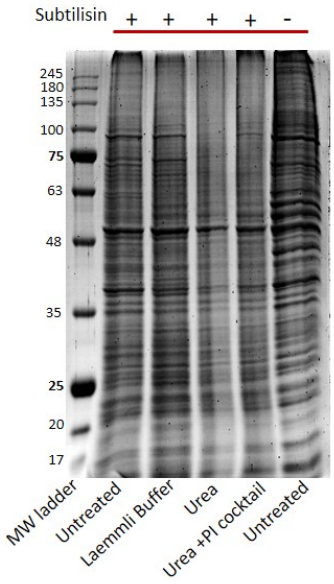


Figure 31: *Subtilisin quenching strategies.*

As can be observed, none of the above presented strategies resulted effective in quenching the subtilisin proteolytic action and, moreover, the digestion extent in the urea processed samples resulted much higher compared to the subtilisin-only exposed one, suggesting that urea-induced denaturation helps the enzyme to digest proteins with a faster rate.

Another experiment was then performed, exploiting the serine protease covalent inhibitor phenylmetane-sulfonil-fluoride (PMSF). Briefly, subtilisin was added to HeLa cell lysate aliquots (1:500 w/w), leaving an undigested sample as a reference. As soon as subtilisin was added, the samples were treated with PMSF at the final concentrations of 0.1mM, 0.5mM and 1mM respectively, and shaked for 30 minutes at 25°C. As for the previous experiment, one of

the subtilisin-treated aliquots was not supplemented with PMSF to be kept as a digestion control (Ctrl sample). The samples were then submitted to SDS-PAGE and a densitometric analysis was carried out to assess PMSF quenching rate (Figure 32).

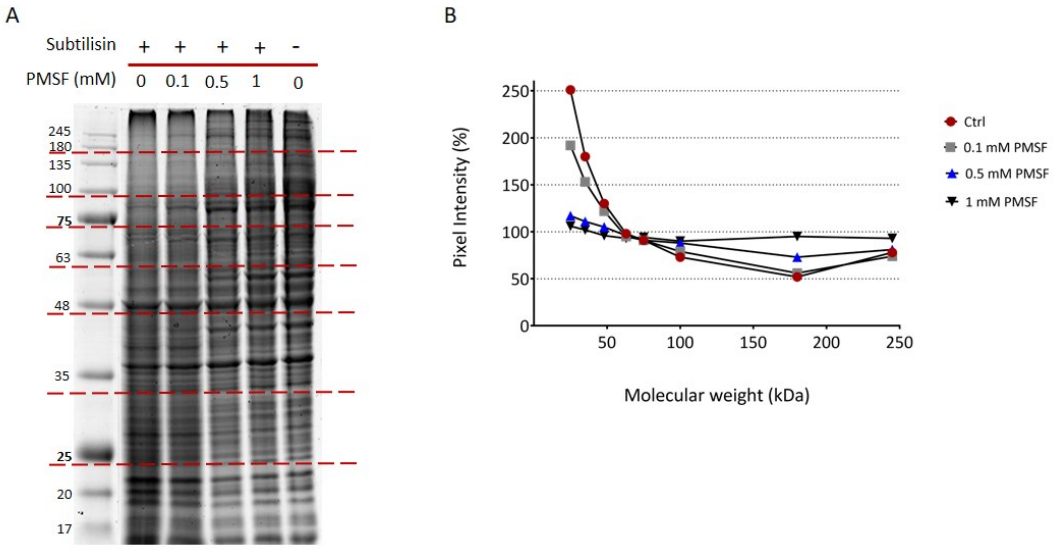


Figure 32: *Subtilisin quenching with PMSF. (A) SDS-PAGE Coomassie stained gel run. Red dashed lines define the molecular weight ranges exploited for the densitometric analysis. (B) Densitometric analysis output. Ctrl: PMSF-unexposed and subtilisin-digested sample (first lane of the gel). Undigested sample is not depicted in the graph, since its pixel intensities have been set at 100%.*

As can be observed by the densitometric analysis output, 0.1 mM PSMF was not enough to quench subtilisin, as the obtained curve resembled the control sample one, whereas 0.5 mM PMSF resulted more effective, even if the digestion rate at molecular weights higher than 100 kDa was still too pronounced. Conversely, the 1 mM PMSF curve showed the best results resembling, along all molecular weight ranges, the undigested sample one. Thus, 1 mM PMSF was chosen as the best subtilisin quenching strategy for the following DARTS experiments.

2.2.3. Testing the optimized strategy: the Radicicol/Hsp90s system

Once optimized DARTS most critical features, we moved on to test this strategy on the well-known radicicol/Hsp90 system.

Radicicol is a macrocyclic antifungal antibiotic (Figure 33) which targets the N-terminal ATP/ADP binding site of Hsp90, thus being a potent and specific Hsp90 ATPase activity inhibitor, with nanomolar affinity[160,161].

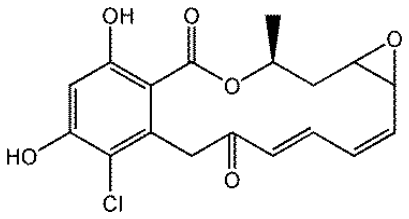


Figure 33: *Radicicol.*

To test the efficacy of our workflow, a DARTS experiment was performed treating a HeLa cell lysate with radicicol at two different concentrations. More in details, the lysate was prepared in non-denaturing conditions, suspending a HeLa pellet in PBS 0.1% Igepal supplemented with an endogenous protease inhibitors cocktail and lysing it on ice through a Dounce manual homogenizer. The obtained lysate was then centrifuged to remove the debris, and the protein concentration was determined and adjusted to 3 μg/μL with PBS. 300 μg aliquots of the obtained lysate were incubated with either radicicol (at the final concentrations of 0.8 μM or 8 μM) or just the vehicle (DMSO) for 1 hour at room temperature and under continuous shaking. Each sample was then divided into three aliquots: one of them was kept as reference, whereas the other two were treated with 1:1000 w/w and 1:500 w/w subtilisin. The limited proteolysis was carried out at 25°C, for 30 minutes and under continuous shaking.

Subtilisin was then quenched with 1mM PMSF and the samples submitted to SDS-PAGE on a 12% polyacrylamide gel. After Coomassie staining, the gel was both inspected *by eye* and submitted to a densitometric analysis to highlight protein bands whose intensity increased with a direct proportionality to radicicol concentration. The subtilisin to proteins ratio of 1:1000 w/w resulted the best condition to appreciate radicicol effects, with the best protein stabilization effect at the lowest molecule amount around 90 kDa (gel image and densitometric analysis, Figure 34 A and B).

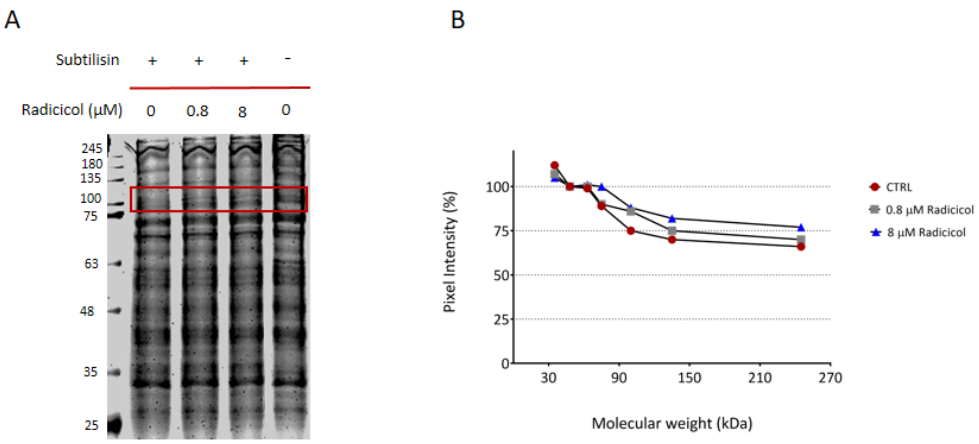


Figure 34: *DARTS experiment carried out with radicicol on a HeLa cell lysate. (A) Coomassie stained gel showing proteins protection from subtilisin (1:1000 w/w) upon radicicol interaction (red box) (B) Densitometric analysis output. Proteins at 90 kDa are the most protected towards proteolysis at the lowest radicicol concentration. Undigested sample is not depicted in the graph, since its pixel intensities have been set at 100%.*

Thus, protein bands whose intensity increased in presence of radicicol were carefully excised from all of the lanes and submitted to an *in situ* tryptic digestion[34] to give peptide mixtures for the following nano-ESI-LC-MS/MS analysis, performed on a nano-flow UPLC system interfaced with the nano-ESI source of a LTQ Orbitrap XL mass spectrometer. The obtained

data, firstly elaborated by the Xcalibur software to give peak lists containing all of the MS information about the peptides mixtures, were then searched against SwissProt Human database through the bio-informatics server MASCOT Daemon. This engine allows to retrieve protein identification from their corresponding peptides, by comparing the experimental digests with the ones generated *in silico* from the SwissProt database proteins. The strength of a peptide match is based on the coincidence of both the precursor peptide mass and the MS/MS experimental fragment masses with the predicted data. A protein with more matching MS/MS scans in the search process is marked with a higher overall score and top ranked proteins are often the most abundant ones in the sample.

MASCOT output comprises a list of proteins, defined by their SwissProt ID, molecular weight and a short description, together with the statistical parameters correlated to the identification process. Among these parameters, as briefly mentioned before, the Score is directly related to the reliability of the identification process: the higher the score, the more robust the identification process. Together with this parameter, MASCOT assigns to each protein a number of Matches, being the number of peptides directly matched to a certain protein, and an emPAI value that takes into account the matched peptides in respect to the total number of peptides present in the database for that entry. The higher their values, the more confident the identification process.

Thus, interpretation of the DARTS results was accomplished through a semi-quantitative analysis, taking into account the matches values, linking the number of assigned peptides to the abundance of a certain protein in a sample. Correlating this concept with a DARTS experiment, the positive control (untreated and undigested) represents the top-matches identification sample, whereas the negative control (untreated and subtilisin exposed) should give the least matches assigned to a determined protein, due to its lower amount consequent to enzymatic hydrolysis.

Nevertheless, a protein interacting with a molecule will give a higher number of matches compared to the negative control: due to its reduced proteolysis susceptibility a major amount of full-length protein can be found.

Thus, to asses radicicol effect on its partners, the matches assigned to the negative control proteins (i.e. untreated with radicicol and submitted to limited proteolysis) were compared with those assigned to proteins upon radicicol treatment and subtilisin action, using the positive control matches as normalizers, to give corrections for different proteins abundance in the lysate. A protection percentage was then calculated for every single protein as follows:

$$\text{Protection (\%)} = [(\text{Matches}_{\text{Radicicol}} - \text{Matches}_{\text{Control}})/\text{Matches}_{\text{Lysate}}]*100.$$

Thus, radicicol protected proteins were identified as those showing both the highest protection percentages at the lowest molecule concentration and protection percentages increasing accordingly with the molecule amount.

As can be observed in Figure 35A, Hsp90-α and Hsp90-β show matches increase in a radicicol concentration-dependent fashion, resulting the most protected proteins at the lowest molecule concentration (Figure 35B).

A

Accession	Mass	Description	ctrl			0.8 µM radicicol			8 µM radicicol			Lysate		
			score	matches	emPAI	score	matches	emPAI	score	matches	emPAI	score	matches	emPAI
P07900	85 006	Heat shock protein HSP 90-alpha	1017	45	0.71	3172	102	1.74	4662	149	2.26	52112	150	3.32
P08238	83 554	Heat shock protein HSP 90-beta	621	24	0.31	2124	68	1.27	2031	77	1.2	2026	86	0.58

B

Accession	Mass	Description	0.8 µM	8 µM
P55786	103 895	Puromycin-sensitive aminopeptidase	0	0
O14787	102 862	Transportin-2	0	0
P11586	102 180	C-1-tetrahydrofolate synthase, cytoplasmic	0	0
Q14974	98 420	Importin subunit beta-1	8	38
Q92598	97 716	Heat shock protein 105 kDa	0	0
P11217	97 487	Glycogen phosphorylase, muscle form	0	0
P13639	96 246	Elongation factor 2	24	12
P34932	95 127	Heat shock 70 kDa protein 4	25	31
P14625	92 696	Endoplasmin	5	5
P55072	89 950	Transitional endoplasmic reticulum ATPase	24	32
Q9NR30	87 804	Nucleolar RNA helicase 2	0	0
P07900	***85 006***	***Heat shock protein HSP 90-alpha***	***68***	***100***
P08238	***83 554***	***Heat shock protein HSP 90-beta***	***100***	***100***
P19338	76 625	Nucleolin	0	0
Q92945	73 355	Far upstream element-binding protein 2	0	0
Q9BZQ2	73 327	Testicular spindle-associated protein SHCBP1L	0	0

Figure 35: *DARTS experiment output. (A) Hsp90-α and Hsp90-β MASCOT identification parameters. (B) Shortened list of identified proteins, reported with their Accessions, Molecular Weights and protection percentages at both radicicol concentrations.*

To further validate the mass-spectrometric DARTS data, the same samples were submitted to a Western blotting analysis with an anti-Hsp90s antibody (Figure 36A). Since GAPDH (i.e. Glyceraldehyde 3-Phosphate Dehydrogenase) proved to be subtilisin resistant under these experimental conditions, it was used as a loading normalizer for the subsequent densitometric analysis (Figure 36B).

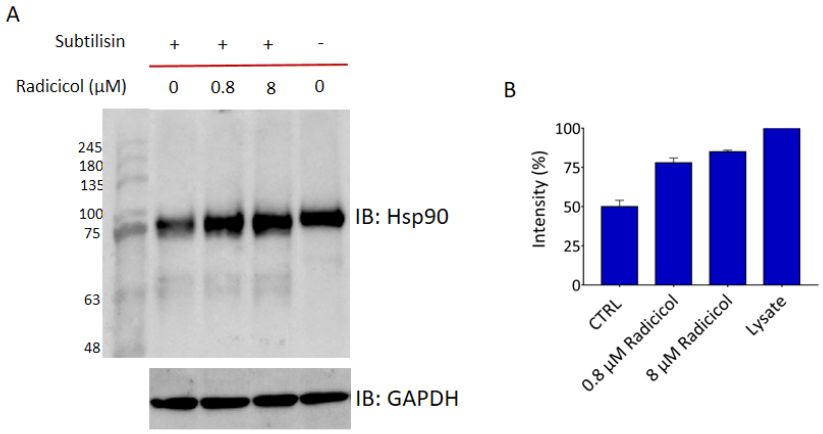


Figure 36: *(A) Western blotting analysis carried out on the DARTS samples reveals Hsp90s protection upon radicicol interaction. GAPDH is exploited as a loading normalizer. (B) Western blotting densitometric analysis performed through ImageJ. The histogram is the result of two independent experiments (standard deviations are reported). Undigested Hsp90s (i.e. Lysate sample) were rated as 100%.*

As can be observed, radicicol exerts a protective effect upon subtilisin towards Hsp90s, thus validating the mass spectrometric data and the reliability of our DARTS workflow.

2.3 T-LiP-MRM strategy optimization

The second step of our Drug Target Deconvolution strategy was the optimization of the t-LiP-MRM approach, with the aim of identifying proteins domains directly or distally affected by the interaction with a small molecule by measuring abundance changes of their fully tryptic peptides.

As already mentioned in paragraph 2.1, DARTS and t-LiP-MRM share the same experimental features in the initial steps. Thus, we took advantage of the same conditions optimized for DARTS in our t-LiP-MRM workflow, incubating lysate samples with either the molecule or just the vehicle, submitting them, for 30 minutes, to limited proteolysis with different subtilisin amounts and then quenching the proteolysis with 1 mM PMSF.

Then, as opposite to DARTS, t-LiP-MRM samples have to be denatured (i.e. adding urea to a final concentration of 4M) to carry out an extensive *in-solution* tryptic digestion. The tryptic peptide mixtures are then submitted to a desalting step (i.e. through C18 packed cartridges) and analyzed through LC-MRM-MS. Measurement of the fully tryptic peptides areas corresponding to the target protein(s) gives a map of the different proteolytic patterns generated as a consequence of the ligand binding.

Being the most crucial part of such a workflow, several experiments were carried out to optimize the targeted LC-MRM-MS analysis, and preliminary studies were focused on finding the best LC conditions to exploit. Other important issues, such as the MRM-MS method building and a first strategy for data analysis will be also discussed later on in this chapter (paragraphs 2.3.2.1, 2.3.2.2 and 2.3.2.4).

2.3.1 LC parameters optimization: tackling sensitivity issues

Due to instrumental issues, we were forced to optimize the LC-MRM-MS on a UPLC system instead of a nano-flow one. For this reason, we needed to find LC parameters being a good compromise between a satisfactory sensitivity for the quantification the lower abundant proteins and reasonable analysis times.

For this purpose, a tryptic digest of Bovine Serum Albumin (BSA) was prepared through *in solution* digestion and desalting, and 35 ng were analyzed through nano-ESI LC-MS/MS in DDA mode, on a nanoACQUITY system coupled with an LTQ Orbitrap XL mass spectrometer.

The LC-MS run was then exploited to retrieve the parameters for the subsequent MRM method. More in details, the six most intense precursor ions (Q1_*m/z*), as well as their three most intense y type daughter ions (Q3_*m/z*), were selected for the following analysis, performed on the same BSA digest injected on a Shimadzu LC system coupled with a 6500 Q-TRAP.

We started working, in UPLC conditions, at a flow rate of 400 µL/min, to first identify the best transition for each peptide, which was selected as the one giving the most intense signal and the lowest noise level out of the three reported in the method. For this purpose, 500 ng of BSA digest were analyzed: peptides were separated through a 15 minutes gradient at the previously stated flow rate.

The selected transitions are reported in Figure 37B and were exploited for the following LC-MRM-MS runs, focused to disclose the limit of detection for these UPLC conditions. Thus, 100ng, 200ng and 500ng of BSA digest were injected in the same LC conditions and with a MRM method made of the previously selected transitions (the LC-MRM-MS trace for the 500ng sample is reported in Figure 37A), revealing 100ng of tryptic digest as our lowest detection limit (Figure 37C), but also enlightening a good correspondence between the amount of injected digest and the obtained signal (R^2 values higher than 0.95, data not shown).

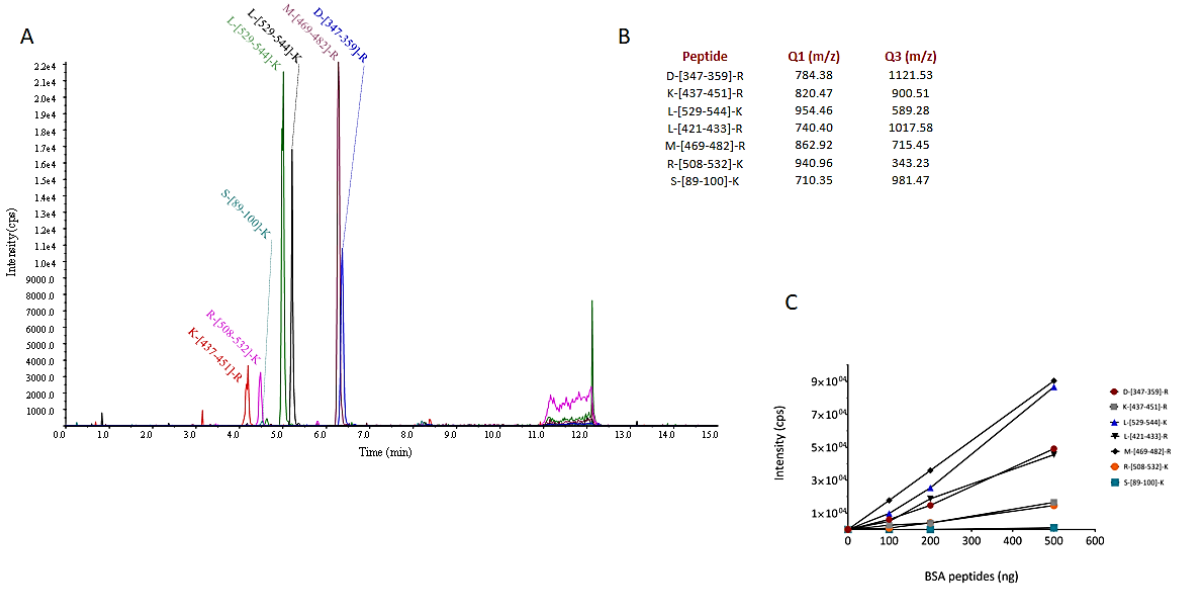


Figure 37: *BSA tryptic digest analysis in UPLC conditions. (A) LC-MRM-MS trace obtained injecting 500ng of BSA digest at a 400 µL/min flow rate. (B) BSA selected transitions. (C) BSA detection limit evaluation.*

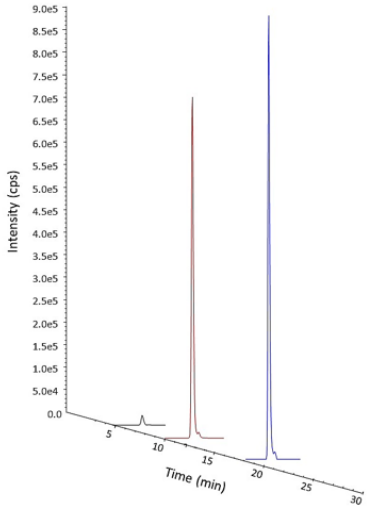
We then moved to apply lower flow rates, to gain sensitivity with the aim of adapting our t-LiP-MRM strategy to the analysis of a broadest range of proteins, avoiding the need of an enormous amount of peptides for an acceptable quantification of the low abundant proteins.

For this purpose, 100ng of BSA digest were analyzed at the flow rates of both 200 µL/min and 100 µL/min, with 30 minutes and 1 hour gradients, respectively.

An example trace (extracted ion current, XIC) obtained for the peptide L-[421-433]-R is reported in Figure 38A, showing a more then 700 fold increase in signal intensity when shifting from a flow rate of 400 µL/min to a flow rate of 200 µL/min, whereas only a slight increase can be observed when shifting from a 200 µL/min to a 100 µL/min flow rate.

To better estimate the sensitivity gain, the area of each peptide was then quantified and ratios were calculated in respect to the previous conditions, as reported in Figure 38B.

A



B

ID	$UPLC_{200}/UPLC_{400}$	$UPLC_{100}/UPLC_{400}$	$UPLC_{100}/UPLC_{200}$
D-[347-359]-R	250	560	2.24
K-[437-451]-R	140	250	1.79
L-[529-544]-K	25	45	1.81
L-[421-433]-R	910	2300	2.53
M-[469-482]-R	70	140	2.00
R-[508-532]-K	60	100	1.67
S-[89-100]-K	715	1400	1.96
Average	**310**	**685**	**2**

Figure 38: *BSA analysis at different flow rates. (A) XICs of the L-[421-433]-R BSA tryptic peptide at the flow rates of 400 µL/min (black trace), 200 µL/min (red trace) and 100 µL/min (blue trace), reported at their retention time and with their corresponding signal intensity. (B) Ratios calculated dividing the area of each BSA peptide at the flow rate of 200 µL/min (first column of the table) and 100 µL/min (second column of the table) for the area of the corresponding peptide obtained at 400 µL/min. Last column reports the comparison of the outputs obtained at the flow rates of 100 µL/min and 200 µL/min.*

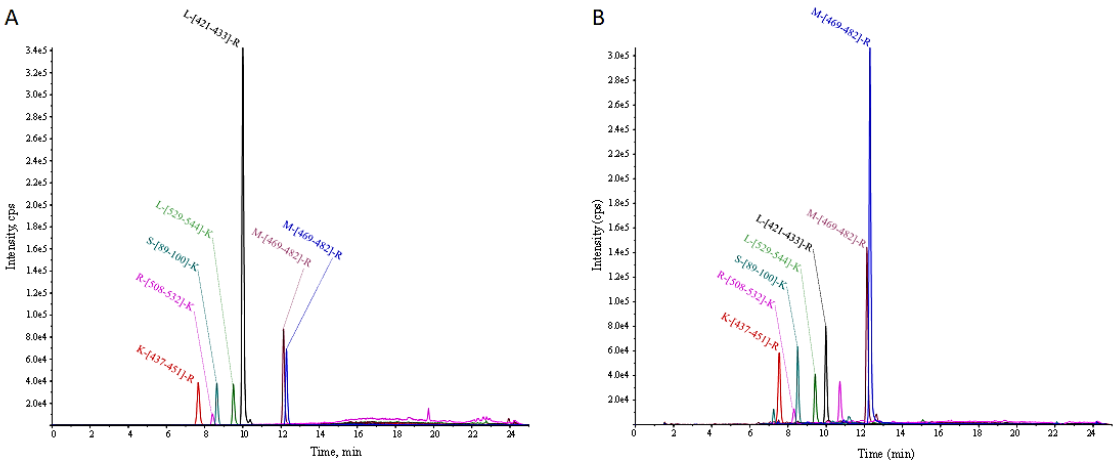
More in details, an average intensity gain was calculated from all the analyzed peptides. As can be appreciated in Figure 38B, more than 300 folds increase of the peaks area was obtained shifting from a flow rate of 400 µL/min to a flow rate of 200 µL/min (first column of the table), whereas a further slight increase of only 2 folds could be appreciated shifting from 200 µL/min to 100 µL/min (last column of the table).

Thus, we decided to asset the flow rate at 200 µL/min with a 30 minutes gradient, which seemed to be a good compromise between an acceptable sensitivity and reasonable run times.

We then proceeded to evaluate a more complex system to test the background effects, as in a cellular lysate, on both the tryptic digestion pattern and detection sensitivity.

Thus, BSA was spiked in a HeLa cell lysate, obtained in non-denaturing conditions, at 1:100 and 1:500 w/w ratios, respectively. The samples were then denatured and submitted to *in solution* tryptic digestion and desalting and an amount equivalent to 100 ng BSA was analyzed through LC-MRM-MS with the previously defined conditions. The areas of all the BSA tryptic peptides were then measured and compared with a standard made of the same pure BSA tryptic digest exploited for the previous experiment.

As can be observed by the LC-MRM traces shown in Figure 39, the tryptic digestion patterns between pure BSA (panel A) and the BSA spiked in the HeLa lysate (BSA to proteins ratio of 1:100 w/w, panel B) appeared to be different in terms of the preferentially produced peptides. Nevertheless, when summing the peptide intensities of each sample to obtain an overall BSA intensity, no significant differences could be observed between the two samples (panel C), suggesting that background proteins do not affect the sensitivity of our proposed method.



C

ID	$BSA_{standard}$	BSA_{lysate}
D-[347-359]-R	4.83E+05	2.15E+06
K-[437-451]-R	3.69E+05	5.25E+05
L-[529-544]-K	2.83E+05	3.19E+05
L-[421-433]-R	2.43E+06	5.81E+05
M-[469-482]-R	6.13E+05	9.84E+05
R-[508-532]-K	9.01E+04	1.08E+05
S-[89-100]-K	2.88E+05	4.48E+05
BSA total intensity	4.55E+06	5.11E+06
$BSA_{lysate}/BSA_{standard}$ ratio		1.12

Figure 39: *Comparison of the MRM-MS traces of the tryptic digestion profiles of pure BSA (A) and of BSA spiked in a HeLa cell lysate (B, BSA to proteins ratio of 1:100 w/w). Peptide areas are reported in (C), together with the BSA overall intensity in the lysate (second column) and the pure system (first column) and their ratio.*

The HeLa sample in which BSA was spiked at 1:500 w/w gave the same results of the 1:100 w/w experiment, ensuring an efficient digestion and detection even at lower BSA amount (data not shown).

2.3.2 t-LiP-MRM testing: the Radicicol/Hsp90s system

Once optimized the UPLC conditions, we moved on performing t-LiP-MRM on the radicicol/Hsp90s system, to address and delineate MRM-related issues, such as the build-up of the mass-spectrometric method and some preliminary data analysis features.

Radicicol is known to interact with the N-terminal ATPase domain of Hsp90s, and this interaction site has been enlightened through both molecular docking studies on the N-terminal domains of the human α and β isoforms[162] and co-cristallization analyses on the *Saccharomyces cerevisiae* protein[160] (68% sequence homology with the human Hsp90s).

2.3.2.1 MRM-MS method build-up: prediction of Hsp90s tryptic peptides and their related fragments through PeptideAtlas and SRMAtlas

To build-up a reliable MRM method for a t-LiP experiment, an accurate analysis should be performed to identify the fully tryptic peptides of the protein(s) of interest. More in details, for an unbiased evaluation of the interaction features between a molecule and its target protein(s), proteotypic tryptic peptides should be selected[142,143]. For these peptides, additional analyses have to be carried out to identify their most intense fragment ions, to select the best transitions to use in the t-LiP experiments.

When the recombinant protein is not available, its proteotypic tryptic peptides should be obtained and tested in dedicated MRM assays to evaluate their fragmentation behavior and to define the best conditions to measure each transition. These optimization steps would require time and high costs.

Luckily, as already mentioned in chapter 1, this issue can be overcomed using the data repository *PeptideAtlas*[144], an on-line collection of observed peptides and their related information, retrieved by a large number of experimental data sets and disposable for browsing or downloading purposes.

When searching a protein, *PeptideAtlas* will automatically link to a dynamic page full of all its available information. The first section of this page provides information on alternative protein and gene name(s), linking to gene ontology and showing how many distinct peptides have been observed and their number of spectra. The *Sequence Motif* section summarizes the

peptide coverage of a protein through a color-coded diagram that shows all the peptides that map, uniquely or redundantly, in the protein sequence, providing also information on the peptides unlikely to be observed through mass spectrometry. The *Sequence* section depicts in red the peptides already observed for that protein, estimating the protein coverage.

The *distinct observed peptides* table presents all the distinct observed peptides that map to a particular protein, together with several information such as the number of times a peptide has been observed in a certain number of samples and the best probability for it to be observed. It also enlightens several scoring parameters, such as the *Empirical Observability Score* (EOS), which reflects the rate of detection of a protein via a particular peptide and the *Empirical Suitability Score* (ESS), which ranks how suitable the peptide is as a proteotypic one, on the basis of the total number of observations and the EOS. It also includes penalties if the peptide is not fully tryptic, contains missed cleavages or has undesirable residues that could influence its suitability for targeted proteomics experiments.

The subsequent two tables rank the tryptic peptides of a given protein as either PABST peptides or *predicted highly observable peptides*.

The PABST (*PeptideAtlas* Best SRM Transition) peptides are ranked based on the PABST algorithm, which reports the most suitable peptides for targeted proteomics experiments, listing in which organism(s) they were observed by *shotgun* proteomics and if they were exploited or not in MRM experiments reported in prior publications.

Low abundant proteins or proteins that have not yet been observed by MS can be addressed in the *predicted highly observable peptides* table, which reports the theoretical peptides of a protein obtained from an *in silico* tryptic digestion and utilizes five predictive algorithms[166-169] to express the likelihood of a peptide to be observed.

When a protein has peptides ranked in both the PABST and the *predicted highly observable* tables, these peptides are usually superimposable.

Thus, to identify the tryptic peptides to analyze in our first t-LiP-MRM experiment, we searched both Hsp90-α and Hsp90-β in the *PeptideAtlas* Human build: PABST peptides were disposable for Hsp90-α and were completely superimposable with the *predicted highly observable peptides*, whereas only one PABST peptide was ranked for Hsp90-β.

Thus, we retrieved the peptides to be analyzed from the PABST ranking system for Hsp90-α and from the *predicted highly observable peptides* table for Hsp90-β, for a total of 52 peptides among which 11 mapping for both of the isoforms, as could be expected by their 85.8 % sequence homology.

We then selected the best transitions for each peptide, searching for highly probable fragment ions providing the highest signal/noise intensity. To carry out such a selection, we searched the *PeptideAtlas* retrieved proteotypic peptides in the *SRMAtlas* Human build (*Complete Human SRMAtlas*), a publicly available resource of SRM (or MRM) assays, generated for the human proteome by measuring known synthetic peptides on different quadrupole-bearing mass spectrometers.

The search parameters were delineated as follows:

- number of highest intensity fragment ions to keep (i.e. how many fragment ions should be reported from each peptide) was set to 8;
- target instrument (i.e. the mass spectrometer on which the actual experiment would be performed, to determine in which order MRM assays from different transition sources will be reported in the query output) was set to QTRAP 5500, very similar to our QTRAP 6500;
- transitions source (i.e. the instruments on which the MRM assays were generated and that have to be considered to build the research output) was left unchanged from the default status, in which all sources (QTOF, Agilent QQQ, Qtrap5500, Ion Trap, Predicted) are considered;

- precursor exclusion range (option that allows to exclude fragment ions whose m/z window is around the precursor m/z) was kept blank, with 5 Da being enforced by default by the query engine;
- search proteins form (i.e. the target proteome against with the queried peptides should be mapped, described by a certain type of accession numbers) was set to SwissProt;
- duplicate peptides (i.e. peptides that map to multiple proteins) was set as *Unique in results*. Indeed, protein isoforms like the cytosolic Hsp90s do not have only unique peptides, thus selecting this parameter allows to retain multi-mapping peptides in the query result: the number of proteins in the target proteome for which a particular peptide maps will be indicated in the output *N_map* column;
- heavy label was kept blank, since no labeling procedure was exploited in our experiment;
- labeled transitions was kept as *default*, since unlabeled peptides have been exploited to generate the *SRMAtlas* assays in current builds and our samples were also unlabeled;
- maximum m/z (i.e. maximum m/z values to be reported, which apply to both precursor and fragment ions) was set at 1100 Da, to further operate in the QTRAP 6500 low masses profile;
- minimum m/z was kept blank;
- allowed ions types (i.e. common fragment ions types) were set as b-ions and y-ions;
- allowed peptide modification was set as the carbamidomethylation of cysteines (C[160]).

Thus, all of the previously selected Hsp90s peptides were queried to obtain the MS-related parameters to be used to build up the MRM methods (i.e. the mass to charge ratios of the Q1

precursor and Q3 fragment ions and their charges, the ion series of the fragments, their rank order and relative intensity in a CID spectrum, their sequence specific calculated retention time and the *N_map* value).

The obtained result table was then subsequently refined as follows:

- peptides whose following aminoacid at the C-terminus was either K or R were removed, to avoid biases in the quantification step related to variable trypsin missed cleavages patterns in different samples;
- the 8 fragment ions reported for each peptide were investigated and only the 3 most intense were selected, preferring when possible the y-series over the b-series ions;
- peptides having a *N_map* higher than 1 were annotated to identify the Hsp90s isoforms they mapped to, in order to evaluate the homology sequence levels with the originally queried Hsp90-α and Hsp90-β proteins. More in details, the cytosolic Hsp90-α isoforms Hsp90-α-2, Hsp90-α-A2, putative Hsp90-α-A4 and putative Hsp90-α-A5 and the cytosolic Hsp90-β isoforms putative Hsp90-β2, Hsp90-β3 and Hsp90-β4 were also identified. All of these isoforms share sequence homology levels higher than 65% with Hsp90-α and Hsp90-β, especially in the ATPase domain targeted by radicicol. Thus, these multi-mapping peptides were still considered significative for the evaluation of radicicol-induced Hsp90s conformational modifications and subtilisin resistance. With such a high sequence homology in the N-terminus, it would be unlikely that radicicol might specifically target one isoform over the other.

Thus, a comprehensive list consisting of 52 peptides, each presenting the three previously selected transitions, was obtained.

2.3.2.2 MRM-MS method build-up: retention time assignment, selection of the best transition and Hsp90s mapping

After the selection of Hsp90-α and Hsp90-β tryptic peptides and their three best daughter ions to analyze, we moved to test the retrieved transitions (Q1/Q3) onto a HeLa tryptic digest, in order to select the best Q1/Q3 to analyze and to annotate the retention time for each peptide.

Thus, a HeLa cell lysate was submitted to an *in solution* digestion protocol: proteins were first denatured with urea, then disulfide bonds were reduced with dithiothreitol (DTT) and carboxyamidomethylated with iodoacetamide (IAA), prior to dilute the urea and add trypsin and LysC to carry out an extensive overnight digestion. The enzymes action was then quenched and the obtained peptides mixture was desalted and submitted to the LC-MRM-MS analysis.

For each peptide, the XICs of all of the three transitions were inspected: at first, the retention time was assigned taking into account that peaks related to the same precursor exhibit the same chromatographic behavior, meaning that all of the XICs should present a peak eluting at the same time. Then, the best transition was selected as the one showing the most intense peak at the previously selected retention time and the best signal to noise ratio (Figure 40).

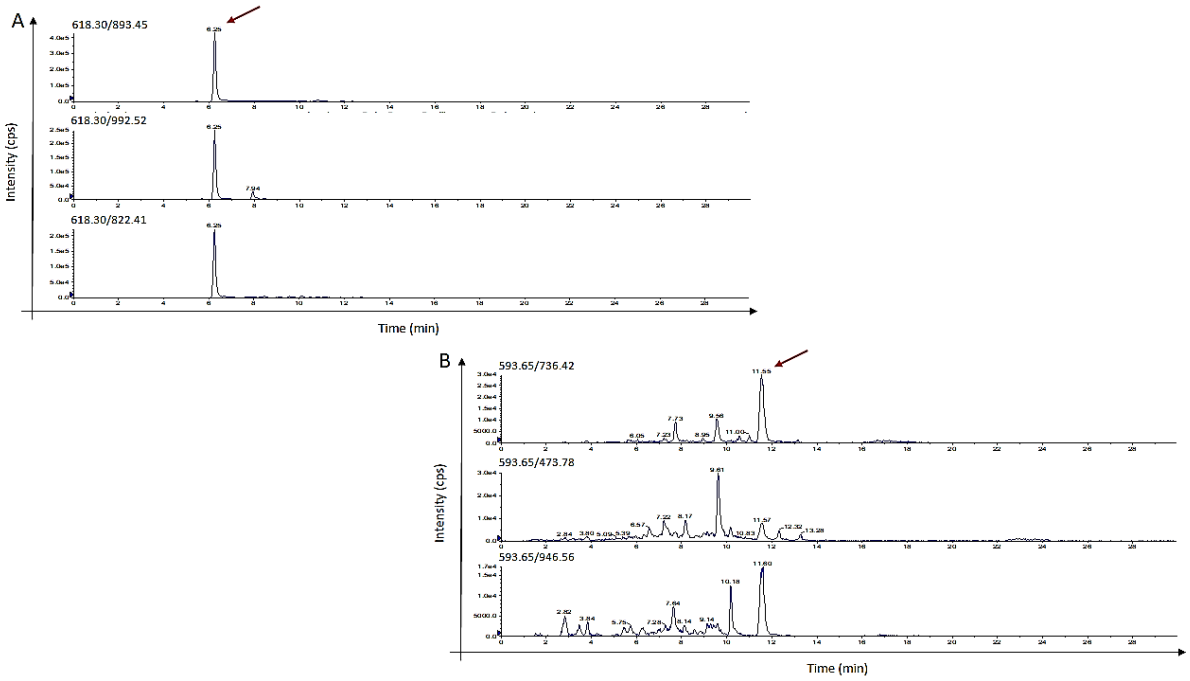


Figure 40: *Exemplificative XICs for the transition selection step. (A) All of the three transitions show a base peak at 6.25 min, which was unambiguously assigned as the peptide retention time. The signal to noise ratio is optimal for all the transitions, but 618.30/893.45 was selected due to its highest signal intensity (red arrow). (B) Two out of three transitions show a base peak at 11.55 min (top and bottom XICs) and a low-intensity peak at the same retention time is also present in the middle panel. Thus, 11.55 min was assigned as the peptide retention time and the transition 593.65/736.42 was selected due to its highest signal intensity and signal to noise ratio (red arrow).*

This selection process was repeated for all the precursors and a global MRM method was obtained, consisting of 42 transitions related to the experimentally observed peptides.

These peptides were then mapped onto Hsp90-α and Hsp90-β sequences, to assess the sequence coverage and make sure all of the Hsp90s domains were traced, for an unbiased analysis of radicicol effects.

More in details, Hsp90-α and Hsp90-β sequences were retrieved from UniProt and analyzed through *PAWS*, a free programs that allows to visualize proteins digestion patterns with a specified enzyme. Thus, trypsin was set as digesting enzyme and cystein carboxyamidomethylation as fixed modification, then the experimentally observed peptides were queried as singly-charged species, with a tolerance of 300 ppm on their *m/z* ratios.

As can be noticed in Figure 41, peptides mapping for all of the Hsp90s domains could be observed (tracked peptides are highlighted in yellow), obtaining a more than satisfactory coverage, taking into account the complex matrix we worked with, of 41% for Hsp90-α and of 42% for Hsp90-β.

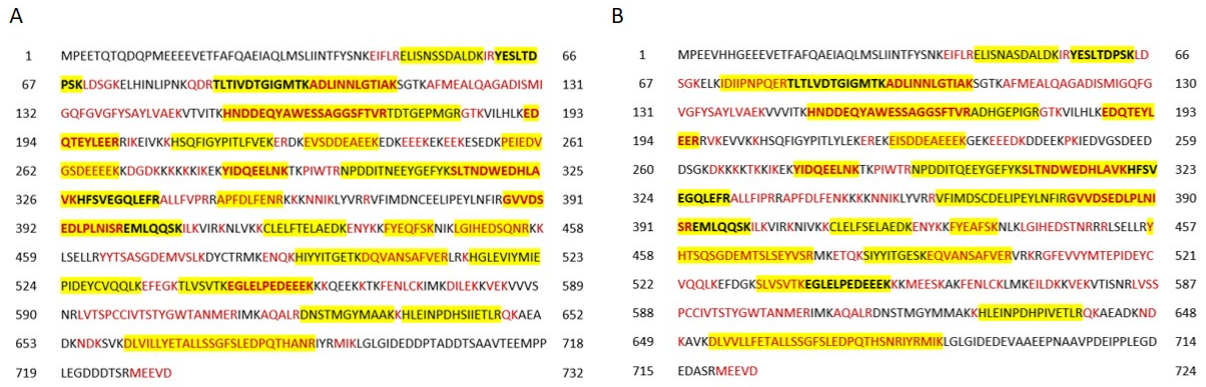


Figure 41: *Hsp90-α (A) and Hsp90-β (B) aminoacidic sequences reporting trypsin cleavage sites. Adjacent tryptic peptides are represented with different colors, whereas the experimentally mapped peptides are highlighted in yellow, depicting in bold the ones shared by the two isoforms.*

2.3.2.3 T-LiP-MRM: experimental procedure

As briefly explained in paragraph 2.3, 300 µg HeLa lysate aliquots were prepared in non-denaturing conditions and incubated with either radicicol (0.8 µM and 8 µM final

concentrations) or the vehicle (DMSO) for 1 hour at room temperature and under continuous agitation. The obtained samples were then divided into three aliquots each and submitted to limited proteolysis with subtilisin (1:2500 w/w, 1:1000 w/w and 1:500 w/w ratios, respectively) for 30 minutes at 25°C under native conditions, so that subtilisin preferential cleavage sites would be uniquely dependent on Hsp90s structural features in the DMSO-treated sample. Radicicol binding, stabilizing Hsp90s, should instead induce structural changes, making the proteins regions directly or distally involved in the interaction less exposed to the aspecific protease.

Furthermore, 300 µg of the DMSO-treated sample were submitted to a mock proteolysis to be kept as a positive control.

Subtilisin was then quenched with 1mM PMSF and the samples shifted to denaturing conditions adding urea (4M final concentration) to carry out an extensive *in solution* tryptic digestion, producing complex peptides mixtures subsequently submitted to a desalting step. The obtained samples were then analyzed, in triplicate, through LC-MRM-MS, exploiting the previously optimized chromatographic conditions and MRM method.

2.3.2.4 T-LiP-MRM: data analysis

To perceive radicicol induced proteolytic resistance on Hsp90-α and Hsp90-β, changes in abundance of their fully tryptic peptides should be observed as a consequence of the molecule-induced subtilisin resistance. Indeed, in comparison with the DMSO-treated lysate, in which no Hsp90s stabilization occurs, the tryptic peptides coming from the radicicol-treated sample and corresponding to the proteins region(s) directly interacting with the molecule or distally affected by the binding event, should be significantly more intense. Conversely, a similar intensity should be detectable for peptides belonging to regions not involved in the radicicol-induced Hsp90s structural fluctuations.

Thus, our data analysis has been divided in two fundamental steps: at first, Hsp90s regions reliably undergoing subtilisin cleavages were identified, comparing the DMSO-treated samples submitted or not to limited proteolysis, then the selected peptides were examined comparing the DMSO and radicicol-treated samples, at a given subtilisin to proteins ratio.

Regarding the first step, we reasoned that tryptic peptides mapping for Hsp90s regions undergoing subtilisin processing in a non-random fashion (i.e. *LiP peptides*) should respond to increasing subtilisin amounts with a reverse proportionality, meaning that their areas should decrease the more subtilisin was exploited for the limited proteolysis step (Figure 42).

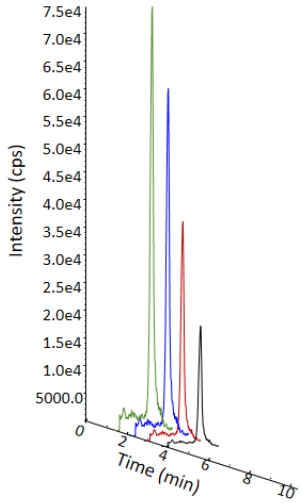


Figure 42: *Exemplificative XICs of a tryptic peptide mapping for an Hsp90 region reliably processed by subtilisin (i.e. a LiP peptide). The area of the peak decreases with increasing subtilisin amounts (green peak: no subtilisin, blue peak: subtilisin to proteins ratio of 1:2500 w/w, red peak: subtilisin to proteins ratio of 1:1000 w/w, black peak: subtilisin to proteins ratio of 1:500 w/w).*

Thus, peptide areas from three injection replicates were averaged, to obtain values for the fold changes evaluation. More in details, we exploited the DMSO-exposed and subtilisin-untreated sample (i.e. lysate) as a reference, bearing only fully tryptic peptides with the highest possible intensity, and calculated fold changes in respect to the subtilisin-treated samples as follows:

$$Fc_{Lysate/2500} = Area_{Lysate}/Area_{DMSO\_1:2500},$$

$$Fc_{Lysate/1000} = Area_{Lysate}/Area_{DMSO\_1:1000},$$

$$Fc_{Lysate/500} = Area_{Lysate}/Area_{DMSO\_1:500}.$$

All the transitions with significant fold changes, higher than 2 in at least two differentially subtilisin treated samples, were subsequently compared: peptides mapping for Hsp90s regions reliably cleaved by subtilisin and thus presenting the following trend

$$Fc_{Lysate/2500} < Fc_{Lysate/1000} < Fc_{Lysate/500}$$

were carried onto the subsequent data analysis step. Of the 42 peptides originally quantified, only 20 were selected (Figure 43).

Hsp90-α	Hsp90-β	$Fc_{Lysate/2500}$	$Fc_{Lysate/1000}$	$Fc_{Lysate/500}$
-	E-[42-53]-K	1.89	20.50	20.55
E-[47-58]-K	-	1.76	16.15	45.88
Y-[61-69]-K	Y-[56-64]-K	2.20	32.52	NaN
-	I-[73-82]-R	1.32	2.57	3.85
T-[88-100]-K	T-[83-95]-K	1.34	8.69	30.65
A-[101-112]-K	A-[96-107]-K	3.25	123.55	NaN
H-[154-173]-R	H-[149-168]-R	7.66	25.60	28.56
H-[210-224]-K	-	3.49	10.64	42.44
E-[251-269]-K	-	78.33	215.79	NaN
H-[328-338]-R	H-[320-330]-R	2.03	NaN	NaN
A-[347-355]-R	-	1.34	3.21	14.89
-	V-[360-378]-R	2.35	82.03	NaN
F-[437-443]-K	-	2.46	NaN	NaN
-	Y-[457-475]-R	171.90	NaN	NaN
-	S-[482-491]-K	7.73	315.95	316.48
H-[490-499]-K	-	3.43	13.92	36.26
D-[500-510]-R	-	4.72	45.05	97.24
D-[621-631]-K	-	6.23	14.76	NaN
-	H-[625-639]-R	2.37	11.87	27.22
H-[633-647]-R	-	1.41	2.78	8.96

Figure 43: *Selected LiP peptides, reported with their calculated fold changes. The fold change increases when increasing amount of subtilisin have been used for the limited proteolysis step. 'NaN' stands for Not a Number, as no tryptic peptide could be quantified in the subtilisin treated sample and no fold change could be calculated.*

.

Assessment of the Hsp90s regions undergoing conformational changes upon radicicol binding was done by comparing the radicicol-treated with the corresponding untreated samples, at a given subtilisin amount. Fold changes were calculated, for all the subtilisin amounts, as follows:

$$Fc_{0.8\mu M/CTRL} = Area_{Radicicol\_0.8\mu M}/Area_{DMSO},$$

$$Fc_{8\mu M/CTRL} = Area_{Radicicol\_8\mu M}/Area_{DMSO}.$$

Peptides presenting significant fold changes, higher than 2 at the lowest molecule concentration, in at least one of the three differentially subtilisin-treated samples, were considered symptomatic of Hsp90s conformational changes due to radicicol binding (Figure 44).

		1_1000 w/w		1_500 w/w	
Hsp90-α	Hsp90-β	$Fc_{0.8\mu M/CTRL}$	$Fc_{8\mu M/CTRL}$	$Fc_{0.8\mu M/CTRL}$	$Fc_{8\mu M/CTRL}$
-	**E-[42-53]-K**	**3.92**	**52.21**	0.98	1.52
E-[47-58]-K	-	1.02	1.13	**2.09**	**2.82**
Y-[61-69]-K	**Y-[56-64]-K**	**5.63**	**5.84**	-	-
-	**I-[73-82]-R**	**2.00**	**2.01**	0.97	0.99
T-[88-100]-K	**T-[83-95]-K**	**2.67**	**2.99**	1.09	1.34
A-[101-112]-K	**A-[96-107]-K**	1.60	1.81	**3.08**	**3.65**
H-[154-173]-R	H-[149-168]-R	0.84	1.23	1.48	1.16
H-[210-224]-K	-	0.97	1.00	1.23	2.02
E-[251-269]-K	-	0.95	1.00	-	-
H-[328-338]-R	H-[320-330]-R	0.00	-	-	-
A-[347-355]-R	-	1.34	1.70	1.63	1.87
-	**V-[360-378]-R**	**6.24**	**8.55**	-	-
F-[437-443]-K	-	-	-	-	-
-	**Y-[457-475]-R**	-	-	-	-
-	S-[482-491]-K	1.30	1.47	0.98	0.95
H-[490-499]-K	-	1.94	2.00	1.81	1.36
D-[500-510]-R	-	0.88	1.05	1.53	1.25
D-[621-631]-K	-	0.75	1.08	-	-
-	**H-[625-639]-R**	**2.06**	**2.37**	0.93	1.43
H-[633-647]-R	-	1.29	1.42	1.90	2.24

Figure 44: *LiP peptides reported with their significant ($p<0.05$) radicicol vs control fold change values. Hsp90s aminoacid numbering is comprehensive of the initial methionine. Peptides mapping for proteins regions affected by radicicol binding (i.e. fold change ≥ 2 at the lowest molecule concentration in at least one of the three experiments) are reported in bold. '-' means the peptide was not detectable in both radicicol-treated untreated and samples.*

As can be observed from the reported data, the highest number of peptides whose area increases in presence of radicicol is located along the N-terminal Hsp90s ATPase domain, which is indeed the known natural product binding site.

Nevertheless, to fully validate our t-LiP-MRM method, the peptides identified in Hsp90s N-terminal region were compared with the outputs of molecular docking studies performed on the

same human domains[162], taking also into account previous co-cristallization analyses carried out on the *Saccharomyces cerevisiae* isoform[160].

As for the docking analysis, Khandelwal and co-workers report a picture in which a water-mediated network of hydrogen bonds at the bottom of Hsp90s N-terminal pocket surrounds the resorcinol ring, the carbonyl and 4-phenol functions of radicicol (Figure 45).

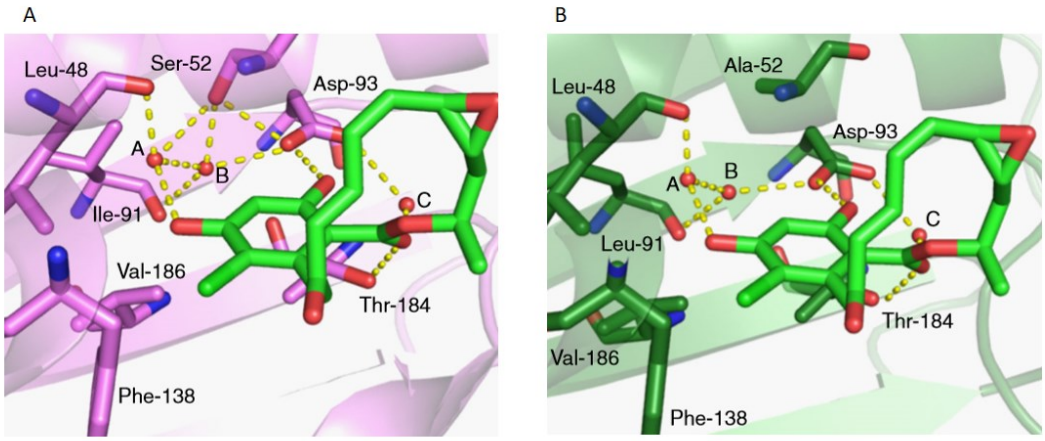


Figure 45: *Modeling of radicicol into Hsp90-α (A, PDB code2XAB) and d Hsp90-β (B, PDB code 1UYM) N-terminal ATP-binding site. Aminoacid numbering is given on the basis of Hsp90-β. Adapted from Khandelwal et al.*[162].

As can be observed, Leu48, Ser52, Ile91, Asp93 and Thr184 are directly involved in water-mediated hydrogen bonds interactions with radicicol for Hsp90-α, whereas Lue48, Leu91, Asp93 and Thr184 are the involved aminoacids for the β isoform. These data are largely superimposable with the co-crystallization studies conducted by Roe and co-workers in 1999 on the *S. cervisiae* Hsp90 N-terminal domain.

For the clearest possible comparison of our t-LiP-MRM results with the molecular docking data, the same PDB Hsp09s N-terminal domains exploited by Khandelwal *et al.*[162] have been

used. Indeed, we visualized these PDB structures on Swiss-PDB viewer (SPDB viewer), highlighting in red t-LiP-MRM identified peptides mapping for Hsp90s regions involved in the interaction with radicicol (Figure 46). As shown, the correspondence between our results and the literature-reported ones is more than satisfactory, with four out of five and three out of four aminoacids, for Hsp90-α and Hsp90-β respectively, embedded in our t-LiP-identified modulated peptides.

More in details, as previously reported in Figure 44, we could observe the highest intensity fold change (i.e. 52 fold at 8μM radicicol) for the Hsp90-β E-[42-53]-K peptide embedding Leu48, which engages in a water-mediated hydrogen bond with radidicol, whereas the peptide T-[83-95]-K, embedding both Leu91 and Asp93, was detected with a more modest but still significant fold change, as well as the Hsp90-α corresponding peptides E-[47-58]-K and T-[88-100]-K were.

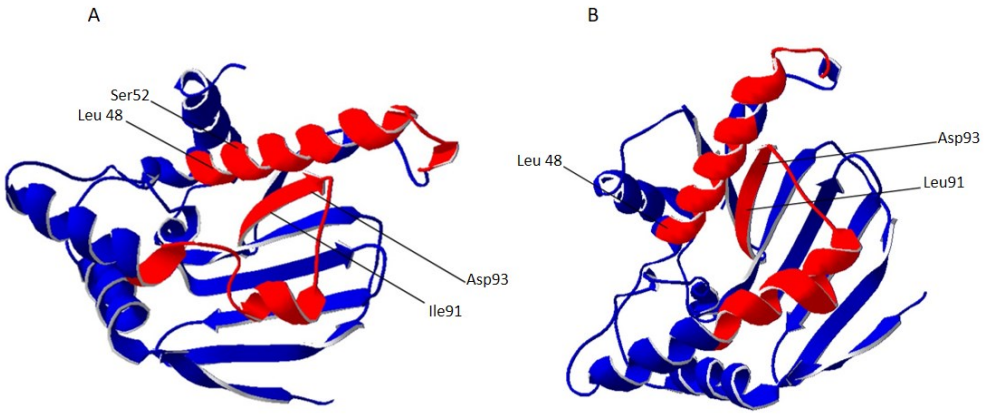


Figure 46: *Hsp90-α (A, PDB code 2XAB) and d Hsp90-β (B, PDB code 1UYM) N-terminal ATPase sites visualized through SPDB viewer and represented through sticks and ribbons. Hsp90s backbone is depicted in blue, whereas radicicol protected LiP peptides are highlighted in red. Aminoacidic residues identified as being involved in the interaction with the molecule by both our t-LiP-MRM experiment and Khandelwal et al. point out from the proteins structures.*

Interestingly, tryptic peptides embedding aminoacids directly involved in Hsp90s/radicicol interactions were detected together with adjacent peptides, which showed remarkable fold changes, neighboring with the molecule binding sites and thus being strongly involved in radicicol-induced Hsp90s structural rearrangements. As an example, Hsp90-β peptide Y-[56-64]-K showed a more than 5-fold intensity increase at the lowest radicicol concentration, as a result of it being adjacent to the most protected E-[42-53]-K peptide.

Thus the t-LiP-MRM experiment, carried out to probe Hsp90s structural changes following radicicol binding in the complex environment of a cell lysate, gave us a highly positive feedback, highlighting the strength of our workflow and ensuring the reliability of the obtained results, in terms of mapping protein regions directly or distally involved in the interaction with a ligand.

CHAPTER 3

Joining forces: a label-free functional proteomics platform paired with conventional affinity chromatography discloses actin proteins as main 5-epi-Sinuleptolide biological targets[163].

Adapted from

Morretta, E., Esposito, R., Festa, C., Riccio, R., Casapullo, A. and Monti, M. C. "Discovering the Biological Target of 5-epi-Sinuleptolide Using a Combination of Proteomic Approaches." *Marine Drugs* **2017**, *15*, 312.

3.1 Background

Comprising approximately half of the world's biodiversity, the marine environment is a massive source of structurally multifaceted compounds, showing unique chemical features (i.e. not found in terrestrial organisms) and endowed with a range of biologically relevant activities.

The majority of marine products is generated by invertebrates, such as sponges, tunicates, mollusks, bryozoans and soft corals, the latter being a group of colonial invertebrates (phylum Coelentrata, class Anthozoa, subclass octacorallia) constituting a dominant portion of the tropical reef biomass throughout the world[164,165] (Figure 47).

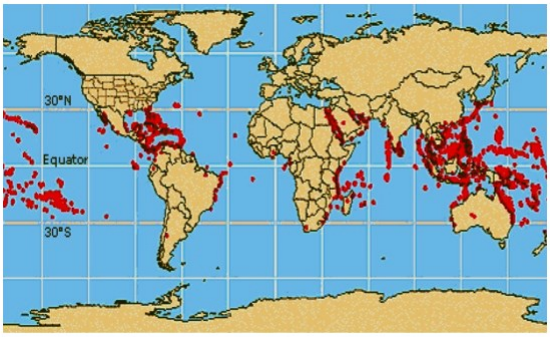


Figure 47: *Soft coral reefs (red dots) distribution.*

Among the soft corals, the *Sinularia* genus (Figure 48) is one of the most widely distributed and prolific, elaborating a rich and unique group of secondary metabolites (e.g. sesquiterpenes, diterpenes, sterols/steroidal glycosides, sphingosine derivatives, glycolipids and spermidine derivatives), known for their versatile chemical features and anti-cytomegalovirus, anti-inflammatory and cytotoxic activities[166–169].

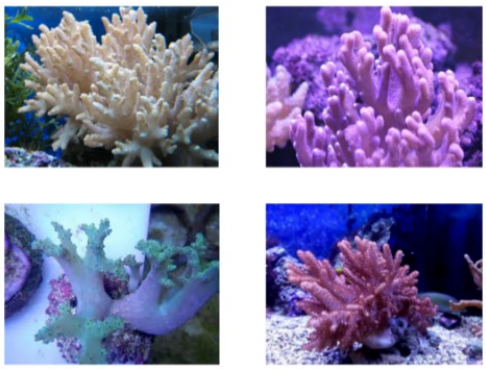


***Figure 48**: Sinularia genus soft corals specimens.*

Several norditerpenoids have been isolated and structurally elucidated from *Sinularia* species to date[164] among which two diasteroisomers, namely Sinuleptolide (SNEP) and 5-epi-Sinuleptolide (5-epi-SNEP), resulted two of the most interesting metabolites (Figure 49).

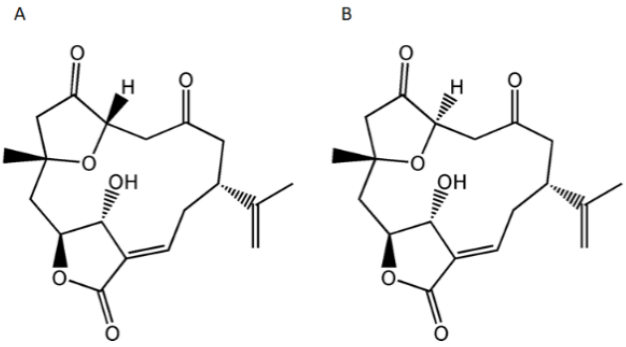


***Figure 49:** 5-epi-SNEP (A) and SNEP (B).*

More in details, SNEP was found to inhibit lipopolysaccharide-induced TNF-α and nitric oxide production in murine macrophage-like RAW 264.7 cells[170] and to exert a selective anti-proliferative effect on the gingival squamous cell carcinomas Ca9-22[171], compared to normal

human gingival fibroblasts. This latter phenomenon was reflected by cell cycle redistribution (subG1 accumulation and G2/M arrest), apoptosis induction, reactive oxygen species (ROS) generation, mitochondrial dysfunction and DNA damage.

Besides, 5-epi-SNEP was also found to inhibit lipopolysaccharide-induced TNF-α production and the subsequent nitrogen oxide production in RAW 264.7[170], but it also decreased the proliferation of mouse lymphocytic leukemia cells[172].

Furthermore, both SNEP and 5-epi-SNEP showed to significantly reduce cell viability in human epidermoid carcinoma A431 cells, in head and neck SCC9 and SCC25 cells and in premalignant keratinocytic HaCaT cells, but 5-epi-SNEP cytotoxicity was much more pronounced than the SNEP exerted one, especially on SCC25 cells.

More in details, treatment of these cells with 5-epi-SNEP increased apoptotic bodies formation and induced cell-cycle arrest during the G2/M phase, up-regulating p53 and p21 and inhibiting cyclin B1 and cyclin-dependent kinease 1 expression. Additionally, both the death receptor- and mitochondria-mediated caspase pathways resulted to be critical in 5-epi-SNEP-induced apoptosis of SCC25 skin cancer cells[173].

Thus, since 5-epi-SNEP showed an interesting and multi-faceted biological activity profile, and since no studies on its interacting proteins were disposable, we selected this marine metabolite as an interesting case of study for a functional proteomics-based Drug-Target Deconvolution analysis, in order to shed light on its mechanism of action at a molecular level.

3.2 5-epi-SNEP main cellular target(s) identification through *label-free* and *label-based* Functional Proteomics

With the aim of identifying 5-epi-SNEP principal protein partner(s), the functional proteomics platform described in the previous chapter has been carried out and coupled to conventional AP-MS, for a clearer picture of the marine metabolite interactomic profile.

More in details, 5-epi-SNEP target(s) deconvolution process can be summarized as follow: (1) application of the *label-free* DARTS approach to identify the molecule partner protein(s); (2) application of the *label-based* AP-MS strategy to corroborate and confirm the DARTS results and (3) t-LiP-MRM employment for the characterization of 5-epi-SNEP interaction features with its cellular target(s).

3.2.1 Identification of 5-epi-SNEP cellular partners through DARTS

As a first step, the choice of the optimal subtilisin to proteins ratio for the best visualization of different proteins stability at several 5-epi-SNEP concentrations was required.

For this purpose, a HeLa lysate was prepared in non-denaturing conditions, as already reported, and 300 µg aliquots were incubated with either the un-modified 5-epi-SNEP (5 µM or 50 µM final concentrations) or just the vehicle (DMSO). Each sample was then submitted to limited proteolysis in native condition with subtilisin at 1:1000 w/w and 1:500 w/w, respectively, keeping an aliquot of the DMSO-treated sample undigested, as a reference. Subtilisin was then quenched with PMSF and the samples separated on 12% 1D-SDS-PAGE and visualized through Coomassie staining.

As can be observed in Figure 50, the 1:500 w/w subtilisin to proteins ratio was the best condition to appreciate 5-epi-SNEP effect, with a huge intensity increase in a broad molecular weight range already at the lowest 5 µM 5-epi-SNEP concentration, when compared with the untreated sample (red box in Figure 50).

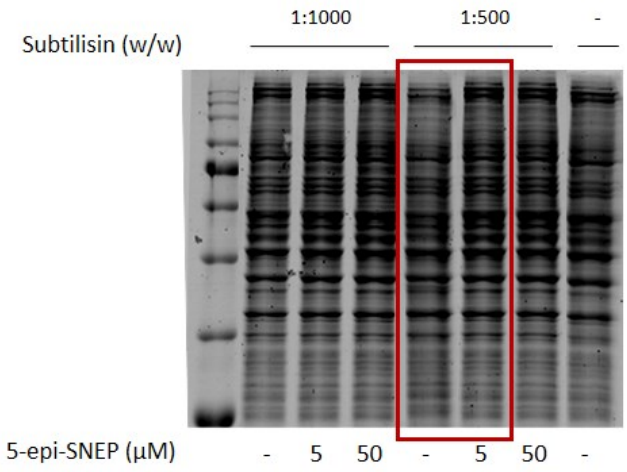


Figure 50: *Preliminary DARTS experiment carried out with two 5-epi-SNEP concentrations and at two subtilisin amounts, showing the best protease to proteins ratio to appreciate the molecule exerted protection is 1:500 w/w. Red box indicates 5-epi-SNEP exerted protection in a broad molecular weight range when compared to the untreated sample.*

On this basis, in the following DARTS experiment a 10-folds lower 5-epi-SNEP concentration was added and the resulting samples resolved through SDS-PAGE.

The gel runs were then both inspected *by eye* and submitted to a densitometric analysis, to identify protein bands whose intensities increased with a direct proportionality in respect to 5-epi-SNEP concentration. As can be observed in Figure 51, even at its lowest concentration the marine metabolite still exerted a protective effect on a broad range of molecular weights.

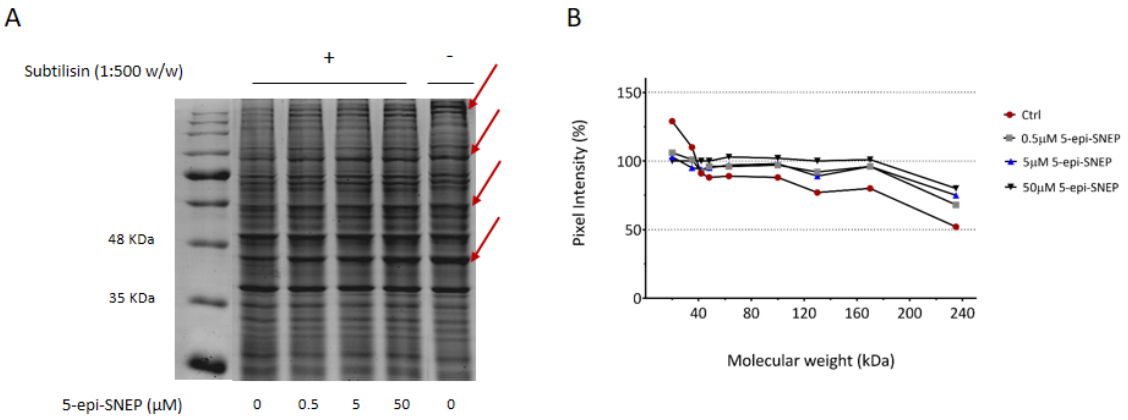


Figure 51: *12% SDS-PAGE (A) and the relative densitometric analysis (B) showing 5-epi-SNEP induced proteolytic resistance on a broad range of molecular weights at the lowest 0.5 µM concentration. Red arrows in (A) point just some of the bands whose intensity increases accordingly with the molecule concentrations, exemplificative of the marine metabolite exerted protection.*

Thus, gel bands clearly corresponding to proteins protected towards the proteolysis were excised only from the lanes of the 0.5 µM 5-epi-SNEP treated sample and the two control experiments (lysate treated or not with subtilisin) and submitted to an *in situ* tryptic digestion[34] to obtain peptide mixtures for nano-flow RP-UPLC MS/MS analysis.

5-epi-SNEP protected proteins were then identified submitting the obtained MS data to MASCOT analysis, by directly comparing the matches of the three different samples and calculating a protection percentage for each protein in correspondence of its molecular weight, as already described in chapter 2 (paragraph 2.2.3).

As can be observed in Figure 52A, the matches of the cytoplasmic actin isoform 1 increased by 10-folds in the sample treated with 5-epi-SNEP prior to the limited proteolysis step, as a result of the molecule exerted proteolytic protection upon this protein.

For a more confident identification of the marine metabolite interacting proteins, the DARTS experiment was carried out in duplicate: 5-epi-SNEP protected proteins were identified as the ones being the most protected at the lowest molecule concentration and showing a reproducible protection extent between the two biological replicates.

A

	DMSO + subtilisn			0.5 µM 5-epi-SNEP +subtilisin			DMSO - subtilisn		
Description	Score	Matches	emPAI	Score	Matches	emPAI	Score	Matches	emPAI
Actin, cytoplasmic 1	50	16	0.28	1986	117	4.11	4186	240	7.45

B

Accession	Mass	Description	Averaged Protection (%)
ACTB_HUMAN	***42 052***	***Actin, cytoplasmic 1***	***42***
ARP2_HUMAN	*45 017*	*Actin-related protein 2*	*40*
PGK1_HUMAN	*44 985*	*Phosphoglycerate kinase 1*	*40*
BACH_HUMAN	*42 454*	*Cytosolic acyl coenzyme A thioester hydrolase*	*40*
TBB5_HUMAN	*50 095*	*Tubulin beta chain*	*38*
MYH10_HUMAN	*229 827*	*Myosin-10*	*36*
PYR1_HUMAN	*245 167*	*CAD protein*	*35*
AATM_HUMAN	47 886	Aspartate aminotransferase, mitochondrial	33
ENOA_HUMAN	47 481	Alpha-enolase	30
DDX21_HUMAN	87 804	Nucleolar RNA helicase 2	29
PUR6_HUMAN	47 790	Multifunctional protein ADE2	25
ALDOA_HUMAN	39 851	Fructose-bisphosphate aldolase A	23
RL4_HUMAN	47 953	60S ribosomal protein L4	22
GDIB_HUMAN	51 087	Rab GDP dissociation inhibitor beta	20
TBA1B_HUMAN	50 804	Tubulin alpha-1B chain	20
SPTB2_HUMAN	275 237	Spectrin beta chain, non-erythrocytic 1	20
OXDD_HUMAN	38 024	D-aspartate oxidase	17
SYTC_HUMAN	84 294	Threonine--tRNA ligase, cytoplasmic	17
IF4A1_HUMAN	46 353	Eukaryotic initiation factor 4A-I	16
TLN1_HUMAN	271 766	Talin-1	16
HS90B_HUMAN	83 554	Heat shock protein HSP 90-beta	15

Figure 52: *(A) MASCOT parameters related to the isoform 1 of cytoplasmic Actin identification. All of the parameters significantly increase when the HeLa lysate was treated with 5-epi-SNEP prior to the limited proteolysis step. (B) List of putative 5-epi-SNEP interacting proteins, reported with the averaged protection percentage resulting from two independent experiments.*

Figure 52B reports a list of putative 5-epi-SNEP interacting proteins together with the averaged protection percentages resulting from the two independent DARTS experiments. As can be observed, the isoform 1 of cytoplasmic actin is the most protected protein at the lowest molecule concentration (red and bold in Figure 52B). Nevertheless, several actin-related and

un-related proteins also result stabilized towards the proteolysis by the marine metabolite (blue and Italic proteins in Figure 52B).

Thus, to corroborate DARTS-driven hypothesis of 5-epi-SNEP interacting with actin, we sought to exploit the AP-MS functional proteomics strategy to test whether or not the same targets could still be identified.

3.2.2 Identification of 5-epi-SNEP cellular partners through AP-MS

3.2.2.1 Analysis of 5-epi-SNEP alcohol function reactivity

Given 5-epi-SNEP interesting structural features, summarized in Figure 53, we sought to exploit the molecule alcoholic function to achieve a straightforward functionalization scheme.

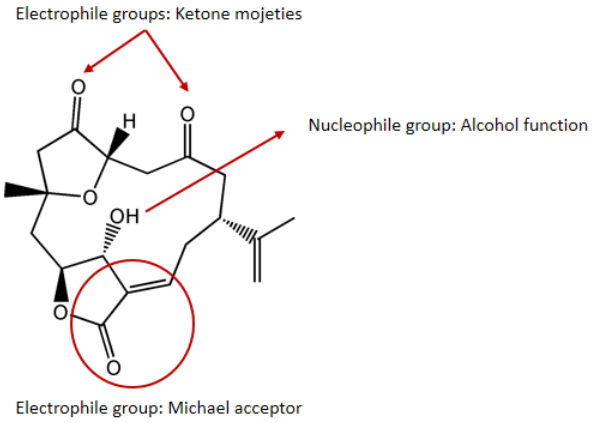


Figure 53: *5-epi-SNEP structural features.*

Thus, we analyzed the reactivity profile of the 5-epi-SNEP alcohol function to subsequently use this group for the molecule immobilization onto an epoxy-activated matrix.

For this purpose, 5-epi-SNEP was treated with a 250 w:w excess of glycidol (2-hydroxymethyl oxiran) in CH_3CN/50 mM $NaHCO_3$/1% triethyl-amine. The reaction was

monitored through RP-HPLC-UV observing the formation of a new peak at 20 minutes r.t. after 1 hour (Figure 54B, blue trace). Formation of the 5-epi-SNEP/glycidol adduct (compound 1 in Figure 54A) was assessed through HR-MS and HR-MS/MS. The HR-MS analysis showed a single ion at 427.1722 u.m.a. (Figure 54C), most likely related to the sodiated dehydrated-derivatives of the reaction product, namely compounds 2A and 2B (Figure 54A). Indeed, the HR-MS/MS analysis showed fragmentation patterns compatible with either the loss of one H_2O molecule from sodiated compound 2B (409.1755 u.m.a.) and the loss of one CO molecule from sodiated compound 2A (399.1773 u.m.a.), pointing out that only these two products were formed. Moreover, the fragments at 371.1462 u.m.a. and 353.1357 u.m.a., compatible with the loss of the glycidol-derived side chain with or without the original 5-epi-SNEP alcohol group, confirmed the reaction mechanism postulated in Figure 54A.

.

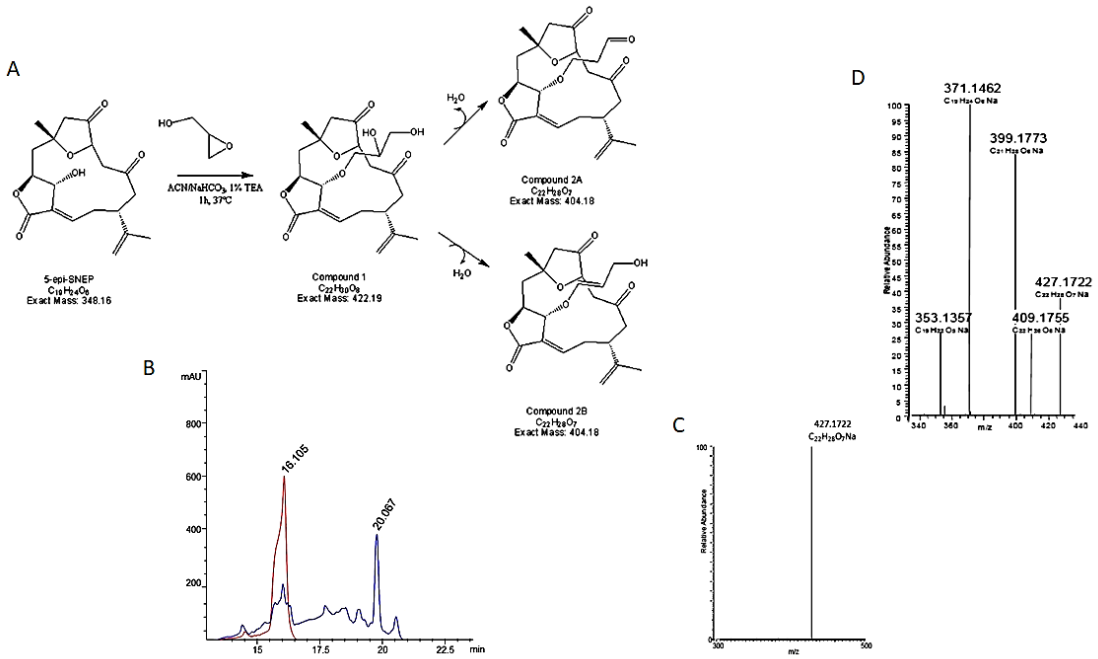


Figure 54: *(A) 5-epi-SNEP reaction scheme with glycidol: each compound is reported with its molecular formula and its exact mass. (B) RP-HPLC-UV traces of the reaction at its t_0, where only 5-epi-SNEP can be detected (red trace, retention time= 16.105 min), and after*

one hour, where the product can be detected (blue trace, retention time= 20.067 minutes). (C) HR-MS analysis of the reaction products. (D) HR-MS/MS analysis of the reaction products.

3.2.2.2 Generation of modified functional matrices

To generate the functional matrices for the AP-MS strategy, epoxy-activated Sepharose™ 6B resin was selected, bearing a long hydrophilic spacer arm particularly suitable for the immobilization of small molecules, and functionalized with epoxide moieties, useful for the reaction with 5-epi-SNEP.

Thus, two aliquots of resin were swelled and suspended in a coupling buffer consisting of CH_3CN/50 mM $NaHCO_3$/1% triethyl-amine and containing or not 1 µmol of the marine metabolite, to obtain both empty beads (i.e. negative control) and 5-epi-SNEP-bearing beads.

5-epi-SNEP immobilization occurred via the nucleophile attack of its alcoholic function to the beads epoxy groups (Figure 55A) and it was monitored by RP-HPLC-UV, injecting iso-volumetric aliquots of 5-epi-SNEP coupling solution at different incubation times and integrating the obtained peaks areas of the free compound before (t_0) and after 24 hours of reaction (t_{24}). Furthermore, a reference solution made of the same proportions of marine metabolite and coupling buffer was also analyzed, to ascertain 5-epi-SNEP stability over time for a less biased evaluation of its immobilization extent.

Since the compound resulted stable at each incubation time, the immobilization yield was simply calculated by the area decrement of 5-epi-SNEP corresponding t_0 and t_{24} chromatographic peaks, as shown in Figure 55B.

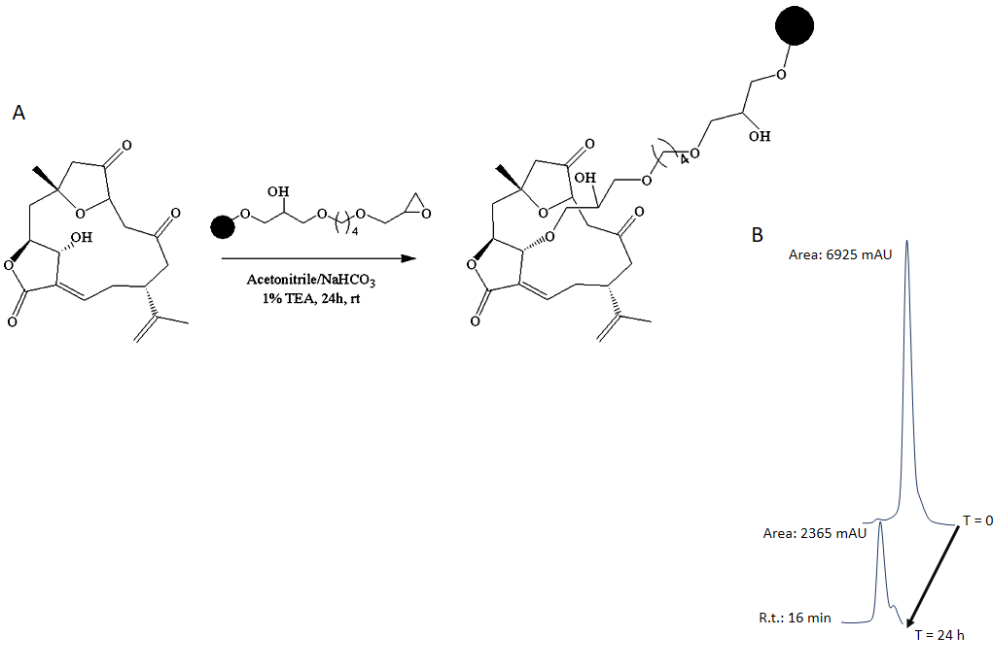


Figure 55: *(A) 5-epi-SNEP immobilization scheme onto the epoxy-activated resin beads. (B) RP-HPLC-UV analysis showing 5-epi-SNEP immobilization after 24 hours reaction with the matrix.*

Thus, an immobilization rate of 66% was calculated after 24 hours: 660 nmol of 5-epi-SNEP covalently bounded to around 100 µL of resin (i.e. 1.9-4 µmol of active epoxy groups), corresponding to a matrix derivatization extent ranging from 17% to 35%.

After the immobilization step, both the resin samples were extensively washed to remove the molecule non-specifically bounded to the beads or the DMSO in the case of the negative control. The amount of 5-epi-SNEP non-specific binding was also evalued as 1% by RP-HPLC-UV analysis of the washing waste solution.

A deactivation step with H_2O:isopropanol (1:2 vol/vol) was then performed to block all remaining epoxy groups on the resins.

Finally, the two resins were extensively washed with PBS to remove isopropanol traces and to be prepared for the subsequent affinity chromatography step.

3.2.2.3 Affinity Chromatography

HeLa cells pellets were lysed as previously reported, then proteins concentration estimated and adjusted to 3 μg/μL. The obtained lysate was divided into two aliquots of 1 mg of proteins each, which were incubated with the 5-epi-SNEP functionalized beads and the negative control matrix, respectively, to promote the interaction between the immobilized compound and its potential partner(s).

The beads were then isolated and washed three times with PBS, to reduce the amount of non-specific ligands weakly bound to the solid matrix, whereas the tightly bound proteins, most likely specifically *fished-out* by 5-epi-SNEP, were released treating the beads with a highly denaturant buffer (i.e. Laemmli buffer) for 5 minutes at 95°C.

Thus, the protein mixtures eluted from the 5-epi-SNEP-bearing and the negative control beads were resolved by 1D-SDS-PAGE on a 12% polyacrylamide gel, which was then fixed and submitted to Coomassie staining, in order to visualize the marine metabolite interacting proteins.

As can be observed in Figure 56, as for the DARTS experiment, the AP-MS output revealed a strong protein enrichment in a broad range of molecular weights when compared to the negative control beads, with the highest intensity increase around 42 kDa.

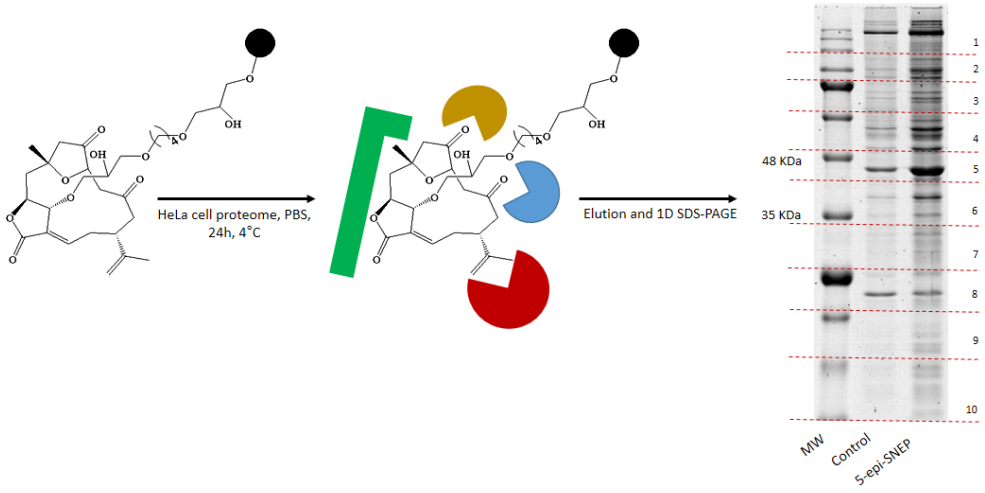


Figure 56: *5-epi-SNEP affinity chromatography incubation scheme and output on a 12% 1D-SDS-PAGE. Red dashed lines represent the cut areas for the in situ tryptic digestion.*

Thus, identification of 5-epi-SNEP *fished-out* proteins was accomplished by cutting the entire gel lanes of the control and the molecule-bearing beads into 10 pieces and submitting them to *in situ* tryptic digestion[34], to give peptide mixtures suitable for the subsequent HR-MS analysis.

3.2.2.4 Identification of 5-epi-SNEP interacting proteins through MS and bioinformatics analysis

The tryptic peptide mixtures coming from each gel slice were then analyzed through nano-flow RP-UPLC MS/MS: once converted into peak lists, the MS/MS data were submitted to bioinformatics analysis through the server MASCOT, for the protein identification step.

As opposed to DARTS, in which a semi-quantitative analysis is necessary to identify the molecule proteolysis protected and thus interacting proteins, the data analysis of an AP-MS experiment is usually simpler and faster: the list of 5-epi-SNEP potential interactors was refined

subtracting all the proteins shared with the negative control. Furthermore, the results obtained from two independent experiments were superimposed for a more confident list of putative and specific 5-epi-SNEP cellular targets.

As a result, even if the AP-MS workflow led to the identification of a quite big list of putative 5-epi-SNEP partner proteins, the cytoplasmic actin isoforms 1 and 2 were by far the best identified interacting targets, as reported in Figure 57.

Accession	Score	Mass	Matches	emPAI	Description
ACTG_HUMAN	***28445***	***42 108***	***849***	***32.56***	***Actin, cytoplasmic 2***
ACTB_HUMAN	***11533***	***42 052***	***407***	***1.12***	***Actin, cytoplasmic 1***
ACTBL_HUMAN	1697	42 318	104	1.11	Beta-actin-like protein 2
TBB4B_HUMAN	1623	50 255	73	5.63	Tubulin beta-4B chain
TBA1C_HUMAN	1118	50 548	53	1.85	Tubulin alpha-1C chain
1433E_HUMAN	1064	29 326	29	3.22	14-3-3 protein epsilon
ANXA1_HUMAN	996	38 918	28	1.96	Annexin A1
VDAC1_HUMAN	768	30 868	27	1.79	Voltage-dependent anion-selective channel protein 1
LDHB_HUMAN	693	36 900	32	0.54	L-lactate dehydrogenase B chain
EF2_HUMAN	685	96 246	37	0.73	Elongation factor 2
CAZA1_HUMAN	604	33 073	28	0.89	F-actin-capping protein subunit alpha-1
1433Z_HUMAN	518	27 899	22	1.58	14-3-3 protein zeta/delta
TERA_HUMAN	456	89 950	22	0.51	Transitional endoplasmic reticulum ATPase
ARF1_HUMAN	431	20 741	11	1.76	ADP-ribosylation factor 1
TPM3_HUMAN	405	32 987	28	1.23	Tropomyosin alpha-3 chain
COF2_HUMAN	297	18 839	19	1.31	Cofilin-2
PGAM1_HUMAN	287	28 900	20	1.08	Phosphoglycerate mutase 1
VDAC3_HUMAN	244	30 981	44	0.98	Voltage-dependent anion-selective channel protein 3
ADT3_HUMAN	242	33 073	13	0.61	ADP/ATP translocase 3
TBB6_HUMAN	226	50 281	12	0.52	Tubulin beta-6 chain
VDAC2_HUMAN	220	32 060	10	0.64	Voltage-dependent anion-selective channel protein 2
PRDX4_HUMAN	213	30 749	12	0.67	Peroxiredoxin-4
PPIB_HUMAN	180	23 785	16	0.94	Peptidyl-prolyl cis-trans isomerase B
ANXA5_HUMAN	177	35 971	17	1.08	Annexin A5
RS10_HUMAN	116	18 886	13	0.75	40S ribosomal protein S10

Figure 57: *List of putative 5-epi-SNEP protein partners identified through the AP-MS strategy. The list is the result of two independent experiments.*

In fact, MASCOT reported score and matches for both of the actin isoforms are exceptionally high if compared to the other proteins of the list.

Thus, AP-MS led to a confident identification of cytoplasmic actin isoforms as main 5-epi-SNEP partners, corroborating the previously obtained DARTS results.

It is worthy to notice that the DARTS and AP-MS experiments shared the identification of the isoform 1 of cytoplasmic actin as one of the main 5-epi-SNEP partners, whereas the isoform 2 was identified by AP-MS only. Even if these results could seem peculiar of the experimental procedures exploited, these two isoforms share extremely high sequence homology (i.e. 98.9%), differing from each other only by four aminoacids at their N-terminus peptide.

Thus, since it would be unlikely to state that 5-epi-SNEP could interact specifically with one isoform over the other, as also strongly suggested by AP-MS, in the subsequent paragraphs 5-epi-SNEP main cellular partners will be generically referred as *actins*.

3.2.3 Validation of DARTS and Affinity Chromatography obtained MS data: Western Blotting analyses

Validation of the mass spectrometric data obtained by both the functional proteomics approaches about 5-epi-SNEP interaction with actins was achieved by Western blotting.

Thus, both the DARTS and AP-MS samples were separated through SDS-PAGE and then transferred onto nitrocellulose membranes which were first blocked with a milk solution and then hybridized with a primary antibody against actin isoforms. The antigen/antibody reaction was monitored by incubating the membranes with a proper secondary antibody conjugated with horse radish peroxidase: in presence of H_2O_2 and activated luminol, a chemi-luminescent signal is detected when the secondary antibody recognizes the primary antibody/target protein complex.

As for the AP-MS samples, a significant antigen/antibody reaction could be monitored in the 5-epi-SNEP gel lane compared to the control, providing a clear indication of a specific actins enrichment induced by the marine compound (Figure 58A).

Concerning the DARTS experiment, as shown in Figure 58B, increasing concentrations of 5-epi-SNEP undoubtedly protected actins from the subtilisin action, when compared to the

control samples exposed or not to the enzyme (first and last lane of the membrane, respectively). Furthermore, since GAPDH proved to be subtilisin-resistant under these experimental conditions, it was used as a loading normalizer for the subsequent densitometric analysis.

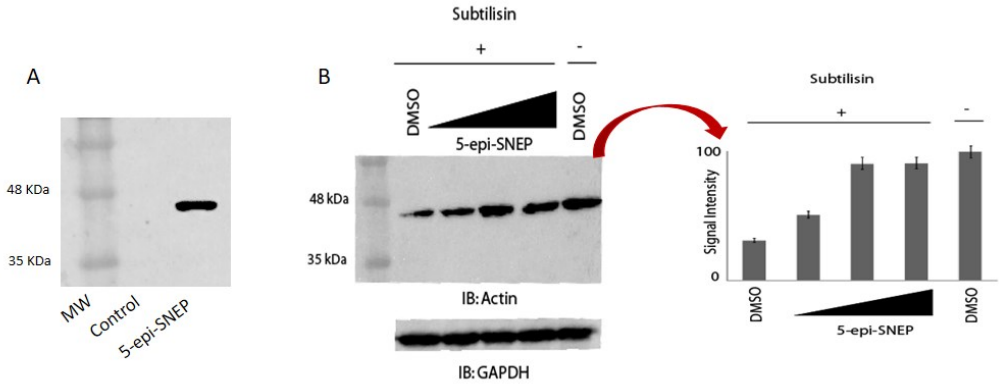


Figure 58: *Western Blot analyses performed with an actins-recognizing antibody on the AP-MS (A) and DARTS (B) obtained samples. Regarding DARTS, the densitometric analysis was performed through ImageJ and the histogram is the result of the signals quantitation from two independent experiments (standard deviations are reported). The undigested actin was rated as 100% and GAPDH was exploited as a loading normalizer.*

3.3 Actin proteins

Actin is the most abundant protein in eukaryotic cells, being highly conserved and participating in the biggest and most complex network of protein-protein interactions. These properties, along with its ability to transition between monomeric (G-actin) and filamentous (F-actin) states, make actin a critical player in many cellular functions, including the elaboration of filaments in the leading edge of migrating cells, the maintenance of membrane integrity, the contraction of muscle cells and the formation of cleavage furrows during cell division.

Vertebrates express six actin isoforms: α-cardiac, α-skeletal, α-smooth and γ-smooth muscle actins are mainly expressed in muscle cells, while the cytoplasmic β- and γ-actin isoforms (i.e. cytoplasmic isoforms 1 and 2, respectively) are ubiquitous. Despite all of these proteins differ by only a few amino acids, the β- and γ-actin isoforms are the two most homologous ones, differing by only 4 aminoacid residues at their N-terminal end (positions 1, 2, 3 and 9).

The 375-aminoacid polypeptide chain of actin folds into two major α/β domains, which are further classified into a total of four-subdomains: subdomains 1 and 3 are structurally related, whereas subdomains 2 and 4 can be viewed as large insertions into the former ones.

Two clefts are formed between the two major actin domains: the upper cleft binds the nucleotide (ATP/ADP) and associates Mg^{2+}, whereas the lower cleft between subdomains 1 and 3 constitutes the major binding site for most actin binding proteins (ABPs) and also mediates important longitudinal contacts between actin subunits in the filamentous F-actin[174,175] (Figure 59).

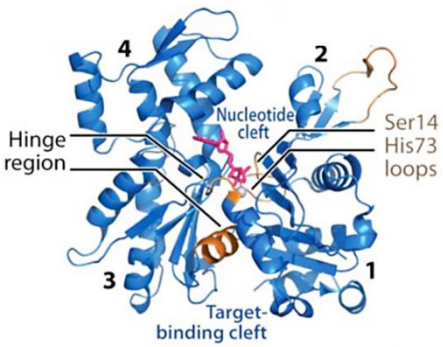


Figure 59: *Classical view of the actin monomer structure, in which subdomains 1 to 4 are labeled. Ser14 and His73 loops, the hinge between domains and the nucleotide- and target-binding clefts are reported. Adapted from Dominiguez et al.*[176].

Monomeric G-actin is continuously involved, together with filamentous F-actin, in a steady-state polymerization/depolymerization mechanism known as *treadmilling*[177]: G actin

monomers join the fast-growing barbed (or +) end of F-actin in the ATP-bound state, then ATP is hydrolyzed into ADP by the hinge helix Gln-137 residue[174,175] and the ADP-bound G-actin dissociates from F-actin pointed (or –) end.

The subdomain 1 Ser-14 β-hairpin loop and the His-73 sensor loop[178] are primarily involved into the structural rearrangements following ATP/ADP exchange. Indeed Ser-14 loop, together with the equivalent subdomain 3 Asp-157 loop, engulfs the phosphates of the nucleotide, making hydrogen bonding contacts with the ATP γ-phosphate and with the main chain of His-73 loop. After ATP hydrolysis and γ-phosphate release, Ser-14 changes orientation to form a hydrogen-bonding contact with the β-phosphate of the nucleotide, freeing space for the His-73 loop, which moves towards the nucleotide and thus appears to sense its state. Furthermore, because ATP sits at the interface between the two major actin domains, another important consequence of its hydrolysis is a weakening of the linkage between domains, which can then rotate more freely with respect to one another.

When polymerized, F-actins interact to form micro-filaments, which are essential cytoskeleton constituents. The major contact points between neighboring F-actins, reported by by Fujii *et al.*[174], are shown in Figure 60, with the stronger bonds being engaged between subdomains 3 and 4.

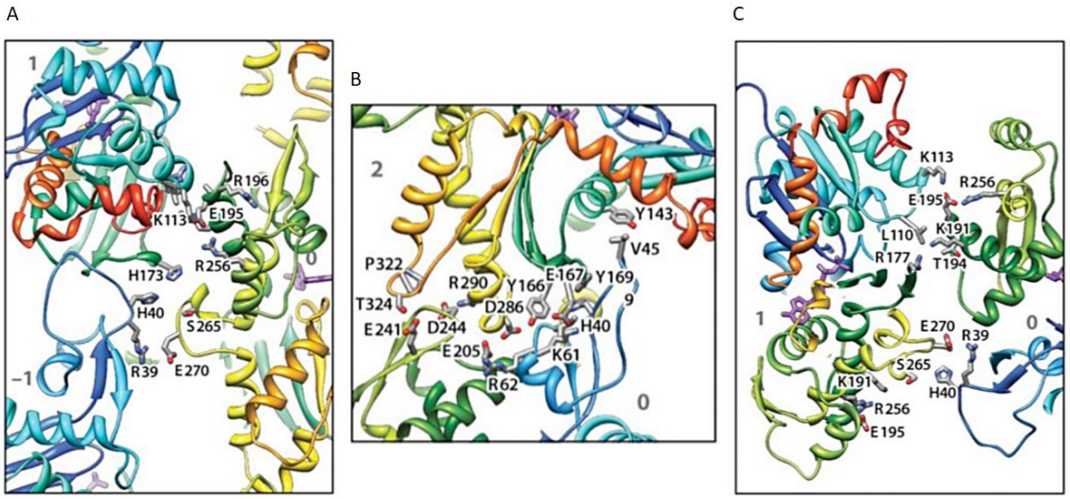


Figure 60: *Neighboring F-actins interactions: four F- actin molecules are shown, labeled in gray from –1 to +2. The proteins are represented with their cartoon secondary structure, color-coded from blue (N terminus) to red (C terminus). Interacting aminoacidic side chains are shown as gray sticks. Adapted from Dominiguez et al.*[176].

As can be observed, residues like Arg-39, His-40, Val-45, Lys-61, Arg-62, Arg-177, Lys-191, Thr-194 and Glu-195 are strongly involved in F-actin intermolecular interactions.

3.4 Analysis of 5-epi-SNEP/actins interaction features through t-LiP-MRM

After identification of actins as main 5-epi-SNEP cellular partners, we moved onto the characterization of the interaction features shared by the marine metabolite with cytoplasmic actins.

As already explained in chapter 2 (paragraph 2.3.2.1), the first step in a t-LiP-MRM experiment is a computational analysis performed through *PeptideAtlas* and *SRMAtlas* to build-up the MRM method for tracking the fully tryptic peptides of the target protein(s) in a complex mixture.

Thus, actins PABST peptides were retrieved from the *PeptideAtlas* Human build, for a total of 23 peptides among which only the one located at the N-terminus was different between the two isoforms, as could be expected by their 98.9% sequence homology.

The best transitions for each of these peptides were then selected through the Complete Human *SRMAtlas* build, searching for highly probable fragment ions providing the highest signal/noise intensity. The results list was refined as already done before (paragraph 2.3.2.1).

Thus, a comprehensive method listing 21 actins peptides and their three best transitions was obtained and tested onto a HeLa lysate tryptic digest.

As previously reported, the XICs of all the transitions of each precursor were inspected to (1) identify actins tryptic peptides which could be experimentally observed, (2) assign the retention time to all of the peptides and (3) identify the best transition for each peptide (the one showing the most intense peak and the best signal to noise ratio). This preliminary experiment led to a global MRM method comprising 16 transitions (Figure 61A).

Furthermore, the actually observed peptides were mapped onto actin 1 and 2 isoforms through the PAWS program, leading to a much satisfactory 63% coverage and to a uniform mapping of the proteins throughout all their length (Figure 61B).

A

Q1_m/z	Q2_m/z	Actin 1	Actin 2
1108.04	628.34	D-[292-312]-R	D-[292-312]-R
1085.21	753.39	M-[119-147]-R	M-[119-147]-R
998.49	391.24	D-[184-191]-K	D-[184-191]-K
983.95	868.33	-	M-[1-18]-K
955.92	967.4	M-[1-18]-K	-
868.12	234.15	Y-[337-359]-K	Y-[337-359]-K
850.73	808.41	L-[216-238]-K	L-[216-238]-K
796.66	890.56	T-[148-177]-R	T-[148-177]-R
644.43	361.25	G-[63-68]-K	G-[63-68]-K
597.64	689.36	S-[234-259]-R	S-[234-259]-R
581.31	634.32	E-[316-326]-K	E-[316-326]-K
566.77	912.44	G-[197-206]-R	G-[197-206]-R
505.92	694.35	I-[85-95]-R	I-[85-95]-R
488.73	630.28	A-[19-28]-R	A-[19-28]-R
400.24	294.52	A-[29-29]-R	A-[29-29]-R
391.2	332.19	H-[40-50]-K	H-[40-50]-K

B

Actin 2

1 MEEEIAALVIDNGSGMCK 18

1 MDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYV 54

55 GDEAQSKRGILTLKYPIEHGIVTNWDDMEKIWHHTFYNELRVAPEEHPVLLTEAPLNP 112

113 KANREKMTQIMFETFNTPAMYVAIQAVLSLYASGRTTGIVMDSGDGVTHTVPIYEGY 169

170 ALPHAILRLDLAGRDLTDYLMKILTERGYSFTTTAEREIVRDIKEKLCYVALDFEQEM 227

228 ATAASSSSLEKSYELPDGQVITIGNERFRCPEALFQPSFLGMESCGIHETTFNSIMK 284

285 CDVDIRKDLYANTVLSGGTTMYPGIADRMQKEITALAPSTMKIKIIAPPERKYSVWIG 342

343 GSILASLSTFQQMWISKQEYDESGPSIVHRKCF 375

Actin 1

Figure 61: *(A) Global MRM method transitions, reported with the corresponding peptides identifiers. (B) Cytoplasmic actin 1 and 2 aminoacidic sequences. Since their 98.9% homology, the isoform 1 is represented in all of its length, whereas only the N-terminal isoform 2 peptide is reported, with the four and only aminoacids differing between the two isoforms being in bold. Adjacent tryptic peptides are represented with different colors and the experimentally mapped ones are highlighted in yellow.*

Regarding the actual t-LiP-MRM experiment, 300 µg aliquots of a HeLa lysate prepared in non-denaturing conditions were incubated with either 5-epi-SNEP (5 µM and 50 µM final concentrations) or DMSO. The obtained samples were then divided into two aliquots and submitted to limited proteolysis with subtilitin (1:1000 w/w and 1:500 w/w ratios, respectively). This step was carried out in native conditions, so that subtilisin preferential cleavage sites would be uniquely dependent on actins structural features for the DMSO-treated sample, whereas 5-epi-SNEP binding should induce structural changes making actins regions directly or distally involved in the interaction less exposed to the un-specific protease.

Furthermore, 300 µg of the DMSO-treated sample were submitted to a mock proteolysis to be kept as a positive control.

Subtilisin was then quenched with PMSF and the samples shifted to denaturing conditions to carry out an extensive *in solution* tryptic digestion and produce peptides mixtures suitable to the LC-MRM-MS analysis, performed exploiting the previously optimized MRM method.

Actins *LiP peptides* were identified comparing the reference samples (submitted or not to limited proteolysis) and searching for tryptic peptides whose areas decreased in dependence on the subtilisin quantity (reverse proportionality), then the selected peptides were examined comparing the DMSO and 5-epi-SNEP-treated samples, to identify peptides whose area increased in presence of the marine metabolite, as previously reported (pharagraph 2.3.2.4).

Among the 16 actins peptides analyzed throughout the global MRM method, 12 were identified as being LiP responsive peptides, and are reported in Figure 62A, together with their fold change values. When comparing their behavior with the 5-epi-SNEP-treated samples, only 4 *LiP peptides* showed a significant area increase, directly proportional to the marine metabolite concentration (Figure 62B).

A

Actin 1	Actin 2	$Fc_{Lysate/1000}$	$Fc_{Lysate/500}$
D-[292-312]-R	D-[292-312]-R	Inf	Inf
M-[119-147]-R	M-[119-147]-R	7.02	9.23
D-[184-191]-K	D-[184-191]-K	1.75	3.77
-	M-[1-18]-K	NaN	NaN
T-[148-177]-R	T-[148-177]-R	465.90	591.85
G-[63-68]-K	G-[63-68]-K	4.06	15.17
S-[234-259]-R	S-[234-259]-R	2.28	5.84
G-[197-206]-R	G-[197-206]-R	NaN	NaN
A-[19-28]-R	A-[19-28]-R	3.61	8.09
A-[29-39]-R	A-[29-39]-R	3.27	7.17
H-[40-50]-K	H-[40-50]-K	NaN	NaN

B

		1:500 w/w	
Actin 1	Actin 2	$Fc_{5\mu M/CTRL}$	$Fc_{50\mu M/CTRL}$
A-[19-28]-R	A-[19-28]-R	2.32	2.35
A-[29-39]-R	A-[29-39]-R	2.00	2.15
G-[63-68]-K	G-[63-68]-K	3.86	3.98
D-[184-191]-K	D-[184-191]-K	2.32	2.50

C

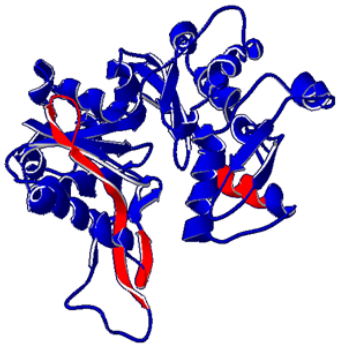


Figure 62: *(A) Selected LiP peptides, reported with the calculated fold changes. 'NaN' stands for Not a Number, as no tryptic peptide could be quantified in the subtilisin treated sample and no fold change could be calculated. (B) 5-epi-SNEP significantly altered ($p<0.05$) LiP peptides, reported with their molecule vs control fold change values (subtilisin to proteins ratio of 1:500 w/w). Actins aminoacid numbering is comprehensive of the initial methionine. (C) Actin 3D structure visualized through SPDB viewer and represented through sticks and ribbons. The protein backbone is depicted in blue, whereas 5-epi-SNEP protected LiP peptides are highlighted in red.*

As can be observed both in Figure 62B and in the 3D actin 1 structure visualized through SPDB viewer in Figure 62C, these peptides map for actins N-terminal ATPase domain, crucial for G-actin polymerization into its filamentous form. Furhtermore, they also embed or neighbor to Arg-39, His-40, Val-45, Lys-61, Arg-62, Arg-177, Lys-191, Thr-194 and Glu-195 which, as reported in the previous paragraph, are strongly involved in F-actin intermolecular interactions.

3.5 The biological consequences of an interaction: *in cell* assays

Encouraged by the results obtained by both *label-free* and *label-based* functional proteomics, we sought to evaluate 5-epi-SNEP interaction with G- and F-actins in a biological context, to both confirm its occurrence and characterize the *in vivo* consequences of such an event.

3.5.1 HeLa cells viability assay

A preliminary MTT assay was carried out, (collaboration with Prof. Tosco lab, Department of Pharmacy, University of Salerno), to select non-cytotoxic 5-epi-SNEP concentrations to be used in the subsequent *in cell* assays.

The MTT assay is a colorimetric test for assessing cell metabolic activity and, as a consequence, their viability. Indeed, mitochondrial NADPH-dependent oxidoreductases are capable of reducing, in living cells, the yellow tetrazolium dye MTT (i.e. 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) to its insoluble formazan, which has a purple color. When solubilized, the formazan produces a colored solution, whose absorbance can be measured at 550 nm to determine cell viability.

Thus, HeLa cells were cultured in a 96-well plate and exposed for 24 h to multiple 5-epi-SNEP concentrations prior to add MTT in each well. After 1 h, the supernatants were removed and the resultant formazan crystals dissolved in DMSO, to further measure the absorbance intensity at 550 nm.

Furthermore, even though all the subsequent *in cell* assays would have been performed on 5-epi-SNEP only, the epoxidated derivatives of the compound (paragraph 3.2.2.1) were also submitted to the MTT assay, to ensure the marine metabolite functionalization in the AP-MS workflow unaffected its biological behavior.

As reported in Figure 63, neither 5-epi-SNEP nor its derivatives with a short epoxide-alkyl chain affected cell viability up to 50 µM, ensuring 5-epi-SNEP blocking through its alcohol function on the resin for AP-MS did not alter its behavior, and allowing us to choose 5-epi-SNEP concentrations of 2.5 and 10 µM for the following *in cell* experiments.

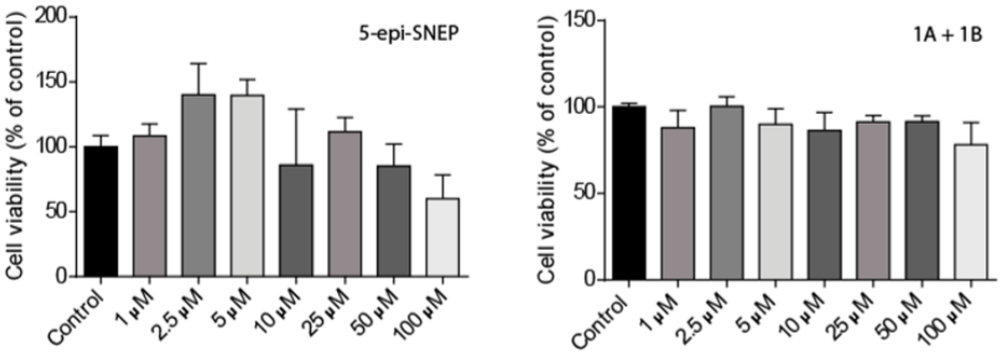


Figure 63: *MTT assay performed on HeLa cells treated with the reported concentrations of 5-epi-SNEP (left panel) or its derivatives (compound 1A and 1B, right panel). Histograms are the result of four independent experiments and the cell viability of each sample is reported, as a percentage, with its standard deviation values. Untreated cell viability (control) was rated as 100%.*

3.5.2 In cell DARTS

To evaluate 5-epi-SNEP *in cell* interaction with cytoplasmic G-actins, an *in cell* DARTS experiments was performed.

More in details, HeLa cells were incubated with either 5-epi-SNEP (2.5 and 10 µM final concentrations), the known F-actin depolymerizing agent Cytocalasin D[179] (CyD, 0.5 µM final concentration) or DMSO as a vehicle for 1 hour. Cells were then washed and harvested.

The obtained pellets were lysed through sonication and proteins concentration was determined and adjusted to 3 µg/µL. 100 µg proteins aliquots from each cell lysate were then submitted to limited proteolysis with subtilisin at the enzyme to substrate ratio of 1:1000 w/w for 30 minutes and in native conditions. Furthermore, 100 µg proteins aliquots from each sample were submitted to a mock proteolysis to be kept as a reference for the evaluation of a potential direct effect of 5-epi-SNEP and CyD on cytoplasmic actins. The resulting samples were then treated with PMSF and boiled in Laemmli buffer to be submitted to SDS-PAGE and Western blotting with an anti-actins antibody (GAPDH was exploited as a loading normalizer).

The Western Blot and the densitometric analysis (Figure 64A and B) pointed out that both 5-epi-SNEP and CyD protected G-actins from the subtilisin proteolytic action without significantly altering its undigested levels. Furthermore, even if to a lesser extent in respect to CyD, 5-epi-SNEP stabilizing effect resulted to be concentration-dependent.

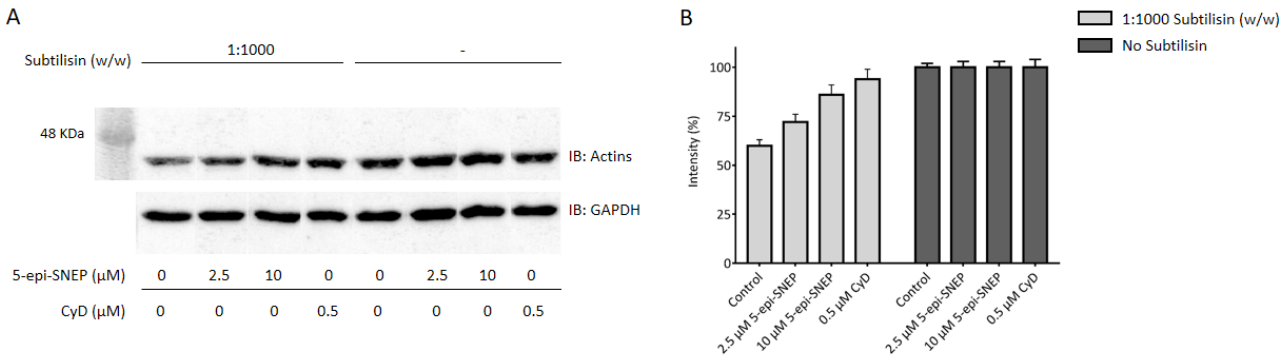


***Figure 64**: (A) Western Blotting analysis of the in cell DARTS experiment. (B) Densitometric analysis of the Western Blot performed through ImageJ. Histogram reports the averaged values from three independent replicates, together with the associated standard deviations. Values are reported as a percentage using GAPDH as a loading normalizer and rating, for each sample, its undigested actin intensity as 100%.*

These results corroborated the *in vitro* obtained data, confirming 5-epi-SNEP interaction with cytoplasmic actins.

3.5.3 Analysis of Actins polymerization state through confocal microscopy

A deeper investigation on the consequence of the 5-epi-SNEP induced structural changes in actin domains crucial for the polymerization in F-actin and its intermolecular interactions was done *in cell* through confocal microscopy (collaboration with Prof. Tosco lab, Department of Pharmacy, University of Salerno).

Thus, HeLa cells were seeded on cover slips and either treated with 5-epi-SNEP (2.5 or 10 μM final concentrations), CyD (0.5 μM) as positive control for actin filament depolymerization or DMSO as a vehicle. Cells were then fixed, permeabilized and blocked and actin microfilaments and nuclei were stained, respectively, with tetramethylrhodamine (TRITC)-phalloidin and 4′,6-diamidino-2-phenylindole (DAPI).

Fluorescence micrographs of the samples actin cytoskeletons and nuclei were then obtained by scanning confocal microscopy and are reported in Figure 65, where actin is visualized in red and nuclei in blue.

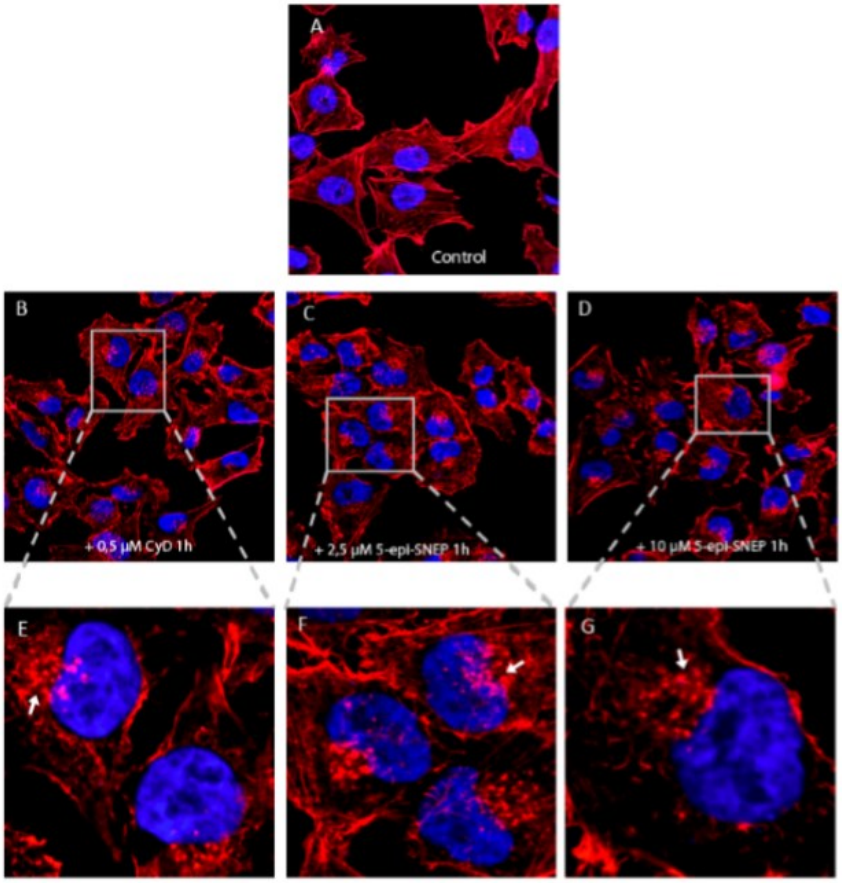


Figure 65: *Effect of 5-epi-SNEP on F-actin. HeLa cells were treated with DMSO (panel A), two different concentrations of 5-epi-SNEP (panels C and D and magnification in panels F and G) or CyD (panel B and magnification in panel E) for 1h. Actin was labeled with TRITC-phalloidin, nuclei with DAPI and fluorescence micrographs were acquired by confocal microscopy. The images (63× magnification. red: F-actin; blue: nuclei) are representative of multiple fields and triplicate cover slips per experiment.*

As can be observed in panels C and D and in their enlarged boxes (panels F and G), 5-epi-SNEP induced a partial disruption of the actin cytoskeleton at its lowest concentration, in respect to the control (panel A). Indeed, as evidenced by with arrows in the expanded panels, a decrease of actin fibers and formation of F-actin amorphous aggregates occurred at all the marine metabolite concentrations, where small actin aggregates co-existed with actin

microfilament bundles and diffuse cortical F-actin. As a confirmation, a similar behavior was detected for the known actin depolymerizing agent CyD[179] (panels B and E).

3.6 5-epi-SNEP: conclusive remarks

5-epi-SNEP is a *Sinularia* soft coral norcembranoid endowed with an interesting and multi-faceted biological activity profile[170,172,173]. Since no studies elucidating its mechanism of action at a molecular level were disposable, we sought to identify this marine metabolite interacting proteins through a functional proteomics-based Drug-Target Deconvolution analysis.

5-epi-SNEP interactome resulted complex and rather promiscuous and thus needed to be studied through both DARTS and the conventional AP-MS approach: the coupling of these two strategies allowed a confident MS-based identification of cytoplasmic actins as the most reliable marine metabolite interactors, as further confirmed by Western Blotting. Indeed, actins resulted the most protected proteins at the lowest 5-epi-SNEP concentration in DARTS experiments as well as the top-identified ones among the marine metabolite *fished-out* proteins in the conventional AP-MS approach.

T-LiP-MRM was then carried out to pinpoint actins regions involved in the interaction with 5-epi-SNEP: 4 peptides mapping for actins N-terminal ATPase domain were identified as being limited proteolysis-sheltered by the molecule.

Since this domain is pivotal for G-actin polymerization into F-actin as well as strongly involved in F-actin intermolecular interactions, *in cell* experiments were then performed to evaluate an eventual 5-epi-SNEP-exerted polymerization hampering.

In cell DARTS showed both the marine metabolite and the known depolymerizing agent CyD exerted proteolytic protection on G-actin, whereas confocal microscopy actin staining experiments enlightened that, as CyD, 5-epi-SNEP induced a partial decrease of actin fibers and formation of F-actin amorphous aggregates.

Thus, a general picture could be highlighted in which 5-epi-SNEP induces G-actins stabilization, interacting with protein domains crucial for both its polymerization in F-actin and for F-actin inter-protein interactions. As a consequence, the marine metabolite exerts a strong depolymerizing effect on the actin filamentous form, thus altering the *treadmilling* process biological balance.

Cytoskeleton structure and dynamics alterations are considered a remarkable symptom of several diseases, such as cancer[180] and neurodegenerative disorders[181], in which cytoskeleton disassembly usually leads to protein abnormal aggregation (i.e. inclusion bodies production) and misfolding. To date, one of the most efficient manners of perturbing actin involves the use of natural toxins which significantly depolymerize F-actin, such as the lactone macrolides cytochalasins and latrunculins, which are widely used to test the contribution of actin in cellular events, such as endo/exocytosis, cell polarity and cell motility and migration. Since these compounds have no specificity for different actin isoforms (i.e. cardiac, smooth muscle, muscle and cytoskeletal isoforms), their use as potential drugs is impaired due to undesirable off-target effects. However, actin-targeting molecules are still considered suitable tools for shedding light on this complex part of cells internal machinery[182].

In this scenario, 5-epi-SNEP could be proposed as a natural compound bearing a novel chemical structure capable of interfering with actin polymerization dynamics, and thus endowed with a a great potential for the study of cytoskeletal alterations at a molecular level.

CHAPTER 4

Disclosing Poly [ADP-Ribose] Polymerase 1 as the main Crellastatin A target through a combination of *label-free* functional proteomics approaches[183].

Adapted from

Morretta, E., Tosco, A., Festa, C., Mozzicafreddo, M., Monti, M. C., Casapullo, A. Crellastatin A, a PARP-1 Inhibitor Discovered by Complementary Proteomic Approaches. *ChemMedChem* **2020**, *15*, 317–323.

4.1 Background

Sponges (Porifera phylum) represent the richest and most variegated source of marine bioactive compounds, comprising around 30% of all the marine products identified until last decade[184]. Indeed, peculiar nucleosides, sterols, alkaloids, peroxides, terpenes, fatty acids, amino acid derivatives and cyclic peptides have been discovered as sponges active compounds so far, with several of them showing chemo-preventive and/or anticancer potentials[185].

Among all of these compounds, sponge sterols are regarded as an important class of bioactive molecules, often containing unusual side-chains and modified ring systems, ranging in carbon content from C_{24} to C_{31}[186–188]. In particular, steroid dimers form a significant group of pharmacologically active compounds, with some of them being among the most potent natural cytotoxic agents[189–191] (e.g. cephalostatins and ritterazines).

Steroid dimers can be classified into (1) acyclic/linear dimers, which involve connections between rings or via C-19 and form the major group of steroid dimers and into (2) cyclic dimers, in which the dimerization leads to formation of new ring systems or macrocyclic structures. Both of these classes can be further distinguished in symmetrical and unsymmetrical dimers: if a compound is composed of two identical steroid monomeric units, it is called a symmetrical dimer, whereas if two different monomeric steroid units are involved or two identical monomeric units are joined in a way that there is no symmetry in the resulting molecule, then the dimer is known as an unsymmetrical one.

Cyclic unsymmetrical steroid dimers show some of the most complex and intriguing chemical features, often related to interesting bioactivity profiles.

In this scenario, Crellastatin A (CreA, Figure 66A), isolated from the Vanuatu Island *Crella* sponge (Figure 66B), is a sulfated dimeric-steroid with a particularly unusual dioxabicyclo system joining its side chains, never before found in natural compounds[192].

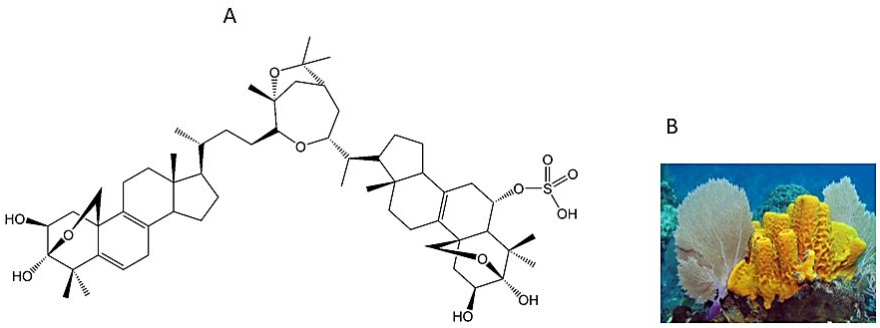


Figure 66: *(A) Crellastatin A. (B) Crella sp.*

CreA exhibited an interesting *in vitro* cytotoxic activity against a human bronchopulmonary non-small-cell lung carcinoma cell line (i.e. NSCLC-N6), showing an IC_{50} of 1.5 µg/mL[192]. Furthermore, it was possible to appreciate, in flow cytometry assays performed on the NSCLC-N6-C98 clone, that a dose of 18 µg/mL CreA induced cells accumulation in the G1 phase of the cell cycle with a concomitant decrease of the cells fraction in the S and G2/M phases.

Moreover, the weaker cytotoxic activities exhibited by the corresponding 2,2′-diacetylated (IC_{50} 9.1 µg/ mL) and desulfated (IC_{50} 9.3 µg/mL) derivatives indicated the hydroxyl and the sulfate groups were mandatory for CreA to exert its biological activity[192].

Thus, since no molecular-level studies were reported on the CreA mechanism of action, we sought to identify and characterize CreA protein partner(s) using proteomic approaches able to preserve its chemical structure (i.e. the essentials hydroxyl and sulfate groups), taking advantage of our optimized *label-free* functional proteomics platform.

4.2 CreA main cellular target(s) identification through *label-free* Functional Proteomics

In our CreA target deconvolution strategy plan, we sought to first exploit DARTS to identify the main marine metabolite cellular partner(s) and then to take advantage of t-LiP-MRM to shed light on the interaction features shared by CreA and its target protein, in order to retrieve information upon the putative interaction site, undergoing structural changes upon the molecule binding.

4.2.1 Identification of CreA cellular partners through DARTS

As previously done (paragraph 3.2.1) a pivotal experiment was performed, which led to the selection of the 1:500 w/w as the best subtilisin to proteins ratio to be used, allowing the clearest visualization of the different proteins stabilities in response to several CreA concentrations.

Thus, a HeLa lysate was prepared in non-denaturing conditions and the protein concentration was determined and adjusted to 3 μg/μL.

300 μg aliquots of the obtained lysate were incubated with CreA (0.3 μM or 3 μM final concentrations) to allow the interaction between the small molecule and its partner(s) and each sample was then submitted to limited proteolysis, in native conditions, with the previously chosen subtilisin amount. Two cell lysate samples were solely treated with DMSO and only one of them with subtilisin, to respectively obtain positive and negative controls.

Subtilisin was then quenched with PMSF and the samples separated through 1D-SDS-PAGE on a 4%-12% polyacrilamide gradient gel for a better Coomassie visualization, in respect to a fixed polyacrylamide percentage gel, of CreA stabilized proteins.

As described in the previous chapters, CreA binding shelters proteins from subtilisin proteolytic action, making possible to monitor their reduced protease sensitivity by detecting

gel bands with increased intensity in response to increasing concentrations of the marine compound.

Thus, the Coomassie stained gel presenting two DARTS biological replicates was both inspected *by eye* and submitted to a densitometric analysis, to identify protein bands whose intensities increased with a direct proportionality in respect to CreA concentrations. As can be observed by both the SDS-PAGE and the densitometric analysis in Figure 67A and B respectively, CreA-exerted proteolytic protection resulted largely superimposable throughout two biological replicates (red arrows in panel A and low standard deviation values reported in the graph in panel B), with the highest intensity increase (at the lowest CreA concentration) at around 120 kDa. Furthermore, CreA protein stabilization resulted to be strongly dependent upon the marine metabolite concentration.

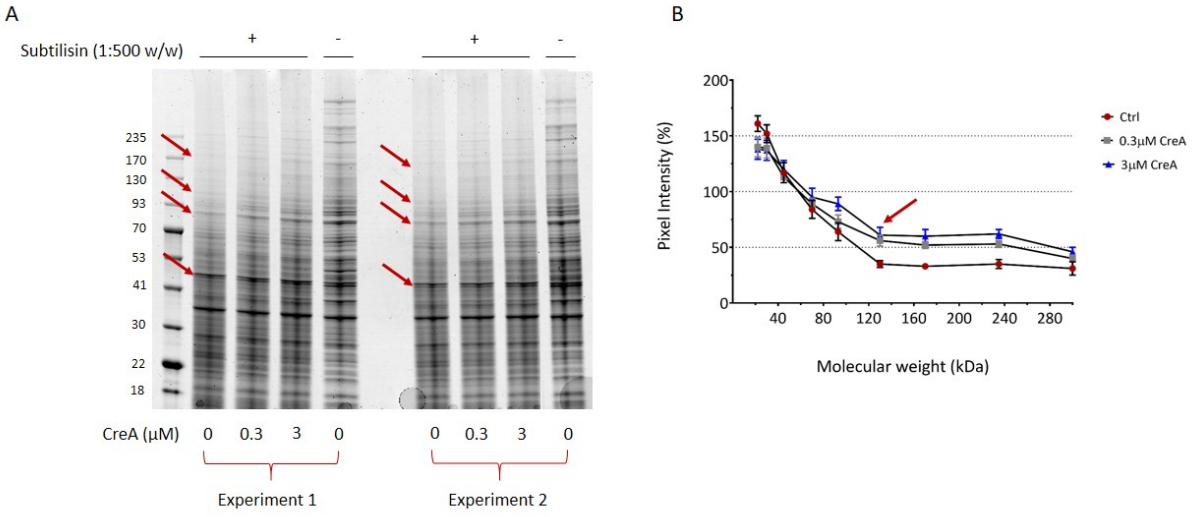


Figure 67: *(A) 4%-12% SDS-PAGE of two DARTS biological replicates showing CreA exerted protection from subtilisin in the same molecular weight ranges (red arrows). (B) Densitometric analysis of the two biological replicates performed through ImageJ. Intensities (%) are reported as averaged values of the two experiments ± standard deviations, rating the undigested sample averaged value (not reported in the graph) as 100%.*

The gel bands whose intensity increased with the molecule concentration were then excised and submitted to *in situ* tryptic digestion[34] and the peptide mixtures were analyzed by nano-flow RP-UPLC MS/MS.

The MS data were submitted to bioinformatics analysis (MASCOT server) for proteins identification and a semi-quantitative analysis was performed by comparing MASCOT retrieved matches among the treated and untreated samples (submitted or not to limited proteolysis), for each biological replicate. A protection percentage was then calculated for each protein in correspondence of its molecular weight, as already described in chapter 2 (paragraph 2.2.3).

Among all of the identified stabilized proteins, Poly [ADP-ribose] Polymerase 1 (PARP 1) resulted the most sheltered at the lowest CreA amount and its protection was both strongly dependent on the marine metabolite concentration and particularly reproducible in the two biological replicates (Figure 68A).

A

			Protection (%)			
			Experiment 1		*Experiment 2*	
Accession	**Mass**	**Description**	**0.3 µM CreA**	**3 µM CreA**	**0.3 µM CreA**	**3 µM CreA**
CLH1	193 260	Clathrin heavy chain 1	31	50	24	55
IQGA1	189 761	Ras GTPase-activating-like protein IQGAP1	35	24	11	26
UBA1	118 858	Ubiquitin-like modifier-activating enzyme 1	12	37	46	54
PARP1	**113 811**	**Poly [ADP-ribose] polymerase 1**	**55**	**73**	**57**	**100**
SYAC	107 484	Alanine--tRNA ligase, cytoplasmic	0	46	28	89
C1TC	102 180	C-1-tetrahydrofolate synthase, cytoplasmic	14	50	17	33
HNRPU	91 269	Heterogeneous nuclear ribonucleoprotein U	11	14	24	28
ENPL	92 696	Endoplasmin	30	74	20	36
GELS	86 043	Gelsolin	0	27	13	13
HS90A	85 006	Heat shock protein HSP 90-alpha	15	30	24	78
HS90B	83 554	Heat shock protein HSP 90-beta	23	38	26	69
HS71A	70 294	Heat shock 70 kDa protein 1A	6	35	4	28
DDX5	69 618	Probable ATP-dependent RNA helicase DDX5	20	20	0	33
ACTC	42 334	Actin, alpha cardiac muscle 1	17	26	3	30
ANM1	42 059	Protein arginine N-methyltransferase 1	50	58	0	13
ACTB	42 052	Actin, cytoplasmic 1	26	30	7	30
HNRPD	38 581	Heterogeneous nuclear ribonucleoprotein D0	11	21	13	20
MAGA4	35 106	Melanoma-associated antigen 4	15	15	6	22

B

			Experiment 1								*Experiment 2*							
			CTRL		0.3 µM CreA		3 µM CreA		Lysate		CTRL		0.3 µM CreA		3 µM CreA		Lysate	
Accession	**Mass**	**Description**	**Score**	**Matches**	**Score**	**Matches**	**Score**	**Matches**	**Score**	**Matches**	**Score**	**Matches**	**Score**	**Matches**	**Score**	**Matches**	**Score**	**Matches**
PARP1_HUMAN	113 811	Poly [ADP-ribose] polymerase 1	-	-	93	6	285	8	253	11	-	-	67	4	204	7	334	7

Figure 68: *(A) List of putative CreA interacting proteins retrieved from two biological replicates. Proteins are reported with the corresponding protection percentages. (B)*

MASCOT parameters related to PARP 1 identification. No undigested PARP 1 was identified in the CTRL experiments, as opposite to the CreA treated samples, in which Mascot parameters almost resembled the Lysate retrieved ones.

Indeed, in both the DARTS replicates, PARP 1 was fully digested in the negative control samples (i.e. molecule-untreated and subtilisin-exposed proteins, CTRL in Figure 68B) as a result of this protein high proteolytic susceptibility, whereas a high PARP 1 amount was detected in the CreA treated samples, almost resembling the positive control (i.e. sample without molecule and subtilisin, Lysate in Figure 68B), as a result of a relevant CreA-exerted protection on PARP 1.

4.2.2 Validation of the DARTS obtained MS data: Western Blotting analysis

Validation of the MS-obtained DARTS data, strongly suggesting PARP 1 as the main CreA partner, was achieved by Western Blotting.

Thus, the DARTS samples were separated through a 12% SDS-PAGE and transferred onto a nitrocellulose membrane, which was firstly blocked with a milk solution and then incubated with a primary antibody raised against PARP 1. To monitor the antigen/antibody reaction, the membrane was subsequently incubated with a proper horse radish peroxidase-conjugated secondary antibody. GAPDH was used as a loading normalizer for the following densitometric analysis.

As shown by both the membranes and the densitometric analysis performed on the *full-length* protein in Figure 69A and B respectively, in both DARTS experiments CreA clearly protected the *full-length* PARP 1 (signal in between the 135 kDa and the 100 kDa molecular

weight markers in Figure 69A) from the protease action, when compared to the negative control samples in which no undigested PARP 1 can be detected.

Furthermore, detection of subtilisin produced a PARP 1 fragment of about 44 kDa whose intensity decreased with a reverse proportionality to CreA amount (Figure 69A), was an additional evidence of the concentration dependent PARP 1 stabilization by the marine metabolite.

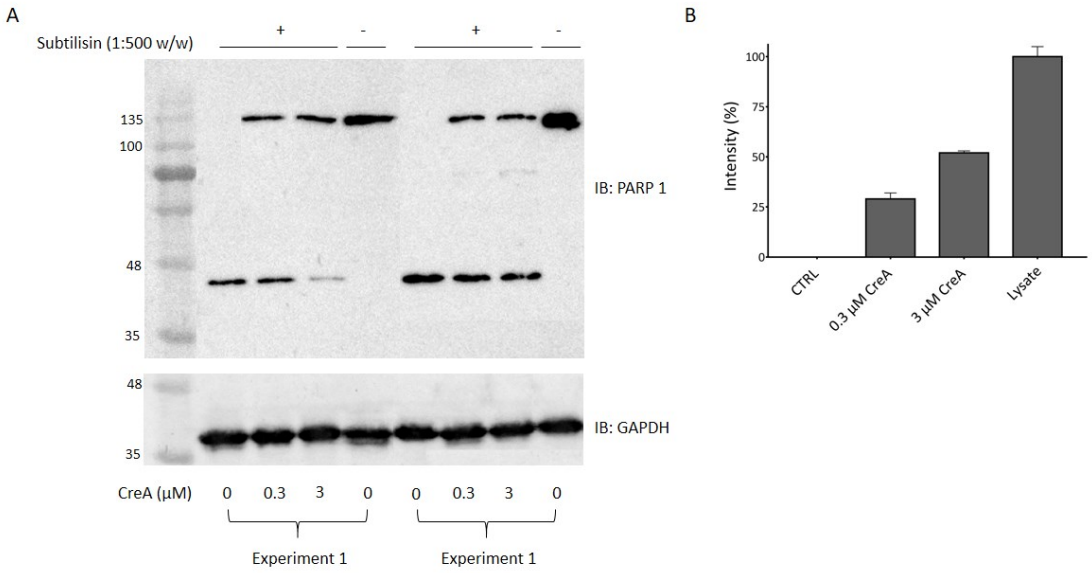


Figure 69: *(A) Western Blot analysis performed with a PARP 1 recognizing antibody on both of the DARTS experiments. (B) Densitometric analysis performed through ImageJ on the full-length PARP 1. The histogram is the result of the signals quantitation from the two independent experiments: pixel intensities are reported as percentages with the related standard deviations, rating the undigested PARP 1 intensity as 100%. GAPDH was exploited as a loading normalizer.*

4.3 Poly [ADP-Ribose] Polymerase 1 (PARP 1)

PARP 1 is a highly conserved, multi-domain and multi-functional nuclear enzyme playing a key role in both DNA repair pathways and transcription regulation through chromatin structure remodeling[193].

The main role of PARP 1 is to produce poly-ADP-ribose units (PAR) from the ADP donor NAD^+ and attaching them, either in a linear or a branched form, to Glu, Asp or Lys residues of acceptor proteins (i.e. itself or other target proteins) by a trans-esterification reaction[194] (Figure 70).

The generated PAR is afterwards rapidly catabolized by proteins like PAR glycohydrolase (PARG), ADP-ribosylhydrolase 3 (ARH3) and O-acyl-ADP-ribose deacylase 1 (OARD1).

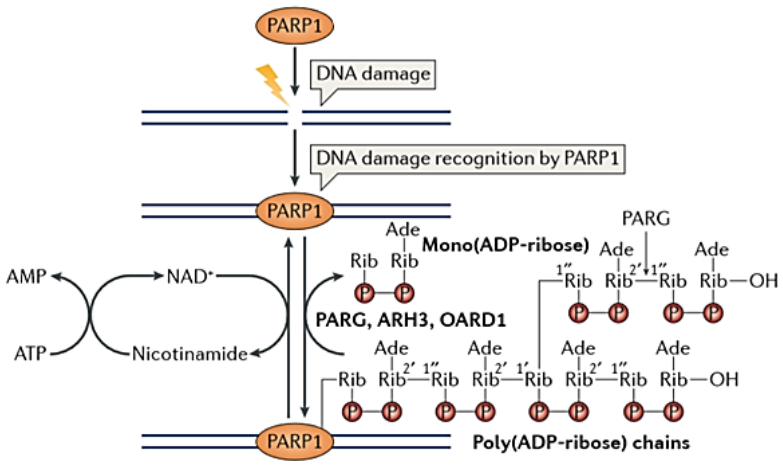


Figure 70: *PARP 1 enzymatic activity following damaged DNA binding. Adapted from Ray Chaudhuri et al.*[194].

Proteins involved in DNA repair pathways and histones related to chromatin structure remodeling can bind to the PAR residues PARP 1 attaches on itself, an event that results in their recruitment to DNA damage/binding sites[195]. The PAR-binding modules in these target proteins

frequently overlap with important functional domains, which are activated through PAR binding and mediate biological functions such as DNA targeting, protein/protein interactions and nuclear localization[196–199].

As for the DNA reparation pathways, PARP 1 plays a crucial role in the recovery of single-stranded DNA breaks and nicks, double-stranded DNA breaks and DNA bulky lesions and in the stabilization of replications forks[194], thus assuring the maintenance of genome integrity in various pathological contexts (e.g. after metabolic, oxidative, oncogenic or genotoxic stresses).

As targeting the DNA damage response (DDR) machinery is an attractive strategy for designing novel chemotherapeutics, the inhibition of PARP 1 catalytic activity in cancers that are defective in DDR-involved genes is emerging as an interesting therapeutic strategy, since it gives rise to synthetic lethality (i.e. a phenomenon in which mutations or perturbations of two genes together result in a loss of cell viability, as opposite to what happens if only one gene is involved). PARP 1 inhibitors are indeed exploited, either in clinics or in trials, for the treatment of breast and ovarian cancers showing BRCA2 mutations or in metastatic prostate cancers characterized by defects in DNA repair genes[194].

In addition to its role in DNA repair, PARP 1 also controls the functions of various components of the transcription machinery. Indeed, apart from nicked DNA, its activity can also be induced by the interaction with histone H4 in a phosphorylated H2Av-histone-bearing nucleosome. When enzymatically active, PARP 1 covalently modifies itself and surrounding nuclear proteins by synthesizing PAR strands that render histones more electronegative, shifting them away from the negatively-charged DNA molecule, thus initiating chromatin loosening and allowing transcription activation[200,201]. Nevertheless, PARP 1 binding at genes regulatory regions can also correlate with transcription repression[202], which may potentially occur to prevent interferences of the transcription machinery with the DNA repair process in

actively transcribing regions. Notably, PARP 1 involvement in the transcription machinery control does not always strictly require its enzymatic activity.

Being involved in such complex networks, PARP 1 has also a complex structure, showing a modular architecture organized in three fundamental domains: the N-terminal DNA binding domain (DBD), the middle auto-modification domain (AD) and the highly conserved C-terminal catalytic domain (CAT)[194,203,204]. DBD and CAT are further constituted of 3 principal subdomains each (Figure 71A).

The DNA-binding DBD comprises three *zinc finger* motifs, namely Zn1, Zn2 and Zn3, and a nuclear localization signal (NLS, not labeled in Figure 71A). Zn1 and Zn2 bind, in a sequence independent manner, to a variety of DNA structures including double-strand breaks (DSB), single-strand breaks (SSB), extensions, hairpins and cruciforms[205–209], whereas Zn3 contains a structurally unique *zinc ribbon* fold mandatory for the mediation of PARP 1 inter-domain contacts critical for the protein DNA damage-dependent catalytic activity[210–212].

The AD consists of the BRCA1 C-terminus (BRCT) region comprising the major sites targeted by PARP 1 for its auto-modification[213,214], critical for the protein recruiting activity in the DNA damage response, as well as in the chromatin remodeling process: an auto-modified PARP 1 loses its ability to interact with the DNA[215,216] and serves as a *shuttle* to enroll proteins required in both of the pathways[194,217].

The CAT domain encompasses three subdomains: the Tryptophan-Glycine-Arginine (WGR) domain, a helical domain (HD) and the ADP-ribosyltransferase (ART) domain[194,218].

The WGR domain participates in DNA binding and forms inter-domain contacts essential for PARP 1 DNA damage-dependent structural rearrangement (i.e. *priming*) and activation[219], whereas the ART domain, that contains the catalytic site and a fold conserved in all PARP family members[220], is directly responsible for PARP 1 poly-ADP ribosylasion activity.

Of all of PARP 1 subdomains, Zn1, Zn3, WGR and ART are strictly required for the protein DNA damage-dependent catalytic activity[209,210]: Zn1, Zn3, and WGR collectively bind to the damaged DNA, organizing PARP 1 into a collapsed conformation (Figure 71B) that creates a network of essential inter-domain contacts[219] stimulating the ART activity, with the consequent auto poly-ADP ribosylation of the AD and the recruitment of several proteins involved in DNA repair.

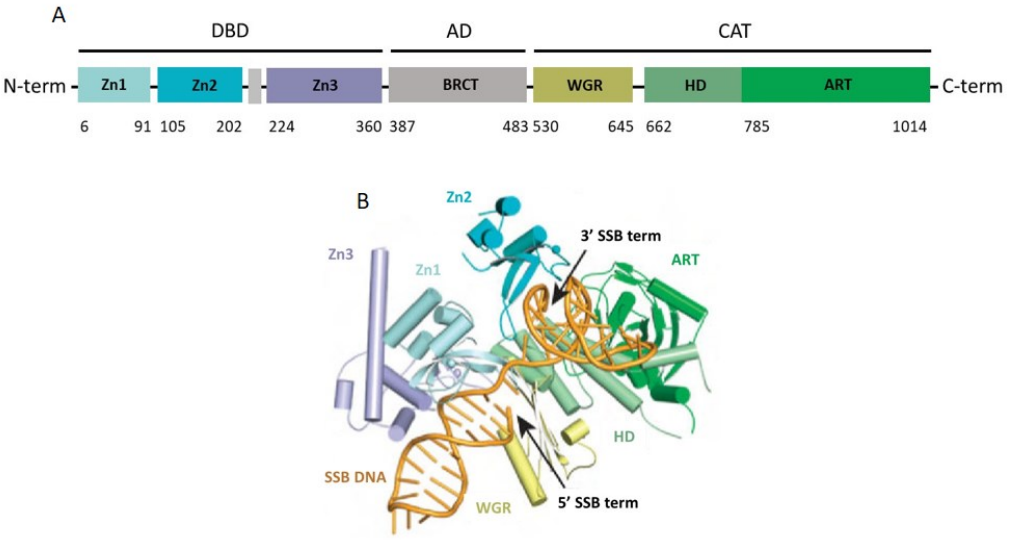


Figure 71*: (A) PARP 1 domains and subdomains architecture. The NLS in DBD is not labeled, being only represented by a gray rectangle. (B) Principal PARP 1 subdomains 3D organization after binding a SSB-DNA. The reported subdomains are represented with the same color code in panels A and B. Adapted from Dawicki-McKenna et al.*[218].

More in details, Zn1, Zn3 and WGR domains primarily contact the DNA ribose-phosphate backbone through sequence-independent interactions made of hydrophobic contacts with exposed nucleotide bases, a common feature of damaged DNA structures. Zn1 binds to DNA through the *backbone grip* and the *base-stacking loop* conserved regions, whereas Zn3 through its N-terminal α-helical region, spanning DNA minor groove. WGR binds to the 5'-terminus of

one DNA strand, holding the DNA backbone between its central β-sheet and the α-helix, with Trp589 stacking against the ribose sugars of nucleotides located at the end of the 5' terminated DNA strand[211].

Upon DNA binding, a network of inter-domain contacts is formed: Zn1 and Zn3 contact each other and both provide a two-point interaction surface for engaging WGR. Indeed, Zn1 and Zn3 are located next to each other on the damaged DNA, both interacting with one WGR face, whereas the opposite WGR face contacts the HD.

In particular, Zn3 *extended loop* contacts both WGR and HD (Figure 72A), whereas a key Zn1-WGR contact is achieved through a salt bridge formed between Zn1 Asp45 and WGR Arg591 (Figure 72B), which also interacts with HD. This latter phenomenon is responsible for the ART catalytic activation following damaged DNA binding.

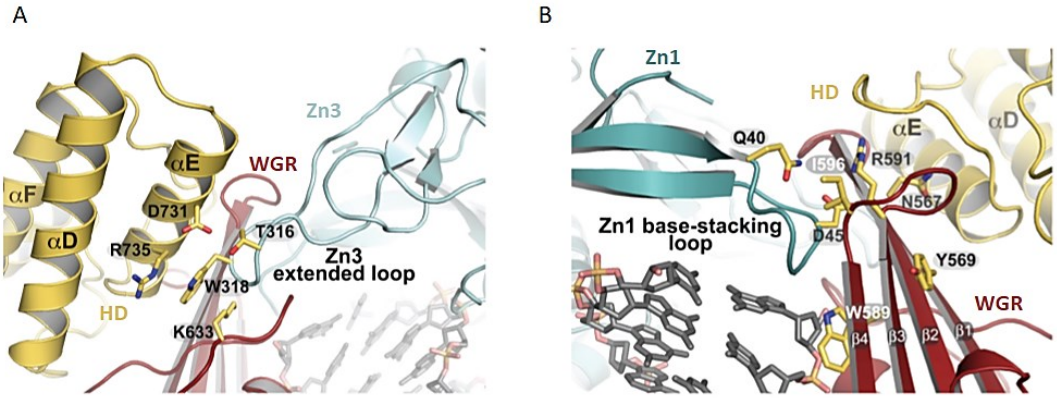


Figure 72: *3D structural representations of HD-WGR–Zn3 (A) and of Zn1–WGR–HD (B) interfaces. Aminoacidic residues majorly involved in the contact points between the represented subdomains are reported in both of the panels. Adapted from Langelier et al.*[211].

Thus, bridging among the damaged DNA, the DBD and the CAT domain, WGR is regarded as a crucial PARP 1 component playing a key role in its autocatalytic activation, a fundamental process in the enzyme orchestrated DNA-repairing pathways.

4.4 Analysis of CreA/PARP-1 interaction features through t-LiP-MRM

Fascinated by the multifaceted roles PARP 1 encompasses in such fundamental cellular pathways, we performed t-LiP-MRM experiments to characterize CreA/PARP 1 interaction features.

We first performed a computational analysis through *PeptideAtlas* and *SRMAtlas*, in order to build-up the MRM method for tracking PARP 1 tryptic peptides in a complex HeLa lysate tryptic digest.

Thus, PARP 1 PABST peptides were retrieved from the *PeptideAtlas* Human build and the best transitions for each of them were selected through the Complete Human *SRMAtlas* build, with the same search parameters and criteria used to refine the results as the ones reported in paragraph 2.3.2.1, except that we removed peptides whose *N_map* was higher than 1.

Thus, a comprehensive method listing 39 PARP 1 peptides and their 3 best daughter ions, for a total of 117 transitions, was obtained and subsequently tested onto a HeLa lysate tryptic digest.

As previously reported, the XICs of all the transitions for each precursor were inspected to accomplish the identification of the experimentally observed PARP 1 tryptic peptides, assign the corresponding retention times and select the best transition for each precursor. This preliminary experiment led to a global MRM method comprising 24 transitions (Figure 73A). Furthermore, the observed peptides were mapped onto PARP 1 sequence through the *PAWS* program, leading to an uniform mapping of the protein throughout all of its domains and to a satisfactory 33% coverage, considering that PARP 1 is a nuclear protein not abundant in a total

cell lysate and that no enrichment or prefractionation steps were performed prior to the LC-MRM analysis (Figure 73B).

A

Q1_m/z	Q2_m/z	Peptide
812.91	707.36	V-[453-467]-K
782.64	799.4	T-[88-119]-K
734.7	407.27	N-[820-838]-K
713.39	503.28	S-[468-486]-K
706.86	335.19	M-[443-467]-K
700.36	1044.53	Q-[270-282]-R
689.38	673.4	T-[866-878]-R
681.41	743.47	V-[552-564]-K
669.83	908.47	A-[685-695]-K
585.3	511.76	E-[168-182]-K
540.68	262.14	G-[183-192]-K
533.8	670.39	A-[487-496]-R
521.77	781.42	Q-[198-207]-K
498.95	532.31	K-[748-761]-K
487.26	506.21	L-[572-579]-K
438.76	204.62	E-[263-269]-K
424.2	333.19	V-[48-59]-K
416.72	331.2	L-[601-607]-K
416.7	674.32	G-[565-571]-K
410.7	244.17	C-[24-30]-K
381.19	291.17	G-[98-105]-K
373.84	346.17	E-[608-616]-K
352.7	261.16	L-[419-425]-K
300.82	234.15	T-[622-629]-K

B

```
1    MAESSDKLYRVEYAKSGRASCKKCSESIPKDSLRMAIMVQSPMFDGKVPHW  51
52   YHFSCFWKVGHSIRHPDVEVDGFSELRWDDQQKVKKTAEAGGVTGKGQDG  101
102  IGSKAEKTLGDFAAEYAKSNRSTCKGCMEKIEKGQVRLSKKMVDPEKPQLG  152
153  MIDRWYHPGCFVKNREELGFRPEYSASQLKGFSLLATEDKEALKKQLPGVK  203
204  SEGKRKGDEVDGVDEVAKKKSKKEKDKDSKLEKALKAQNDLIWNIKDELK  253
254  KVCSTNDLKELLIFNKQQVPSGESAILDRVADGMVFGALLPCEECSGQLVFK  305
306  SDAYYCTGDVTAWTKCMVKTQTPNRKEWVTPKEFREISYLKKLKVKKQDR  355
356  IFPPETSASVAATPPPSTASAPAAVNSSASADKPLSNMKILTLGKLSRNKDEVK  409
410  AMIEKLGGKLTGTANKASLCISTKKEVEKMNKKMEEVKEANIRVVSEDFLQ  460
461  DVSASTKSLQELFLAHILSPWGAEVKAEPVEVVAPRGKSGAALSKKSKGQVK  511
512  EEGINKSEKRMKLTLKGGAAVDPDSGLEHSAHVLEKGGKVFSATLGLVDIV  563
564  KGTNSYYKLQLLEDDKENRYWIFRSWGRVGTVIGSNKLEQMPSKEDAIEHF  614
615  MKLYEEKTGNAWHSKNFTKYPKKFYPLEIDYGQDEEAVKKLTVNPGTKSK  664
665  LPKPVQDLIKMIFDVESMKKAMVEYEIDLQKMPLGKLSKRQIQAAYSILSEV  716
717  QQAVSQGSSDSQILDLSNRFYTLIPHDFGMKKPPLLNNADSVQAKVEMLDN  767
768  LLNNADSVQAKVEMLDNLLDIEVAYSLLRGGSDDSSKDPIDVNYEKLKTDI  801
802  KVVDRDSEEAEIIRKYVKNTHATTHNAYDLEVIDIFKIEREGECQRYKPFKQ  853
854  LHNRRLLWHGSRTTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMVSKS  904
905  ANYCHTSQGDPIGLILLGEVALGNMYELKHASHISKLPKGKHSVKGLGKTT  955
956  PDPSANISLDGVDVPLGTGISSGVNDTSLLYNEYIVYDIAQVNLKYLLKLK  1000
1006 FNFKTSLW  1014
```

Figure 73: *(A) Global MRM method transitions, reported with the corresponding peptides identifiers. (B) PARP 1 aminoacidic sequence. Adjacent tryptic peptides are represented with different colors and the experimentally mapped ones are highlighted in yellow.*

Thus, the actual t-LiP-MRM experiment was performed incubating 300 µg aliquots of a HeLa lysate, prepared in non-denaturing conditions, with either CreA (3 µM final concentration) or DMSO. The samples were then divided into two aliquots and submitted to limited proteolysis with subtilitin (1:500 w/w and 1:250 w/w ratios, respectively) in native conditions. Furthermore, 300 µg of the DMSO-treated sample were submitted to a mock proteolysis to be kept as a positive control.

Subtilisin was then quenched with PMSF and the samples shifted to denaturing conditions to perform an extensive *in solution* tryptic digestion. The obtained peptides mixtures were desalted and submitted to LC-MRM analysis with the previously established MRM method.

The subsequent data analysis step was performed with an alternative strategy not involving fold changes evaluation, in order to consider and rank some peptides that were not quantified into the negative control experiments in respect to the CreA-treated samples (i.e. for which no CreA/CTRL fold change is calculable) and to also evaluate the extent of the molecule-exerted protection taking the positive control sample as a normalizing factor.

More in details, PARP 1 *LiP peptides* were identified comparing the DMSO-treated samples submitted (i.e. negative controls) or not (i.e. positive control or lysate) to limited proteolysis, searching for tryptic peptides responsive to subtilisin in a concentration dependent fashion. Instead of evaluating this behavior through fold changes, the intensity differences between each positive control and the corresponding negative control peptide were calculated at a given subtilisin amount, as follows:

$$\Delta_{Area\_Lys\text{-}250} = Area_{Lysate} - Area_{CTRL250};$$

$$\Delta_{Area\_Lys\text{-}500} = Area_{Lysate} - Area_{CTRL500}.$$

Thus, if a given tryptic peptide maps for a PARP 1 region reliably cleaved by subtilisin, its Δ_{Area} should decrease the less subtilisn has been used for the limited proteolysis step: all PARP 1 tryptic peptides presenting $\Delta_{Area\_Lys\text{-}500} < \Delta_{Area\_Lys\text{-}250}$ were identified as being *LiP peptides* and were selected for the following data analysis step. On this basis, among the 24 analyzed PARP 1 tryptic peptides, only 18 were selected.

For the sake of a straightforward data visualization, Figure 74A reports the selected LiP peptides with their associated Δ_{Areas} shown as percentages (i.e. $Norm\_\Delta_{AreaLys\text{-}CTRL}$), calculated exploiting the positive control peptides as normalizing factors, as follows:

$$Norm\_\Delta_{Area\_Lys\text{-}CTRL250} = (\Delta_{Area\_Lys\text{-}250}/Area_{\_positive\ control})*100;$$

$$Norm\_\ \Delta_{Area\_Lys\text{-}CTRL500} = (\Delta_{Area\_Lys\text{-}500}/Area_{\_positive\ control})*100.$$

These LiP peptides were then examined comparing the DMSO and CreA-treated samples, searching for possible area increases induced by the marine metabolite. Thus, intensity differences were calculated as follows:

$$\Delta_{Area\_peptide250} = Area_{CreA250} - Area_{CTRL250};$$

$$\Delta_{Area\_peptide500} = Area_{CreA500} - Area_{CTRL500}.$$

To evaluate reliable intensity increases (due to the marine compound interaction with PARP 1), we then calculated a PARP 1 overall intensity difference between the treated and untreated samples ($\Delta_{Area\_PARP\ 1}$) averaging all the $\Delta_{Area\_peptides}$ at each subtilisin amount. A 95% confidence value was then calculated and exploited to construct a confidence interval related to the $\Delta_{Area\_PARP\ 1}$ values.

All of those peptides showing an area increase exceeding the confidence interval in at least one subtilisin amount and in both the biological replicates were selected: only 8 of the 18 analyzed peptides satisfied the selection criteria, identifying the PARP 1 regions directly or distally involved in the interaction with CreA.

For a straightforward data visualization, percentages were calculated, as previously reported, for all of the peptides (i.e. Norm_$\Delta_{AreaCreA-CTRL}$, Figure 74B).

A

	Norm Δ_{Area}Lys-CTRL	
Peptide	**1:250_subtilisin**	**1:500_subtilisin**
C-[24-30]-K	30	20
G-[98-105]-K	86	78
T-[88-119]-K	92	70
E-[168-182]-K	71	51
G-[183-192]-K	86	61
E-[263-269]-K	100	93
L-[419-425]-K	19	13
M-[443-467]-K	59	5
V-[453-467]-K	92	69
S-[468-486]-K	100	67
V-[552-564]-K	78	69
G-[565-571]-K	83	31
L-[601-607]-K	90	69
E-[608-616]-K	32	23
T-[622-629]-K	53	48
A-[685-695]-K	71	51
N-[820-838]-K	90	62
T-[866-878]-R	100	82

B

	Norm Δ_{Area}CreA-CTRL	
Peptide	**1:250_subtilisin**	**1:500_subtilisin**
C-[24-30]-K	5	**14**
G-[98-105]-K	1	1
T-[88-119]-K	2	1
E-[168-182]-K	3	6
G-[183-192]-K	2	9
E-[263-269]-K	**11**	6
L-[419-425]-K	**11**	6
M-[443-467]-K	4	3
V-[453-467]-K	3	10
S-[468-486]-K	0	1
V-[552-564]-K	**16**	7
G-[565-571]-K	1	5
L-[601-607]-K	4	10
E-[608-616]-K	5	**16**
T-[622-629]-K	**17**	2
A-[685-695]-K	**15**	6
N-[820-838]-K	1	4
T-[866-878]-R	0	**36**

***Figure 74**: (A) PARP 1 selected LiP peptides, reported with their Norm_Δ_{Area} values obtained comparing the positive and negative controls. For each of them, the Norm_Δ_{Area} decreases when increasing subtilisin amounts have been used for LiP. (B) Selected LiP peptides reported with their Norm_$\Delta Area$ values obtained comparing CreA treated samples with the corresponding negative controls at each subtilisin amount. Only 8 peptides are protected by the molecule and are depicted in bold red. PARP 1 aminoacid numbering is comprehensive of the initial methionine.*

These peptides were mapped on a PARP 1 schematic cartoon (Figure 75A) and on the protein 3D structure visualized through SPDB viewer (Figure 75B): as can be observed, the major part of them maps for the protein CAT domain, being especially located in its WGR subdomain.

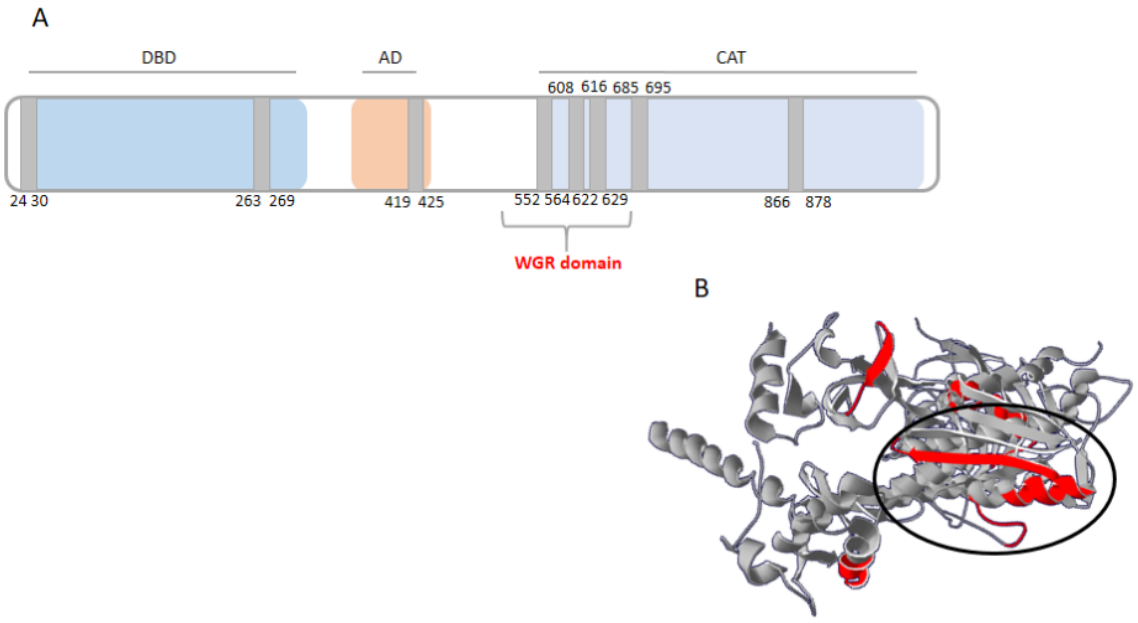


Figure 75: *(A) Schematic cartoon representing the three PARP 1 principal domains in three different colors. CreA significantly altered (95% confidence) LiP peptides are depicted as gray stripes. (B) 3D PARP 1 stick and ribbon representation obtained through SPDB viewer. The protein backbone is depicted in gray, whereas CreA protected LiP peptides are highlighted in red, with the WGR ones being indicated by the black circle.*

As mentioned in the previous paragraph, WGR bridges between PARP 1 DBD and CAT domains, providing essential contacts for the protein autocatalytic activation in response to damaged DNA binding. Interestingly, most of our t-LiP-identified CreA protected peptides lye in close proximity to WGR interacting points with both Zn3 (i.e. T-[622-629]-K is adjacent to Lys633 in the WGR subdomain) and Zn1 (i.e. C-[24-30]-K is in close proximity with Gln40 in Zn1 subdomain and V-[552-564]-K is adjacent to Asn567 and Tyr569 in WGR subdomain), suggesting a picture in which CreA could interfere with WGR/DBD interaction, thus hampering the subsequent PARP 1 autocatalytic activation.

4.5 T-LiP-MRM data corroboration: blind molecular docking analysis

Intrigued by the t-LiP-MRM results and given the complex PARP 1 modular domain architecture, we sought to corroborate our data with a blind molecular docking analysis, (collaboration with Dr. Matteo Mozzicafreddo, University of Camerino).

PARP 1 3D resolved crystallographic structure was retrieved form the Protein Data Bank[221] and used to predict the feasible binding mode and strength (i.e. the predicted equilibrium dissociation constant $K_{D,pred}$) with the designed and minimized CreA[222], considering a docking zone comprising the whole PARP 1.

On the basis of the predicted affinity, CreA produced its best interaction poses into the protein binding site involving the amino acids Met43, Phe44, Leu559, Thr566, Ser568, Tyr570, Arg587, Trp589, Glu620, Pro635, Lys636, Lys637 and Phe638 (Figure 76A). Moreover, Met43 and Arg587 resulted involved in the formation of H-bonds with oxygen atoms of CreA, whereas the sulfate group seemed to be exposed to the solvent.

The $K_{D,pred}$ value related to the above reported binding site was estimated to be of 20.29 ± 7.63 nM.

Looking at their positions, most of the amino acid residues involved in CreA/PARP 1 interaction fit in the region 559-638, corresponding to the protein WGR subdomain previously identified by t-LiP-MRM.

Thus, we sought to report our t-LiP-MRM results on the CreA/PARP 1 complex obtained from the molecular docking analysis, visualizing it through SPDB viewer.

Figure 76B shows the obtained output, where PARP 1 backbone is depicted in gray, CreA in light blue and the t-LiP-MRM identified protected peptides are highlighted in red. As can be observed, CreA produces its best interaction poses in a WGR region prevalently mapped by the peptides V-[552-564]-K, E-[608-616]-K and T-[622-629]-K but also comprising Zn1 subdomain residues (i.e. the t-LiP-MRM identified C-[24-30]-K peptide and the docking

enlightened Met43 and Phe44 aminoacids), which is in line with WGR subdomain establishing contacts with the Zn1 subdomain to achieve enzymatic autocatalytic activation.

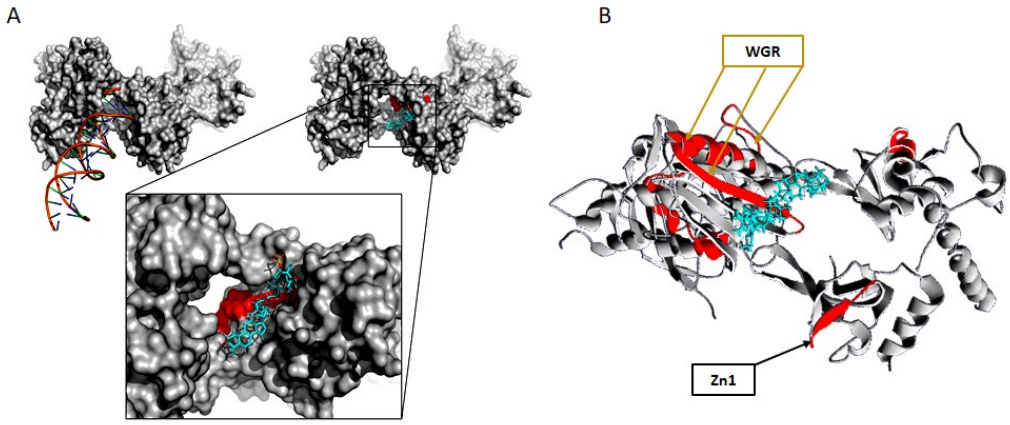


Figure 76: *(A) Molecular docking analysis of the CreA/PARP 1 complex, showing PARP 1 interaction with both damaged DNA (left panel) and CreA (right panel and magnification). CreA best interaction poses mostly involve the WGR subdomain (highlighted in red in the magnification) and to some extent the Zn1 one. The complex is rendered through PyMOL Molecular Graphics System, Version 2.0.4. (B) PARP 1/CreA complex obtained from the docking analysis. T-LiP MRM identified CreA protected peptides are highlighted in red: the WGR and Zn1 mapping peptides, also identified in the docking analysis, are signaled by arrows.*

Thus, blind molecular docking and t-LiP-MRM pointed to the same direction, strongly suggesting a CreA preferential interaction with PARP 1 WGR domain. Furthermore, molecular docking enlightened other aminoacidic residues, not mapped in our LiP experiments, which are directly involved in the inter-domain contacts WGR makes with the DBD Zn1 and Zn3 subdomains and with the CAT subdomain HD.

Indeed, as previously exposed, one WGR face contacts Zn1 and Zn3, whereas the opposite interacts with the HD.

In particular, aminoacids surrounding Lys633, which interacts with HD, have been identified through molecular docking (i.e. Glu620, Pro635, Lys636, Lys637 and Phe638), as well as the ones surrounding Asn567 and Tyr569 (i.e. Thr566, Ser568, Tyr570) and their interacting Zn1 Gln40 and Asp45 residues (i.e. Met43 and Phe44). Remarkably though, WGR Arg587, Trp589 were also identified by molecular docking: these aminoacidic residues lye in close proximity with WGR Arg591, responsible for a key contact with Zn1 Asp45, achieved through the formation of a salt bridge.

Thus, molecular docking not only corroborated t-LiP-MRM, but also enlightened other WGR and Zn1 interesting regions putatively involved in the interaction with CreA, which are indispensable for PARP 1 inter-domain interactions and activity.

4.6 The consequences of PARP 1/CreA interaction: *in vitro* activity assays

To investigate whether CreA could hamper PARP 1 inter-domain contacts and its subsequent autocatalytic activation following damaged DNA binding, we sought to build-up an UPLC-MRM-MS-based *in vitro* assay, to evaluate the enzyme activity in presence of the marine compound.

To achieve this purpose, we took advantage of PARP 1 ART activity, responsible for PAR production from NAD^+, which is exploited as an ADP donor and is consequently hydrolyzed into NAM: a potential CreA-induced enzymatic inhibition could be inferred by a reduced NAM production in presence of the marine compound, when compared with the free enzyme.

Mass spectrometric parameters were optimized by directly infusing NAM and were subsequently exploited for its MRM-based quantitation. Then, a preliminary time course experiment was carried out to tune the assay conditions, selecting the best incubation time at

which the enzyme produced a satisfactory NAM amount. For this purpose, PARP-1 was incubated with a mixture of a damaged DNA and a histone H3 peptide as a substrate, then NAD^+ was added to start the reaction. Reaction aliquots were quenched at different times with 10 mM final concentration of 3-aminobenzamide (3-ABA), a known PARP 1 inhibitor, and NAM production was monitored through UPLC-MRM-MS (Figure 77).

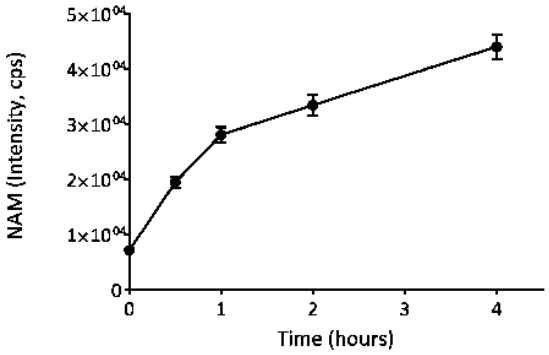


Figure 77: *Time-course preliminary PARP 1 activity assay.*

The obtained results allowed us to select 2 hours as the best reaction time to monitor in a subsequent assay, performed with six different CreA concentrations and with a fixed 3-ABA amount, exploited as a positive control of PARP 1 inhibition.

Thus, PARP 1 was pre-incubated with either CreA (5 µM, 10 µM, 25 µM, 50 µM, 100 µM and 150 µM final concentrations), 3-ABA (100 µM final concentration) and DMSO as a negative control (i.e. CTRL), then the damaged DNA, histone H3 peptide and NAD^+ mixture was added. Reaction aliquots were quenched with 10 mM 3-ABA after 2 hours and the samples were subjected to UPLC-MRM-MS to measure PARP 1 produced NAM.

As shown in Figure 78A, 100 µM 3-ABA retained around 75% of PARP 1 activity, whereas CreA hampered it more than 50% already at 50 µM and its inhibition profile resulted

concentration-dependent. Thus, the marine compound IC_{50} was calculated from the sigmoid curve (Figure 78B) and estimated as 32.96 µM.

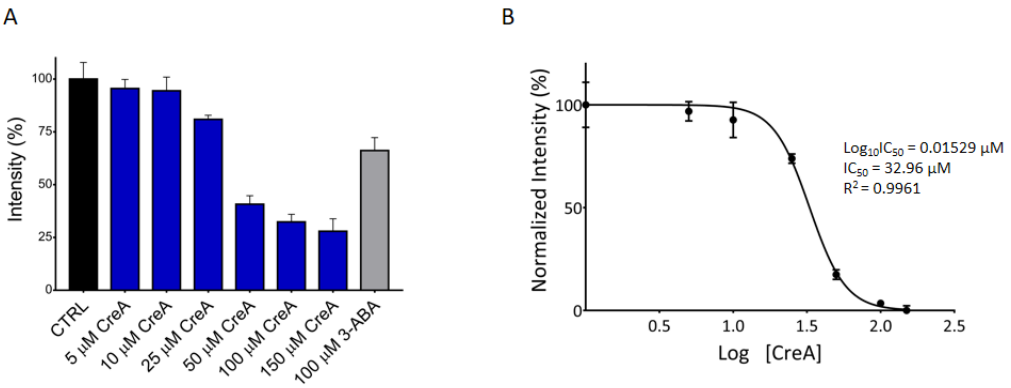


Figure 78: *(A) Intensity (%) of the NAM produced after 2 hours in presence of several CreA concentrations, reported rating the CTRL NAM intensity at 100%. 3-ABA has been used as a positive control of PARP 1 inhibition. Values are the result of three injection replicates and are reported with the corresponding standard deviations. (B) Sigmoidal curve reporting NAM normalized intensity versus logarithmic CreA concentration, with the relative retrieved* IC_{50} *values (logarithmic or not) and* R^2*, calculated through GraphPad Prism 7.00.*

4.7 CreA: conclusive remarks

CreA is a sulfated dimeric steroid isolated from the Vanuatu Island *Crella* sponge, which exhibited an interesting *in vitro* cytotoxic activity against human NSCLC-N6 cells and induced their C98 clone accumulation in the cell cycle G1 phase[192].

Since the hydroxyl and the sulfate groups were discovered to be mandatory for CreA biological activity[192], we sought to identify and characterize the unknown protein partner(s) of the marine compound taking advantage of our optimized *label-free* functional proteomics platform.

MS-based DARTS experiments led to the identification of PARP 1 as the most subtilisin-sheltered protein at the lowest CreA amount: its protection was both strongly dependent on the marine metabolite concentration and particularly reproducible in the two biological replicates, as validated by Western Blotting.

T-LiP-MRM experiments subsequently allowed the characterization of CreA/PARP 1 interaction features: the major part of the molecule protected peptides was discovered to map for the protein CAT domain, being especially located in its WGR subdomain.

As previously mentioned (paragraph 4.3), WGR provides contact points between PARP 1 CAT and DBD, essential for the protein autocatalytic activation in response to damaged DNA binding. Interestingly, most of the t-LiP-identified CreA protected peptides lyes in close proximity to WGR interacting points with DBD, suggesting a picture in which CreA could interfere with these two domains interaction, thus hampering PARP 1 autocatalytic activation.

This hypothesis was corroborated by blind molecular docking: CreA best interaction poses in PARP 1 WGR domain was enlightened, with a $K_{D,pred}$ value of 20.29 ± 7.63 nM.

Furthermore, t-LiP-MRM and blind molecular docking findings well fitted with the *in vitro* inhibition of PARP 1 enzymatic activity, evaluated through an UPLC-MRM-MS-based assay.

Due to its biochemical activity, PARP 1 is involved in several pathways and is strictly connected to the pathogenesis of different diseases, such as inflammation[223], neurodegenerative disorder[224,225], diabetes[226] and, mostly, cancer[194,227–229]. Consequently, the research for new PARP 1 inhibitors is a multifaceted area, particularly relevant for the treatment of cancer deseases lacking in DNA damage response-involved genes[194].

On the basis of our findings, CreA appears to be an interesting starting point for the development of simplified analogs as new PARP 1 inhibitors playing a (poly-) pharmacological role in the therapy of such daunting human diseases.

CHAPTER 5

Label-free functional proteomics reveals cytosolic HSP71A and mitochondrial GRP75 as the main cellular partners of a *Mycale rotalis* acetogenin.

5.1 Background

Sponges of the genera *Mycale* (*Arenochalina*) are rich producers of diverse and physiologically active secondary metabolites (e.g. alkaloids, polyketides, terpenoids and lipids, Figure 79) showing a wide range of biological properties which are reflected in their cytotoxic, anti-malarial, anti-HIV, anti-inflammatory, anti-fungal and anti-bacterial activities[230].

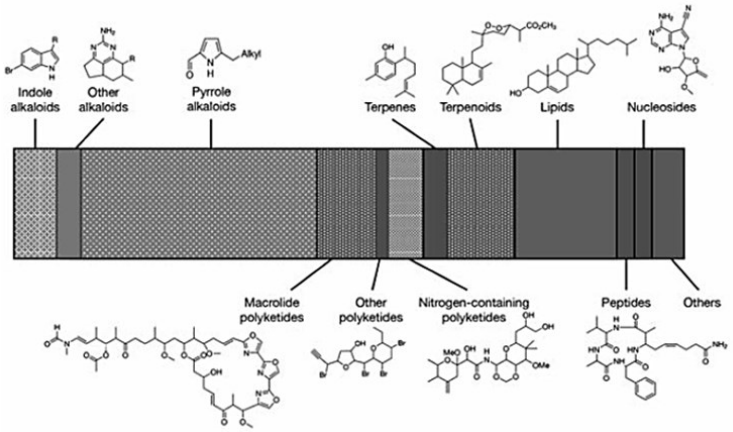


***Figure 79**: Mycale compound classes structural diversity. Shaded regions are related to the number of compounds within each class. The presented structures are representative of each class compounds. Adapted from Habener et al.*[230].

Among the *Mycale* sponges secondary metabolites, an interesting group of molecules, commonly found in the red algae *Laurencia* genus, have been identified in *Mycale rotalis* by Giordano *et al.*[231], as a result of preferential growth of this sponge on and engulf of *Laurencia* species. These compounds are structurally characterized by the presence of oxane rings of various sizes, an enyne or allenic side chain and at least one halogen atom.

In particular, polybrominated acetogenins have been found as non-terpenoid molecules originating in the polyketide pathway from a common fatty acid-derived C_{15} precursor[232] and, among them, we selected BrACG[231] (Figure 80) as the most interesting.

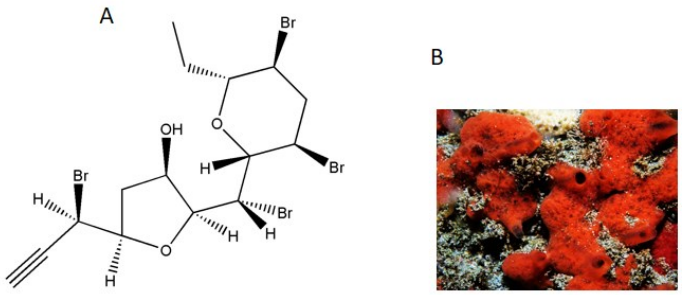


Figure 80: *(A) BrACG. (B) Mycale rotalis sponge.*

Given the huge range of bioactivities shown by *Mycale* sp. metabolites and since no data have still been reported about BrACG biological properties, we sought to shed light on this poly-brominated compound activity profile (collaboration with Professor Vincenzo Piccialli lab, Department of Chemical Sciences, University of Naples "Federico II").

5.2 BrACG main protein partner(s) identification through label-free functional proteomics

BrACG bioactivity profile characterization started with the identification of its preferential protein partner(s). As previously reported (chapter 4), BrACG target(s) deconvolution was achieved taking advantage of our optimized *label-free* functional proteomics platform, based on DARTS and t-LiP-MRM.

5.2.1 Identification of BrACG cellular partners through DARTS

The usual pivotal experiment was first performed (pharagraph 3.2.1), allowing to select 1:500 w/w subtilisin to proteins ratio as the best condition to achieve a clear visualization of proteins stabilization induced by BrACG.

Thus, a HeLa lysate was prepared in non-denaturing conditions and the protein concentration was measured and adjusted to 3 μg/μL.

300 μg lysate aliquots were then incubated with either BrACG (0.5 μM, 5 μM or 50 μM final concentrations) or the vehicle (DMSO) to allow the interaction between the small molecule and its partner(s) to occur. Each sample was then submitted to limited proteolysis in native conditions with the previously chosen subtilisin amount, keeping an undigested DMSO-treated lysate aliquot as a positive control. Subtilisin was then quenched with PMSF and the samples separated through 1D-SDS-PAGE on a 4%-12% polyacrilamide gradient gel, which was then Coomassie-stained.

BrACG interacting protein(s) were recognized by detecting gel bands whose intensity increased, in respect to the digested control, inspecting the stained gel *by eye* and after a densitometric analysis. Looking at the SDS-PAGE and densitometric plot in Figure 81A and B respectively, BrACG exerted proteolytic protection produced the highest intensity increase around 75 kDa. Furthermore, this stabilizing effect was strongly dependent upon the marine metabolite concentration.

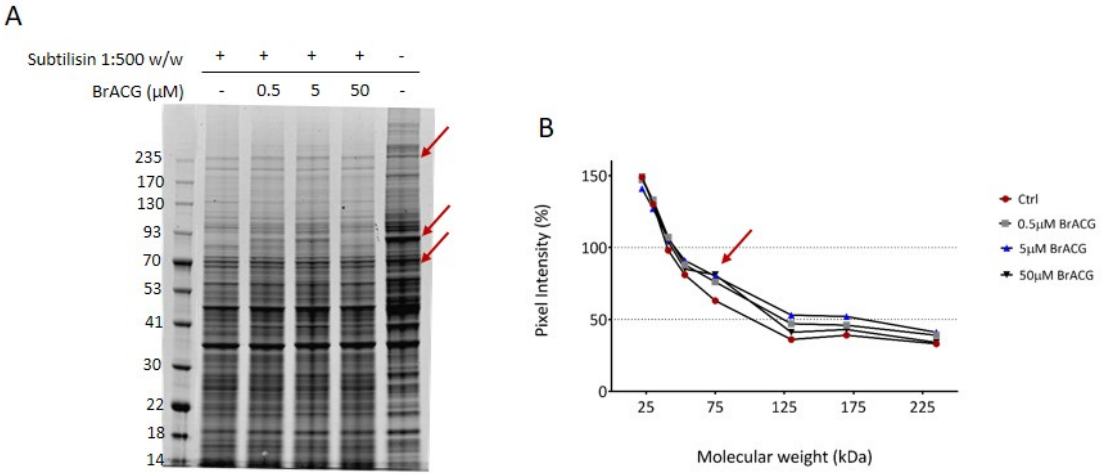


Figure 81: *(A) 4%-12% SDS-PAGE showing BrACG protein sheltering from subtilisin (red arrows). (B) Densitometric analysis of the DARTS experiment (ImageJ) showing the bottommost BrACG concentration induced the highest pixel intensity increase around 75kDa, where the protection is also dependent on the molecule concentration. The undigested sample intensities are not depicted in the graph, as they are rated as 100%.*

Thus, gel bands whose intensity increased in presence of BrACG were excised and submitted to *in situ* tryptic digestion[34] to give peptide mixtures for the subsequent nano-flow RP-UPLC MS/MS analysis.

The MS data were subsequently submitted to MASCOT server for protein identification and semi-quantitative analysis.

Thus, BrACG protected proteins were identified by directly comparing the MASCOT matches with those of the positive control experiment (BrACG-untreated and undigested sample), used as a normalizing factor, and of the negative control one (BrACG-untreated and subtilisin-exposed sample), which values were an indication of the proteins maximal response to the protease treatment. Thus, a protection percentage was calculated for each protein identified in

correspondence of its molecular weight. To obtain a more confident list of putative BrACG binding proteins, this experiment was carried out in duplicate.

On the basis of the MASCOT parameters reported in Figure 82B, two Hsp70 isoforms were identified as the most reliable BrACG partners: the mitochondrial Stress-70 protein (GRP75) and the cytosolic Heat shock 70 kDa protein 1A (HS71A). Indeed, they were the most protected proteins at the lowest molecule concentration in both of the DARTS experiments and their protection was directly proportional to BrACG concentrations and mostly reproducible between the two replicates (Figure 82A).

A

			Protection (%)					
			Experiment 1			Experiment 2		
Accession	Mass	Description	0.5 µM	5 µM	50 µM	0.5 µM	5 µM	50 µM
P38646	***73 920***	***Stress-70 protein, mitochondrial***	***19***	***34***	***47***	***17***	***36***	***45***
P0DMV8	***70 294***	***Heat shock 70 kDa protein 1A***	***11***	***11***	***22***	***13***	***60***	***68***
P29401	68 519	Transketolase	0	6	16	29	55	60
P30101	57 146	Protein disulfide-isomerase A3	6	11	22	3	28	30

B

			Experiment 1										*Experiment 2*									
			CTRL		0.5 µM BrACG		5 µM BrACG		50 µM BrACG		Lysate		CTRL		0.5 µM BrACG		5 µM BrACG		50 µM BrACG		Lysate	
Accession	Mass	Description	Score	Matches	Score	Matches	Score	Matches	Score	Matches	Score	Matches	Score	Matches	Score	Matches	Score	Matches	Score	Matches	Score	Matches
P38646	73 920	Stress-70 protein, mitochondrial	465	10	702	16	773	21	1281	25	1147	32	820	25	1167	34	1519	44	1589	49	1977	53
P0DMV8	70 294	Heat shock 70 kDa protein 1A	-	-	23	1	40	1	58	2	430	9	530	15	705	21	1423	43	1452	47	1270	47

Figure 82: *(A) List of putative BrACG interacting proteins retrieved from two biological replicates. Proteins are reported with the corresponding protection percentages. (B) MASCOT parameters related to GRP75 and HS71A (i.e. P38646 and P0DMV8 respectively) identification.*

5.2.2 Validation of the DARTS obtained MS data: Western Blotting analysis

Validation of DARTS obtained MS data, outlining GRP75 and HS71A as the most reliable BrACG cellular partners, was obtained by Western Blotting using antibodies for each of the two Hsp70 isoforms.

Regarding HS71A, as shown by both the membrane (Figure 83B) and the densitometric analysis output (Figure 83C, gray bars), increasing BrACG concentrations clearly protected this protein from subtilisin cleavage, compared to the negative control sample (first lane of the membrane), in which very low amount of intact *full-length* HS71A can still be detected.

Concerning GRP75, the exploited antibody recognizes an epitope made of its C-terminal aminoacids (i.e. residues 525-679). As can be observed by the membrane in Figure 83A, apart from the higher molecular weight signal related to the *full-length* protein (the only signal present in the BrACG and subtilisin-untreated sample, last membrane lane) it is possible to detect another signal, related to a subtilisin fragment produced by the removal of a relatively small GRP75 N-terminal peptide. As can be observed (Figure 83A), increasing BrACG concentrations protect both *full-length* GRP75 and its fragment, reducing the production of secondary cleavages. The densitometric analysis shown in Figure 83C (blue bars) reports BrACG induced proteolysis sheltering of the *full-length* protein.

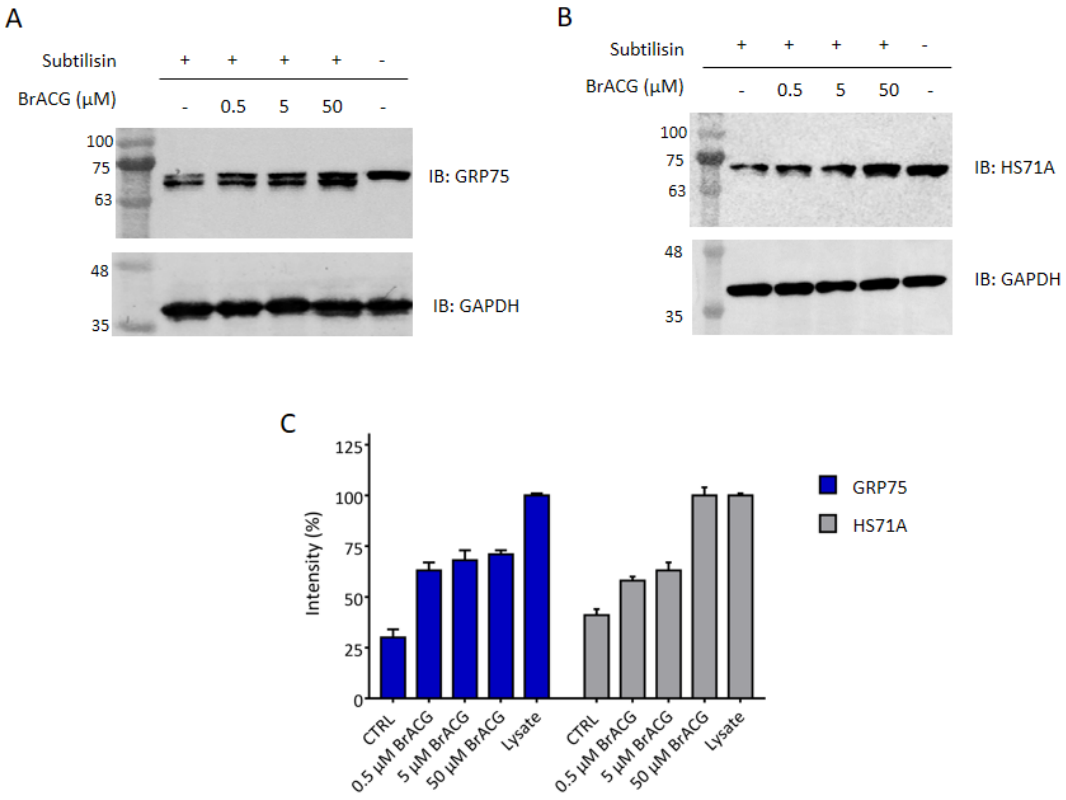


Figure 83: *Western Blot analysis performed with anti-GRP75 (A) and anti-HS71A (B) antibodies showing increasing BrACG concentrations exerting proteolytic protection on both Hsp70 isoforms. GAPDH is used as a loading normalizer. (C) ImageJ performed densitometric analysis of the Western Blots (full-length GRP75 has been quantified). The histogram is the result of the analysis of two independent experiments: pixel intensities are reported as percentages with the related standard deviations, rating the undigested GRP75 and HS71A intensities as 100%.*

5.3 70 kDa heat shock proteins

The balance of protein synthesis, folding, trafficking, assembly and degradation, commonly referred to as *protein homeostasis* or *proteostasis*, is a central process for a correct cellular

functioning. Indeed, cells have developed a number of strategies to control it under physio/pathological conditions.

The ubiquitous and highly conserved molecular chaperones of the 70 kDa heat shock protein (Hsp70) family are key players in protein homeostasis not only during stress-associated conditions, but also in optimal growth circumstances. Members of the Hsp70 family are indeed involved in a wide range of cellular *housekeeping* activities, including the folding of newly synthesized proteins, the translocation of polypeptides into mitochondria and endoplasmic reticulum, the disassembly of protein complexes and the regulation of protein activity. Furthermore, Hsp70s prevent the aggregation and promote the refolding of misfolded denatured proteins, solubilize aggregated proteins and cooperate with cellular degradation machineries (e.g. proteasome and autophagy pathways) to clear aberrant proteins and aggregates[233]. Additionally, Hsp70s are implicated in regulation of the heat shock response[234] and of apoptosis[235].

Eukaryotes are characterized by several Hsp70 isoforms, with 13 homologues being expressed in distinct compartments of human cells. Among them, the most important isoforms are the constitutively and highly expressed cytosolic Hsc70 (HSPC), the cytosolic heat-inducible Hsp70 (HS71A and HS71B), the endoplasmic reticulum resident BiP chaperone (GRP78) and the mitochondrial mortalin (GRP75). Although having some grade of functional redundancy, Hsp70s exhibit a high degree of specialization, accomplished by differences in substrate recognition and allosteric regulation, target site localization and cooperation with distinct co-chaperones and nucleotide exchange factors[236].

Among all of the aforementioned isoforms, apart from the regulation of protein folding and quality control[237] GRP75 plays crucial roles in mitochondrial homeostasis, being fundamental for the iron-sulfur cluster-biogenesis and insertion into Fe-S apoproteins[238] as well as for the mitochondrial protein import[239]. Furthermore, GRP75 regulates mitochondrial ATP levels, membrane potential and permeability and the response to reactive oxygen species[240].

Interestingly, a GRP75 fraction also localizes to the cytosol, where it associates with proteins involved in signaling, apoptosis, or senescence. Significantly, GRP75 binds to p53, negatively regulating this tumor suppressor[241] and also blocking its mediated arrest of centrosome duplication[242].

Therefore, together with the cytosolic Hsp70s whose expression greatly increases in cancer cells, GRP75 plays an anti-apoptotic role in cancers, but its expression levels and function are also altered in several neurodegenerative diseases, characterized by neuronal mitochondrial dysfunction (e.g. Alzheimer's and Parkinson's diseases)[243,244].

Thus, heat inducible and mitochondrial Hsp70s are considered as valuable protein targets in the defeat of several human pathologies.

All the Hsp70 family members share at least two of the following four structural features: an N-terminal 45-kDa nucleotide binding domain (NBD), a 15-kDa substrate binding domain (SBDβ), a 10-kDa helical lid domain (SBDα) and a disordered C-terminal tail (Figure 84A).

The NBD is organized into two lobes (composed in total by the four IA, IB, IIA and IIB subdomains) separated by a deep cleft at whose bottom lies the catalytic center, where ATP binds[245] (Figure 84B). NDB is connected with SBDβ by a conserved flexible linker, essential for the allosteric mechanism coupling ATP hydrolysis in NBD to substrate binding in SBDβ[246,247]. Indeed SBDβ, made of an eight-stranded β-sandwich, contains the polypeptide binding cavity within its central hydrophobic pocket[248,249], fully enclosed by SBDα helical lid in absence of ATP (Figure 84C).

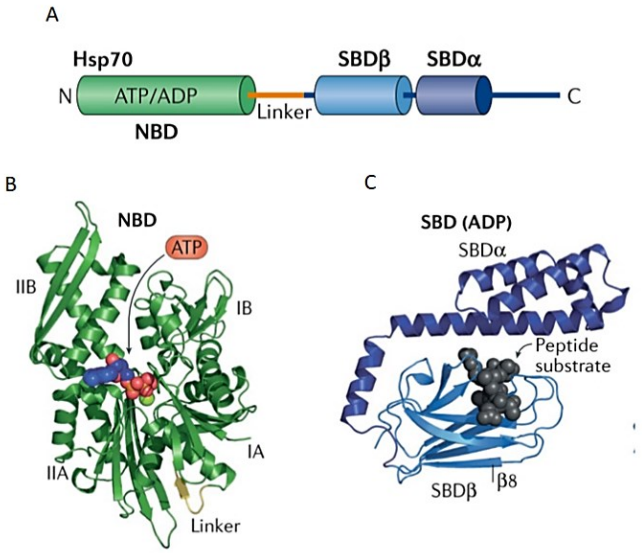


Figure 84: *(A) Hsp70s domain organization. (B) Nucleotide binding domain (NBD) subdomain architecture. (C) Substrate binding domain (SBD) base (β) and lid (α) spatial organization around a substrate. Adapted from Rosenzweig et al.*[236].

NDB and SDB are thus reciprocally controlled by a *bidirectional heterotrophic allostery*, dependent on both the presence of ATP/ADP on the NBD and of a client protein bound to the PBD. Indeed, NBD ATP-bound state leads the PBD to achieve a low-affinity state with client proteins, whereas peptide binding to PBD stimulates a weak NBD ATPase activity, which leads to conformational changes that enhance PDB affinity for client proteins. ADP/ATP exchange in NBD returns the PBD to a low-affinity state for substrates, leading to their release.

More in details, in the ADP-bound state, the SBD binds to peptide substrates with high affinity but with very low association and dissociation rates. ATP binding to the catalytic center weakens SBD/substrate interactions, inducing the NBD lobes to rotate and bind to the interdomain linker, with the consequent SBDβ association (Figure 85). The SBDβ then immobilizers the NBD lobes, leaving the catalytic residues in a conformation unsuitable for ATP hydrolysis[241], which explains the low basal ATPase activity of Hsp70s. Substrate binding

into SBDβ hydrophobic pocket triggers the release of the SBDα/β from the NBD, allowing its back-rotation towards a position suitable to promote ATP hydrolysis.

Furthermore, upon ATP hydrolysis, the dissociated α-helical lid docks onto the substrate binding pocket of the SBDβ, preventing substrate dissociation[250]. The substrate-induced SBDβ release from the NBD though, leads the linker to lose its ability of keeping the NBD lobes in the optimal position for ATP hydrolysis: in this scenario, Hsp70s interaction with the J-domain of a co-chaperone prevents the latter phenomena, thus enabling NDB to transition to the *fully competent* ATP hydrolysis state.

Regarding substrates release, in the nucleotide-free state, NBD and SBD are largely independent, but upon ATP binding SBDα and SBDβ separate from each other, allowing SBDβ to interacts with the NBD. This rearrangement opens up the substrate binding pocket, thus facilitating substrate release (Figure 85).

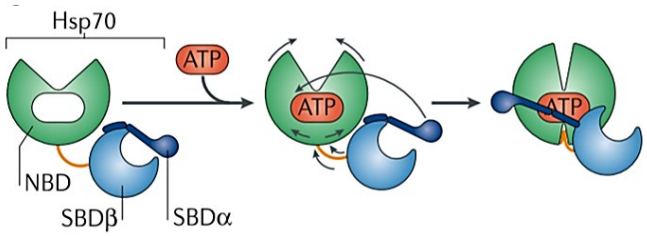


Figure 85: *Schematic representation of ATP-induced Hsp70s conformational changes. When binding to the NBD, ATP leads to rotation of the NBD lobes, insertion of the linker, SBDα detachment from SBDβ and docking of SBDα and SBDβ on the NBD. Adapted from Rosenzweig et al.*[236].

For a better comprehension of the t-LiP-MRM and molecular docking data reported in the following paragraphs, information about HS71A NBD and GRP75 NBD and SBD will be provided in the subsequent sections.

5.3.1 HS71A NBD structural features

HS71A NBD exists in equilibrium between a *closed* and an *open form*, and this equilibrium is likely to be shifted towards the *open form* upon NBD interaction with SBD.

In the closed form, subdomain IIB Glu268 forms a hydrogen bond with subdomain IA Tyr15 and an electrostatic interaction with subdomain IB Lys56 (Figure 86A), whereas a hydrogen bond between Tyr15 and Lys56 is also formed following nucleotide binding (Figure 86B).

In contrast, when nucleotide-free NBD interacts with SBD, the open-form is achieved in which Glu268 is shifted too far away to interact with either Tyr15 or Lys56, while Tyr15 and Lys56 still interact with each other[251] (Figure 86C).

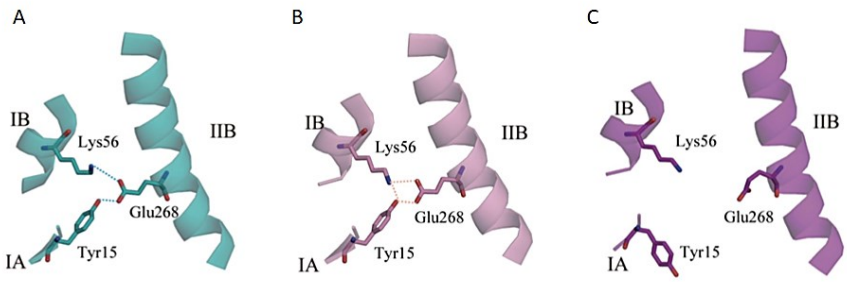


Figure 86: *Glu268, Tyr15 and Lys56 interaction features in (A) nucleotide-free NBD, (B) nucleotide-bound NBD and (C) nucleotide-free NBD-SBD. Adapted from Shida et al.*[251].

Therefore, Tyr15, Lys56 and Glu268 play a key role in controlling the relative position of subdomains IIB, IA and IB and in stabilizing NBD closed form, fixing its binding pocket in a form suitable for ATP binding. When ATP binds to the NBD, its γ-phosphate group strongly interacts with Thr13, Thr14, Gly203 and Thr204 (Figure 87) and NBD closed form becomes stable making, on the other hand, the conformational transition towards the open form unlikely unless NBD binds with SBD[251].

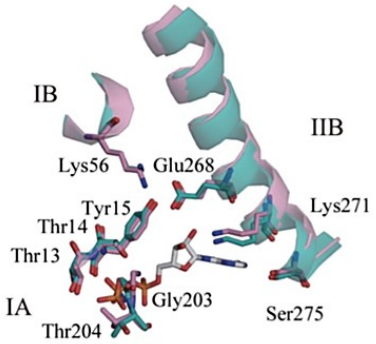


Figure 87*: NBD central part in complex with a non-hydrolysable ATP analog (AMPPNP, white-based stick model). Thr13, Thr14, Tyr15, Gly203, Thr204, Lys271 and Ser275 interact with AMPPNP. Lys56 belongs to subdomain IB and Glu268 belongs to subdomain IIB. The picture has been obtained by superimposing the nucleotide-free NBD (cyan) and the AMPPNP-bound NBD (pink). Adapted from Shida et al.*[251].

Indeed, when SBDα interacts with NBD subdomains IA and IIA, the NBD open form is produced in which subdomain IIA β-sheet rotates clockwise around an α-helix from subdomain IA (Figure 88A and B), helped by their high-content in hydrophobic amino-acid residues. This rotation is limited by Asp199 and Asp206, located next to the hydrophobic surface of subdomain IIA β-sheet. Thus, a correct rotation separates subdomain IIB from subdomains IA and IB making Glu268 unable to interact with Tyr15 and Lys56, respectively. On the other hand, subdomain IIA rotation also poses its Ser340 close enough to subdomain IIB Gly230, allowing the formation of a hydrogen bond that fastens subdomain IIB on subdomain IIA and stabilizes the NBD open form[251] (Figure 88C).

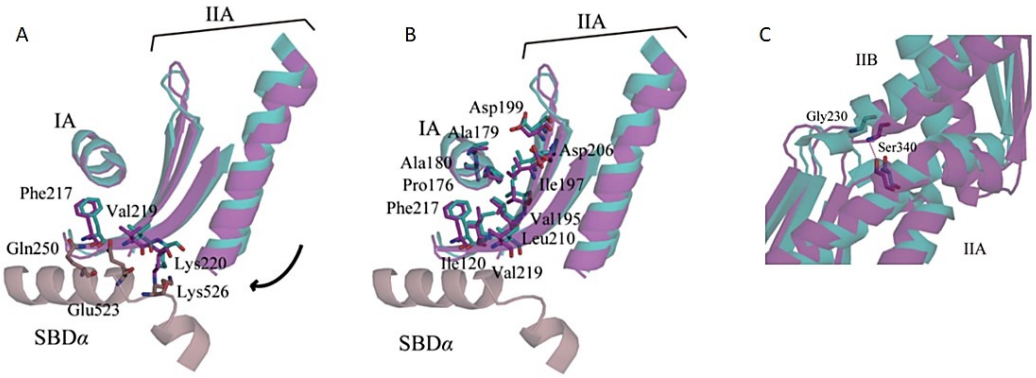


***Figure 88**: (A, B) Superimposition of the nucleotide-free NBD (cyan) and the nucleotide-free NBD-SBD (purple) showing subdomain IIA shifts towards the SBD in the open-form NBD. Panel A highlights Subdomain IIA/SBD interface and the SBD, while B subdomains IA/IIA interface. (C) Superimposition of the nucleotide-free NBD (cyan) and the nucleotide-free NBD-SBD (purple), showing the interaction between Gly230 and Ser340. The interacting residues are shown as stick models. Adapted from Shida et al.*[251].

Taken together, these evidences show how the residues Tyr15, Lys56 and Glu268, fixing the NBD closed form and making it suitable for ATP binding, and how Gly230 and Ser340, interacting between subdomains IIA and IIB to perfectly fix their relative positions in the NBD open form, play a crucial role in HS71A allosteric cycle and activity.

5.3.2 GRP75 NBD structural features

As all the other Hsp70s, GRP75 nucleotide-binding pocket is located at the center of NBD, which is subdivided into four subdomains (i.e. IA, IB, IIA, and IIB, Figure 89) showing conformational flexibility and variable orientations depending on nucleotide occupancy and binding to co-chaperone proteins.

GRP75 NBD structure resembles the closed Hsp70-NBD one, in which subdomains IB and IIB contact each other through a salt bridge between Lys106 and Glu313 (Figure 89, magnification). Furthermore, the aliphatic portions of subdomain IB Val110 and Thr111 interact with Asn302, Met303, Leu305, Gln306, and Arg309[252].

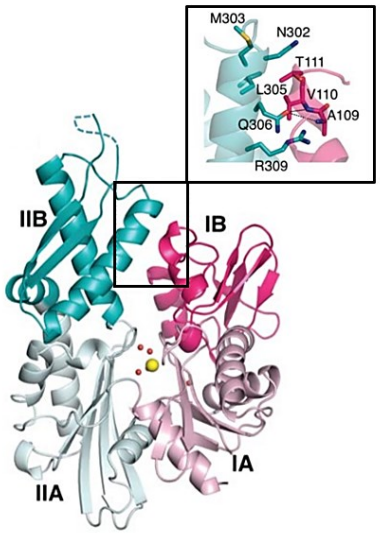


Figure 89: *Overall structure of GRP75 NBD. Black box delimited magnification shows a close-up view of subdomains IB/IIB interactions at the top of the interdomain cleft. Adapted from Amick et al.*[252].

Even though NBD nucleotide-binding pocket is largely conserved between GRP75 and other Hsp70-family members, some exceptions can be still observed: Met389 (Ser in Hsp70), which interacts with the side of the ADP adenine moiety, Cys317 (Arg in Hsp70), which packs against adenine top and forms a polar interaction with its N7 atom and Asn64 (Tyr in Hsp70), which interacts with ADP α-phosphate[252] (Figure 90).

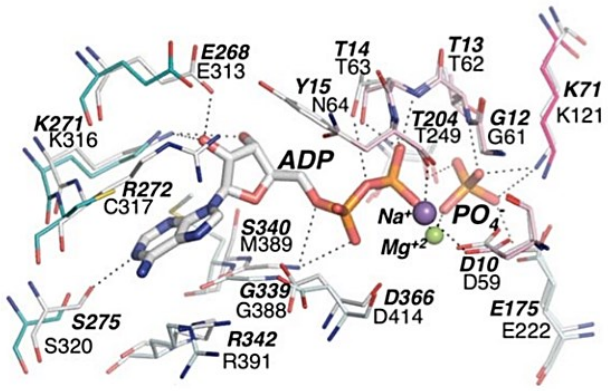


Figure 90: *Comparison of GRP75 and HS71B-ADP-Pi NBD. HS71B residues are colored with white carbons and the residues labeled in bold italics. Plain labels designate GRP75 counterparts. Adapted from Amick et al.*[252].

Apart from interacting with ATP/ADP and with co-chaperones, GRP75 also interacts with the tumor suppressor p53, in a site encompassing residues 253–282[253] (Figure 91A), which define a surface formed by two β-strands and a hairpin turn in subdomain IIA and extend into a helix from subdomain IIB. Furthermore, a recent *in vitro* study suggests that also 260–288 NBD aminoacidic residues might form a binding site for the p53 C-terminal negative regulatory domain[254]. In addition, GRP75/p53 association might also be regulated GRP75 J-domain co-chaperone Tid1[255], interacting with the NBD region encompassed by residues 253–282[256].

Furthermore, the known GRP75 inhibitor MKT-077, which down regulates p53 sequestration, also interacts with the protein NBD in a region mapped by the residues 252–310, overlapping with the putative p53/Tid-1-interacting site[257]. More in details, Amick *et al.* mapped MKT-077 between residues 267–271 and the side chain of Tyr196 (Figure 91B): in this position these molecule might both disrupt p53 docking and, as the Tyr196 containing loop approaches residues 267–271 in GRP75 ATP-bound conformation[258,259], MKT-077 might also destabilize it, indirectly preventing Tid1 association.

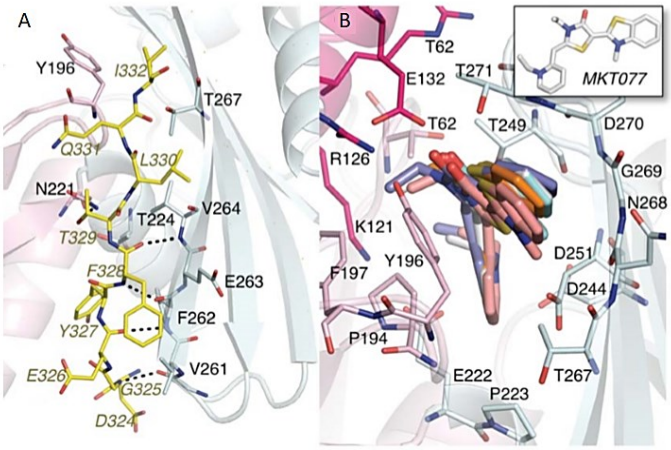


Figure 91: *(A) Interaction between p53 N-terminal domain and the subdomain IIA β-strand. p53 sequence is shown as yellow sticks and labeled in italics. (B) Five high-scoring MKT-077 (thick sticks) poses docked to GRP75-NBD. Selected GRP75 residues within 4Å of MKT-077 are shown and labeled. Adapted from Amick et al.*[252].

Thus, GRP75 NBD results essential not only for the protein allosteric cycle involved in mitochondrial proteostasis, but also for the interaction with the tumor suppressor p53, revealing itself as an interesting site to be targeted by specific inhibitors, preventing p53 sequestration and the GRP75-induced anti-apoptotic effect.

5.3.3 GRP75 SBD structural features

The 18-kDa GRP75 SBD is composed of two sets of four-stranded anti-parallel β-sheets forming a twisted sandwich. Its 10-kDa *substrate lid* that *swallows and seals* substrates in the ADP-bound state is composed of a helix, kinked in the middle and bent upwards[260].

GRP75 SBD preferentially binds to peptides containing internal aliphatic and basic residues, which are buried deep in the α-strands of folded proteins but result exposed if the protein is misfolded.

Remarkably, one of the four SBD β-sheets includes Gly489, a crucial residue for GRP75 interaction with proteins involved in the Iron-Sulfur Cluster (Figure 92). Indeed, GRP75 SBD interacts with Frataxin (i.e. FXN), with the Iron-Sulfur Cluster assembly enzyme ISCU (i.e. ISCU) and with Cysteine desulfurase (i.e. NFS1).

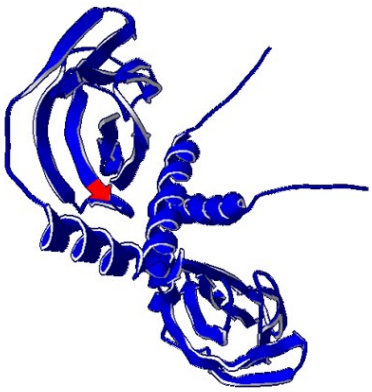


Figure 92: *GRP75 SBD 3D structure visualized through SPDBviewer. The protein backbone is in blue, with Gly489 highlighted in red.*

More in details, ISCU is a scaffold protein essential for the *de novo* synthesis of iron-sulfur (Fe-S) clusters required for the maturation of both mitochondrial and cytoplasmic [2Fe-2S] and [4Fe-4S] proteins. Cluster assembly on ISCU depends on the function of both NFS1, which serves as the sulfur donor for cluster synthesis, and FXN, which acts as the putative iron donor.

After being transiently assembled on the scaffold protein ISCU, the [2Fe-2S] cluster is then released and transferred to the glutaredoxin GLRX5 for the subsequent formation of mitochondrial [2Fe-2S] proteins, synthesis of [4Fe-4S] clusters and their target-specific insertion into the recipient apo-proteins, aided by GRP75.

5.4 Analysis of BrACG/GRP75 and of BrACG/HS71A interaction features through t-LiP-MRM

To shed light on the interactions engaged by BrACG with such central components of the molecular chaperones cellular network, t-LiP-MRM experiments were performed to identify both GRP75 and HS71A regions undergoing structural changes upon the marine metabolite binding.

As ususal, a computational analysis was performed to build-up the MRM method for tracking GRP75 and HS71A tryptic peptides in a complex peptide mixture.

Thus, both Hsp70 isoforms tryptic peptides were retrieved from the *PeptideAtlas* Human build and their best transitions selected through the Complete Human *SRMAtlas* build. The obtained list was then refined so that to each precursor corresponded the three best fragment ions. The *SRMAtlas* exploited search parameters and the criteria used to refine the results are reported in paragraph 2.3.2.1.

Thus, a comprehensive methods listing 78 and 51 transitions, for GRP75 and HS71A respectively, was obtained and subsequently tested onto a HeLa lysate tryptic digest.

More in details, the XICs of all the transitions for each precursor were inspected to evaluate the experimentally observable peptides, assign them the corresponding retention times and subsequently select the best transition for each precursor. This preliminary experiment led to a global MRM method comprising 20 transitions for GRP75 and 11 for HS71A (Figure 93A).

Furthermore, the observed peptides were mapped onto both isoforms sequences through *PAWS*: as shown in Figure 93B and C, for both of them the experimentally observed peptides mapped to all the domains, but the obtained coverage percentages were strictly different (i.e. a satisfactory 40% for GRSP75 and an acceptable 20% for HS71A).

A

Q1_m/z	Q2_m/z	GRP75	HS71A
904.96	902.49	S-[469-485]-K	-
847.93	1097.5	N-[188-202]-R	-
823.44	666.35	V-[219-234]-K	-
777.42	278.15	L-[160-173]-K	-
770.41	305.18	G-[514-535]-K	-
737.4	876.48	E-[542-555]-K	-
725.86	866.4	T-[86-99]-R	-
645.84	1063.55	V-[395-405]-R	-
621.84	758.45	D-[207-218]-R	-
619.64	672.83	V-[579-595]-K	-
616.34	674.38	Q-[635-646]-K	-
562.25	275.17	E-[660-671]-K	-
531.66	269.16	L-[499-513]-R	-
510.73	689.36	D-[626-634]-R	-
497.23	715.36	Y-[128-135]-K	-
489.73	234.15	A-[369-377]-K	-
479.75	746.34	V-[77-85]-R	-
445.22	513.28	E-[176-187]-K	-
421.7	322.19	H-[26-39]-R	-
382.67	261.16	E-[492-498]-K	-
807.91	248.16	-	A-[113-126]-K
744.35	643.31	-	T-[37-49]-R
644.31	989.44	-	N-[540-550]-K
599.35	742.46	-	D-[160-171]-R
592.33	537.31	-	F-[459-469]-R
559.25	272.17	-	A-[221-236]-R
489.28	288.66	-	A-[329-342]-R
451.75	375.22	-	S-[312-319]-K
423.7	201.63	-	A-[527-533]-R
402.73	376.18	-	I-[501-507]-K
344.21	218.15	-	L-[50-56]-K

B

1 MISASRAAAARLVGAAASRGPTAARHQDSWNGLSHEAFRLVSRRDYASEAIKGAVVGIDLGTTNSC 66
67 VAVMEGKQAKVLENAEGARTTPSVVAFTADGERLVGMPAKRQAVTNPNNTFYATKRLIGRRYDDPE 132
133 VQKDIKNVPFKIVRASNGDAWVEAHGKLYSPSQIGAFVLMKMKETAENYLGHTAKNAVITVPAYFND 199
200 SQRQATKDAGQISGLNVLRVINEPTAAALAYGLDKSEDKVIAVYDLGGGTFDISILEIQKGVFEVKSTN 268
269 GDTFLGGEDFDQALLRHIVKEFKRETGVDLTKDNMALQRVREAAEKAKCELSSSVQTDINLPYLTMD 335
336 SSGPKHLNMKLTRAQFEGIVTDLIRRTIAPCQKAMQDAEVSKSDIGEVILVGGMTRMPKVQQTVQDL 402
403 FGRAPSKAVNPDEAVAIGAAIQGGVLAGDVTDVLLLDVTPLSLGIETLGGVFTKLINRNTTIPTKKSQV 472
473 FSTAADGQTQVEIKVCQGEREMAGDNKLLGQFTLIGIPPAPRGVPQIEVTFDIDANGIVHVSAKDKGT 539
540 GREQQIVIQSSGGLSKDDIENMVKNAEKYAEEDRRKKERVEAVNMAEGIIHDTETKMEEFKDQLPAD 606
607 ECNKLKEEISKMRELLARKDSETGENIRQAASSLQQASLKLFEMAYKKMASEREGSGSSSGTGEQKE 672
673 DQKEEKQ 679

C

1 MAKAAAIGIDLGTTYSCVGVFQHGKVEIIANDQGNRTTPSYVAFTDTERLIGDAAKNQVALNPQN 65
66 TVFDAKRLIGRKFGDPVVQSDMKHWPFQVINDGDKPKVQVSYKGETKAFYPEEISSMVLTKMKEI 130
131 AEAYLGYPVTNAVITVPAYFNDSQRQATKDAGVIAGLNVLRIINEPTAAAIAYGLDRTGKGERNVLI 197
198 FDLGGGTFDVSILTIDDGIFEVKATAGDTHLGGEDFDNRLVNHFVEEFKRKHKKDISQNKRAVRRLR 264
265 RRLRTACERAKRTLSSSTQASLEIDSLFEGIDFYTSITRARFEELCSDLFRSTLEPVEKALRDAKLDKAQ 330
331 DLKAQIHDLVLVGGSTRIPKVQKLLQDFFNGRDLNKSINPDEAVAYGAAVQAAILMGDKSENVQD 390
391 LLLLDVAPLSLGLETAGGVMTALIKRNSTIPTKQTQIFTTYSDNQPGVLIQVYEGERAMTKDNNLLGR 458
459 FELSGIPPAPRGVPQIEVTFDIDANGILNVTATDKSTGKANKITITNDKGRLSKEEIERMVQEAEKYKAE 528
529 DEVQRERVSAKNALESYAFNMKSAVEDEGLKGKISEADKKKVLDKCQEVISWLDANTLAEKDEFEHK 595
596 RKELEQVCNPIISGLYQGAGGPGPGGFGAQGPKGGSGSGPTIEEVD 641

Figure 93: *(A) Global MRM method transitions, reported with the corresponding peptides identifiers for both GRP75 and HS71A. GRP75 (B) and HS71A (C) aminoacidic sequences: adjacent tryptic peptides are represented with different colors and the experimentally mapped ones are highlighted in yellow.*

This could seem an unexpected finding: since GRP75 is a mitochondrial protein, less abundant in a total lysate, it should be mapped with a much higher difficulty compared to its cytosolic counterpart. Nevertheless, the observed coverages are not so surprising. First of all, the low sequence homology between GRP75 and HS71A (49.8%) could give different covering responses. Furthermore and more importantly, GRP75 has more proteotypic peptides compared to HS71A, and this accounts for its better mapping. Indeed, the Hsp70 mitochondrial isoform shares less than 50% sequence homology with other UniProt fully annotated isoforms, whereas HS71A shows homology sequence levels higher than 85% with at least other 7 well known Hsp70 isoforms.

Thus, the mapping of both the proteins was considered to be satisfactory and t-LiP-MRM was then performed.

For this purpose, 300 µg aliquots of a HeLa lysate prepared in non-denaturing conditions were incubated with either BrACG (5 µM and 50 µM final concentrations) or DMSO to allow the interaction between the marine metabolite and its targets to occur. The obtained samples were then divided into two aliquots and submitted to limited proteolysis with subtilitin (1:1000 w/w and 1:500 w/w ratios, respectively). Furthermore, 300 µg of the DMSO-treated sample were submitted to a mock proteolysis, to be kept as a positive control.

Subtilisin was then quenched with PMSF and the samples shifted to denaturing conditions for an extensive *in solution* tryptic digestion. The produced peptides mixtures were then desalted and submitted to the LC-MRM-MS analysis, carried out with the previously built MRM method. For a reliable identification of GRP75 and HS71A tryptic peptides protected by BrACG, the LiP experiment was carried out in duplicate.

The obtained MS data were then analyzed as previously reported (paragraph 4.4). Briefly, Hsp70s *LiP peptides* were identified comparing the DMSO-treated samples submitted (i.e. negative controls) or not (i.e. positive control) to limited proteolysis, searching for tryptic peptides responsive to subtilisin in a concentration dependent fashion, by evaluating intensity differences between each positive control peptide and the corresponding negative control one, at a given subtilisin amount (i.e. $\Delta_{Area\_Lys\text{-}500}$ and $\Delta_{Area\_Lys\text{-}1000}$). Thus, all the Hsp70s tryptic peptides whose Δ_{Area} was higher the more subtilisin was exploited were selected for the following data analysis step. 12 peptides were considered as *LiP peptides* for GRP75 and 7 for HS71A.

For a clear data visualization, Figure 94A reports the selected *LiP peptides* with their associated Δ_{Areas} shown as percentages (i.e. Norm_$\Delta_{AreaLys-CTRL}$), calculated as previously reported (paragraph 4.4).

These peptides were examined comparing the DMSO and BrACG-treated samples, evaluating the differential peak area modulation among the treated and untreated samples, at a given subtilisin amount (i.e. $\Delta_{Area\_peptide500}$ and $\Delta_{Area\_peptide1000}$).

To evaluate reliable intensity increases following the marine compound interaction with GRP75 and HS71A, their overall intensity difference between the treated and untreated samples (i.e. Δ_{Area_GRP75} and Δ_{Area_HS71A}) was then calculated, averaging all the $\Delta_{Area\_peptides}$ at each subtilisin amount. A 95% confidence value was then calculated and exploited to construct a confidence interval related to both Δ_{Area_GRP75} and Δ_{Area_HS71A} values.

The 1:1000 w/w subtilisin amount resulted the best condition to appreciate the structural changes induced by BrACG on Hsp70s: all of the peptides showing an area increment exceeding the confidence interval in both the biological replicates were selected. More in details, 3 peptides for GRP75 and 2 for HS71A satisfied the selection criteria, identifying GRP75 and HS71A regions directly or distally involved in the interaction with BrACG.

For a straightforward data visualization, percentages were calculated for all the peptides (i.e. Norm_$\Delta_{AreaBrACG-CTRL}$), as shown in Figure 94B, where BrACG protected *LiP peptides* are reported in red and bold.

A

		Norm Δ_{Area}Lys-CTRL	
GRP75	HS71A	1:500_subtilisin	1:1000_subtilisin
V-[77-85]-R	-	69	46
T-[86-99]-R	-	92	90
L-[160-173]-K	-	86	47
E-[176-187]-K	-	54	35
N-[188-202]-R	-	29	5
D-[207-218]-R	-	98	97
V-[219-234]-K	-	58	22
S-[469-485]-K	-	89	83
V-[579-595]-K	-	45	11
E-[492-498]-K	-	72	70
D-[626-634]-R	-	61	17
Q-[635-646]-K	-	100	87
E-[660-671]-K	-	30	7
-	A-[113-126]-K	9	35
-	N-[540-550]-K	86	87
-	A-[221-236]-R	61	8
-	A-[329-342]-R	57	62
-	S-[312-319]-K	54	63
-	I-[501-507]-K	90	85
-	L-[50-56]-K	5	24

B

		Norm Δ_{Area}BrACG-CTRL	
GRP75	HS71A	5µM-CTRL	50µM-CTRL
V-[77-85]-R	-	33	25
T-[86-99]-R	-	1	2
L-[160-173]-K	-	3	8
E-[176-187]-K	-	7	20
N-[188-202]-R	-	**18**	**47**
D-[207-218]-R	-	0	0
V-[219-234]-K	-	**39**	**60**
S-[469-485]-K	-	3	9
V-[579-595]-K	-	7	19
E-[492-498]-K	-	**45**	**78**
D-[626-634]-R	-	12	30
Q-[635-646]-K	-	0	0
E-[660-671]-K	-	6	7
-	**L-[50-56]-K**	**24**	**28**
-	**A-[113-126]-K**	**6**	**28**
-	A-[221-236]-R	0	11
-	S-[312-319]-K	0	0
-	A-[329-342]-R	0	0
-	I-[501-507]-K	5	3
-	N-[540-550]-K	0	0

Figure 94: *(A) GRP75 and HS71A selected LiP peptides, reported with their Norm_Δ_{Area} values obtained comparing the positive and negative controls. For each of them, the Norm-_Δ_{Area} decreases when less subtilisin has been used for LiP. (B) Selected LiP peptides reported with their Norm_Δ_{Area} values obtained comparing BrACG treated samples with the corresponding negative controls at 1:1000 w/w subtilisin amount. BrACG protected peptides are depicted in bold red. Hsp70s aminoacid numbering is comprehensive of the initial methionine.*

As can be noticed, the major part of the identified protected LiP peptides maps for the proteins highly conserved N-terminal nucleotide-binding domain, essential for the ATPase and chaperone activity of these proteins. Thus, to evaluate the spatial organization of these peptides in the proteins N-termini, we mapped them onto the corresponding GRP75 and HS71A 3D NBD structures: these isoforms domains, in their apo-forms, were retrieved from the Protein Data Bank and visualized through SPDB viewer as shown in Figure 95A and B, respectively.

More in details, GRP75 backbone is depicted in blue, HS71A in gray and BrACG protected peptides are reported in red. As can be observed, the protected peptides are in close proximity in the 3D structures of the NBDs in both Hsp70 isoforms.

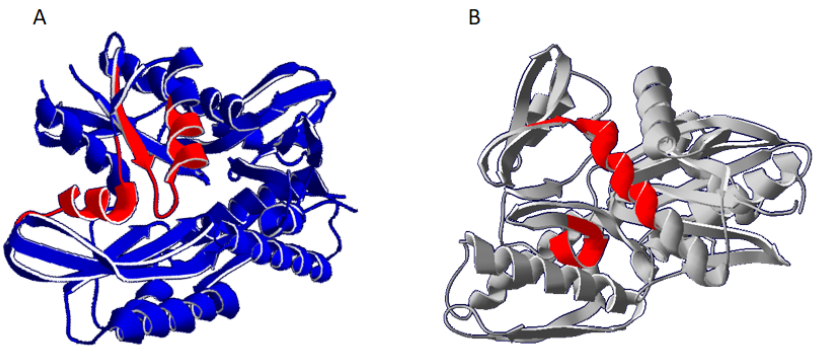


Figure 95: *3D GRP75 (A) and HS71A (B) NBD stick and ribbon representations obtained through SPDB viewer. BrACG protected LiP peptides are highlighted in red.*

Moreover, looking at their positions and comparing them to the previously reported data (paragraphs 5.3.1 and 5.3.2), we noticed that the protected NBD peptides are located in Hsp70s regions involved in interesting intra- and inter-proteins interaction patterns.

Starting with HS71A, t-LiP-MRM identified BrACG protected L-[50-56]-K peptide embeds subdomain IB Lys56. This aminoacid is involved, in the NDB closed form, in an electrostatic interaction with subdomain IIB Glu268 and in a hydrogen bond with subdomain IB Tyr15. Thus Lys56 plays a key role in controlling subdomains IIB, IA and IB relative position and in stabilizing NBD closed form, fixing its binding pocket in an ATP-binding suitable form.

Regarding GRP75, the t-LiP-MRM identified BrACG protected N-[188-202]-R peptide comprising Tyr196, directly involved in the protein NBD interaction with both p53 and the known inhibitor MKT-077 which, as previously stated, reduces GRP75-mediated p53 sequestration and anti-apoptotic behavior. Furthermore, Tyr196 containing loop approaches residues 267–271 in GRP75 ATP-bound conformation, thus its impairment could also destabilize the protein ATP-bound conformation and indirectly prevent the association of its co-chaperone Tid1. Furthermore, the other protected identified peptide V-[219-234]-K includes Asn221 and Thr224, which are also involved in p53 interaction.

On the other hand, V-[219-234]-K also includes Glu222 which is responsible for the interaction with the ADP associated PO_4^-.

Furthermore, in our t-LiP-MRM experiment we also identified the additional SBD protected peptide E-[492-498]-K, lying in close proximity with Gly489, involved in GRP75 interaction with the Iron-Sulfur Cluster involved FXN, ISCU and NFS1 proteins (Figure 96).

Figure 96: *GRP75 SDB 3D structure visualized through SPDBviewer. The protein backbone is in black, Gly489 in orange and the t-LiP-MRM identified protected peptide in red.*

Taken together, these data suggest a picture in which BrACG interaction involves aminoacidic residues crucial for both HS71A and GRP75 allosteric cycle (i.e. HS71A Lys56 and GRP75 Tyr196 and Glu222) and essential, for the latter protein, to make interaction with the tumor suppressor p53 (i.e. Tyr196, Asn221 and Thr224) and with ISC proteins.

5.5 T-LiP-MRM data corroboration: blind molecular docking analysis

Fascinated by the t-LiP-MRM results, we moved to a blind molecular docking analysis, (collaboration with Dr. Matteo Mozzicafreddo, University of Camerino).

Thus, BrACG was designed (including the addition of tautomeric states, partial charges and protonation) and minimized using the Avogadro software, as previously reported[222]. The proteins 3D structures (i.e. both of the isoforms apo-form nucleotide- and substrate-binding domains) were obtained from the Protein Data Bank[221] and prepared incorporating the partial charges, adding polar protons and removing crystal waters and extra co-crystallized ligands. Then, GOLD[261] was performed to achieve the molecular docking using the ChemScore as scoring function and the predicted equilibrium dissociation constant (i.e. $K_{D,pred}$) was calculated. The best complex geometry was rendered using PyMol software and a 2D representation was also created using the PoseView server.

The obtained BrACG/GRP75 and BrACG/HS71A complexes showed dissociation constants in the micromolar or sub-micromolar range (Figure 97).

More in details, BrACG binds GRP75 NBD ($K_{D,pred}$= 5.254 μM) very close to the ADP binding site (the amino acids involved in the interaction are reported in the 2D scheme in Figure 97A), but it shows a higher affinity for the protein SBD ($K_{D,pred}$= 445.6 ± nM, Figure 97B), forming an H-bond with Ser35 (Ser473 of the entire protein) and being thus located in close proximity to the beta strand including Gly51 (Gly489 of the entire protein), involved in the interaction with FXN, ISCU and NFS1.

Regarding HS71A, BrACG binds it in its NBD very close to the ADP binding site ($K_{D,pred}$= 471.9 nM, Figure 97C), as already observed for GRP75 and thus highly expected, given that NDBs are the most conserved domains across all Hsp70s isoforms.

On the other hand, BrACG interaction with HS71A SBD did not show a particularly interesting affinity ($K_{D,pred}$= 27.95 μM), since no particular interaction was enlightened.

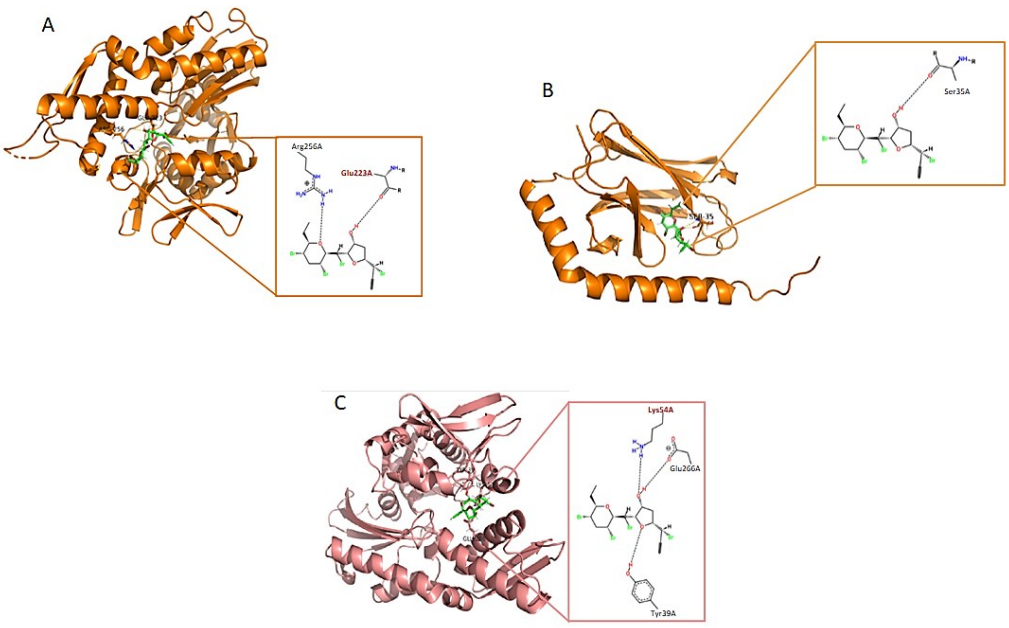


Figure 97: *Molecular docking analysis of (A) BrACG/GRP75 NBD, (B) BrACG/GRP75 SBD and (C) BrACG/HS71A NBD, showing BrACG best interaction poses and the aminoacidic residues involved, highlighted in red if identified by t-LiP-MRM.*

Interestingly the peptides V-[219-234]-K and L-[50-56]-K, embedding molecular docking enlightened aminoacids involved in the best interaction poses engaged by BrACG, respectively resulted as the most GRP75 and HSP71A protected at the lowest BrACG concentration in our t-LiP-MRM experiments. Thus, we sought to report the latter results on the BrACG/GRP75 and BrACG/HS71A complexes obtained from the molecular docking analysis, visualizing them through SPDB viewer.

Figure 98 shows the obtained outputs, where GRP75 and HS71A backbones are represented in orange and pink respectively and BrACG is depicted in green, as for the previously reported molecular docking results (Figure 97). Moreover, the t-LiP-MRM identified BrACG protected peptides are represented in gray and the molecular docking-retrieved aminoacidic residues involved in the interaction with the marine metabolite in red.

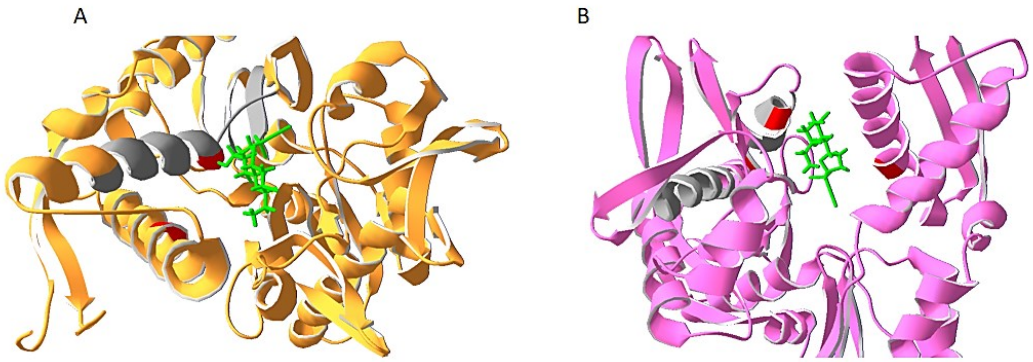


Figure 98: *BrACG/GRP75 (A) and BrACG/HS71A (B) complexes obtained from the molecular docking analysis and visualized through SPDB viewer. T-LiP MRM identified BrACG protected peptides are represented in gray, whereas the molecular docking aminoacids involved in the best BrACG interaction poses are highlighted in red.*

As can be observed, t-LiP-MRM peptides identified as being protected by BrACG include aminoacids involved in the best interaction poses engaged by the molecule with its target proteins, namely Glu-223 for GRP75 (Figure 98A) and Lys54 for HS71A (Figure 98B).

Thus, blind molecular docking and t-LiP-MRM pointed towards the same direction, strongly suggesting that BrACG interacts with both of the Hsp70 isoforms N-terminal ATPase domains, even if showing a higher affinity for HS71A.

Moreover, the LiP data enlightened an additional BrACG interaction site within GRP75 SBD (i.e. peptide E-[492-498]-K), which was also corroborated by the molecular docking calculated higher affinity of the marine metabolite for this GRP75 binding site.

5.6 BrACG: conclusive remarks and future experimental plans

BrACG is a polybrominated acetogenin isolated from the *Mycale rotalis* sponge by Giordano *et al.*[231]. Given the huge range of bioactivities shown by *Mycale* sp. metabolites[230], we sought to shed light on this marine compound activity profile through the characterization of its target protein(s) by our *label-free* functional proteomics platform.

MS-based DARTS experiments led to the identification of both GRP75 and HS71A as the most reliable BrACG cellular partners: these Hsp70 isoforms were indeed the most protected proteins at the lowest molecule amount in both of the DARTS experiments and their protection was directly proportional to BrACG concentrations and mostly reproducible between two replicates.

Western Blotting analyses, performed with both GRP75 and HS71A specific antibodies, successfully validated the MS-obtained data.

T-LiP-MRM experiments were then carried out to characterize BrACG interaction features with both of the proteins. The main protein regions involved in the interaction with BrACG belong to the highly conserved N-terminal nucleotide-binding domain, essential for the ATPase and chaperone activity, as well as involved in interesting intra- and inter-proteins interactions (paragraphs 5.3.1 and 5.3.2). Furhtermore, we also identified BrACG protected a GRP75 SBD peptide, mapping for a region involved in the interaction with Iron-Sulfur Cluster involved proteins. These results were corroborated by a blind molecular docking analysis.

On the basis of the interesting inter- and intra-molecular interaction patterns involving BrACG putatively targeted HS71A and GRP75 regions, and given the key role of these proteins in *proteostasis* during both stress-associated and optimal growth conditions (paragraph 5.3), *in vitro* activity assays are now under investigation to evaluate BrACG effects on both these Hsp70 isoforms.

CHAPTER 6

Conclusions and Future Perspectives.

Natural products (NPs) have long been explored as invaluable sources for drug discovery: their chemical diversity and complexity offers a range of *chemotypes* frequently showing biologically relevant scaffolds and pharmacophore patterns naturally evolved as preferred ligand–protein binding motifs[59,60].

Indeed, most of the therapeutically employed NPs and NP-derived drugs act on protein targets, physically and functionally interconnected in cellular pathways and networks: the one-drug-one-target paradigm is far away from reality; indeed, poly-pharmacology is achieved through molecules capable of acting at multiple cellular levels.

Therefore, it is important to identify a drug target spectrum as thoroughly as possible, to fully exploit its therapeutic potential dictated by the molecule primary target(s) and minimize the toxicity caused by its off-target(s).

In this scenario, the functional proteomics AP-MS approach has been the top-choice technique in the last 15 years, with brilliant results in the targetome profiling of bioactive compounds. Briefly, this strategy relies on a pull-down experiment in which a chemical functionalized NP version is used as a bait to *fish out* its interacting partners from a complex mixture.

Nevertheless, despite being a well consolidated, robust and straightforward approach having brought undeniable results, this strategy is not universally applicable, being limited by the chemical features of the molecule to be characterized for the mandatory *on-beads* immobilization step. This strictly requirement makes AP-MS a non-universally applicable strategy: compounds that do not exhibit properly reactive functional group cannot be taken into analysis, as well as compounds showing a promiscuous and difficult-to-handle reactivity. Furthermore, if few (or no) SAR information on the analyte is available, as frequent for NPs, the modification of a functional group would leave some concerns about the preservation of its original bioactivity. Moreover, AP-MS is not capable of providing information about NPs

covalent *interactors*, nor to shed light on NP/target protein(s) interaction features (e.g. interaction site or protein regions influenced by the molecule binding).

Thus, to overcome AP-MS major drawbacks, a new toolbox of mass-spectrometry based techniques has been established, over the past decade, avoiding any chemical modification of the molecules for a proteome-wide profiling of their *targetome*. These new approaches can thus be virtually exploited to analyze any kind of molecule, regardless of its chemical features, enabling an *in-solution* study of protein/ligand binding through the evaluation of ligand-induced protein conformational changes, which translate into a major stability of the target to several perturbation events.

The molecule-induced conformational changes can be perceived either by globally characterizing the target protein (i.e. protein-centered methods) or by analyzing the peptides deriving from its extensive digestion (i.e. peptide-centered methods).

In this scenario, my PhD work has been focused on the design and optimization of a *label-free* functional proteomics platform, providing a universally applicable and more comprehensive Drug-Target Deconvolution strategy, aimed to the analysis of the targetome profile of a wide variety of molecules, with no limitations posed by their chemical features.

Among all of the *label-free* strategies reported in the Chapter 1 of this thesis, the limited proteolysis based approaches seemed the most appealing for my purpose, endowed with the undeniable advantage of being fast, not requiring any peculiar or expensive reactive to be performed and whose application could lead to satisfactory results on the mass spectrometers disposable in my department.

Furthermore, to characterize several aspects of a molecule targetome, I sought to couple two different limited proteolysis-based approaches having complementary readouts, focusing my work on the protein-centered Drug Affinity Responsive Target Stability (DARTS) and on the

peptide-centered targeted Limited Proteolysis Coupled to Multiple Reaction Monitoring Mass Spectrometry (t-LiP-MRM). Indeed, in my Drug Target Deconvolution strategy plan, I intended DARTS as the strategy to be used when first approaching a bioactive molecule, with the aim of identifying its protein targets, whereas I reserved t-LiP-MRM as a follow-up method for the investigation of molecules/protein partners interaction features, to overcome the classical functional proteomics limitations in providing insights into putative binding site(s) and/or protein region(s) strongly influenced by the molecule binding.

Thus, several preliminary tryouts and optimization steps were necessary to finely tune DARTS and t-LiP-MRM experimental features to ensure a smooth and consistent workflow. This involved the choice and the refinement of the best handling conditions and quenching strategies of the non-specific protease for both DARTS and t-LiP-MRM and the optimization of the chromatographic conditions and the outlining of the best strategy to identify the transitions for each target protein for the latter procedure.

Then, *a proof of concept* seemed mandatory to test the optimized platform, and the radicicol/Hsp90s system seemed ideal since Hsp90s are highly abundant cytosolic proteins, whose detection through DARTS, as well as whose tryptic peptides quantitation through t-LiP-MRM, should not have posed any abundance-related issues. Furthermore, Hsp90s showed a consistent number of tryptic peptides placed on *PeptideAtlas*/*SRMAtlas*, thus their mapping did not seem to pose any particular inconvenience.

DARTS experiments led to the identification of both Hsp90-α and Hsp90-β as the main radicicol targets, as also validated by the bio-orthogonal Western Blotting approach. As for t-LiP-MRM, the individuation of the Hsp90s tryptic peptides (and daughter ions) to be analyzed in the actual experiment led to a much satisfactory 41% coverage and to an even mapping of the proteins alongside all of their domains, obtained analyzing a complex HeLa tryptic digest without any pre-fractionation or enrichment step. Then, t-LiP-MRM experiments gave evidence

of 5 peptides for Hsp90-α and 8 for Hsp90-β whose intensity significantly increased with a radicicol concentration dependent fashion, mostly mapping for the proteins N-terminal ATPase domain, in full agreement with the known radicicol binding site. Among them, the one showing the highest intensity increases (52 folds) embeds Leu48, an aminoacidic residue involved in an H-bond with radicicol. Furthermore, other three peptides mapping for the molecule binding residues were identified, even if with lower intensity increases (2 or 3 folds).

Thus, this *proof of concept* showed the reliability of our optimized proteomics platform, pinpointing the undeniable t-LiP-MRM capability of identifying proteins regions involved in the interaction with a small molecule.

Encouraged from these results, I moved on to study three marine metabolites showing characteristic and distinctive chemical features, to shed light on the wide applicability of this *label-free* platform to the analysis of bioactive compounds, regardless of their structural features.

The first case of study was the norcembranoid 5-epi-sinuleptolide (5-epi-SNEP), whose target profile resulted complex and quite promiscuous, on the basis of the DARTS experiments: in this case the classical AP-MS approach was also performed, to give additional information for refining the small molecule complex interactome.

Indeed, DARTS experiments showed that 5-epi-SNEP exerted a protective effect on subtilisin action on a broad range of molecular weights, giving rise to a wide list of putative interactors, among which cytoplasmic actins resulted as the most protected from proteolysis, but alongside them at least other six actin-related and un-related proteins were identified with lower but still similar protection percentages. On the other hand, AP-MS gave a clearer picture of 5-epi-SNEP interactomic profile, confirming the molecule *fished-out* a complex network of proteins, with the major part of them being related to the cytoskeleton, and with the cytoplasmic

actins as main interactors. This evidence enlightened the importance to couple DARTS and AP-MS approaches in solving complex proteiomics issues.

Actins MRM method *build-up* and mapping led to the highest coverage obtained in my PhD work, with 63% of the proteins sequences being mapped for an even tracking of all actins domains. Subsequently, t-LiP-MRM led to the identification of 4 peptides mapping for actins N-terminal ATPase domain, identified as the region involved in the interaction with the molecule. Since this domain is fundamental for G-actin polymerization into F-actin and some of the identified peptides embedded aminoacids strongly involved in F-actin intermolecular interactions, we moved on to perform *in cell* experiments to evaluate possible effects of 5-epi-SNEP on actin polymerization.

In cell DARTS showed a proteolytic protection on G-actin by both 5-epi-SNEP and the known F-actin depolymerizing agent CyD, whereas confocal microscopy actin staining experiments enlightened that, as CyD, our marine metabolite induced a partial decrease of actin fibers and formation of F-actin amorphous aggregates.

The second studied marine metabolite was Crellastatin A (CreA), a sulfated bis-steroid. As opposite to 5-epi-SNEP, DARTS experiments with CreA led to a straightforward result, being highly reproducible and unambiguously identifying Poly [ADP Ribose] Polymerase 1 as the main molecule target. Indeed, this protein resulted by far the most protected at the lowest CreA amount and its protection was dependent on the molecule concentration and also highly reproducible between the biological replicates, as further confirmed by Western Blotting.

Regarding PARP 1 transitions prediction and mapping, a uniform mapping of the protein throughout all of its domains was achieved and a satisfactory 33% coverage was reached. Considering that PARP 1 is a nuclear protein, and thus low abundant in a total cell lysate compared to Hsp90s and actins, and that no enrichment or prefractionation steps were performed before the LC-MRM analysis, these results were considered much satisfactory, even

if greater quantity of the starting material was required. Thus, this appears to be one of t-LiP-MRM drawbacks: if the target is not fairly abundant, more cell lysate and proteases are needed to perform the protocol or prefractionation of the sample (at the protein or peptide level) and/or enrichment steps might be needed for a reliable quantitation of the target tryptic peptides.

The following t-LiP-MRM experiments performed to analyze CreA/PARP 1 interaction features led to interesting results, pinpointing CreA preferentially targets the protein WGR domain, which bridges between the DNA-binding and the catalytic domains (DBD and CAT), providing essential contacts for the protein autocatalytic activation in response to damaged DNA recognition. Interestingly, the other CreA sheltered peptides were identified in both DBD and CAT, providing us with a further insight in t-LiP-MRM potential: not only did this technique allow the identification of a molecule binding site on its target protein (e.g. the radicicol case, for which the t-LiP-MRM identified protected peptides matched the binding site, known by both crystallographic and computational analysis) as supposed for PARP 1 WGR domain, but it also gives information on the long-range consequences of a molecule binding, mediated by allosteric regulations. Indeed, WGR domain contacts both DBD and CAT and induces conformational changes in the whole PARP 1 which are responsible for its activation.

The t-LiP-MRM results were confirmed by a blind molecular docking analysis, which pointed towards the same CreA/PARP 1 binding site: CreA engaged its best interaction poses with the protein WGR domain, with a predicted dissociation constant of 20.29 ± 7.63 nM.

Driven by the importance of WGR for PARP 1 functions, an enzymatic *in vitro* assay was performed to evaluate whether CreA could reduce PARP 1 activity: as expected, CreA inhibited PARP 1 enzymatic activity, with an estimated IC_{50} of 32.96 μM.

The last molecule I worked on was a poly-brominated acetogenin (BrACG).

As for CreA, BrACG DARTS experiments led to a fast and straightforward identification of the molecule target proteins. Interestingly, both the cytoplasmic and the mitochondrial Hsp70

isoforms (HS71A and GRP75 respectively) resulted as the molecule most reliable cellular partners, as further confirmed by Western Blotting analyses.

When it came to the proteins transitions prediction and mapping, I faced a peculiar case. Indeed, GRP75 coverage was of 40% whereas HS71A one only of 20%. This could seem an unexpected finding: as a mitochondrial protein, GRP75 should be less abundant in a total lysate, and thus mapped with a higher difficulty compared to its cytosolic counterparts. Nevertheless, GRP75 has more proteotypic peptides compared to HS71A, which has no PABST peptide ranked and this phenomenon accounts for the mitochondrial isoform better mapping. Thus, this appeares as another t-LiP-MRM shortcoming: proteins like HS71A, sharing high homology sequence levels with several other proteins, do not have a consistent number of proteotypic peptides to be tracked and their mapping could be difficult to achieve and even require the *in house* development and validation of MRM assays through the analysis of synthetic peptides, if no previous deposited information is available.

Nevertheless, I proceeded with the t-LiP-MRM experiment on both GRP75 and HS71A identifying, respectively, 3 and 2 peptides mapping for these proteins regions directly or distally involved in the interaction with BrACG. Globally, 4 out of 5 peptides mapped for the proteins highly conserved N-terminal nucleotide-binding domain, essential for their ATPase and chaperone activity.

Blind molecular docking analyses performed on both BrACG/GRP75 and BrACG/HS71A complexes showed an optimal correspondence with the t-LiP-MRM results: t-LiP-MRM peptides identified as being protected by BrACG include aminoacids involved in the best interaction poses engaged by the molecule with its target proteins, whereas peptides not embedding molecular docking identified residues are anyway protending in the BrACG interaction pocket.

Thus, once again blind molecular docking and t-LiP-MRM pointed towards the same direction and the t-LiP-MRM data resulted in a reliable picture of BrACG interaction features with its targets.

In conclusion, a *label-free* functional proteomics platform coupling DARTS and t-LiP-MRM, avoiding any chemical modification of the compound to be studied, has been developed and extensively optimized across the three years of my PhD.

Starting from the analysis of the known radicicol/Hsp90s system and proceeding to the *interactome* characterization of three marine metabolites showing different chemical characteristics and whose proteins partners and interaction features were unknown, the platform was applied and implemented, when necessary, based on the quality of the results obtained in each case of study. Several parameters were taken into account, starting from the DARTS output visualization and proteins identification through semi-quantitative mass spectrometric analyses until the t-LiP-MRM methods building and data handling.

DARTS demonstrated to be a fast and straightforward approach for the interactome characterization of virtually any bioactive molecule, regardless of its chemical features, as demonstrated in the 5-epi-SNEP, CreA and BrACG cases and also through the reference compound radicicol. DARTS workflow is indeed easy-flowing and in one day multiple experiments can be performed in parallel. Nevertheless, to disclose as thoroughly as possible a molecule target, it is necessary to analyze several experimental conditions, in which increasing amounts of the bioactive compounds need to be incubated with a cell lysate. Thus, a long time of mass spectrometric analysis might often been required, counterbalanced by the relatively short times needed to perform semi-quantitative bio-informatics analysis.

DARTS has shown to be a fast analytical approach with molecules characterized by small inetractomes (i.e. CreA and BrACG), whereas molecules with a wider interacting profile might

need to be studied with mixed approaches (i.e. the 5-epi-SNEP case) and a higher multiplexing of the DARTS experimental conditions could be necessary, building more experiments with different molecule amounts and subtilisin quantities: a reliably protected protein should be sheltered by subtilisin at several enzyme amounts and in a molecule concentration dependent fashion.

Apart from this drawback, DARTS suffers from the necessity of visualizing, through the low dynamic range SDS-PAGE technique, the proteins sheltered by proteolysis in presence of a given molecule: low abundant proteins and proteins whose molecular weight is lesser than 30 kDa, co-migrating with the peptides generated during the limited proteolysis step, are difficult to visualize and identify.

T-LiP-MRM proved to be a powerful technique to pinpoint protein regions involved in the interaction with a molecule, either identifying putative binding sites or regions distally influenced by the interaction. As for DARTS, t-LiP-MRM workflow is fast, requiring only two days for the samples preparation and less than one day for the preliminary computational analysis. Based on the predicted transitions, the selection of the best one for each identified tryptic peptide could require more or less than one day but, when it comes to the analysis of the actual t-LiP experiments, a quite long mass spectrometric analysis time could be required for a robust identification and quantitation of LiP peptides whose intensity increases in presence of a bioactive compound. Given the huge amount of spectra acquired, the data analysis step is also tricky.

Nevertheless, t-LiP-MRM gave satisfactory results for all of the molecules I worked on, being corroborated by blind molecular docking analyses and giving insights into their mechanisms of action: 5-epi-SNEP depolymerizing effect could be supposed from its actin binding site, as well as CreA inhibitory activity on PARP 1, two hypothesis that were subsequently evaluated through appropriate assays.

The undeniable strength of such an approach surely relies on its ability to pinpoint the protein regions involved in the interaction with a molecule, directly in a complex mixture. Nevertheless, complementary approaches are needed to give additional validation.

When it comes to its shortcoming, the abundance of the protein to be analyzed could be an issue and enrichment or pre-fractionation steps could be necessary, requiring longer experimental times and potentially leading to sample loss. Furthermore, if no validated MRM assay is deposited onto data repositories, alternative prediction strategies must be employed, relying on particular targeted proteomics programs or on an *in house* build-up and validation of proper assays on synthetic peptides. This last eventuality would, of course, increase the time and the costs associated with such an experiment.

Thus, relying on the analyzed molecules, DARTS coupling with t-LiP-MRM successfully provided a fast, straightforward and unexpansive *label-free* platform for the characterization of bioactive compounds interactomic profile. Nevertheless, future analyses could be performed with the intent of optimizing a *gel-free* DARTS approach: this requirement seems to meet the fascinating possibility of shifting LiP from a targeted to an untargeted readout, for the broadest and less biased possible identification of molecules target proteins and of their putative interaction sites within a single experimental procedure.

CHAPTER 7

Experimental Procedures.

7.1 A label-free functional proteomics platform: optimization and coupling of Drug Affinity Responsive Target Stability (DARTS) and targeted Limited Proteolysis-Multiple Reaction Monitoring Mass Spectrometry (t-LiP-MRM).

7.1.1 DARTS strategy optimization

7.1.1.1 Proteome from HeLa cells

The human uterine cervical cancer cells HeLa were obtained from the American Type Culture Collection (ATCC) and grown (37°C, 5% CO_2 atmosphere) in Dulbecco's modified Eagle medium (DMEM, Euroclone, Milan, Italy) supplemented with 10% (vol/vol) fetal bovine serum (Euroclone, Milan, Italy), 100 U/mL penicillin and 100 mg/mL streptomycin (Euroclone, Milan, Italy). Cells were harvested, collected by centrifugation (1000×g, 5 min) and washed three times with phosphate saline buffer (PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4).

Their proteome was extracted by mechanical lysis (Dounce homogenizer), suspending the pellet in ice cooled PBS containing 0.1% Igepal and a cocktail of protease inhibitors (Sigma Aldrich - Merk) and carrying out alternative cycles of friction and rest (4°C). The obtained suspension was submitted to centrifugation (10000 rpm, 4°C for 5 minutes, Centrifuge 5424 R, Eppendorf) to remove the protein solution from the pelleted debris. Proteins concentration of

the cleared lysate supernatant was determined by the spectrophotometric Bradford assay (Bio-Rad, Hercules, CA, USA) and subsequently adjusted to 3 μg/μL with PBS.

7.1.1.2 Limited Proteolysis step optimization

50 μg aliquots of the previously obtained HeLa lysate were treated with subtilisin (enzyme to lysate proteins ratios of 1:2500, 1:1500, 1:1000 and 1:500 w/w respectively) and submitted to limited proteolysis for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). A 50 μg aliquot was kept undigested, submitting it to a mock proteolysis carried out adding H_2O and incubating the sample for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf).

10 μg of the obtained mixtures were then added of Laemmli buffer[28] (60 mM Tris-HCl pH 6.8, 2% SDS; 0.001% bromophenol blue; 1'% glycerol; 2% β-mercaptoethanol) and heated at 95°C for 5 minutes to be subsequently loaded onto a 12% poly-acrilamide gel for a 1D-SDS-PAGE analysis. The gel was then fixed for 15 minutes (fixing solution: 50% H_2O, 40% MeOH, 10% AcOH), washed three times (10 minutes each) with H_2O and then submitted to Coomassie staining (Coomassie G-250, Bio-Rad, Hercules, CA, USA) for 1 hour at room temperature under continuous shaking. The excess dye was removed by extensively washing the gel with H_2O and a scan image of the resulting gel was then obtained through the LabScan software and submitted to a densitometric analysis through ImageJ (molecular weight ranges: 245-135 kDa, 135-100 kDa, 100-75 kDa, 75-63 kDa, 63-48 kDa, 48-35 kDa, 35-25 kDa and 25-17 kDa).

Data were represented, as percentages of the mock-proteolyzed sample, through GraphPad Prism 7.

A subsequent experiment was then performed treating 50 µg HeLa lysate aliquots with subtilisin (enzyme to lysate proteins ratios of 1:1000, 1:500 and 1:100 w/w respectively) and incubating them for either 30 minutes or 1 h at 25°C and 500 rpm (Termomixer, Eppendorf). A 50 µg aliquot was kept undigested, as already reported.

10 µg of the obtained mixtures were then added of Laemmli buffer[28] and heated at 95°C for 5 minutes to be subsequently loaded onto a 12% poly-acrilamide gel for a 1D-SDS-PAGE analysis. The gel was fixed for 15 minutes, washed three times with H_2O and then submitted to Coomassie staining (Coomassie G-250, Bio-Rad, Hercules, CA, USA) for 1 hour at room temperature under continuous shaking. The excess dye was removed by extensively washing the gel with H_2O and a scan image of the resulting gel was then obtained through LabScan and submitted to a densitometric analysis through ImageJ (molecular weight ranges: 245-135 kDa, 135-100 kDa, 100-75 kDa, 75-63 kDa, 63-48 kDa, 48-35 kDa, 35-25 kDa and 25-17 kDa). Data were represented, as percentages of the mock-proteolyzed sample, through GraphPad Prism 7.

7.1.1.3 Subtilisin quenching strategy

50 µg HeLa lysate aliquots were treated with subtilisin (enzyme to lysate proteins ratio of 1:500 w/w) and immediately after with either H_2O (positive control of subtilisin digestion) Laemmli buffer[28], 8 M Urea (7.5 M final concentration) or 8 M Urea (7.5 M final concentration) supplemented with 10x final concentration of a protease inhibitors cocktail (Sigma Aldrich - Merk) and leaved shaking for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). A 50 µg aliquot was kept undigested.

After 30 minutes, 10 µg of the obtained mixtures were added of Laemmli buffer[28] and heated at 95°C for 5 minutes to be subsequently loaded onto a 12% poly-acrilamide gel for a 1D-SDS-PAGE analysis. The gel was then fixed, washed and submitted to Coomassie staining (Coomassie G-250, Bio-Rad, Hercules, CA, USA) for 1 hour at room temperature under continuous shaking. The excess dye was removed by extensively washing the gel with H_2O and a scan image was then obtained through LabScan.

A subsequent experiment was then performed treating 50 µg HeLa lysate aliquots with subtilisin (enzyme to lysate proteins ratios of 1:500) and only one of them with H_2O (mock proteolysis). Immediately after, samples were treated with phenylmetane-sulfonil-fluoride (PMSF, Sigma Aldrich - Merck) at the final concentrations of 0.1 mM, 0.5 mM and 1 mM respectively, and leaved shaking for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). One subtilisin-treated aliquot was not added of PMSF to be kept as a proteolysis control.

After 30 minutes, 10 µg of the obtained mixtures were added of Laemmli buffer[28], heated at 95°C for 5 minutes and loaded onto a 12% poly-acrilamide gel for a 1D-SDS-PAGE analysis. The gel was then fixed, washed and submitted to Coomassie staining (Coomassie G-250, Bio-Rad, Hercules, CA, USA) for 1 hour at room temperature under continuous shaking. The excess dye was removed by extensively washing the gel with H_2O and a scan image was then obtained through LabScan and submitted to a densitometric analysis through ImageJ (molecular weight ranges: 245-135 kDa, 135-100 kDa, 100-75 kDa, 75-63 kDa, 63-48 kDa, 48-35 kDa, 35-25 kDa and 25-17 kDa). Data were represented, as percentages of the mock-proteolyzed sample, through GraphPad Prism 7.

7.1.2 Optimized strategy testing: Radicicol/Hsp90s system

7.1.2.1 DARTS

300 µg aliquots of a 3 µg/µL HeLa cell lysate obtained as previously described were incubated with either 0.8 µM or 8 µM Radicicol final concentrations (Sigma Aldrich) or DMSO for 1 hour at room temperature (final DMSO amount: 1% vol/vol) and under continuous shaking (Bio RS-24 Mini-Rotator, BioSan). The samples were then splitted in 100 µg aliquots and submitted to limited proteolysis with 1:1000 w/w and 1:500 w/w subtilisin to proteins ratios, respectively. Limited proteolysis was carried out for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). 100 µg of the DMSO-treated lysate were submitted to a mock

proteolysis, carried out adding H_2O and incubating the sample as reported for the enzyme-treated ones. Subtilisin was then quenched with 1mM PMSF, incubating the samples for 10 minutes at 25°C and 500 rpm (Termomixer, Eppendorf).

10 μg of each sample were treated with Laemmli buffer[28] and heated at 95°C for 5 minutes. 1D-SDS-PAGE was then performed on a 12% polyacrylamide gel which was then fixed, washed and *Coomassie* stained (Coomassie G-250, Bio-Rad, Hercules, CA, USA). The excess dye was then removed by extensively washing the gel with H_2O.

The scan image of the resulting gel (LabScan) was submitted to a densitometric analysis through ImageJ (molecular weight ranges: 245-180 kDa, 180-135 kDa, 135-90 kDa, 90-63 kDa, 63-50 kDa, 50-40 kDa, 40-35 kDa and 35-25 kDa). Data were represented, as percentages of the mock-proteolyzed sample, through GraphPad Prism 7.

7.1.2.2 Samples preparation for the mass spectrometric analysis: in situ tryptic digestion

Shevchenko *in situ* digestion protocol was performed as described by the authors[34] on the 1:1000 w/w subtilisin treated samples. Briefly, gel bands whose intensity increased in presence of radicicol were excided form all of the lanes by chopping them in smaller pieces, which were then washed by shrinking/swelling cycles using CH_3CN and ammonium bicarbonate (AmBic, 50 mM, pH 8.5), alternatively. Then, disulphide bonds were reduced by treating the gel pieces with 1,4-dithiothreitol (DTT, 6.5 mM in 50 mM AmBic, 60 minutes, 60°C) and the formed thiols were carboxyamidomethylated with iodoacetamide (IAA, 54 mM in 50 mM AmBic, 30

minutes, room temperature, in the dark). Residual reagents were removed by shrinking/swelling cycles and gel pieces rehydrated in a 12 ng/µL trypsin/LysC solution (Promega, Madison, Wisconsin) on ice for 1 h. The enzymes excess was then removed and 40 µL of 50 mM AmBic were added to allow protein digestion to proceed overnight at 37°C (Thermomixer, Eppendorf). The supernatant was then collected and peptides were extracted from the gel slices shrinking them twice with 100% CH_3CN. All of the supernatants were collected and combined to be then dried out under vacuum (SpeedVac Concentrator Plus, Eppendorf) and solubilized in 15 µL of 10% Formic Acid (FA) for the subsequent nano-flow RP-UPLC MS/MS analysis.

7.1.2.3 Nano-flow RP-UPLC MS/MS analysis

The peptide mixtures (5 µL) were injected into a nano-ACQUITY UPLC system (Waters, Milford, MA, USA) and separated on a 1.7-µm BEH C18 column (Waters) at a flow rate of 280 nL/min. Peptides elution was achieved using a linear gradient of B from 20% to 90% over 55 min (solution A: 95% H_2O, 5% CH_3CN, 0.1% acetic acid; solution B: 95% CH_3CN, 5% H_2O, 0.1% acetic acid). MS and MS/MS data were acquired on an LTQ Orbitrap XL high-performance liquid chromatography MS system (Thermo-Scientific, Waltham, MA, USA) equipped with an electrospray source (ESI). The ten most intense doubly and triply charged peptide ions were chosen and fragmented. The resulting MS data were processed by MS Converter General User Interface software (ProteoWizard;

http://proteowizard.sourceforge.net/project.shtml) to generate peak lists for protein identifications.

7.1.2.4 Proteins identification and semi-quantitative analysis

The obtained raw files (.raw) were converted into Mascot Generic Format data files (.mgf) through MSConvert and loaded onto the Mascot Daemon graphical user interface (MatrixScience, London, UK) to achieve proteins identification. The SwissProt database (release January 2017, 553474 sequences, 198069095 residues) was employed to retrieve *in silico* proteins digestion and the following settings were exploited: trypsin as the enzyme, two missed cleavages allowed; carbamidomethyl (C) as fixed modification; oxidation (M) and phosphorylation (ST) as variable modifications; 80 ppm as peptide tolerance; 0.8 Da as MS/MS tolerance.

The obtained data were filtered by molecular weight ranges accordingly with the gel cutting patterns and a semi-quantitative analysis was then performed comparing MASCOT matches among the analyzed compounds. Protection percentages were thus calculated, for each radicicol amount, as follows:

$$\text{Protection (\%)} = [(\text{Matches}_{\text{Radicicol}} - \text{Matches}_{\text{Control}})/\text{Matches}_{\text{Lysate}}]*100.$$

7.1.2.5 Validation of DARTS obtained MS data: Western Blotting analysis

15 μg of the 1:1000 w/w DARTS obtained protein mixtures were treated with Laemmli buffer[28], boiled at 95°C for 5 minutes, resolved on a 12% SDS-PAGE gel and transferred onto a nitrocellulose membrane. The membrane was blocked, for 1 h at room temperature, in a 5% non-fat dried milk containing TBS-t solution (31 mM Tris pH 8, 170 mM NaCl, 3.35 mM KCl, 0.05% Tween 20) and then incubated overnight at 4°C and under continuous shaking with a primary monoclonal antibody raised against Hsp90-α/β (1:1000, Santa Cruz Biotechnology, Inc., Dallas, TX, USA). The antibody excess was then removed, membranes were washed three times with TBS-t and incubated, for 1 h at room temperature and under shaking, with a mouse peroxidase-conjugated secondary antibody (1:2500; Thermo-Scientific). The signal was detected using an enhanced chemiluminescent substrate and LAS 4000 (GE Healthcare, Waukesha, WI, USA) digital imaging system.

Afterwards, the membrane was also hybridized with an anti-Glyceraldehyde 3-Phosphate Dehidrogenase antibody (GAPDH, 1:2500, mouse, Invitrogen) and the signal was detected as already described.

DARTS Western Blottings were repeated twice and a densitometric analysis was also performed through ImageJ. Data were elaborated averaging the values from the independent replicates, using GAPDH as a loading normalizer and rating undigested Hsp90-α/β intensity as 100%. The obtained intensity values were represented through GraphPad Prism 7.

7.1.3 T-LiP-MRM strategy optimization

7.1.3.1 Bovine Serum Albumin (BSA) tryptic digest preparation

1 mg of BSA was dissolved in 8 M urea/50 mM AmBic and submitted to an *in solution* digestion protocol. Briefly, disulphide bonds were reduced with 10 mM DTT (in 50 mM AmBic) for 1h at 25°C and 800 rpm (Thermomixer, Eppendorf) and then alkylated with 20 mM IAA (in 50 mM AmBic) for 30 minutes, at 25°C and 800 rpm, in the dark (Thermomixer, Eppendorf). IAA was then quenched with 10 mM DTT (10 minutes, 25°C, 800 rpm) and urea was diluted up to 1 M with 50 mM AmBic before adding the trypsin/LysC solution (Promega, Madison, Wisconsin) at the enzyme to BSA ratio of 1:100 w/w. Digestion was allowed to proceed overnight at 37°C under continuous shaking (800 rpm, Thermomixer, Eppendorf) and then quenched adding FA to lower the pH to 3. The peptides mixture was then dried under vacuum (SpeedVac Concentrator Plus, Eppendorf), dissolved in 1 mL 5% FA and desalted through a Sep-Pak C18 1 cc (50 mg) cartridge (Waters, Milford, MA, USA).

Briefly, the cartridge was activated flushing 3 mL of 100% CH_3CN and then conditioned with 3 mL of 0.1% FA. The sample was then loaded, desalted flushing the cartridge with 3 mL of 0.1% FA and finally eluted flushing two times 500 µL of 80% CH_3CN, 20% H_2O, 0,1% FA. For the subsequent MS analysis, the peptide mixture was dried under vacuum (SpeedVac Concentrator Plus, Eppendorf) and re-dissolved in 10% FA.

7.1.3.2 Nano-flow RP-UPLC MS/MS analysis

35 ng of the BSA tryptic digest were injected into a nano-ACQUITY UPLC system (Waters, Milford, MA, USA) and separated on a 1.7-µm BEH C18 column (Waters) at a flow rate of 280 nL/min. Peptides elution was achieved using a linear gradient of B from 20% to 90% over 55 min (solution A: 95% H_2O, 5% CH_3CN, 0.1% acetic acid; solution B: 95% CH_3CN, 5% H_2O, 0.1% acetic acid). MS and MS/MS data were acquired on an LTQ Orbitrap XL high-performance liquid chromatography MS system (Thermo-Scientific, Waltham, MA, USA) equipped with an electrospray source (ESI). The six most intense doubly and triply charged peptide ions were chosen and fragmented. The obtained raw file (.raw) was inspected through Xcalibur and analyzed through MASCOT, as previously reported. BSA was identified with a Score of 11946, 571 matches, an emPAI of 10.41 and a 61% coverage. The DDA retrieved six most intense BSA precursor ions (Q1_m/z), as well as their three most intense daughter ions of the y type (Q3_m/z), were selected for the further UPLC-MRM-MS analysis.

7.1.3.3 UPLC-MRM-MS analysis

UPLC–ESI-MRM-MS analyses were performed on a 6500 Q-TRAP from AB Sciex equipped with Shimadzu LC-20A and Auto Sampler systems. Q-TRAP 6500 was operated in positive MRM scanning mode, with declustering potential (DP) set at 80V, entrance potential (EP) at 10V, collision energy (CE) at 35V and cell exit potential (CXP) at 22V.

First UPLC separations were performed injecting 500 ng of BSA on a Kinetex 5 μm C18 100 Å column (50 × 2.10 mm, Phenomenex, Torrance, USA) at a flow rate of 400 μL/min using 0.1% FA in H_2O (A) and 0.1% FA in CH_3CN (B) as mobile phases, and a linear gradient from 5 to 95% of B over 15 min. The best transition for each BSA peptide was selected as the one giving the most intense signal out of the three reported in the method and with the lower noise. The selected transitions were exploited for the subsequent analysis, performed injecting 100 ng, 200 ng and 500 ng of BSA digest.

Subsequently, 100 ng of BSA digest were injected on an Aeris Widepore XB C18 column (150 × 2.10 mm, 3.6 μm XB, Phenomenex, Torrance, USA), using 0.1% FA in H_2O (A) and 0.1% FA in CH_3CN (B) as mobile phases, and a linear gradient from 5 to 95% of B over both 30 min (flow rate: 200 μL/min) and 1 hour (flow rate: 100 μL/min).

For all of the previously reported experimental conditions, the area of each BSA tryptic peptide peak was measured using the Analyst Software from AB Sciex and intensity gains were calculated as ratios over the values recorded for the 100 ng BSA analyzed at 400 μL/min.

7.1.3.4 Background effect evaluation: sample preparation

A HeLa cell lysate was obtained as described before. Proteins concentration was determined by the spectrophotometric Bradford assay (Bio-Rad, Hercules, CA, USA) and subsequently adjusted to 3 μg/μL with PBS.

300 µg lysate aliquots were then spiked with BSA in the 1:500 w/w and 1:100 w/w BSA to proteins ratios. The obtained samples were shifted to denaturing conditions adding urea (4 M final concentration) to perform *in solution* digestion and desalting, as previously described. The obtained peptides mixtures where dissolved in 10% FA for the following UPLC-MRM-MS analysis.

7.1.3.5 UPLC-MRM-MS analysis

UPLC–ESI-MRM-MS analyses were performed on a 6500 Q-TRAP from AB Sciex equipped with Shimadzu LC-20A and Auto Sampler systems. Q-TRAP 6500 was operated in positive MRM scanning mode, with declustering potential (DP) set at 80V, entrance potential (EP) at 10V, collision energy (CE) at 35V and cell exit potential (CXP) at 22V.

100 ng of pure BSA tryptic digest were analyzed alongside HeLa tryptic digests amounts corresponding to 100 ng of spiked BSA.

Chromatographic separation was achieved injecting the samples onto an Aeris Widepore XB C18 column (150 × 2.10 mm, 3.6 µm XB, Phenomenex, Torrance, USA), using 0.1% FA in H_2O (A) and 0.1% FA in CH_3CN (B) as mobile phases, and a linear gradient from 5 to 95% of B over 30 min (flow rate: 200 µL/min).

The area of each BSA tryptic peptide peak was measured using the Analyst Software from AB Sciex and an overall BSA intensity was then calculated for all of the samples by summing all the BSA tryptic peptides retrieved areas.

7.1.4 t-LiP-MRM testing: the Radicicol/Hsp90s system

7.1.4.1 Hsp90-α/β transitions computational prediction

Hsp90-α (UniProt accession number P07900) and Hsp90-β (UniProt accession number P08238) tryptic peptides previously detected by MS where selected through the proteomics data resource *Peptide Atlas* on its Human build (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas): PABST peptides were selected for Hsp90-α and *Predicted Highly Observable Peptides* for Hsp90-β. All the retrieved peptides were queried into the complete Human *SRM Atlas* build (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/GetTransitions), to retrieve their best daughter ions.

The *SRMAtlas* query parameters were set as follows:

- number of highest intensity fragment ions to keep: 8;
- target instrument: QTRAP 5500;
- transitions source: QTOF, Agilent QQQ, Qtrap5500, Ion Trap, Predicted;
- precursor exclusion range: kept blank;
- search proteins form: SwissProt;
- duplicate peptides: unique in results;
- heavy label: kept blank;
- labeled transitions: kept as default;

- maximum m/z: 1100 Da;
- minimum m/z: kept blank;
- allowed ions types: b-ions and y-ions;
- allowed peptides modification: carbamidomethylation of cysteines (C[160]).

The obtained list, containing 52 precursors and 416 fragments, was subsequently refined as follows:

- peptides whose following C-terminus aminoacid was either K or R were removed;
- among the 8 fragment ions reported for each precursor, only the three most intense were selected preferring, when possible, the y-series ions over the b-series ones.

Thus, a comprehensive method listing 52 Hsp90-α/β peptides and their three best transitions was obtained and subsequently tested onto a HeLa lysate tryptic digest.

7.1.4.2 Hsp90-α/β best transitions selection

300 µg of HeLa cell lysate were submitted to an *in solution* digestion protocol. Briefly, proteins were denatured using 8 M urea/50 mM AmBic (4 M final urea concentration), disulphide bonds were reduced with 10 mM DTT for 1h at 25°C and 800 rpm (Thermomixer, Eppendorf) and then alkylated with 20 mM IAA for 30 minutes, at 25°C and 800 rpm, in the dark (Thermomixer, Eppendorf). IAA was then quenched with 10 mM DTT (10 minutes, 25°C, 800 rpm) and urea was diluted up to 1 M with 50 mM AmBic before adding the trypsin/LysC

solution (Promega, Madison, Wisconsin) at the enzyme to proteins ratio of 1:100 w/w. Digestion was allowed to proceed overnight at 37°C under continuous shaking (800 rpm, Thermomixer, Eppendorf) and then quenched adding FA to lower the pH to 3. The peptides mixture was then dried under vacuum (SpeedVac Concentrator Plus, Eppendorf), dissolved in 1 mL 5% FA and desalted through a Sep-Pak C18 1 cc (50 mg) cartridge (Waters, Milford, MA, USA).

More in details, the cartridge was activated flushing 3 mL of 100% CH_3CN and then conditioned with 3 mL of 0.1% FA. The sample was then loaded, desalted flushing the cartridge with 3 mL of 0.1% FA and finally eluted flushing two times 500 µL of 80% CH_3CN, 20% H_2O, 0,1% FA. For the subsequent MS analysis, the peptides mixture was dried under vacuum (SpeedVac Concentrator Plus, Eppendorf) and re-dissolved in 10% FA.

UPLC–ESI-MRM-MS analyses were performed on a 6500 Q-TRAP from AB Sciex equipped with Shimadzu LC-20A and Auto Sampler systems. UPLC separation was performed on an Aeris Widepore XB C18 column (150 × 2.10 mm, 3.6 µm XB, Phenomenex, Torrance, USA), using 0.1% FA in H_2O (A) and 0.1% FA in CH_3CN (B) as mobile phases, and a linear gradient from 5 to 95% of B over 30 min (flow rate: 200 µL/min). Q-TRAP 6500 was operated in positive MRM scanning mode, with declustering potential (DP) set at 80V, entrance potential (EP) at 10V, collision energy (CE) at 35V and cell exit potential (CXP) at 22V.

UPLC–ESI-MRM/MS runs were performed injecting 15 µg of the peptide mixture: the XICs of all the transitions of each precursor were inspected to (1) identify Hsp90-α/β tryptic peptides which could actually be experimentally observed, (2) assign the retention time to all of the

peptides and (3) identify the best transition as the one showing the most intense peak and the best signal to noise ratio.

This preliminary experiment led to a global MRM method comprising 42 transitions.

The observed peptides were then mapped onto both of the Hsp90 isoforms UniProt-retrieved sequences through the PAWS program, with the following parameters:

- cleave: K/R;
- mass: average;
- fixed modification: cysteins carbamidomethylation;
- search tolerance: 300 ppm.

Both of the isoforms coverage (%) was also calculated as follows:

Coverage (%) = (Experimentally mapped aminoacids/total aminoacids)*100.

7.1.4.3 T-LiP-MRM experiment

300 µg of HeLa cells proteome, extracted as previously reported, were incubated with DMSO or Radicicol (0.8 µM and 8 µM final concentrations), for 1 hour at room temperature and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then splitted and submitted to limited proteolysis with 1:2500 w/w, 1:1000 w/w and 1:500 w/w subtilisin to proteins ratios, respectively. Limited proteolysis was carried out for 30 minutes at

25°C and 500 rpm (Termomixer, Eppendorf). 300 μg of a DMSO-treated lysate aliquot were submitted to a mock proteolysis, carried out adding H_2O instead of the enzyme.

Subtilisin was then quenched with PMSF (1 mM final concentration) and the samples shifted to denaturing condition adding urea (4 M final concentration) to perform *in solution* digestion and desalting, as previously described.

15 ug of each sample were then injected in the LC–ESI-MRM-MS system as already described and analyzed through the previously optimized 42-transitions MRM method. The area of each Hsp90-α/β isoform tryptic peptide peak was then measured using the Analyst Software from AB Sciex.

Each sample was analyzed in triplicate and GraphPad Prism 7 was exploited for the data analysis step to calculate fold changes and the corresponding p-values.

Radicicol protected Hsp90s LiP peptides (i.e. fold change ≥ 2 at the lowest radicicol concentration and $p < 0.05$) were mapped onto the following PDB structures: Hsp90-α pdbID 2XAB[262] and Hsp90-β pdbID 1UYM[263].

7.2 Joining forces: a label-free functional proteomics platform paired with conventional affinity chromatography discloses actin proteins as main 5-epi-Sinuleptolide biological targets.

7.2.1 Preliminary quality control: 5-epi-SNEP HPLC-MS analysis

5-epi-Sinuleptolide (5-epi-SNEP) was isolated from the Soft Coral *Sinularia leptoclados* and characterized as reported by Bowden *et al.*[264]. Its relative stereochemistry, derived by NMR spectroscopic data, was subsequently confirmed by X-ray diffraction analysis, as reported by Turner *et al*[265].

The purity of the gifted compound was also verified through reverse phase HPLC-UV (RP-HPLC-UV) and mass spectrometry, as follows.

5-epi-SNEP was dissolved in DMSO and 3 μg were injected in a 1100 Series Chromatographer (Agilent, Santa Clara, California, USA), equipped with a UV detector set at 220 nm. HPLC runs were carried out on a C18 column (Luna Omega 5 μm Polar C18 150× 2.1mm, Phenomenex, Torrance, CA, USA) at a flow rate of 0.200 mL/min. Elution was achieved by means of a linear gradient of B from 10% to 95% over 20 min (solution A: H_2O and 0.1% trifluoroacetic acid; solution B: 95% CH_3CN, 5% H_2O and 0.07% trifluoroacetic acid).

The HPLC peak at 16.105 minutes (Figure 99A) was analyzed in positive ionization mode on a LTQ-Orbitrap-XL mass spectrometer from Thermo, giving a base peak with a m/z value of 371.09 Da, corresponding to the sodiated 5-epi-SNEP ion (Figure 99B).

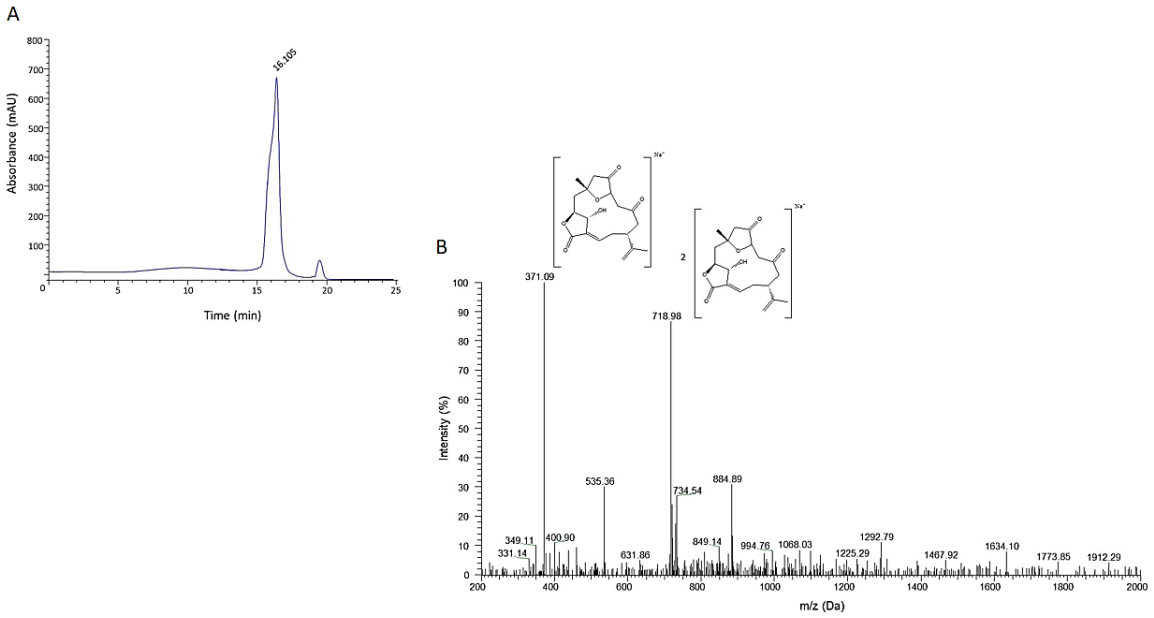


Figure 99: *5-epi-SNEP purity check through PR-HPLC-UV showing a single peak at 16.105 minutes. (B) Mass spectrometric analysis of the 16.105 minutes chromatographic peak.*

7.2.2 DARTS: subtilisin titration experiment

300 µg aliquots of HeLa lysate (obtained as described in pharagraph 7.1.1.1) were incubated with either 5-epi-SNEP (5 µM or 50 µM final concentrations) or DMSO (final DMSO amount:

1% vol/vol), for 1 hour at room temperature and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then splitted in 100 μg aliquots and submitted to limited proteolysis with 1:1000 w/w and 1:500 w/w subtilisin to proteins ratios, respectively. Limited proteolysis was carried out for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). 100 μg of the DMSO-treated lysate were submitted to a mock proteolysis, carried out adding H_2O and incubating the sample as reported for the enzyme-treated ones.

Subtilisin was then quenched with phenylmetane sulfonil fluoride (PMSF, 1 mM final concentration), incubating the samples for 10 minutes at 25°C and 500 rpm.

10 μg of the obtained mixtures were then added of Laemmli buffer[28] (60 mM Tris-HCl pH 6.8, 2% SDS; 0.001% bromophenol blue; 1'% glycerol; 2% β-mercaptoethanol) and heated at 95°C for 5 minutes to be subsequently loaded onto a 12% poly-acrilamide gel for a 1D-SDS-PAGE analysis. The gel was then fixed for 15 minutes (fixing solution: 50% H_2O, 40% MeOH, 10% AcOH), washed three times (10 minutes each) with H_2O and then submitted to Coomassie staining (Coomassie G-250, Bio-Rad, Hercules, CA, USA) for 1 hour at room temperature under continuous shaking. The excess dye was removed by extensively washing the gel with H_2O and a scan image of the resulting gel was then obtained through LabScan.

7.2.3 DARTS: complete experiment

300 μg aliquots of a HeLa cell lysate obtained as previously described (pharagraph 7.1.1.1) were incubated with either 5-epi-SNEP (0.5 μM, 5 μM and 50 μM final concentrations) or

DMSO for 1 hour at room temperature (final DMSO amount: 1% vol/vol) and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then submitted to limited proteolysis, as above reported, with 1:500 w/w subtilisin to proteins ratio. A cell lysate aliquot underwent mock proteolysis. Subtilisin was then quenched with 1 mM PMSF and 10 μg of each sample were treated with Laemmli buffer and heated at 95°C for 5 minutes. 1D-SDS-PAGE was then performed and the obtained gel fixed, washed and Coomassie stained (Coomassie G-250, Bio-Rad, Hercules, CA, USA). The excess dye was then removed by extensively washing the gel with H_2O.

The scan image of the resulting gel (LabScan) was submitted to a densitometric analysis through ImageJ (molecular weight ranges: 245-180 kDa, 180-135 kDa, 135-100 kDa, 100-63 kDa, 63-50 kDa, 50-40 kDa, 40-35 kDa and 35-20 kDa). Data were represented, as percentages of the mock-proteolyzed sample, through GraphPad Prism 7.

7.2.4 DARTS: samples preparation for the mass spectrometric analysis: in situ tryptic digestion

Shevchenko *in situ* digestion protocol was performed as described by the authors[34]. Briefly, gel bands whose intensity increased in presence of 0.5 μM 5-epi-SNEP were excided form the negative control, the 0.5 μM 5-epi-SNEP and the positive control lanes by chopping them in smaller pieces, which were then washed by shrinking/swelling cycles using CH_3CN and ammonium bicarbonate (AmBic, 50 mM, pH 8.5), alternatively. Then, disulphide bonds were

reduced by treating the gel pieces with 1,4-dithiothreitol (DTT, 6.5 mM in 50 mM AmBic, 60 minutes, 60°C) and the formed thiols were carboxyamidomethylated with iodoacetamide (IAA, 54 mM in 50 mM AmBic, 30 minutes, room temperature, in the dark). Residual reagents were removed by shrinking/swelling cycles and gel pieces rehydrated in a 12 ng/µL trypsin/LysC solution (Promega, Madison, Wisconsin) on ice for 1 h. The enzymes excess was then removed and 40 µL of 50 mM AmBic were added to allow protein digestion to proceed overnight at 37°C (Thermomixer, Eppendorf). The supernatant was then collected and peptides were extracted from the gel slices shrinking them twice with 100% CH_3CN. All of the supernatants were collected and combined to be then dried out under vacuum (SpeedVac Concentrator Plus, Eppendorf) and solubilized in 12 µL of 10% Formic Acid (FA) for the subsequent nano-flow RP-UPLC MS/MS analysis.

7.2.5 DARTS: nano-flow RP-UPLC MS/MS analysis

The peptides mixtures (5 µL) were injected into a nano-ACQUITY UPLC system (Waters, Milford, MA, USA) and separated on a 1.7-µm BEH C18 column (Waters) at a flow rate of 280 nL/min. Peptides elution was achieved using a linear gradient of B from 20% to 90% over 55 min (solution A: 95% H_2O, 5% CH_3CN, 0.1% acetic acid; solution B: 95% CH_3CN, 5% H_2O, 0.1% acetic acid). MS and MS/MS data were acquired on an LTQ Orbitrap XL high-performance liquid chromatography MS system (Thermo-Scientific, Waltham, MA, USA) equipped with an electrospray source (ESI). The ten most intense doubly and triply charged

peptide ions were chosen and fragmented. The resulting MS data were processed by MS Converter General User Interface software (ProteoWizard; http://proteowizard.sourceforge.net/project.shtml) to generate peak lists for protein identifications.

7.2.6 DARTS: proteins identification and semi-quantitative analysis

The obtained raw files (.raw) were converted into Mascot Generic Format data files (.mgf) through MSConvert and loaded onto the Mascot Daemon graphical user interface (MatrixScience, London, UK) to achieve proteins identification. The SwissProt database (release January 2017, 553474 sequences, 198069095 residues) was employed to retrieve *in silico* proteins digestion and the following settings were exploited: trypsin as the enzyme, two missed cleavages allowed; carbamidomethyl (C) as fixed modification; oxidation (M) and phosphorylation (ST) as variable modifications; 80 ppm as peptide tolerance; 0.8 Da as MS/MS tolerance.

The obtained data were filtered by molecular weight ranges accordingly with the gel cutting patterns and a semi-quantitative analysis was then performed comparing MASCOT matches among the analyzed compounds. Protection percentages were thus calculated as follows:

$$\text{Protection (\%)} = [(\text{Matches}_{\text{5-epi-SNEP}} - \text{Matches}_{\text{Control}})/\text{Matches}_{\text{Lysate}}]*100.$$

7.2.7 AP-MS: Analysis of 5-epi-SNEP alcohol function reactivity

1 mg of 5-epi-SNEP was incubated with glycidol (250 w:w excess over 5-epi-SNEP) in a buffer made of CH_3CN/50 mM sodium bicarbonate ($NaHCO_3$, 50:50 vol/vol) 1% (vol/vol) triethylamine (TEA). The reaction was carried out for 1 h at 37°C and 500 rpm (Thermomixer, Eppendrf). The obtained products were purified by RP-HPLC on a 1100 Series Chromatographer (Agilent, Santa Clara, CA, USA) equipped with a UV detector set at 220 nm. HPLC runs were carried out on a C18 column (Luna Omega 5 µm Polar C18 150× 2.1 mm, Phenomenex, Torrance, CA, USA) at a flow rate of 0.200 mL/min. Elution was achieved by means of a linear gradient of B from 10% to 95% over 20 min (solution A: H_2O and 0.1% trifluoroacetic acid; solution B: 95% CH_3CN, 5% H_2O and 0.07% trifluoroacetic acid).

The eluted peak was analyzed through HR-MS and HR-MS/MS on an LTQ Orbitrap XL mass spectrometer (Thermo-Scientific).

7.2.9 AP-MS: generation of 5-epi-SNEP modified functional matrix

Epoxy-activated Sepharose™ 6B matrix was swollen with H_2O (200 µL/mg) for 60 min and extensively washed with H_2O. 5-epi-SNEP (1 µmol) was diluted in 150 µL of CH_3CN/100 mM $NaHCO_3$ (50:50 vol/vol) 1% TEA and added to 100 µL of matrix at room temperature for 24 h under continuous shaking (Bio RS-24 Mini-Rotator, BioSan). A control matrix was obtained,

under the same experimental conditions, without the metabolite. The 5-epi-SNEP immobilization reaction was analyzed by RP-HPLC-UV, as reported in the previous paragraph.

In parallel, evaluation of 5-epi-SNEP stability in CH_3CN/100 mM $NaHCO_3$ 1% TEA was also perfumed, diluting 5-epi-SNEP in this buffer to achieve the same metabolite concentration as in the resin coupling buffer. Iso-volumetric aliquots of the obtained solution were analyzed through RP-HPLC-UV.

The compound stability was evaluated by integrating 5-epi-SNEP chromatographic peaks at t = 0 and 24 h.

After 24 h (immobilization yield of 66%), the unbounded 5-epi-SNEP was removed from the functionalized beads, as well as the coupling buffer from the control ones. Both of the resins were washed two times with isopropanol (iPrOH)/H_2O (2:1 vol/vol) and the functionalized resin washes analyzed through RP-HPLC-UV to evaluate the rate of 5-epi-SNEP unspecific binding. Afterwards, both of the matrices were incubated with 150 µL of the same iPrOH/H_2O solution for 3 h at room temperature and under continuous shaking (Bio RS-24 Mini-Rotator, BioSan) to inactivate the free epoxy groups. Then, matrices were washed extensively with PBS to remove iPrOH traces and to be conditioned for the affinity chromatography step.

7.2.8 AP-MS: affinity Chromatography

HeLa cells lysates were obtained as previously described (pharagraph 7.1.1.1), then protein concentration was estimated through the Bradford assay (Bio-Rad, Hercules, CA, USA) and adjusted to 3 μg/μL with PBS.

5-epi-SNEP–functionalized beads suspension (100 μL) and the same amount of the control matrix were separately incubated with 1 mg of the obtained lysate proteins under continuous shaking (24 h, 4°C). The beads were then collected, the unbounded proteins removed and the remaining resins washed three times with PBS to remove the un-specifically interacting proteins. The tightly-bound proteins were then eluted boiling the beads in 20 μL of Laemmli buffer for 5 minutes and by a subsequent centrifugation (5 minutes, 13000g, Centrifuge 5424 R, Eppendorf) for an easier collection of the protein-rich supernatants.

The proteins eluted from the control and 5-epi-SNEP-bearing beads were separated on a 12% SDS-PAGE, which was fixed and stained with Coomassie G-250 (Bio-Rad, Hercules, CA, USA), as previously reported. The experiment was repeated twice.

7.2.9 AP-MS: samples preparation for the mass spectrometric analysis: in situ tryptic digestion

The stained gel control and 5-epi-SNEP lanes were cut into 10 pieces alongside the molecular weight markers as follows: (1) 245-135 kDa, (2) 135-75 kDa, (3) 75-63 kDa, (4) 63-

50 kDa, (5) 50-40 kDa, (6) 40-35 kDa, (7) 35-25 kDa, (8) 25-20 kDa, (9) 20-17 kDa, (10) 17-11 kDa.

Shevchenko *in situ* digestion protocol was performed as described before[34]. Peptides extracted from the gel slices were dried out under vacuum (SpeedVac Concentrator Plus, Eppendorf) and solubilized in 12 µL of 10% Formic Acid (FA) for the subsequent nano-flow RP-UPLC MS/MS analysis.

7.2.10 AP-MS: nano-flow RP-UPLC MS/MS analysis

The peptide mixtures (5 µL) were injected into a nano-ACQUITY UPLC system (Waters, Milford, MA, USA) and separated on a 1.7-µm BEH C18 column (Waters) at a flow rate of 280 nL/min through a linear gradient of B from 20% to 90% over 55 min (solution A: 95% H_2O, 5% CH_3CN, 0.1% acetic acid; solution B: 95% CH_3CN, 5% H_2O, 0.1% acetic acid). MS and MS/MS data were acquired on a LTQ Orbitrap XL high-performance liquid chromatography MS system (Thermo-Scientific, Waltham, MA, USA) equipped with an electrospray source (ESI). The ten most intense doubly and triply charged peptide ions were chosen and fragmented. The resulting MS data were processed by MS Converter General User Interface software (ProteoWizard; http://proteowizard.sourceforge.net/project.shtml) to generate peak lists for protein identifications.

7.2.11 AP-MS: proteins identification

The obtained raw files (.raw) were converted into Mascot Generic Format data files (.mgf) through MSConvert and loaded onto the MASCOT Daemon graphical user interface (MatrixScience, London, UK) to achieve proteins identification. The SwissProt database (release January 2017, 553474 sequences, 198069095 residues) was employed to retrieve *in silico* proteins digestion and the following settings were exploited: trypsin as the enzyme, two missed cleavages allowed; carbamidomethyl (C) as fixed modification; oxidation (M) and phosphorylation (ST) as variable modifications; 80 ppm as peptide tolerance; 0.8 Da as MS/MS tolerance.

The obtained data were filtered by molecular weight ranges and ranked by their Mascot ions scores (cut-off: Mascot score ≥ 100). The two protein lists were then compared to identify the marine metabolite partner proteins by subtracting the control-beads eluted proteins from the 5-epi-SNEP ones. Two independent experiment were superimposed to further shorten the putative partners list.

7.2.12 Validation of DARTS and Affinity Chromatography obtained MS data: Western Blotting analyses

8 µL of the AP-MS eluted samples and 30 µg of the DARTS obtained protein mixtures, both boiled in Laemmli buffer, were separately resolved on a 12% SDS-PAGE gel and transferred

onto a nitrocellulose membrane. The membrane was blocked, for 1 h at room temperature, in a 5% non-fat dried milk containing TBS-t and then incubated overnight at 4°C and under continuous shaking with a primary monoclonal antibody raised against actin (1:500, Santa Cruz Biotechnology, Inc., Dallas, TX, USA). The antibody excess was then removed, membranes were washed three times with TBS-t and incubated, for 1 h at room temperature and under shaking, with a mouse peroxidase-conjugated secondary antibody (1:2500; Thermo-Scientific). The signal was detected using an enhanced chemiluminescent substrate and LAS 4000 (GE Healthcare, Waukesha, WI, USA) digital imaging system.

Regarding the DARTS experiment, after the anti-actin blotting, the membrane was also hybridized with an anti-Glyceraldehyde 3-Phosphate Dehidrogenase antibody (GAPDH, 1:2500, mouse, Invitrogen) and the signal was detected as already described.

Both AP-MS and DARTS Western Blots were repeated twice.

A densitometric analysis of the DARTS Western Blots was also performed through ImageJ. Data were elaborated averaging the values from the independent replicates, using GAPDH as a loading normalizer and rating undigested actin intensity as 100%. The resulting histogram was plotted through GraphPad Prism 7.

7.2.13 T-LiP-MRM: cytoplasmic actins 1 and 2 transitions computational prediction

Cytoplasmic Actin isoforms 1 (UniProt Entry Name: ACTB_HUMAN) and 2 (UniProt Entry Name: ACTG_HUMAN) tryptic peptides previously detected by MS where selected

through the proteomics data resource *Peptide Atlas* (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas) on its Human build and queried into the complete Human *SRM Atlas* build (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/GetTransitions), to retrieve their best daughter ions.

The *SRMAtlas* query parameters were set as follows:

- number of highest intensity fragment ions to keep: 8;
- target instrument: QTRAP 5500;
- transitions source: QTOF, Agilent QQQ, Qtrap5500, Ion Trap, Predicted;
- precursor exclusion range: kept blank;
- search proteins form: SwissProt;
- duplicate peptides: unique in results;
- heavy label: kept blank;
- labeled transitions: kept as default;
- maximum m/z: 1100 Da;
- minimum m/z: kept blank;
- allowed ions types: b-ions and y-ions;
- allowed peptides modification: carbamidomethylation of cysteines (C[160]).

The obtained list, containing 23 precursors and 184 fragments, was subsequently refined as follows:

- peptides whose following C-terminus aminoacid was either K or R were removed;
- among the 8 fragment ions reported for each precursor, only the three most intense were selected preferring, when possible, the y-series ions over the b-series ones.

Thus, a comprehensive method listing 21 actins peptides and their three best transitions was obtained and subsequently tested onto a HeLa lysate tryptic digest.

7.2.14 T-LiP-MRM: cytoplasmic actins 1 and 2 best transitions selection

300 µg of HeLa cell lysate obtained as described before (pharagraph 7.1.1.1), were submitted to an *in solution* digestion protocol. Briefly, proteins were denatured using 8 M urea/50 mM AmBic (4 M final urea concentration), disulphide bonds were reduced with 10 mM DTT for 1h at 25°C and 800 rpm (Thermomixer, Eppendorf) and then alkylated with 20 mM IAA for 30 minutes, at 25°C and 800 rpm, in the dark (Thermomixer, Eppendorf). IAA was then quenched with 10 mM DTT (10 minutes, 25°C, 800 rpm) and urea was diluted up to 1M with 50 mM AmBic before adding the trypsin/LysC solution (Promega, Madison, Wisconsin) at the enzyme to proteins ratio of 1:100 w/w. Digestion was allowed to proceed overnight at 37°C under continuous shaking (800 rpm, Thermomixer, Eppendorf) and then quenched adding FA to lower the pH to 3. The peptides mixture was then dried under vacuum (SpeedVac Concentrator Plus, Eppendorf), dissolved in 1 mL 5% FA and desalted through a Sep-Pak C18 1 cc (50 mg) cartridge (Waters, Milford, MA, USA).

Briefly, the cartridge was activated flushing 3 mL of 100% CH_3CN and then conditioned with 3 mL of 0.1% FA. The sample was then loaded, desalted flushing the cartridge with 3 mL of 0.1% FA and finally eluted flushing two times 500 µL of 80% CH_3CN, 20% H_2O, 0,1% FA. For the subsequent MS analysis, the peptides mixture was dried under vacuum (SpeedVac Concentrator Plus, Eppendorf) and re-dissolved in 10% FA.

UPLC–ESI-MRM-MS analyses were performed on a 6500 Q-TRAP from AB Sciex equipped with Shimadzu LC-20A and Auto Sampler systems. UPLC separation was performed on an Aeris Widepore XB C18 column (150 × 2.10 mm, 3.6 µm XB, Phenomenex, Torrance, USA), using 0.1% FA in H_2O (A) and 0.1% FA in CH_3CN (B) as mobile phases, and a linear gradient from 5 to 95% of B over 30 min (flow rate: 200 µL/min). Q-TRAP 6500 was operated in positive MRM scanning mode, with declustering potential (DP) set at 80V, entrance potential (EP) at 10V, collision energy (CE) at 35V and cell exit potential (CXP) at 22V.

LC–ESI-MRM/MS runs were performed injecting 15 µg of the peptide mixture: the XICs of all the transitions of each precursor were inspected to (1) identify actins tryptic peptides which could actually be experimentally observed, (2) assign the retention time to all of the peptides and (3) identify the best transition as the one showing the most intense peak and the best signal to noise ratio.

This preliminary experiment led to a global MRM method comprising 16 transitions.

The observed peptides were then mapped onto both of the actin isoforms UniProt-retrieved sequences through the PAWS program, with the following parameters:

- cleave: K/R;

- mass: average;
- fixed modification: cysteins carbamidomethylation;
- search tolerance: 300 ppm.

Both of the isoforms coverage (%) was also calculated as follows:

Coverage (%) = (Experimentally mapped aminoacids/total aminoacids)*100.

7.2.15 T-LiP-MRM experiment

300 µg of HeLa cells proteome extracted as previously reported (pharagraph 7.1.1.1) were incubated with DMSO or 5-epi-SNEP (5 µM and 50 µM final concentrations), for 1 hour at room temperature and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then splitted and submitted to limited proteolysis with 1:1000 w/w and 1:500 w/w subtilisin to proteins ratios, respectively. Limited proteolysis was carried out for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). 300 µg of a DMSO-treated lysate aliquot were submitted to a mock proteolysis, carried out adding H_2O instead of the enzyme.

Subtilisin was then quenched with PMSF (1 mM final concentration) and the samples shifted to denaturing condition adding urea (4 M final concentration) to perform in solution digestion and desalting, as described before.

15 µg of each sample were then injected in the LC–ESI-MRM-MS system as already described and analyzed through the previously optimized 16-transitions MRM method. The

area of each actin isoform tryptic peptide peak was then measured using the Analyst Software from AB Sciex.

Each sample was analyzed in triplicate and GraphPad Prism 7 was exploited for the data analysis step to calculate fold changes and the corresponding p-values.

5-epi-SNEP protected actins LiP peptides (i.e. fold change ≥ 2 at the lowest radicicol concentration and $p < 0.05$) were mapped onto the following PDB structure: pdbID 4RWT[266].

7.2.16 In cell assays: preliminary HeLa cells viability assay

The human uterine cervical cancer cells HeLa were obtained from the American Type Culture Collection (ATCC) and were cultured in DMEM medium (Euroclone, Milan, Italy) supplemented with 10% (v/v) fetal bovine serum (Euroclone), 100 U/mL penicillin and 100 µg/µL streptomycin (Euroclone) at 37°C in a 5% CO_2 atmosphere.

The effect of both 5-epi-SNEP and its epoxydated derivatives (7.2.8) on cell viability were evaluated through the colorimetric MTT ((3-4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide) metabolic activity assay.

Briefly, 5×10^4 HeLa cells were cultured in a 96-well plate and, after 24 h, separately exposed to multiple concentrations of the two molecules (100, 50, 25, 10, 5, 2.5 and 1 µM) for 24 h. Then, 10 µL of MTT (5 mg/mL) was added in each well. After 1 h at 37°C, the supernatants were removed and the resultant formazan crystals were dissolved in 100 µL DMSO. The absorbance intensity was measured by a micro-plate reader at 550 nm with a reference

wavelength of 620 nm. All of the experiments were performed in quadruplicate, and the relative cell viability (%) was expressed as a percentage of the untreated control cells.

7.2.17 In cell assays: DARTS

HeLa cells were incubated with either 5-epi-SNEP (2.5 and 10 μM final concentrations), the known F-actin depolymerizing agent Cytocalasin D (CyD, 0.5 μM final concentration) or DMSO as a vehicle for 1 hour. Cells were then washed and harvested.

The obtained pellets were suspended in PBS 0.1% Igepal supplemented with proteases inhibitors (1x final concentration) and lysed through sonication (1 minute, 30% amplitude, 9.9 seconds pulses, Vibra cell, SONICS). The obtained suspensions were then submitted to centrifugation (10000 rpm, 4°C, 5 minutes, Centrifuge 5424 R, Eppendorf) to remove the protein solution from the pelleted debris. Proteins concentration of the cleared lysate supernatant was determined by Bradford assay (Bio-Rad, Hercules, CA, USA) and subsequently adjusted to 3 μg/μL with PBS.

100 μg proteins aliquots from each cell lysate were then submitted (or not) to limited proteolysis with subtilisin at the enzyme to substrate ratio of 1:1000 w/w for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf).

Subtilisin was then quenched with PMSF (1 mM final concentration), incubating the samples for 10 minutes at 25°C and 500 rpm.

30 μg of the obtained samples were then boiled in Laemmli buffer, resolved on a 12% SDS-PAGE gel and transferred onto a nitrocellulose membrane. The membrane was blocked, for 1 h at room temperature, in a 5% non-fat dried milk containing TBS-t solution and then incubated overnight at 4°C and under continuous shaking with a primary anti-actin monoclonal antibody (1:500, Santa Cruz Biotechnology, Inc., Dallas, TX, USA). The antibody excess was then removed, the membrane washed three times with TBS-t and incubated, for 1 h at room temperature and under shaking, with a mouse peroxidase-conjugated secondary antibody (1:2500; Thermo-Scientific). The signal was detected using an enhanced chemiluminescent substrate and LAS 4000 (GE Healthcare, Waukesha, WI, USA) digital imaging system.

Subsequently, the membrane was also hybridized with an anti-GAPDH antibody (1:2500, mouse, Invitrogen) and the signal was detected as already described.

The experiment was repeated three times.

A densitometric analysis of the Western Blots was then performed through ImageJ, averaging the values from the independent replicates and rating, for each sample, its undigested actin intensity as 100%. GAPDH was exploited as a loading normalizer.

7.2.18 In cell assays: analysis of Actins polymerization through confocal microscopy

In order to analyze the effect of 5-epi-SNEP on F-actin, 4×10^4 HeLa cells per well were seeded on cover slips in 24-well plastic plates. After 24 h, cells were treated or not with 5-epi-SNEP (2.5 or 10 μM final concentrations) for 1 h. Cytochalasin D (CyD) treatment (0.5 μM

final concentration, 1 h) was used as positive control for actin filament depolymerization. Then, cells were fixed with 4% paraformaldehyde in PBS for 10 min, permeabilized and blocked in 1% bovine serum albumine for 30 min. Actin microfilaments and nuclei were stained, respectively, with 2 μg/mL tetramethylrhodamine (TRITC)-phalloidin (Sigma-Aldrich, Darmstad, Germany), and 4′,6-diamidino-2-phenylindole (DAPI, Thermo-Scientific Waltham, MA, USA) for 1 h.

The images were collected on a Zeiss LSM 510 confocal microscope (Carl Zeiss Microscopy GmbH, Oberkochen, Germany) using LSM 510 Meta software 4.0 SP2 version. The obtained images are representative of multiple fields and triplicate cover slips per experiment.

7.3 Disclosing Poly [ADP-Ribose] Polymerase 1 as the main Crellastatin A target through a combination of label-free functional proteomics approaches.

7.3.1 Preliminary quality control: CreA HPLC-MS analysis

Crellastatin A (CreA) was isolated from the Vanuatu Island marine sponge *Crella* sp. and its structural assignment was accomplished through 2D NMR spectroscopy, as reported by D'Auria *et al.*[192]. The purity of the gifted compound was also verified through reverse phase HPLC-UV (RP-HPLC-UV) and mass spectrometry, as follows.

CreA was dissolved in DMSO and 15 µg were injected in a 1100 Series Chromatographer equipped with a UV detector set at 220 nm (Agilent, Santa Clara, California, USA). A C18 column (Luna Omega 5 µm Polar C18 150 × 2.1 mm, Phenomenex, Torrance, CA, USA) and a flow rate of 0.200 mL/min were exploited. Elution was achieved by means of a linear gradient of B from 20% to 95% over 25 min (solution A: H_2O; solution B: 95% CH_3CN and 5% H_2O). The HPLC peak at 20.77 minutes (Figure 100A) was analyzed in negative ionization mode on a 6500 Q-TRAP from AB Sciex, giving a m/z value of 1007.4 Da, corresponding to the $[M-H]^-$ CreA ion (Figure 100B).

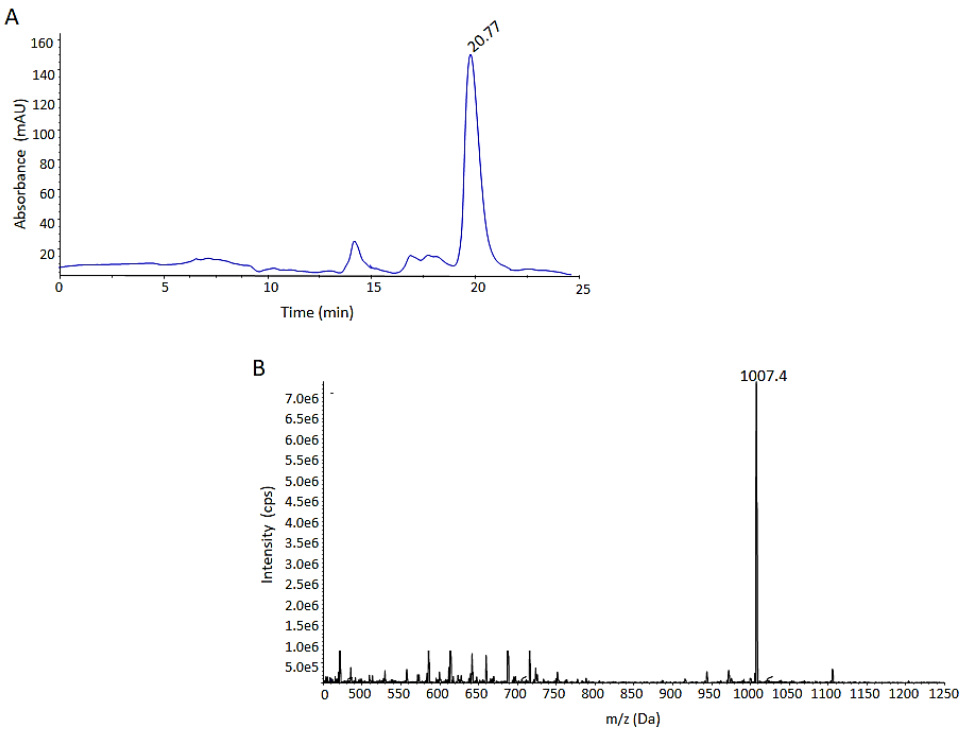


Figure 100: *CreA purity check through PR-HPLC-UV showing a single peak at 20.77 minutes. (B) Mass spectrometric analysis of the 20.77 minutes chromatographic peak.*

7.3.2 DARTS: subtilisin titration experiment

300 µg aliquots of HeLa lysate (pharagraph 7.1.1.1) were incubated with either CreA (0.3 µM or 3 µM final concentrations) or DMSO (final DMSO amount: 1% vol/vol), for 1 hour at room temperature and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then splitted in 100 µg aliquots and submitted to limited proteolysis with 1:1000 w/w and 1:500 w/w subtilisin to proteins ratios, respectively. Limited proteolysis was carried out for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). 100 µg of the DMSO-treated lysate were submitted to a mock proteolysis, carried out adding H_2O and incubating the sample as reported for the enzyme-treated ones.

Subtilisin was then quenched with PMSF (1 mM final concentration), incubating the samples for 10 minutes at 25°C and 500 rpm.

10 µg of the obtained mixtures were then added of Laemmli buffer[28] and heated at 95°C for 5 minutes to be subsequently loaded onto a 12% poly-acrilamide gel for a 1D-SDS-PAGE analysis. The gel was then fixed for 15 minutes (fixing solution: 50% H_2O, 40% MeOH, 10% AcOH), washed three times (10 minutes each) with H_2O and then submitted to Coomassie staining (Coomassie G-250, Bio-Rad, Hercules, CA, USA) for 1 hour at room temperature under continuous shaking. The excess dye was removed by extensively washing the gel with H_2O and a scan image of the resulting gel was then obtained through LabScan.

7.3.3 DARTS: complete experiment

300 μg aliquots of a HeLa cell lysate obtained as previously described (pharagraph 7.1.1.1) were incubated with either CreA (0.3 μM and 3 μM final concentrations) or DMSO for 1 hour at room temperature (final DMSO amount: 1% vol/vol) and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then submitted to limited proteolysis, as above reported, with 1:500 w/w subtilisin to proteins ratio. A cell lysate aliquot underwent mock proteolysis. Subtilisin was then quenched with 1 mM PMSF and 10 μg of each sample were treated with Laemmli buffer and heated at 95°C for 5 minutes. 1D-SDS-PAGE was then performed on a 4%-12% polyacrilamide gradient gel (Criterion^{TM} XT Precast Gel, 4-12% Bis-Tris, 12 +2 well comb, Bio-Rad, Hercules, CA, USA) and the obtained gel fixed, washed and Coomassie stained (Coomassie G-250, Bio-Rad, Hercules, CA, USA). The excess dye was then removed by extensively washing the gel with H_2O.

The scan image of the resulting gel (LabScan) was submitted to a densitometric analysis through ImageJ (molecular weight ranges: higher than 235 kDa, 235-170 kDa, 170-130 kDa, 130-93 kDa, 93-70 kDa, 70-53 kDa, 53-41 kDa, 41-30 kDa and 30-18 kDa). Data were represented, as percentages of the mock-proteolyzed sample, through GraphPad Prism 7.

7.3.4 DARTS: samples preparation for the mass spectrometric analysis

Shevchenko *in situ* digestion protocol was performed as described by the authors[34]. Briefly, gel bands whose intensity increased in presence of CreA were excided form all of the gel lanes by chopping them in smaller pieces, which were then washed by shrinking/swelling cycles using CH_3CN and ammonium bicarbonate (AmBic, 50 mM, pH 8.5), alternatively. Then, disulphide bonds were reduced by treating the gel pieces with 1,4-dithiothreitol (DTT, 6.5 mM in 50 mM AmBic, 60 minutes, 60°C) and the formed thiols were carboxyamidomethylated with iodoacetamide (IAA, 54 mM in 50 mM AmBic, 30 minutes, room temperature, in the dark). Residual reagents were removed by shrinking/swelling cycles and gel pieces rehydrated in a 12 ng/µL trypsin/LysC solution (Promega, Madison, Wisconsin) on ice for 1 h. The enzymes excess was then removed and 40 µL of 50 mM AmBic were added to allow protein digestion to proceed overnight at 37°C (Thermomixer, Eppendorf). The supernatant was then collected and peptides were extracted from the gel slices shrinking them twice with 100% CH_3CN. All of the supernatants were collected and combined to be then dried out under vacuum (SpeedVac Concentrator Plus, Eppendorf) and solubilized in 12 µL of 10% Formic Acid (FA) for the subsequent nano-flow RP-UPLC MS/MS analysis.

7.3.5 DARTS: nano-flow RP-UPLC MS/MS analysis

The peptide mixtures (5 µL) were injected into a nano-ACQUITY UPLC system (Waters, Milford, MA, USA) and separated on a 1.7-µm BEH C18 column (Waters) at a flow rate of 280 nL/min. Peptides elution was achieved using a linear gradient of B from 20% to 90% over 55 min (solution A: 95% H_2O, 5% CH_3CN, 0.1% acetic acid; solution B: 95% CH_3CN, 5% H_2O, 0.1% acetic acid). MS and MS/MS data were acquired on an LTQ Orbitrap XL high-performance liquid chromatography MS system (Thermo-Scientific, Waltham, MA, USA) equipped with an electrospray source (ESI). The ten most intense doubly and triply charged peptide ions were chosen and fragmented. The resulting MS data were processed by MS Converter General User Interface software (ProteoWizard; http://proteowizard.sourceforge.net/project.shtml) to generate peak lists for protein identifications.

7.3.6 DARTS: proteins identification and semi-quantitative analysis

The obtained raw files (.raw) were converted into Mascot Generic Format data files (.mgf) through MSConvert and loaded onto the Mascot Daemon graphical user interface (MatrixScience, London, UK) to achieve proteins identification. The SwissProt database (release January 2017, 553474 sequences, 198069095 residues) was employed to retrieve *in silico* proteins digestion and the following settings were exploited: trypsin as the enzyme, two

missed cleavages allowed; carbamidomethyl (C) as fixed modification; oxidation (M) and phosphorylation (ST) as variable modifications; 80 ppm as peptide tolerance; 0.8 Da as MS/MS tolerance.

The obtained data were filtered by molecular weight ranges accordingly with the gel cutting patterns and a semi-quantitative analysis was then performed comparing MASCOT matches among the analyzed compounds. Protection percentages were thus calculated, for each CreA amount, as follows:

$$\text{Protection (\%)} = [(\text{Matches}_{\text{CreA}} - \text{Matches}_{\text{Control}})/\text{Matches}_{\text{Lysate}}]*100.$$

7.3.7 Validation of DARTS obtained MS data: Western Blotting analysis

15 µg of the DARTS obtained protein mixtures were treated with Laemmli buffer[28], boiled at 95°C for 5 minutes, resolved on a 12% SDS-PAGE gel and transferred onto a nitrocellulose membrane. The membrane was blocked, for 1 h at room temperature, in a 5% non-fat dried milk containing TBS-t solution (31 mM Tris pH 8, 170 mM NaCl, 3.35 mM KCl, 0.05% Tween 20) and then incubated overnight at 4°C and under continuous shaking with a primary monoclonal antibody raised against PARP1 (1:1000, Santa Cruz Biotechnology, Inc., Dallas, TX, USA). The antibody excess was then removed, membranes were washed three times with TBS-t and incubated, for 1 h at room temperature and under shaking, with a mouse peroxidase-conjugated secondary antibody (1:2500; Thermo-Scientific). The signal was detected using an

enhanced chemiluminescent substrate and LAS 4000 (GE Healthcare, Waukesha, WI, USA) digital imaging system.

Afterwards, the membrane was also hybridized with an anti-Glyceraldehyde 3-Phosphate Dehidrogenase antibody (GAPDH, 1:2000, mouse, Invitrogen) and the signal was detected as already described.

DARTS Western Blottings were repeated twice and a densitometric analysis was also performed through ImageJ. Data were elaborated averaging the values from the independent replicates, using GAPDH as a loading normalizer and rating undigested PARP 1 intensity as 100%. The obtained values were represented through GraphPad Prism 7.

7.3.8 T-LiP-MRM: PARP 1 transitions computational prediction

PARP 1 (UniProt accession number: P09874) tryptic peptides previously detected by MS where selected through the proteomics data resource Peptide Atlas (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas) on its Human build and queried into the complete Human *SRM Atlas* build (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/GetTransitions), to retrieve their best daughter ions.

The *SRMAtlas* query parameters were set as follows:

- number of highest intensity fragment ions to keep: 8;
- target instrument: QTRAP 5500;

- transitions source: QTOF, Agilent QQQ, Qtrap5500, Ion Trap, Predicted;
- precursor exclusion range: kept blank;
- search proteins form: SwissProt;
- duplicate peptides: unique in results;
- heavy label: kept blank;
- labeled transitions: kept as default;
- maximum m/z: 1100 Da;
- minimum m/z: kept blank;
- allowed ions types: b-ions and y-ions;
- allowed peptides modification: carbamidomethylation of cysteines (C[160]).

The obtained list, containing 39 precursors and 312 fragments, was subsequently refined as follows:

- peptides whose following C-terminus aminoacid was either K or R were removed;
- among the 8 fragment ions reported for each precursor, only the three most intense were selected preferring, when possible, the y-series ions over the b-series ones;
- peptides presenting a *N_map* value higher than 1 were discarded.

Thus, a comprehensive method listing 39 PARP 1 peptides and their three best transitions was obtained and subsequently tested onto a HeLa lysate tryptic digest.

7.3.9 T-LiP-MRM: PARP 1 best transitions selection

300 µg of HeLa cell lysate obtained as described before (pharagraph 7.1.1.1), were submitted to an *in solution* digestion protocol. Briefly, proteins were denatured using 8 M urea/50 mM AmBic (4 M final urea concentration), disulphide bonds were reduced with 10 mM DTT for 1h at 25°C and 800 rpm (Thermomixer, Eppendorf) and then alkylated with 20 mM IAA for 30 minutes, at 25°C and 800 rpm, in the dark (Thermomixer, Eppendorf). IAA was then quenched with 10 mM DTT (10 minutes, 25°C, 800 rpm) and urea was diluted up to 1 M with 50 mM AmBic before adding the trypsin/LysC solution (Promega, Madison, Wisconsin) at the enzyme to proteins ratio of 1:100 w/w. Digestion was allowed to proceed overnight at 37°C under continuous shaking (800 rpm, Thermomixer, Eppendorf) and then quenched adding FA to lower the pH to 3. The peptides mixture was then dried under vacuum (SpeedVac Concentrator Plus, Eppendorf), dissolved in 1 mL 5% FA and desalted through a Sep-Pak C18 1 cc (50 mg) cartridge (Waters, Milford, MA, USA).

Briefly, the cartridge was activated flushing 3 mL of 100% CH_3CN and then conditioned with 3 mL of 0.1% FA. The sample was then loaded, desalted flushing the cartridge with 3 mL of 0.1% FA and finally eluted flushing two times 500 µL of 80% CH_3CN, 20% H_2O, 0,1% FA. For the subsequent MS analysis, the peptides mixture was dried under vacuum (SpeedVac Concentrator Plus, Eppendorf) and re-dissolved in 10% FA.

UPLC–ESI-MRM-MS analyses were performed on a 6500 Q-TRAP from AB Sciex equipped with Shimadzu LC-20A and Auto Sampler systems. UPLC separation was performed

on an Aeris Widepore XB C18 column (150 × 2.10 mm, 3.6 µm XB, Phenomenex, Torrance, USA), using 0.1% FA in H_2O (A) and 0.1% FA in CH_3CN (B) as mobile phases, and a linear gradient from 5 to 95% of B over 30 min (flow rate: 200 µL/min). Q-TRAP 6500 was operated in positive MRM scanning mode, with declustering potential (DP) set at 80V, entrance potential (EP) at 10V, collision energy (CE) at 35V and cell exit potential (CXP) at 22V.

LC–ESI-MRM/MS runs were performed injecting 30 µg of the peptides mixture: the XICs of all the transitions of each precursor were inspected to (1) identify PARP 1 tryptic peptides which could actually be experimentally observed, (2) assign the retention time to all of the peptides and (3) identify the best transition as the one showing the most intense peak and the best signal to noise ratio.

This preliminary experiment led to a global MRM method comprising 24 transitions.

The observed peptides were then mapped onto PARP 1 UniProt-retrieved sequence through the PAWS program, with the following parameters:

- cleave: K/R;
- mass: average;
- fixed modification: cysteins carbamidomethylation;
- search tolerance: 300 ppm.

PARP 1 coverage (%) was also calculated as follows:

Coverage (%) = (Experimentally mapped aminoacids/total aminoacids)*100.

7.3.10 T-LiP-MRM experiment

300 µg of HeLa cells proteome extracted as previously reported (pharagraph 7.1.1.1) were incubated with DMSO or CreA (3 µM final concentration, final DMSO amount: 1% vol/vol), for 1 hour at room temperature and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then splitted and submitted to limited proteolysis with 1:500 w/w and 1:250 w/w subtilisin to proteins ratios, respectively. Limited proteolysis was carried out for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). 300 µg of a DMSO-treated lysate aliquot were submitted to a mock proteolysis, carried out adding H_2O instead of the enzyme.

Subtilisin was then quenched with PMSF (1 mM final concentration) and the samples shifted to denaturing condition adding urea (4 M final concentration) to perform *in solution* digestion and desalting, as described before.

30 µg of each sample were then injected in the LC–ESI-MRM-MS system as already described and analyzed through the previously optimized 24-transitions MRM method. The area of each PARP 1 tryptic peptide peak was then measured using the Analyst Software from AB Sciex.

Two biological replicates were performed, as well as injection replicates.

CreA protected PARP 1 LiP peptides were mapped onto the following PDB structure: pdbID 4DQY[267].

7.3.11 T-LiP-MRM data corroboration: blind molecular docking analysis

A blind molecular docking analysis was used to predict the feasible binding mode and the binding strength between the CreA, designed and minimized using Avogadro software[268], and the human PARP 1, as previously reported[222]. PARP 1 crystallographic structure was obtained from the pdbID 4DQY[267]. Autodock Vina[269] (version 1.1.2), on an Intel Core i7/Mac OS X 10.14 – based platform, was used considering a docking zone including the entire protein with a grid of 70, 82 and 107 Å in the x, y, and z directions, whereas the NN score 2.0 python script[270] was used to calculate the predicted equilibrium dissociation constant ($K_{D,pred}$). The final complex geometry was rendered by PyMol software (The PyMOL Molecular Graphics System, Version 2.0.4 Schrödinger, LLC).

7.3.12 The consequences of PARP 1/CreA interaction: in vitro activity assays

Nicotinamide (NAM) mass spectrometric parameters were optimized by infusing a 0.1 µg/µL solution made up in 5 mM ammonium acetate (AmAc) in H_2O/0.1% FA, in a QTRAP 6500 system operated in positive ion mode. MS and MS/MS spectra were acquired to determine the transition to analyze and DP, EP, CE and CXP parameters were optimized. The selected parameters are reported in Figure 101.

Q1_m/z	Q3_m/z	Time (msec)	DP (volts)	EP (volts)	CE (volts)	CXP (volts)
123.000	80.000	50	140	5	20	20

***Figure 101**: MRM optimized parameters for NAM detection.*

Then, UPLC conditions were optimized on the Shimadzu LC-20A interfaced with 6500 QTRAP as follows: chromatographic separation was achieved onto a Luna Omega Polar C18 column (50 × 2.1 mm, 1.6 µm, Phenomenex, Torrance, USA), using 0.1% FA in H_2O (A) and 0.1% FA in CH_3CN (B) as mobile phases, and the gradient reported in Figure 102. NAM detection was accomplished with the previously reported parameters, operating 6500 QTRAP in positive MRM ion mode and both Q1 and Q3 at a *Unit* resolution.

A

Time (min)	Module	Event	Parameter
0.00	Pumps	Pump B Conc.	0
0.50	Pumps	Pump B Conc.	0
3.00	Pumps	Pump B Conc.	10
3.10	Pumps	Pump B Conc.	95
5.00	Pumps	Pump B Conc.	95
5.10	Pumps	Pump B Conc.	0
7.00	Pumps	Pump B Conc.	0
7.10	Controller	Stop	

B

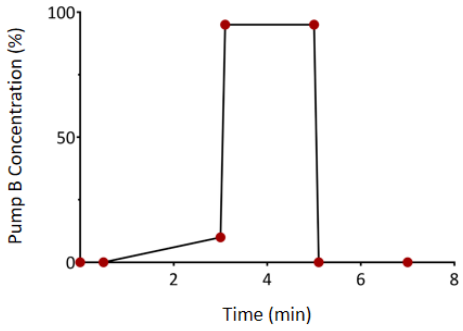


***Figure 102:** Shimadzu LC conditions reported as Pump B gradient on both a table (A) and a graph (B).*

The PARP 1 human recombinant protein (ALX-201-063) was purchased from Enzo Life Sciences. To evaluate its NAM production rate from nicotinamide adenine dinucleotide (NAD^+), 50 ng PARP 1 were incubated, at 30°C and 350 rpm (Termomixer, Eppendorf), with 500 ng of Histone H3 peptide (amino acids 44-63), 50 ng of annealed EcoRI-linker DNA and 500 ng of NAD^+ in 20 µL of a buffer made of 50 mM Tris–HCl, 5 mM $MgCl_2$, 2.5 mM DTT, pH 8.0.

Reaction aliquots (1 µl) were diluted in 59 µl of 5 mM AmAc in H_2O/0.1% FA (1 ng/µl final NAD^+ concentration) and quenched with 10 mM 3-aminobenzamide (3-ABA, a known PARP 1 inhibitor) at different times (0 and 30 minutes, 1, 2 and 4 hours). 3 µl of the obtained mixtures were subjected to UPLC-MRM-MS in the previously optimized conditions. To measure the produced amount of NAM, the area of each peak at different time points was measured using the Analyst Software from AB Sciex. Values from three injection replicates were averaged and plotted, as a function of the incubation time, through GraphPad Prism 7.

Then, to test CreA activity on PARP 1, the same assay has been repeated only monitoring the NAM produced after 2 hours reaction time. Briefly, solutions of PARP 1 containing the desired amount of CreA (5 µM, 10 µM, 25 µM, 50 µM, 100 µM and 150 µM final concentrations) 3-ABA (100 µM final concentration) or just DMSO as the vehicle (final DMSO amount for all of the samples: 1%), were incubated for 15 min at 30°C and 350 rpm (Termomixer, Eppendorf). Subsequently, H3 peptide, DNA and NAD^+ were added and the

reaction was allowed to proceed for 2 hours at 30°C and 350 rpm (Termomixer, Eppendorf). Then, 1 µL of each sample was diluted in 59 µl of 5 mM AmAc in H_2O/0.1% FA (1 ng/µl final NAD^+ concentration) and quenched with 10 mM 3-ABA. 3 µl of the obtained mixtures were subjected to UPLC-MRM-MS and the area of each NAM peak was measured using the Analyst Software from AB Sciex. PARP 1 activity measured in presence of only DMSO was set as 100%, whereas 3 ABA was used as a positive control for PARP 1 inhibition. CreA IC_{50} was then calculated fitting the data, obtained averaging the values of three injections replicates per sample, in a sigmoidal curve (variable slope) through GraphPad Prism 7.

7.4 *Label-free* functional proteomics reveals cytosolic HSP71A and mitochondrial GRP75 as the main cellular partners of a Mycale rotalis acetogenin.

7.4.1 Preliminary quality control: BrACG HPLC-MS analysis

BrACG was isolated from the sponge *Mycale* rotalis collected in the Stagnone di Marsala lagoon (Sicily) and its purification and characterization were achieved as previously reported[231,271].

Briefly, pure BrACG was obtained by filtration on a silica gel pad (eluent n-hexane-EtOAc, 4:6) followed by RP-HPLC using n-hexane-EtOAc (75:25) as eluent (Luna C18 column, 25 cm

× 4.6 mm, 5 µm, Phenomenex Torrance, CA, USA; flow rate 1 mL/min). The compound was then analyzed through MS, IR and NMR.

Upon receipt the pure gifted compound identity was also verified through mass spectrometry, as follows. BrACG was dissolved in DMSO and a 0.1 µg/µL solution was then prepared by diluting an aliquot in H_2O/CH_3CN 50:50 0.1% tryethylamine (TEA). This solution was subsequently analyzed onto a Q-ToF Premiere (Waters) mass spectrometer operated in negative ion mode. The resulting mass spectra is reported in Figure 103.

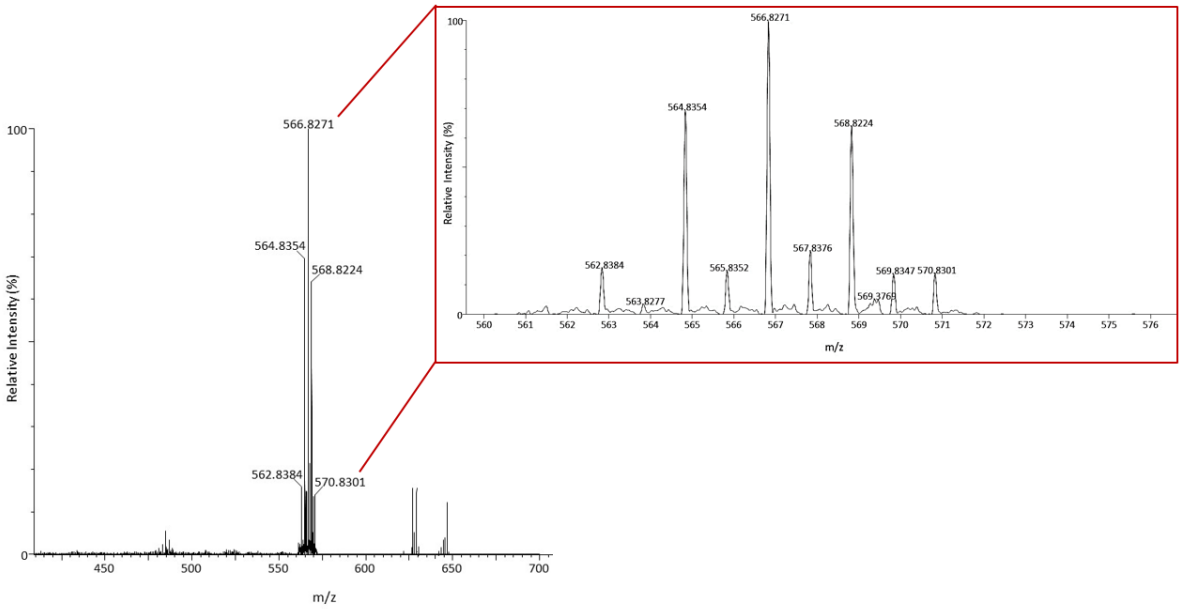


Figure 103: *BrACG MS analysis.*

7.4.2 DARTS: subtilisin titration experiment

300 μg aliquots of HeLa lysate (pharagraph 7.1.1.1) were incubated with either BrACG (5 μM or 50 μM final concentrations) or DMSO (final DMSO amount: 1% vol/vol), for 1 hour at room temperature and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then splitted in 100 μg aliquots and submitted to limited proteolysis with 1:1000 w/w and 1:500 w/w subtilisin to proteins ratios, respectively. Limited proteolysis was carried out for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). 100 μg of the DMSO-treated lysate were submitted to a mock proteolysis, carried out adding H_2O and incubating the sample as reported for the enzyme-treated ones.

Subtilisin was then quenched with PMSF (1 mM final concentration) for 10 minutes at 25°C and 500 rpm.

10 μg of the obtained mixtures were then added of Laemmli buffer[28] and heated at 95°C for 5 minutes to be subsequently loaded onto a 12% poly-acrilamide gel for a 1D-SDS-PAGE analysis. The gel was then fixed for 15 minutes (fixing solution: 50% H_2O, 40% MeOH, 10% AcOH), washed three times (10 minutes each) with H_2O and then submitted to Coomassie staining (Coomassie G-250, Bio-Rad, Hercules, CA, USA) for 1 hour at room temperature under continuous shaking. The excess dye was removed by extensively washing the gel with H_2O and its scan image was then obtained through LabScan.

7.4.3 DARTS: complete experiment

300 µg aliquots of a HeLa cell lysate obtained as previously described (pharagraph 7.1.1.1) were incubated with either BrACG (0.5 µM, 5 µM or 50 µM final concentrations) or DMSO for 1 hour at room temperature (final DMSO amount: 1% vol/vol) and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then submitted to limited proteolysis, as above reported, with 1:500 w/w subtilisin to proteins ratio. A cell lysate aliquot underwent mock proteolysis. Subtilisin was then quenched with 1mM PMSF and 10 µg of each sample were treated with Laemmli buffer and heated at 95°C for 5 minutes. 1D-SDS-PAGE was performed on a 4%-12% polyacrilamide gradient gel (Criterion^{TM} XT Precast Gel, 4-12% Bis-Tris, 12 +2 well comb, Bio-Rad, Hercules, CA, USA) and the obtained gel fixed, washed and *Coomassie* stained (Coomassie G-250, Bio-Rad, Hercules, CA, USA). The excess dye was then removed by extensively washing the gel with H_2O.

The scan image of the resulting gel (LabScan) was submitted to a densitometric analysis through ImageJ (molecular weight ranges: 235-170 kDa, 170-130 kDa, 130-75 kDa, 75-53 kDa, 53-41 kDa, 41-30 kDa and 30-14 kDa). Data were represented, as percentages of the mock-proteolyzed sample, through GraphPad Prism 7.

7.4.4 DARTS: samples preparation for the mass spectrometric analysis

Shevchenko *in situ* digestion protocol was performed as described by the authors[34]. Briefly, gel bands whose intensity increased in presence of increasing BrACG amounts were excided form all of the gel lanes by chopping them in smaller pieces, which were then washed by shrinking/swelling cycles using CH_3CN and ammonium bicarbonate (AmBic, 50 mM, pH 8.5), alternatively. Then, disulphide bonds were reduced by treating the gel pieces with 1,4-dithiothreitol (DTT, 6.5 mM in 50 mM AmBic, 60 minutes, 60°C) and the formed thiols were carboxyamidomethylated with iodoacetamide (IAA, 54 mM in 50 mM AmBic, 30 minutes, room temperature, in the dark). Residual reagents were removed by shrinking/swelling cycles and gel pieces rehydrated in a 12 ng/µL trypsin/LysC solution (Promega, Madison, Wisconsin) on ice for 1 h. The enzymes excess was then removed and 40 µL of 50 mM AmBic were added to allow protein digestion to proceed overnight at 37°C (Thermomixer, Eppendorf). The supernatant was then collected and peptides were extracted from the gel slices shrinking them twice with 100% CH_3CN. All of the supernatants were collected, combined, dried out under vacuum (SpeedVac Concentrator Plus, Eppendorf) and solubilized in 12 µL of 10% Formic Acid (FA) for the subsequent nano-flow RP-UPLC MS/MS analysis.

7.4.5 DARTS: nano-flow RP-UPLC MS/MS analysis

The peptides mixtures (5 µL) were injected into a nano-ACQUITY UPLC system (Waters, Milford, MA, USA) and separated on a 1.7-µm BEH C18 column (Waters) at a flow rate of 280 nL/min. Peptides elution was achieved using a linear gradient of B from 20% to 90% over 55 min (solution A: 95% H_2O, 5% CH_3CN, 0.1% acetic acid; solution B: 95% CH_3CN, 5% H_2O, 0.1% acetic acid). MS and MS/MS data were acquired on an LTQ Orbitrap XL high-performance liquid chromatography MS system (Thermo-Scientific, Waltham, MA, USA) equipped with an electrospray source (ESI). The ten most intense doubly and triply charged peptide ions were chosen and fragmented. The resulting MS data were processed by MS Converter General User Interface software (ProteoWizard; http://proteowizard.sourceforge.net/project.shtml) to generate peak lists for protein identifications.

7.4.6 DARTS: proteins identification and semi-quantitative analysis

The obtained raw files (.raw) were converted into Mascot Generic Format data files (.mgf) through MSConvert and loaded onto the Mascot Daemon graphical user interface (MatrixScience, London, UK) to achieve proteins identification. The SwissProt database (release October 2019, 95934 sequences, 38078700 residues) was employed to retrieve *in silico* proteins digestion and the following settings were exploited: trypsin as the enzyme, two missed

cleavages allowed; carbamidomethyl (C) as fixed modification; oxidation (M) and phosphorylation (ST) as variable modifications; 80 ppm as peptide tolerance; 0.8 Da as MS/MS tolerance.

The obtained data were filtered by molecular weight ranges accordingly with the gel cutting patterns and a semi-quantitative analysis was then performed comparing MASCOT matches among the analyzed compounds. Protection percentages were thus calculated, for each BrACG amount, as follows:

$$\text{Protection (\%)} = [(\text{Matches}_{\text{BrACG}} - \text{Matches}_{\text{Control}})/\text{Matches}_{\text{Lysate}}]*100.$$

7.4.7 Validation of DARTS obtained MS data: Western Blotting analysis

30 µg of the DARTS obtained protein mixtures were treated with Laemmli buffer[28], boiled at 95°C for 5 minutes and loaded two times each on a 12% SDS-PAGE gel to be transferred onto a nitrocellulose membrane. The membrane was blocked, for 1 h at room temperature, in a 5% non-fat dried milk containing TBS-t solution and cut in two halves. The obtained membranes were then incubated overnight at 4°C and under continuous shaking with a primary monoclonal antibody raised against GRP75 and HS71A (1:1000, Santa Cruz Biotechnology, Inc., Dallas, TX, USA), respectively. The antibodies excess was then removed, membranes were washed three times with TBS-t and incubated, for 1 h at room temperature and under shaking, with a mouse peroxidase-conjugated secondary antibody (1:2500; Thermo-Scientific).

The signal was detected using an enhanced chemiluminescent substrate and LAS 4000 (GE Healthcare, Waukesha, WI, USA) digital imaging system.

Afterwards, the membranes were also hybridized with an anti-Glyceraldehyde 3-Phosphate Dehidrogenase antibody (GAPDH, 1:2000, mouse, Invitrogen) and the signal was detected as already described.

DARTS Western Blottings were repeated twice and a densitometric analysis was also performed through ImageJ. Data were elaborated averaging the values from the independent replicates, using GAPDH as a loading normalizer and rating undigested GRP75/HS71A intensity as 100%. The obtained values were represented through GraphPad Prism 7.

7.4.8 T-LiP-MRM: GRP75/HS71A transitions computational prediction

GRP75 (UniProt accession number: P38646) tryptic peptides previously detected by MS where selected through the proteomics data resource *Peptide Atlas* (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas) on its Human build and queried into the complete Human *SRM Atlas* build (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/GetTransitions), to retrieve their best daughter ions.

The *SRMAtlas* query parameters were set as follows:

- number of highest intensity fragment ions to keep: 8;
- target instrument: QTRAP 5500;

- transitions source: QTOF, Agilent QQQ, Qtrap5500, Ion Trap, Predicted;
- precursor exclusion range: kept blank;
- search proteins form: SwissProt;
- duplicate peptides: unique in results;
- heavy label: kept blank;
- labeled transitions: kept as default;
- maximum m/z: 1100 Da;
- minimum m/z: kept blank;
- allowed ions types: b-ions and y-ions;
- allowed peptides modification: carbamidomethylation of cysteines (C[160]).

HS71A (UniProt accession number: P0DMV8) tryptic peptides, not being disposable as PABST, were directly queried into the complete Human *SRM Atlas* build (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/GetTransitions), to retrieve their best daughter ions, as previously reported for GRP75.

The obtained list, containing 35 precursors and 280 fragments for GRP75 and 26 precursors and 208 fragments for HS71A, was subsequently refined as follows:

- peptides whose following C-terminus aminoacid was either K or R were removed;
- among the 8 fragment ions reported for each precursor, only the three most intense were selected preferring, when possible, the y-series ions over the b-series ones.

Thus, a comprehensive method listing 26 and 17 peptides for GRP75 and HS71A respectively, each with their three best transitions, was obtained and subsequently tested onto a HeLa lysate tryptic digest.

7.4.9 T-LiP-MRM: GRP75 and HS71A best transitions selection

300 µg of HeLa cell lysate obtained as described before (pharagraph 7.1.1.1), were submitted to an *in solution* digestion protocol. Briefly, proteins were denatured using 8 M urea/50 mM AmBic (4 M final urea concentration), disulphide bonds were reduced with 10 mM DTT for 1h at 25°C and 800 rpm (Thermomixer, Eppendorf) and then alkylated with 20 mM IAA for 30 minutes, at 25°C and 800 rpm, in the dark (Thermomixer, Eppendorf). IAA was then quenched with 10 mM DTT (10 minutes, 25°C, 800 rpm) and urea was diluted up to 1M with 50 mM AmBic before adding the trypsin/LysC solution (Promega, Madison, Wisconsin) at the enzyme to proteins ratio of 1:100 w/w. Digestion was allowed to proceed overnight at 37°C under continuous shaking (800 rpm, Thermomixer, Eppendorf) and then quenched adding FA to lower the pH to 3. The peptides mixture was then dried under vacuum (SpeedVac Concentrator Plus, Eppendorf), dissolved in 1 mL 5% FA and desalted through a Sep-Pak C18 1 cc (50 mg) cartridge (Waters, Milford, MA, USA).

Briefly, the cartridge was activated flushing 3 mL of 100% CH_3CN and then conditioned with 3 mL of 0.1% FA. The sample was then loaded, desalted flushing the cartridge with 3 mL

of 0.1% FA and finally eluted flushing two times 500 µL of 80% CH_3CN, 20% H_2O, 0,1% FA. For the subsequent MS analysis, the peptides mixture was dried under vacuum (SpeedVac Concentrator Plus, Eppendorf) and re-dissolved in 10% FA.

UPLC–ESI-MRM-MS analyses were performed on a 6500 Q-TRAP from AB Sciex equipped with Shimadzu LC-20A and Auto Sampler systems. UPLC separation was performed on an Aeris Widepore XB C18 column (150 × 2.10 mm, 3.6 µm XB, Phenomenex, Torrance, USA), using 0.1% FA in H_2O (A) and 0.1% FA in CH_3CN (B) as mobile phases, and a linear gradient from 5 to 95% of B over 30 min (flow rate: 200 µL/min). Q-TRAP 6500 was operated in positive MRM scanning mode, with declustering potential (DP) set at 80V, entrance potential (EP) at 10V, collision energy (CE) at 35V and cell exit potential (CXP) at 22V.

LC–ESI-MRM/MS runs were performed injecting 15 µg of the peptide mixture: the XICs of all the transitions of each precursor were inspected to (1) identify GRP75 and HS71A tryptic peptides which could actually be experimentally observed, (2) assign the retention time to all of the peptides and (3) identify the best transition as the one showing the most intense peak and the best signal to noise ratio.

This preliminary experiment led to a global MRM method comprising 20 transitions for GRP75 and 11 transitions for HS71A.

The observed peptides were then mapped onto both of the Hsp70 isoforms UniProt-retrieved sequences through the PAWS program, with the following parameters:

- cleave: K/R;
- mass: average;

- fixed modification: cysteins carbamidomethylation;
- search tolerance: 300 ppm.

Both of the isoforms coverage (%) was also calculated as follows:

Coverage (%) = (Experimentally mapped aminoacids/total aminoacids)*100.

7.4.10 T-LiP-MRM experiment

300 µg of HeLa cells proteome extracted as previously reported (pharagraph 7.1.1.1) were incubated with DMSO or BrACG (5 µM and 50 µM final concentrations, final DMSO amount 1% vol/vol), for 1 hour at room temperature and under continuous agitation (Bio RS-24 Mini-Rotator, BioSan). The samples were then splitted and submitted to limited proteolysis with 1:1000 w/w and 1:500 w/w subtilisin to proteins ratios, respectively. Limited proteolysis was carried out for 30 minutes at 25°C and 500 rpm (Termomixer, Eppendorf). 300 µg of a DMSO-treated lysate aliquot were submitted to a mock proteolysis, carried out adding H_2O instead of the enzyme.

Subtilisin was then quenched with PMSF (1 mM final concentration) and the samples shifted to denaturing condition adding urea (4 M final concentration) to perform in solution digestion and desalting, as described before.

15 μg of each sample were then injected in the LC–ESI-MRM-MS system and analyzed through the previously optimized MRM method. The area of each GRP75 and HS71A tryptic peptide peak was then measured using the Analyst Software from AB Sciex.

Two biological replicates were performed, as well as injection replicates.

BrACG protected peptides were mapped onto the following PDB structures: pdbID 2e88[251] and 5gjj[272] for HS71A and 4kbo[252] and 3n8e[273] for GRP75.

7.4.11 T-LiP-MRM data corroboration: blind molecular docking analysis

The prediction of binding between human HSP70s and BrACG was carried out performing a molecular docking analysis including the ligand and protein preparation, the genetic algorithm (GA) execution, and the data analysis and image preparation. The ligands were designed, including the addition of tautomeric states, partial charges and protonation, and finally minimized using the Avogadro software (version 1.2.0) 1 with a universal force field, UFF, and a conjugate gradient algorithm until a ΔE lower than 0.001 kJ/mol, as previously reported[222]. The protein three-dimensional structures were obtained from the Protein Data Bank[221] (pdbID: 2e88[251] and 5gjj[272] for HS71A and 4kbo[252] and 3n8e[273] for GRP75) and prepared, using the Hermes software (version 1.10.0)[261] incorporating the Gasteiger(-Marsili) partial charges, adding polar protons and removing crystal waters and extra co-crystallized ligands. Moreover, all planar R-NR1R2 were made available for the cis/trans flipping and the tautomeric states of

Asp, Glu and His residues were adjusted. GOLD (version 5.7.0)[261] was performed to achieve the molecular docking using the ChemScore as scoring function (that consists of protein-ligand hydrogen bond energy (external H-bond), protein-ligand VdW energy (external), ligand internal VdW energy, and ligand torsional strain energy (internal torsion), search efficiency at 200% (very flexible), selecting all atoms within 20Å from the centroid (covering almost completely the whole protein), 20 GA runs and other parameters as default. The resulting ChemScore ΔG, the total free energy change of the system upon ligand binding, and the relationship between this score and experimental free energy of binding, previously obtained[274], were used to calculate the predicted equilibrium dissociation constant $K_{D,pred}$. The best complex geometry, on the base of the ChemScore and the ChemScore ΔG, was rendered using PyMol software (The PyMOL Molecular Graphics System, Version 2.0.4 Schrödinger, LLC.) whereas the 2D representation was created using PoseView server[275].

CHAPERT 8: Appendix

Rise of the machines: probing INTERcellular crosstalk with mass spectrometry.

Morretta, E., Demmers, L., Rontogianni, S., Damen, J. M. A., Wu, W.

Presented at

25th biennial congress of the European Association for Cancer Research, Amsterdam, The Netherlands, 30 June- 03 July, 2018.

8.1 Preface

During the second year of my PhD I had the thrilling and highly formative opportunity to spend six months in the Albert Heck *Biomolecular Mass Spectrometry and Proteomics Laboratory* at Utrecht University, under the direct supervision of Professor Wei Wu.

Throughout this period, my principal investigation area was focused onto the analysis of the Plasma Membrane (PM) and Extracellular Matrix (ECM) proteomes of a panel of 10 breast (cancer) cell lines through *label-free* proteomics, to start shedding light onto their intercellular crosstalk features responsible for the first step of the metastatic process, the detachment.

In the following paragraphs, I will give some insights into the results I obtained, particularly focusing on the PM/ECM contacts of two triple negative breast cancer cell lines (TNBC), MDA-MB-231 and the derivative LM2 which is, as opposite to its parental cell line, rapidly and efficiently metastatic to the lungs. MDA-MB-231 and LM2 different composition in both PM and ECM proteins could explain their strictly different metastatic properties and tropism.

8.2 Background

The local microenvironment, or *niche*, of a cancer cell plays pivotal roles in its development, survival and spreading. Indeed, in order to metastasize, cancer cells need a permissive microenvironment allowing them to grow and, eventually, detach from their primary site in order to reach a secondary site, where they continue to grow producing a metastatic mass (Figure 104).

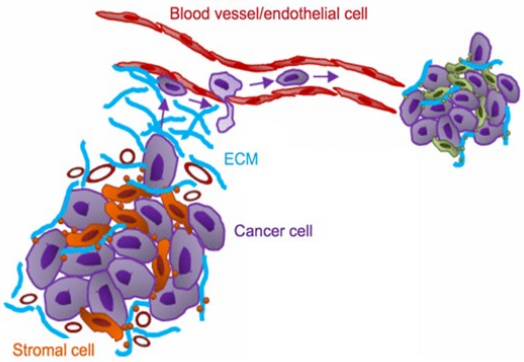


Figure 104: *Schematic and simplified representation of metastatic cells behavior.*

Among the protein equipment of a cell, PM and ECM proteins are easily perceivable as the main docking parties involved in cancer cell metastasis[276–280]. Indeed, PM proteins are directly responsible for the cell communication with the external environment and their interaction with the related ECM counterparts is essential for the definition of cancer cells fate.

As an example, PM integrins are unique multidirectional signaling molecules which activation and binding to the ECM triggers the recruitment of the so-called *adhesome*, a complex array of signaling, scaffolding and cytoskeletal proteins[281–283]. In concert with the PM and ECM proteins, the adhesome forms a complex and highly dynamic machinery responsible for cell survival, migration, polarity and differentiation[284].

Thus, the concomitant analysis of a cancerous cell PM and ECM proteins equipment is a promising strategy to gain functional insights into cells metastatic behavior and tropism.

8.3 PM and ECM isolation strategies from cultured cells

PM and ECM proteins are challenging analytes: compared to their intracellular counterparts, they are less abundant and poorly soluble, thus needing to be both enriched and handled with

shrewdness for a satisfactory analysis through the commonly available mass spectrometric techniques.

Regarding ECM proteins, after collecting cultured cells and thoroughly washing away serum proteins, they can be solubilized and retrieved by scraping the remaining plate content in urea-comprising buffers. The further processing through *in solution* digestion ensures the production of soluble peptides.

PM proteins, on the other hand, can be obtained following several enrichment procedures, among which differential centrifugation, immunoprecipitation and liquid-two-phase partitioning are the most common ones[285].

Differential centrifugation separates subcellular compartments based on size, shape or density. Usually, cell lysates are first centrifuged at a speed that sediments only cell components larger and denser than the desired one, then another centrifugation step at a higher speed is performed to mainly pellet the compartment of interest. This technique is rapid and simple, but the isolated material often results too heterogeneous. To overcome this limitation, differential centrifugation is often combined with density gradient centrifugation, which separates particles mainly due to differences in their *buoyant density*: particles move in a density gradient under the centrifugal field force until their density equals the medium one (i.e. point of isodensity), where further centrifugation will not cause any additional sedimentation. Even if this process should ensure homogenous PM isolation, heterogeneity within different cellular membranous compartments of similar densities is still an issue. Thus, different membranes are not readily separated and several time- and material-consuming combinations of differential and density gradient centrifugations are required, which ultimately results in low yields due to sample loss.

Immunoprecipitation takes advantage of specific PM proteins interactions with a solid phase presenting highly selective antibodies. However, this approach is expensive and nonspecific

adsorption of contaminating membranes can still occur. Furthermore, elution of the PM from the solid support requires harsh conditions, which render most proteins functionally inactive.

Thus, among all of the currently exploited approaches, liquid-two-face partitioning results as the most selective and efficient one, requiring low starting material amounts and ensuring protein structure and biological activity are preserved after the enrichment, due to the mild environment, low interfacial tension, non-denaturing conditions and high water content provided by this system[286].

Liquid-two-face partitioning is based on the evidence that, above a certain concentration (i.e. *critical concentration*), most aqueous mixtures of two structural distinct water-soluble polymers give rise to a two-phase system with each phase enriched in one of the two polymers. The most widely employed two-phase systems consist of poly-(ethylene glycol) (PEG) and dextran, with the top phase enriched in PEG and the bottom phase in dextran.

After adding biomaterial to the system and mixing, phases settle and thereby separate the different biomaterial components on the basis of their affinity for either PEG or dextran, attributed to the phospholipid composition[287] of the various membranes. More in details, PM shows the highest affinity for the more hydrophobic PEG top phase, followed by Golgi vesicles, lysosomes, endoplasmic reticulum and mitochondria[288]. For a higher PM purity, more than one partitioning step is necessary: countercurrent distribution (CD) experiments have to be performed, taking advantage of the fact that the partitioning coefficient of particles remains constant during multiple extractions performed under identical conditions. Therefore, re-extracting both the top and the bottom phases of an already partitioned system with fresh bottom and top phases, respectively, enhances the recovery of PMs as well as their purity over other subcellular membranous components[285]. Once extracted, PM proteins can be retrieved by the PEG phase by precipitation and a following *in solution* digestion can be performed, allowing the production of soluble peptides.

8.4 Experimental setup

Breast cancers (BCs) are the most common malignant disease among Western countries women: with 81% of them being invasive, metastasis is the cause of death in the majority of BC patients.

Thus, to identify signatures in ECM and PM proteins that might help shedding light on different BC subtypes metastatic features, we isolated ECM and PM from nine different breast cancer cell lines: one non-metastatic low grade luminal-type (MCF7), four triple negative (MDA-MB-231, LM2, Hs578T, BT549) and four HER2 positive/ER negative (HCC1954, HCC1419, JIMT1, SKBR3). Furthermore, ECM and PM derived from a cell line resembling normal breast epithelial cells (MCF10A) were also characterized, as a benign control.

Our experimental plan, summarized in Figure 105, was elaborated to obtain from each cell type the pairing PMs and ECMs: to compare their protein contents across the 10 breast (cancer) cell lines, they were digested *in-solution* and analyzed by liquid chromatography-tandem mass spectrometry (nanoESI LC-MS/MS) on a high-resolution mass spectrometer (Q-Exactive HF). The obtained data were submitted to the MaxQuant[149] software for both proteins identification and *label-free* quantification and the subsequent statistical analysis performed through Perseus[289] and several other bio-informatics tools (see next paragraphs).

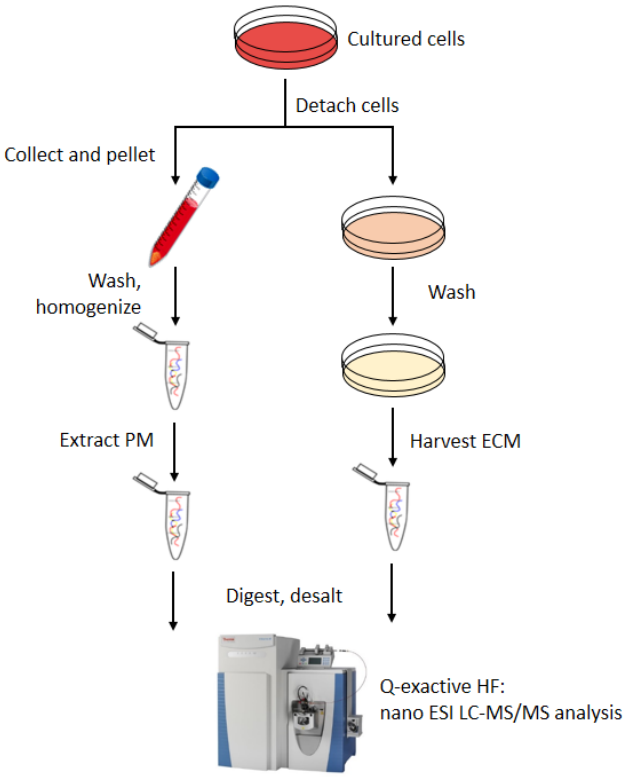


Figure 105: *Schematic representation of the experimental strategy providing the concomitant isolation of ECM and PM proteins.*

8.4.1 PM and ECM proteins isolation and handling

For the concomitant isolation of both ECMs and PMs, the 10 cultured cell lines were detached through trypsinization and the obtained pellets washed in PBS and collected. The remaining plates were then extensively washed to remove serum proteins and then ECMs were obtained by scraping them off from the plates in 8 M Urea/50 mM AmBic.

For the cleanest possible PMs isolation, the 10 cell pellets were subsequently submitted to an extraction procedure based on the consecutive application of differential centrifugation and liquid-two-phase partitioning.

More in details, each cell pellet was suspended in a mild aqueous buffer supplemented by a protease inhibitor cocktail and lysed mechanically in a Dounce homogenizer.

The obtained suspensions were then submitted to differential centrifugation: first they were centrifuged at low speed to pellet nuclei and debris, then the obtained supernatant were centrifuged again at high speed, to pellet both cells inner (i.e. organelle) and plasma membranes. The total membranes pellets were then submitted to liquid-two-phase partitioning through the CD approach: a first partitioning round was performed re-suspending the membranes in the lower density organic phase and extracting it with the higher density one, which was then isolated and re-extracted with fresh organic phase. The two obtained organic phases were combined and extracted again.

Thus PM proteins, enriched in the organic phase, were precipitated and re-suspended in 8 M urea/50 mM AmBic. Protein concentration was then determined for both PM and ECM proteins and *in solution* digestion and desalting were performed to produce tryptic peptides mixtures suitable for the subsequent high resolution nanoESI LC-MS/MS analysis. The obtained MS data were then submitted to *label-free* quantification through the software MaxQuant.

8.4.2 MaxQuant software: PM and ECM proteins label-free quantification

MaxQuant[149] (http://www.biochem.mpg.de/5111795/maxquant) is a freely available quantitative proteomics software package designed for the analysis of large-scale mass-spectrometric data sets, making use of the Andromeda search engine to accomplish proteins identification and supporting all the main quantitative proteomics labeling techniques (e.g. SILAC, SIDL, TMT and iTRAQ) as well as *label-free* quantification approaches.

In a generic data analysis workflow, MaxQuant first corrects for systematic inaccuracies of both measured peptide masses and the corresponding retention times, then peptide peaks mass and intensity are detected and assembled to identify isotope patterns for peptides identification.

For this purpose, peptide and fragment masses are found in an organism specific sequence database and scored by a probability-based approach termed *peptide score*. For restraining a certain number of false positives, the FDR (i.e. false discovery rate) approach is utilized: the organism specific database search not only includes the target sequences, but also their reverse counterparts and contaminants, which helps to determine a statistical *cut-off* for acceptable spectral matches. The last step of the data analysis consists of the assembly of peptide hits into protein hits.

Among all the supported quantification strategies, MaxQuant provides an accurate and robust proteome-wide *label-free* quantification (LFQ) approach, based on peptide peaks intensity determination and normalization and capable of handling a large set of experiments in a manageable computing time.

Thus, the huge amount of high resolution mass spectra obtained from the nanoESI LC-MS/MS analysis of both PM and ECM datasets were separately submitted to a *label-free* quantification (LFQ) process through MaxQuant, allowing the LFQ intensities calculation for each of the identified proteins and thus their comparison among different samples.

8.5 PM proteins dataset

Among the two datasets, we started working on the PM one first. MaxQuant analysis of the PM protein content across the 10 breast (cancer) cell lines allowed the identification of 5438 proteins (1% FDR), differentially distributed and quantified across all of the cell lines (Figure 106A). To check the quality of our data, run-to-run correlation was first evaluated through Perseus to assess the reproducibility of our replicates. As can be observed in Figure 106B, where just an exemplificative extract of our dataset multi-scatterplot is presented, the Pearson correlation coefficient between technical replicates (i.e. a measure of the linear correlation

between two variables, where +1 is total positive linear correlation, 0 is no linear correlation, and −1 is total negative linear correlation) is higher than 0.9, thus pointing out the robustness of our analytical method.

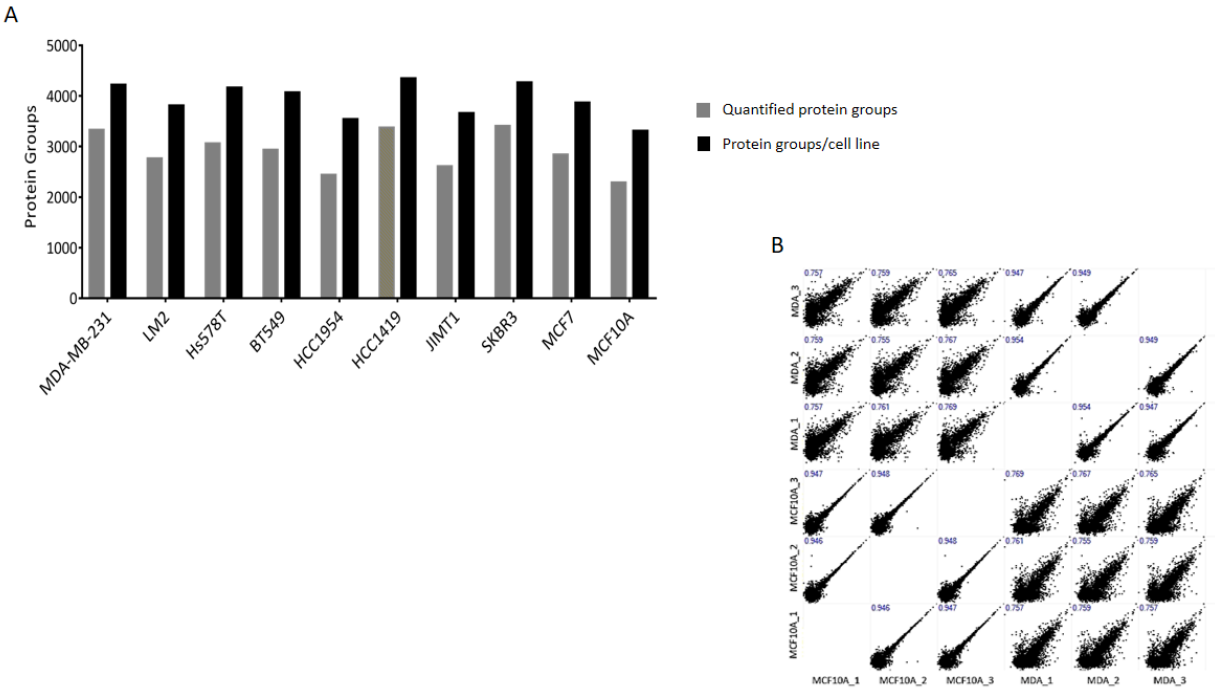


Figure 106: *(A) Bar plot of the total number of identified (black) and quantified (gray) PM protein groups from each cell line. (B) Multi-scatterplot of the mass spectrometric data from the cell lines MDA-MB-231 (i.e. MDA) and MCF10A, showing optimal run-to-run correlation among technical replicates.*

Then, to gain some proofs about the quality of our PM preparations, a Gene Ontology (GO) enrichment analysis was performed through the DAVID database[290,291]. Briefly, this is a system for hierarchically classifying genes or gene products in a graph structure (i.e. an ontology) retrieving, in a ranked list, the set of terms used to describe any gene or gene product and their corresponding p-values.

For a clearer picture of our enrichment quality, the full proteome from the same 10 cell lines, previously analyzed by my supervisor subgroup[292], was exploited as a background proteome, for DAVID to perform the enrichment analysis on the basis of the un-compartmentalized whole protein equipment of these cells. As can be observed in Figure 107, presenting the top 15 GO enriched terms for the *Cellular Component* (A) and *Biological Process* (B) categories, the major part of the most significantly enriched terms ($p<10^{-10}$) are either related to the PM compartment (e.g. membrane, cell periphery, plasma membrane, anchoring junctions) or to biological processes involving PM proteins (e.g. transport, biological adhesion, cell adhesion, membrane organization).

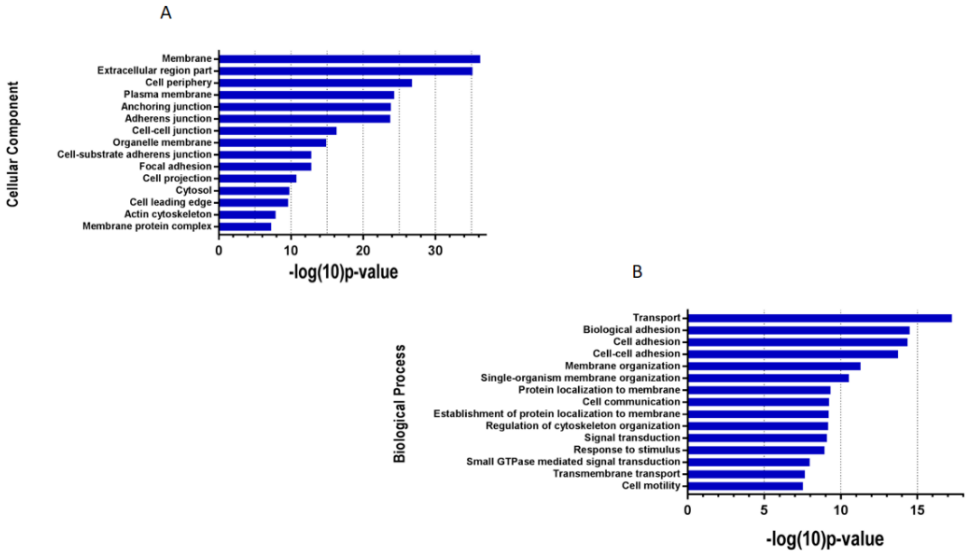


Figure 107: *Gene ontology enrichment analysis of the Cellular Component (A) and Biological Process (B) terms of the PM proteins, performed using the DAVID database and the full proteome of the same cell lines as background.*

Satisfied by our PM preparations, we proceeded to evaluate the breast cancer subtype-specific PM features through hierarchical clustering, with the aim of identifying common features shared among the four triple negative and among the four HER2 positive/ER negative

cells, which could differentiate their behavior from both the non-metastatic low grade luminal-type MCF7 and the normal breast epithelial-resembling MCF10A cells.

As can be observed from the Perseus obtained heatmap in Figure 108A, the triple negative PM proteins showed a good clustering (i.e. same heatmap branching point shared), whereas no clear trend was observed for the HER2 positive ones (i.e. different branching points classification), probably due to their different HER2 amplification grade. Thus, we decided to primarily focus our attention on the triple negative cells, performing a GO enrichment analysis on the PM proteins comprised in the cluster (white box in Figure 108A).

As can be observed in Figure 108B, the term *Focal Adhesion* (FA) resulted the most significantly enriched one ($p<10^{-25}$), suggesting that the FA complex could be responsible for the triple negative cells metastatic properties.

FAs are integrin-containing multi-protein structures found at the end of stress fibers, which form mechanical links between the intracellular actin bundles and both PM receptors and the ECM. Furthermore some FAs components, like FAK and Src, are signaling molecules: both their non-receptor tyrosine kinase activity and adaptor function are fundamental for the integrin signaling pathway, culminating in the reorganization of the actin cytoskeleton, a prerequisite for changes in cell shape and motility. Similar morphological alterations are initiated by the binding of growth factors to their respective PM receptors, emphasizing the considerable crosstalk between adhesion- and growth factor-mediated signaling.

It is thus not difficult to picture that, in cancer cells, FAs regulate the formation of both *invadopodia* and cell-matrix attachment sites, which play essential roles in cell motility, proliferation and survival.

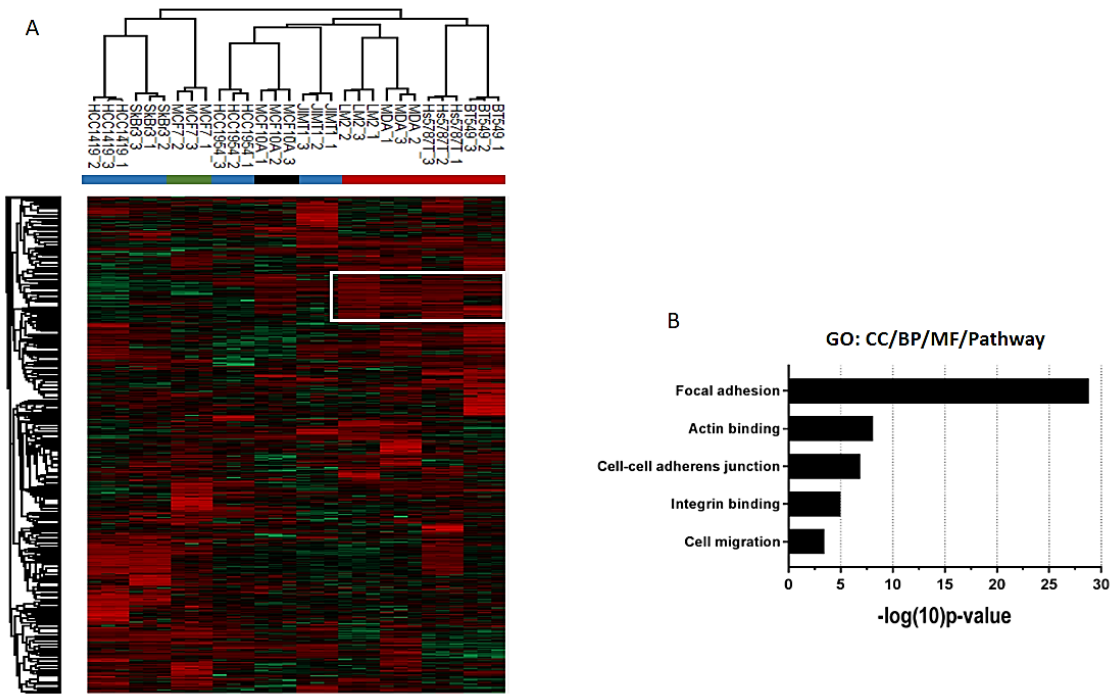


Figure 108: *PM proteins hierarchical clustering, color-coded on top on the basis of different cell subtypes (red: triple negative; blue: HER2 positive/ER negative; green: low grade luminal-type; black: normal breast epithelia resembling cell line). White box shows triple negative cells shared protein cluster. Red: Z-scored intensity higher than zero. Green: Z-scored intensity lower than zero. (B) GO enrichment analysis of the proteins shared by all of the four triple negative cell lines, performed through DAVID. The presented data are related to the terms 'cellular compartment', 'biological process' and 'molecular function'.*

Thus, with the aim of individuating proteins responsible for the different metastatic behavior of these cells, we sought to first restrict our investigation field to the FA complex of the couple MDA-MB-231/LM2. In fact, despite being obtained from MDA-MB-231 by injection of the latter into the tail vein of immunodeficient mice after two rounds of *in vivo* selection, LM2 cells show a very strong tropism for the lungs, as opposite to their parental cells preferentially metastasizing to the bones.

For this purpose, the previously identified proteins were investigated to enlighten their trend in both MDA-MB-231 and LM2, performing a hierarchical clustering in respect to the normal breast epithelia-resembling MCF10A cell line.

As can be observed in Figure 109A, the PM presented proteins are quite different between LM2 and MDA-MB-231, but the major part of these differences is not shared with MCF10A.

Thus, we moved on to directly compare MDA-MB-231 and LM2 PM proteins (Figure 109B): among all the significantly down- and up-regulated proteins, DAVID GO enrichment analysis revealed that 68% were related to cell adhesion (Figure 109C).

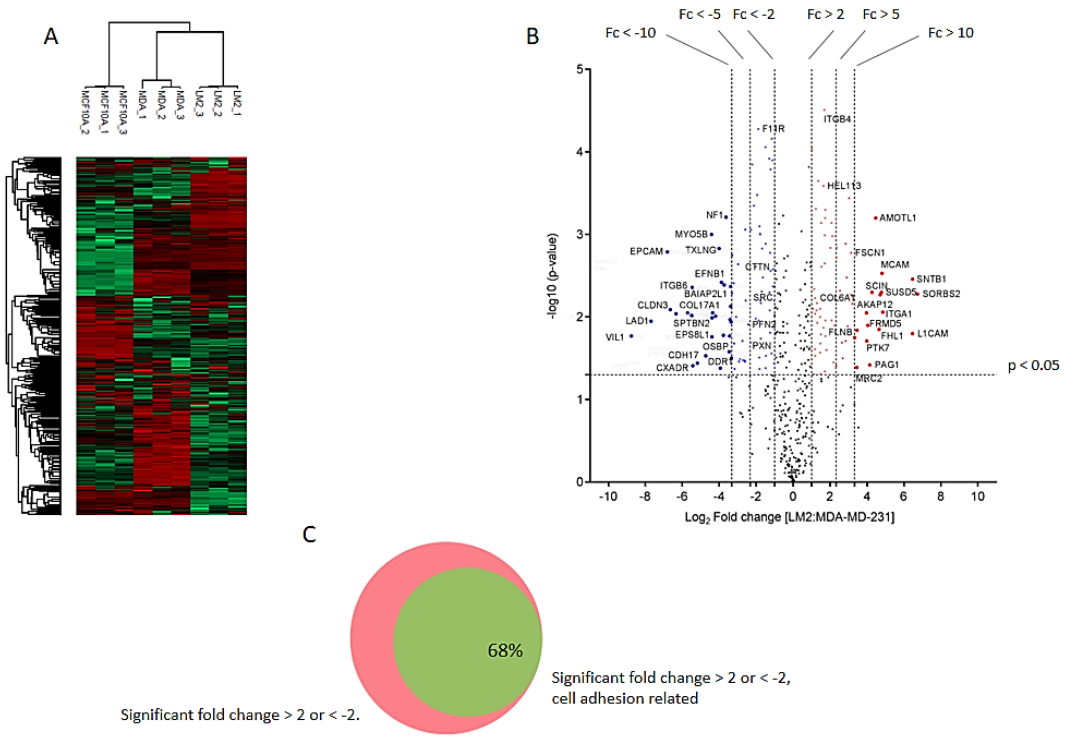


Figure 109: *(A) MDA-MB-231, LM2 and MCF10A PM proteins hierarchical clustering. Red: Z-scored intensity higher than zero. Green: Z-scored intensity lower than zero. (B) Volcano plot of significance versus fold change of LM2 versus MDA-MB-231 PM proteins. Proteins having Fc>2 are represented through red dots, whereas proteins having Fc<-2 through blue dots. Proteins showing a fold change >10 or <-10 are reported through bigger*

dots and with their related gene name. (C) Venn diagram showing the number of cell adhesion-related proteins presenting significant fold changes ($p < 0.05$).

Nevertheless, LM2 up regulated proteins didn't show a high degree of functional correlation when analyzed through the STRING web tool, whereas the down regulated ones did: key proteins involved in the FA maturation process all resulted down regulated in LM2 (Figure 110). More in details integrins α2, β5 and β6, trans-membrane receptors mediating cell-cell and cell-ECM contacts, the organization of the intracellular cytoskeleton and the movement of new receptors to the cell membrane to rapidly respond to events at the cell surface, are down-regulated by 3.1, 4.7 and 11 folds respectively. Furthermore paxilin, a scaffold protein recruiting many structural and regulatory proteins to FAs, is down-regulated by 2.5 folds, its activator Src, a non-receptor tyrosine kinase essential for recruitment of proteins during FA maturation, by 2.6 folds and cortactin, essential for actin polymerization, by 2.2 folds.

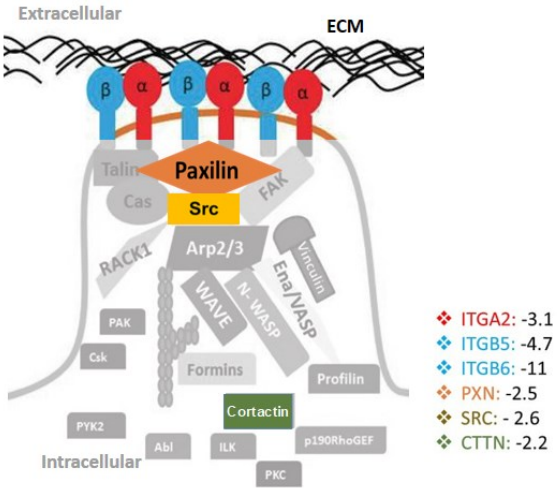


Figure 110: *Schematic representation of FA proteins showing key FA components are down-regulated in LM2 versus MDA-MB-231. ITGA: integrin α; ITGB: integrin β; PXN: paxilin; SRC: proto-oncogene tyrosine-protein kinase Src; CTTN: cortactin.*

Since all of the aforementioned proteins are involved in several key steps of FA maturation and function, their down-regulation should severely impair LM2 cell-ECM contacts. Thus, to address this point we moved onto the analysis of the ECM proteins data set.

8.6 ECM proteins dataset

ECM is a complex and dynamic meshwork of extracellular proteins providing architectural, mechanical and biochemical signals interpreted by cell-surface receptors and other PM poteins, thus orchestrating cell adhesion, migration and proliferation. To achieve a bigger picture of the metastatic properties of our panel of 10 breast (cancer) cell lines, their deposed ECM was also analyzed.

MaxQuant analysis of the ECM proteins across the 10 cell lines and the subsequent filtering to remove serum contaminants allowed the identification of 2780 differentially distributed and quantified proteins (1% FDR, Figure 111A). GO enrichment analysis through DAVID was then performed, exploiting the full proteome of the same cells as a background, to evaluate the quality of our ECM preparations. As can be observed in Figure 111, presenting the top 15 GO enriched terms for the *Cellular Component* (B) and *Biological Process* (C) categories, the major part of the most significantly enriched terms ($p<10^{-15}$) are either related to the ECM compartment (e.g. extracellular region, exosome, vesicle, matrix and space) or to biological processes involving ECM proteins (e.g. biological and cell adhesion, extracellular structure and matrix organization).

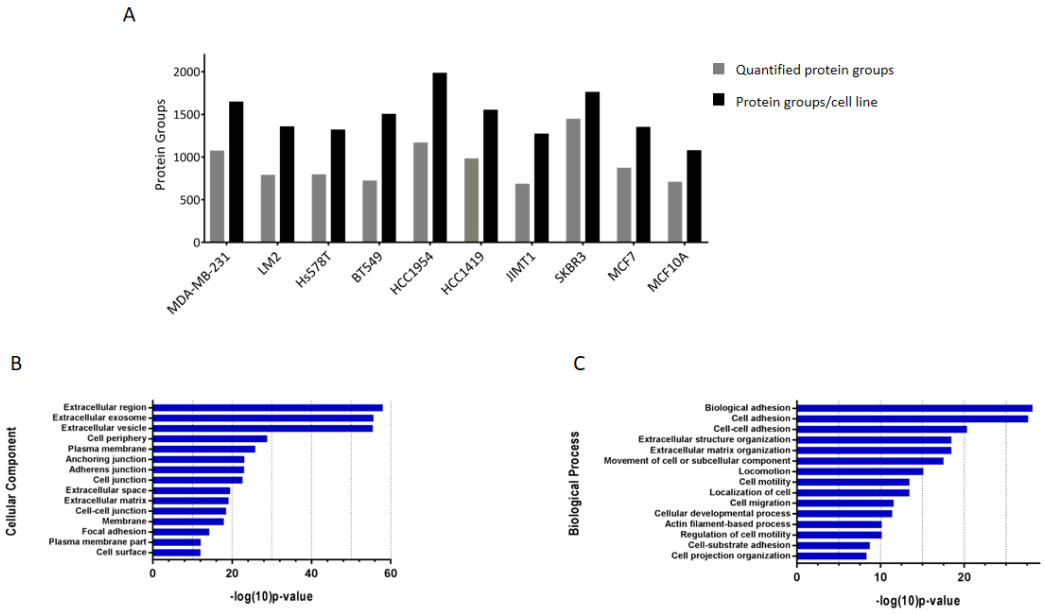


Figure 111: *(A) Bar plot of the total number of identified (black) and quantified (gray) ECM protein groups from each cell line. Gene ontology enrichment analysis in Cellular Component (B) and Biological Process (C) terms of the ECM proteins, performed using the DAVID database and the full proteome of the same cell lines as background.*

We then proceeded to evaluate breast cancer subtype-specific ECM features through hierarchical clustering. As already reported for the PM proteins (Figure 112A) and shown by the ECM-related Perseus heatmap in Figure 112A, a good clustering was revealed for the triple negative ECM proteins (i.e. same heatmap branching point shared), whereas no clear trend could be observed for the HER2 positive ones (i.e. different branching points classification). Furthermore, when closely inspecting the triple negative cells, we noticed that the couple MDA-MB-231/LM2 did not cluster as well as in the PM data set. Indeed, the Perseus obtained PM heatmap (Figure 112A and cropped image in Figure 112B) showed these two cell lines sharing the same branching point and thus clustering the best together, whereas the in the ECM such a

phenomenon could not be observed (red arrows in Figure 112A), suggesting MDA-MB-231 and LM2 depose different ECM proteins.

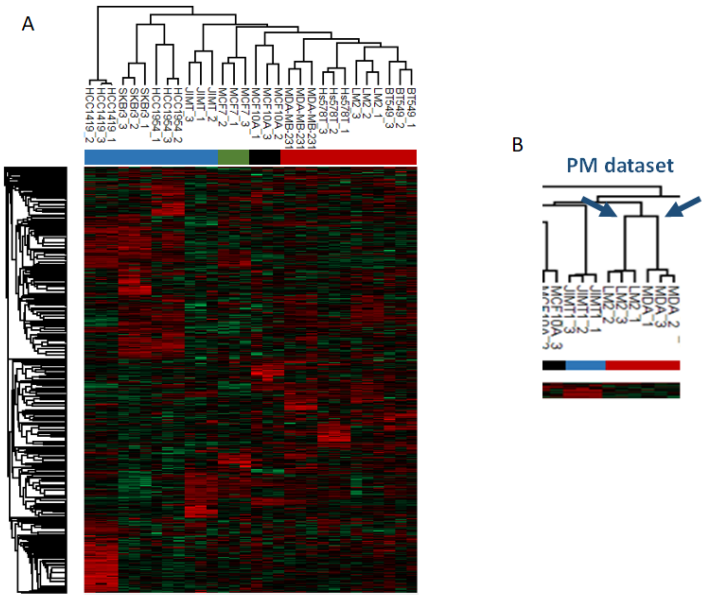


Figure 112: *ECM proteins hierarchical clustering, color-coded on top on the basis of different cell subtypes (red: triple negative; blue: HER2 positive/ER negative; green: low grade luminal-type; black: normal breast epithelia resembling cell line). Red: Z-scored intensity higher than zero. Green: Z-scored intensity lower than zero. Red arrows indicate MDA-MB-231 and LM2 branching points. (B) Cropped Figure 108A: PM proteins related heatmap showing MDA-MB-231 and LM2 share the same branching point (blue arrows).*

Thus, as previously done for the PM data set, we decided to primarily focus our attention on the triple negative cells MDA-MB-231 and LM2, for a deeper analysis of their metastatic properties through the investigation of the strictly different ECM they depose.

As shown in the volcano plot in Figure 113, LM2 ECM up-regulated proteins are fewer than the down-regulated ones. Furthermore, the latter present several ECM key components, among

which laminins (pointed by red arrows in the figure) are down-regulated from 5 to 286 folds, resulting as the overall most down-regulated protein group in LM2 ECM proteins.

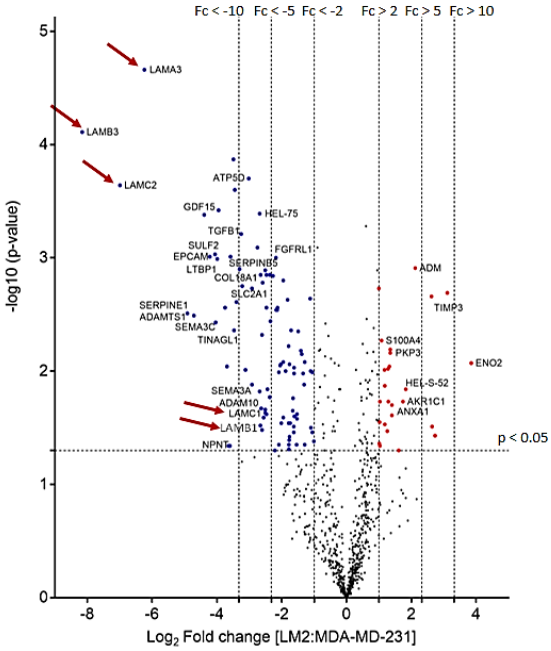


Figure 113: *Volcano plot of significance versus fold change of LM2 versus MDA-MB-231 ECM proteins. Red arrows point at laminin isoforms.*

Laminins are a major component of the ECM basal lamina and are responsible of ECM/ PM contacts, binding to cell membranes through integrin receptors and other PM proteins and being, at the same time, associated with ECM collagens networks. Thus, this data alone seems to suggest that LM2, deposing less basal lamina, should present impaired PM contacts.

For a wider profiling of LM2 ECM/PM contacts, a Gene Set Enrichment Analysis (GSEA) was performed through the homonymous program[293].

More in details, GSEA determines whether an *a priori* defined set of genes or gene products shows statistically significant and concordant differences between two biological states (e.g.

cell types, physio/pathological conditions), analyzing whether the majority of genes falls in the extremes of the set itself: the top and bottom of the set correspond to the largest expression differences between the two biological states.

If the gene set falls at either the top (over-expressed) or bottom (under-expressed), it is thought to be related to phenotypic differences. More in details, GSEA first performs gene ordering by expression difference to than calculate the genes enrichment score (ES) and the related statistical significance. Finally, ES normalization for each set and FDR calculation is achieved.

LM2 ECM proteins were thus submitted to GSEA in comparison with the MDA-MB-231 ones, revealing a negative enrichment (i.e. down-regulation) in the *ECM-Receptor interactions* and *Focal adhesion* pathways, confirming the previously supposed impairment of LM2 PM/ECM contacts (Figure 114).

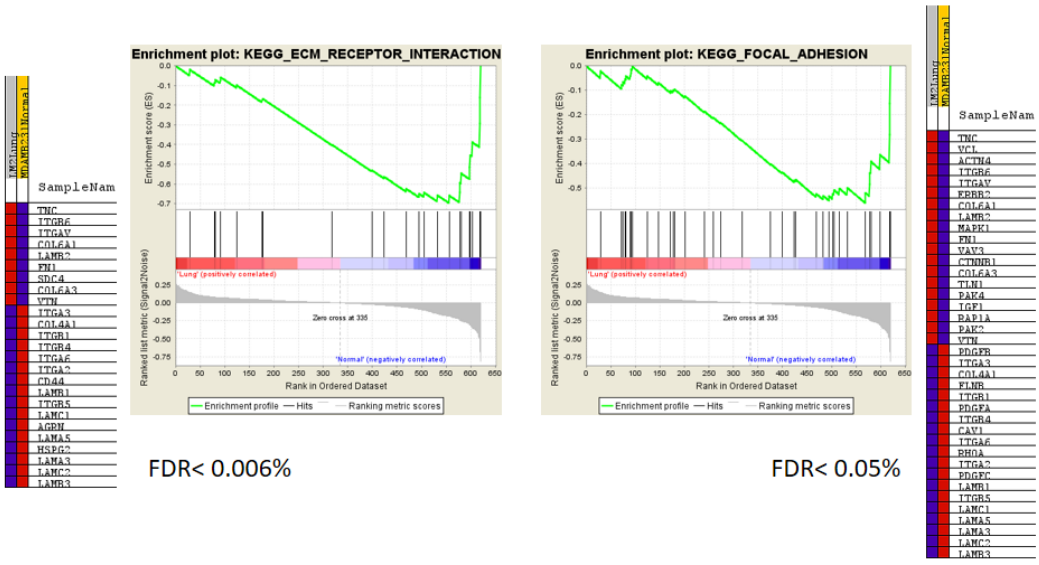


Figure 114: *GSEA output for the KEGG pathways category, reporting the negative enrichment plots with the relative genes and the overall FDR for the Focal adhesion and ECM-Receptor interactions pathways. Red: up-regulated hits; blue: down-regulated hits.*

8.7 MDA-MB-231 and LM2 PM and ECM proteins: rebuilding the puzzle

Both the PM and ECM datasets analysis of MDA-MB-231 and LM2 cells revealed LM2 down-regulates key components of the cell/ECM adhesion machinery, thus we sought to draw one bigger picture comparing the interaction networks of this machinery key proteins identified in both the data sets.

For this purpose, the LM2 significantly down-regulated PM and ECM proteins were analyzed together through the web tool STRING[294], imposing the highest possible confidence (0.900) for the minimum required interaction score, in order to visualize the most reliable interaction network possible. The output was than submitted to a GO annotation for *Biological Process* terms: proteins involved in *cell localization establishment* (red, FDR= 8.82×10^{-17}), *membrane organization* (blue, FDR= 1.44×10^{-7}), *regulation of locomotion* (green, FDR= 5.02×10^{-6}) and of *cell adhesion* (purple, FDR= 4.7×10^{-4}), and in *actin cytoskeleton* (dark green, FDR= $4. \times 10^{-4}$), *extracellular matrix* (light blue, FDR= 9.6×10^{-4}) and *cell junction* organization (yellow, FDR= 9.8×10^{-4}) were highlighted.

As can be observed in Figure 115, a huge interaction network was retrieved from the adhesion machinery related proteins of the two data sets, with the major part of them being involved in the organization of plasma membrane, cytoskeleton and extracellular matrix, and thus being fundamental for correct focal adhesions maturation processes and cell/cell or cell/ECM contacts.

The core part of this interaction network is magnified in the red box in Figure 115 and shows most of the previously identified PM focal adhesion proteins and ECM laminins, together with some of their key *interactors*.

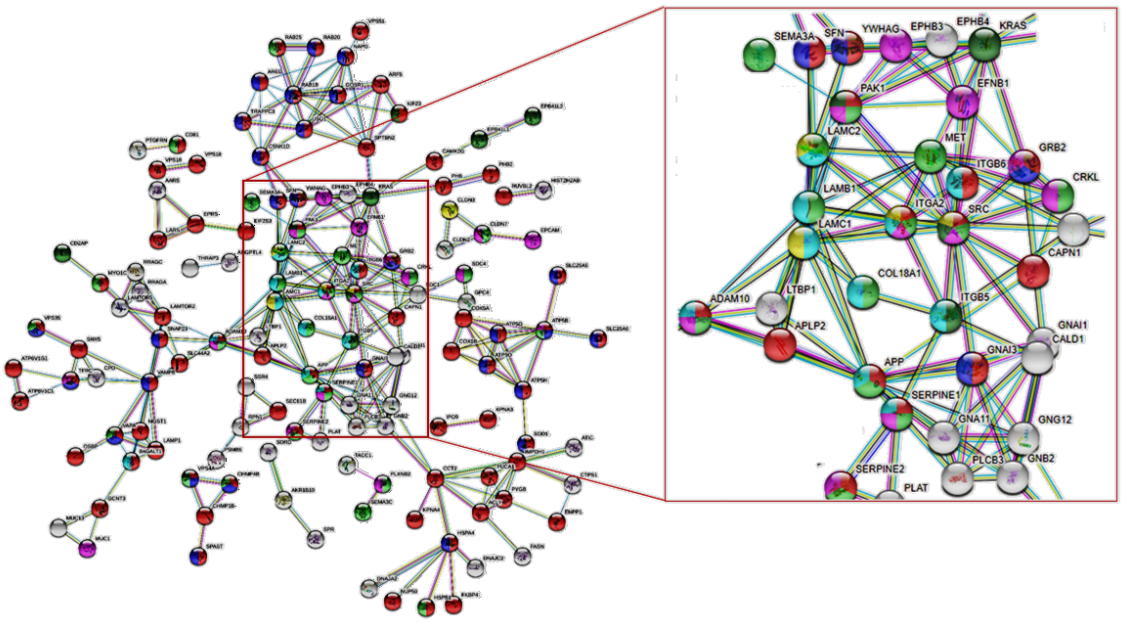


Figure 115: *STRING obtained LM2-down regulated PM and ECM proteins interaction network at the highest confidence level. Color code is explained in the text. Magnification enlightens some of the previously reported FA-related PM proteins and the ECM laminins.*

As previously shown, each of these core proteins is down-regulated to different extents in LM2 compared to MDA-MB-231, ranging from 2 to 11 folds and thus making both the PM and ECM counterparts involved in the cell/ECM contacts severely impaired in this cell line.

Thus, the concomitant analysis of the PM and ECM proteins of the parental MDA-MB-231 and of its derivative highly lung metastatic LM2 cell lines through high resolution *label-free* functional proteomics allowed the identification of LM2 key features differentiating it from its parental cell line. Taken together, these results strongly suggest that LM2 PM/ECM contacts are less efficient: PM FAs seem to be less structured and organized then the MDA-MB-231 ones and LM2, strongly down-regulating its laminin equipment, deposits less basal lamina.

8.8 Experimental Section

Cell lines and cell culture. MDAMB231, BT549, Hs578T, MCF7, SKBR3, HCC1419, HCC1954 (obtained from the ATCC), LM2, and JIMT1 (provided by The Netherlands Cancer Institute, NKI) were cultured in DMEM supplemented with 10% FBS (v/v), 100 U/mL penicillin, 100 µg/mL streptomycin, and 100 µg/mL L-Glutamine. MCF10A cells (ATCC) were grown in DMEM/F12 supplemented with 5% (v/v) horse serum, 20 ng/mL EGF, 0.5 mg/mL hydrocortisone, 10 µg/mL insulin, and 100 ng/mL cholera toxin. All cells were maintained in a humidified incubator at 37°C with 5% CO_2. All cell lines have been tested for mycoplasma contamination.

ECM and PM proteins sample preparation. Cells were harvested through trypsinization, washed twice in PBS and subsequently pelleted and stored at -80°C. The plates were then extensively washed with PBS and the remaining ECM was collected by scraping in 8 M urea/50 mM ammonium bicarbonate (AmBic, pH 8.5) and stored at -80°C.

The PM enrichment was performed through the Plasma Membrane Protein Extraction Kit from abcam (ab65400, abcam, Cambridge, UK) as reported by the manufacturer.

Briefly, cell pellets were re-suspended in 2 mL of the Homogenize Buffer supplemented with 1/500 vol/vol of the Protease Inhibitor Cocktail and submitted to mechanical lysis (Dounce homogenizer), carrying out alternative cycles of friction and rest (4°C) for a total of 10 minutes friction. The obtained suspension was submitted to centrifugation (700 x g, 4°C, 10 minutes) to remove the protein solution from the pelleted debris. The debris were discarded and the obtained suspension centrifuged again (10000 x g, 4°C, 10 minutes) to pellet the total membrane content of the cells. The supernatant, representing the cytosolic fraction, was stored at -80°C.

The total membranes pellets were then submitted to liquid-two-phase partitioning, as reported by the manufacturer. More in details, the total membrane proteins pellets were re-

suspended in 200 µL of the Upper Phase Solution (UPS) and 200 µL of the Lower Phase Solution (LPS) were added to each of them. The samples were mixed and incubated on ice for 5 minutes.

In the meantime, fresh phase tubes were prepared adding 200 µL of the UPS and 200 µL of LPS and incubated on ice (blank). Both of the samples were then centrifuged at 1000 x g for 5 minutes and at 4°C.

The UPS of the membrane-containing samples were collected in a clean tube and stored on ice, whereas the remaining LPS were extracted with fresh UPS (100 µL) from the blank tubes, as previously reported. The obtained UPS were collected and pooled with the first ones for a second UPS extraction round performed adding 100 µL LPS from the blank tubes, incubating the samples on ice and centrifuging at 1000 x g for 5 minutes and at 4°C.

The obtained UPS were then collected, diluted in 5 volumes of water, kept on ice for 5 minutes and centrifuged for 30 minutes at 4°C to allow PM precipitation.

The obtained PM pellet was re-suspended in 8M urea/50 mM AmBic (pH 8.5), supplemented with the Complete mini EDTA-free protease inhibitor cocktail (Roche).

Both PM and ECM samples proteins concentrations were determined through the spectrophotometric Bradford assay (Bio-Rad, Hercules, CA, USA).

Sample preparation for proteomics analysis. Equal amounts of ECM proteins for all of the 10 cell lines, as well as equal PM amounts, were submitted to *in solution* digestion.

Briefly, proteins were reduced (4 mM DTT, 1 hour, 25°C, 800 rpm) and alkylated (8 mM IAA, 30 minutes, 25°C, 800 rpm, in the dark) before being digested with LysC (Wako, Richmond, VA, USA) for 4 h (enzyme/substrate ratio 1:75 w/w) at 37 °C. Samples were then diluted four times with AmBic and further digested by Trypsin (Promega, Madison, WI, USA), overnight at 37°C (enzyme/substrate ratio 1:100 w/w). The digestion was quenched with 5% formic acid and the resulting peptides desalted on SepPak C18 cartridges (Waters Corporation,

Milford, MA). Briefly, the cartridge was activated flushing 3 mL of 100% CH_3CN and then conditioned with 3 mL of 0.1% FA. The sample was then loaded, desalted flushing the cartridge with 3 mL of 0.1% FA and finally eluted flushing two times 500 µL of 80% CH_3CN, 20% H_2O, 0,1% FA. For the subsequent MS analysis, the peptide mixtures were dried under vacuum and re-dissolved in 10% FA, 5% DMSO in H_2O.

Mass spectrometric analysis. For *label-free* quantification, 2 µg of each ECM and PM digest were analyzed by nanoLC-MS/MS on an Orbitrap Q-Exactive HF Mass Spectrometer (ThermoFisher Scientific, Bremen) coupled to an Agilent 1290 Infinity Ultra-High Pressure Liquid Chromatography (UPLC) system (Agilent Technologies), operating in reverse phase (C18) and equipped with a Reprosil pur C18 trap column (100 µm × 2 cm, 3 µm, Dr. Maisch) and a Poroshell 120 EC C18 analytical column (75 µm × 50 cm, 2.7 µm, Agilent Technologies). After trapping for 5 min at a flow rate of 0.05 ml/min in 100% solvent A (0.1% FA in H_2O), peptides elution was achieved with a 160 min gradient from 10 to 40% of solvent B (0.1% FA, 80% CH_3CN) at a flow rate of 300 nL/min. The mass spectrometer was operated in data-dependent acquisition mode, automatically switching between MS and MS2. Full scan MS spectra were acquired using the following settings: full-scan automatic gain control (AGC) target 3e6 at 60000 resolution; scan range 375–1600 m/z; Orbitrap full-scan maximum injection time 20 ms. HCD MS2 spectra were generated for up to 12 precursors (normalized collision energy of 27%) and the fragment ions acquired at a resolution of 30000 with an AGC target value of 1e5 and a maximum injection time of 100 ms.

ECM data processing. All the ECM .raw MS files were searched with the MaxQuant software (version 1.5.3.3). MS/MS spectra were searched by Andromeda against a reviewed Homo sapiens + bovine database (UniProt, November 2017, 161042+32206 entries), using the following parameters: trypsin digestion; maximum of two missed cleavages; cysteine

carbamidomethylation as fixed modification; protein N-terminal acetylation, methionine oxidization, asparagine and glutamine deamidation and proline hydroxylation as variable modifications. Mass tolerance was set to 4.5 and 20 ppm for the MS1 and MS2, respectively. The protein and PSM False Discovery Rate (FDR) were set to 1%. Peptide identifications by MS/MS were transferred between runs to replace missing values for quantification, with a 0.7 min window after retention time alignment.

PM data processing. All the PM .raw MS files were searched with the MaxQuant software (version 1.6.0.1). MS/MS spectra were searched by Andromeda against a reviewed Homo sapiens database (UniProt, November 2017, 161042 entries), using the following parameters: trypsin digestion; maximum of two missed cleavages; cysteine carbamidomethylation as fixed modification; protein N-terminal acetylation, methionine oxidization and asparagine and glutamine as variable modifications. Mass tolerance was set to 4.5 and 20 ppm for the MS1 and MS2, respectively. The protein and PSM False Discovery Rate (FDR) were set to 1%. Peptide identifications by MS/MS were transferred between runs to replace missing values for quantification, with a 0.7 min window after retention time alignment.

Data analysis. All data were analyzed using the Perseus software and Microsoft Excel. MaxQuant LFQ intensities were log2 transformed and missing values were replaced by imputation according to normal distribution with a downshift of 1.8 SDs and a width of 0.3 SDs. To assess the reproducibility of the experiments within the technical replicates (n = 3), Perseus was exploited to both perform principal component analysis (PCA) and calculate Pearson correlation coefficients.

For hierarchical clustering, imputed LFQ intensities were first z-scored and then clustered using Euclidean as a distance measure for column and row clustering.

Gene ontology (GO) analyses were performed with Database for Annotation, Visualization and Integrated Discovery (DAVID), using all the proteins identified by previously performed whole-cell lysate proteomics experiment[292] as background.

To identify the differentially expressed proteins across the PM/ECM subtypes, fold changes were calculated from the averaged LFQ intensities for each protein and t-test was performed imposing a p value lesser than 0.05.

Gene set enrichment analysis (GSEA) was performed through Broad GSEA version 3.0, using gene set collections from the Molecular Signatures database (MSigDB) v6.1.

Bibliography

1. Wilkins, M. R. *et al.* Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it. *Biotechnol. Genet. Eng. Rev.* **13**, 19–50 (1996).
2. Pandey, A. & Mann, M. Proteomics to study genes and genomes. *Nature* **405**, 837–846 (2000).
3. Makarov, A. Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. *Anal. Chem.* **72**, 1156–62 (2000).
4. Makarov, A. *et al.* Performance Evaluation of a Hybrid Linear Ion Trap/Orbitrap Mass Spectrometer. *Anal. Chem.* **78**, 2113–2120 (2006).
5. Chait, B. T. CHEMISTRY: Mass Spectrometry: Bottom-Up or Top-Down? *Science (80-.).* **314**, 65–66 (2006).
6. Han, X., Jin, M., Breuker, K. & McLafferty, F. W. Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons. *Science* **314**, 109–12 (2006).
7. Tran, J. C. *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature* **480**, 254–8 (2011).
8. Chen, B., Brown, K. A., Lin, Z. & Ge, Y. Top-Down Proteomics: Ready for Prime Time? *Anal. Chem.* **90**, 110–127 (2018).
9. Sharon, M. & Robinson, C. V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu. Rev. Biochem.* **76**, 167–93 (2007).
10. Uetrecht, C. *et al.* High-resolution mass spectrometry of viral assemblies: molecular composition and stability of dimorphic hepatitis B virus capsids. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 9216–20 (2008).
11. Uetrecht, C. *et al.* Stability and shape of hepatitis B virus capsids in vacuo. *Angew. Chem.*

Int. Ed. Engl. **47**, 6247–51 (2008).

12. Heck, A. J. R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nat. Methods* **5**, 927–33 (2008).
13. Yates, J. R., Ruse, C. I. & Nakorchevsky, A. Proteomics by Mass Spectrometry: Approaches, Advances, and Applications. *Annu. Rev. Biomed. Eng.* **11**, 49–79 (2009).
14. Laskay, Ü. A., Lobas, A. A., Srzentić, K., Gorshkov, M. V & Tsybin, Y. O. Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments. *J. Proteome Res.* **12**, 5558–69 (2013).
15. Moradian, A., Kalli, A., Sweredoski, M. J. & Hess, S. The top-down, middle-down, and bottom-up mass spectrometry approaches for characterization of histone variants and their post-translational modifications. *Proteomics* **14**, 489–97 (2014).
16. Link, A. J. *et al.* Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* **17**, 676–82 (1999).
17. Wolters, D. A., Washburn, M. P. & Yates, J. R. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**, 5683–90 (2001).
18. Yates, J. R. Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 297–316 (2004).
19. Yates, J. R. Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **33**, 1–19 (1998).
20. Henzel, W. J., Watanabe, C. & Stults, J. T. Protein identification: the origins of peptide mass fingerprinting. *J. Am. Soc. Mass Spectrom.* **14**, 931–42 (2003).
21. Olsen, J. V. & Mann, M. Improved peptide identification in proteomics by two consecutive stages of mass spectrometric fragmentation. *Proc. Natl. Acad. Sci.* **101**, 13417–13422 (2004).
22. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–201 (2004).

23. Gao, J., Opiteck, G. J., Friedrichs, M. S., Dongre, A. R. & Hefta, S. A. Changes in the protein expression of yeast as a function of carbon source. *J. Proteome Res.* **2**, 643–9

24. Butt, A. *et al.* Chromatographic separations as a prelude to two-dimensional electrophoresis in proteomics analysis. *Proteomics* **1**, 42–53 (2001).

25. Giorgianni, F., Desiderio, D. M. & Beranova-Giorgianni, S. Proteome analysis using isoelectric focusing in immobilized pH gradient gels followed by mass spectrometry. *Electrophoresis* **24**, 253–9 (2003).

26. Yan, F. *et al.* A comparison of drug-treated and untreated HCT-116 human colon adenocarcinoma cells using a 2-D liquid separation mapping method based upon chromatofocusing PI fractionation. *Anal. Chem.* **75**, 2299–308 (2003).

27. Lee, C.-L. *et al.* Strategic shotgun proteomics approach for efficient construction of an expression map of targeted protein families in hepatoma cell lines. *Proteomics* **3**, 2472–86 (2003).

28. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–5 (1970).

29. Klose, J. Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues. A novel approach to testing for induced point mutations in mammals. *Humangenetik* **26**, 231–43 (1975).

30. López, J. L. Two-dimensional electrophoresis in proteome expression analysis. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **849**, 190–202 (2007).

31. Evans, C. R. & Jorgenson, J. W. Multidimensional LC-LC and LC-CE for high-resolution separations of biological molecules. *Anal. Bioanal. Chem.* **378**, 1952–61 (2004).

32. Martosella, J., Zolotarjova, N., Liu, H., Nicol, G. & Boyes, B. E. Reversed-phase high-performance liquid chromatographic prefractionation of immunodepleted human serum proteins to enhance mass spectrometry identification of lower-abundant proteins. *J.*

Proteome Res. **4**, 1522–37 (2005).

33. Shevchenko, A., Wilm, M., Vorm, O. & Mann, M. Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68**, 850–8 (1996).

34. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V & Mann, M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **1**, 2856–60 (2006).

35. Viswanathan, S., Unlü, M. & Minden, J. S. Two-dimensional difference gel electrophoresis. *Nat. Protoc.* **1**, 1351–8 (2006).

36. Timms, J. F. & Cramer, R. Difference gel electrophoresis. *Proteomics* **8**, 4886–97 (2008).

37. Larbi, N. Ben & Jefferies, C. 2D-DIGE: Comparative Proteomics of Cellular Signalling Pathways. in *Methods in molecular biology (Clifton, N.J.)* **517**, 105–132 (2009).

38. Dormeyer, W., Mohammed, S., Breukelen, B. van, Krijgsveld, J. & Heck, A. J. R. Targeted analysis of protein termini. *J. Proteome Res.* **6**, 4634–45 (2007).

39. Dai, J. *et al.* Fully automatic separation and identification of phosphopeptides by continuous pH-gradient anion exchange online coupled with reversed-phase liquid chromatography mass spectrometry. *J. Proteome Res.* **8**, 133–41 (2009).

40. Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *J. Chromatogr. A* **499**, 177–196 (1990).

41. Boersema, P. J., Divecha, N., Heck, A. J. R. & Mohammed, S. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J. Proteome Res.* **6**, 937–46 (2007).

42. Wang, H. *et al.* An off-line high pH reversed-phase fractionation and nano-liquid chromatography-mass spectrometry method for global proteomic profiling of cell lines. *J. Chromatogr. B. Analyt. Technol. Biomed. Life Sci.* **974**, 90–5 (2015).

43. Wilm, M. & Mann, M. Analytical properties of the nanoelectrospray ion source. *Anal.*

Chem. **68**, 1–8 (1996).

44. McLuckey, S. A. Principles of collisional activation in analytical mass spectrometry. *J. Am. Soc. Mass Spectrom.* **3**, 599–614 (1992).

45. Sleno, L. & Volmer, D. A. Ion activation methods for tandem mass spectrometry. *J. Mass Spectrom.* **39**, 1091–112 (2004).

46. Paizs, B. & Suhai, S. Fragmentation pathways of protonated peptides. *Mass Spectrom. Rev.* **24**, 508–48 (2005).

47. Roepstorff, P. & Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed. Mass Spectrom.* **11**, 601 (1984).

48. Sobott, F. *et al.* Comparison of CID versus ETD based MS/MS fragmentation for the analysis of protein ubiquitination. *J. Am. Soc. Mass Spectrom.* **20**, 1652–9 (2009).

49. Zubarev, R. A. Electron-capture dissociation tandem mass spectrometry. *Curr. Opin. Biotechnol.* **15**, 12–6 (2004).

50. Jedrychowski, M. P. *et al.* Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics. *Mol. Cell. Proteomics* **10**, M111.009910 (2011).

51. Koenig, T. *et al.* Robust prediction of the MASCOT score for an improved quality assessment in mass spectrometric proteomics. *J. Proteome Res.* **7**, 3708–17 (2008).

52. MacCoss, M. J., Wu, C. C. & Yates, J. R. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal. Chem.* **74**, 5593–9 (2002).

53. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–805 (2011).

54. Frank, A. M. Predicting intensity ranks of peptide fragment ions. *J. Proteome Res.* **8**, 2226–40 (2009).

55. Weatherly, D. B. *et al.* A Heuristic method for assigning a false-discovery rate for protein identifications from Mascot database search results. *Mol. Cell. Proteomics* **4**, 762–72

(2005).

56. Bauer, A. & Brönstrup, M. Industrial natural product chemistry for drug discovery and development. *Nat. Prod. Rep.* **31**, 35–60 (2014).

57. Scheepstra, M. *et al.* A Natural-Product Switch for a Dynamic Protein Interface. *Angew. Chemie Int. Ed.* **53**, 6443–6448 (2014).

58. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.* **14**, 111–29 (2015).

59. Clardy, J. & Walsh, C. Lessons from natural molecules. *Nature* **432**, 829–37 (2004).

60. Carlson, E. E. Natural products as chemical probes. *ACS Chem. Biol.* **5**, 639–53 (2010).

61. Drews, J. Drug Discovery: A Historical Perspective. *Science (80-.).* **287**, 1960–1964 (2000).

62. Sleno, L. & Emili, A. Proteomic methods for drug target discovery. *Curr. Opin. Chem. Biol.* **12**, 46–54 (2008).

63. Bantscheff, M. & Drewes, G. Chemoproteomic approaches to drug target identification and drug profiling. *Bioorg. Med. Chem.* **20**, 1973–1978 (2012).

64. Bondarenko, P. V, Chelius, D. & Shaler, T. A. Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **74**, 4741–9 (2002).

65. Colinge, J., Chiappe, D., Lagache, S., Moniatte, M. & Bougueleret, L. Differential proteomics via probabilistic peptide identification scores. *Anal. Chem.* **77**, 596–606 (2005).

66. Aebersold, R. & Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **537**, 347–355 (2016).

67. Chelius, D. & Bondarenko, P. V. Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *J. Proteome Res.* **1**, 317–23

68. Bateman, R. H. *et al.* A novel precursor ion discovery method on a hybrid quadrupole

orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J. Am. Soc. Mass Spectrom.* **13**, 792–803 (2002).

69. Silva, J. C. *et al.* Quantitative Proteomic Analysis by Accurate Mass Retention Time Pairs. *Anal. Chem.* **77**, 2187–2200 (2005).

70. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017–1031 (2007).

71. Ishihama, Y. *et al.* Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4**, 1265–72 (2005).

72. Ong, S.-E. & Mann, M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). *Nat. Protoc.* **1**, 2650–60 (2006).

73. Ong, S.-E. & Mann, M. Stable Isotope Labeling by Amino Acids in Cell Culture for Quantitative Proteomics. in 37–52 (2007). doi:10.1007/978-1-59745-255-7_3

74. Shiio, Y. & Aebersold, R. Quantitative proteome analysis using isotope-coded affinity tags and mass spectrometry. *Nat. Protoc.* **1**, 139–145 (2006).

75. Hsu, J.-L., Huang, S.-Y., Chow, N.-H. & Chen, S.-H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **75**, 6843–52 (2003).

76. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* **4**, 484–494 (2009).

77. Ross, P. L. *et al.* Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–69 (2004).

78. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–904 (2003).

79. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–

7 (2006).

80. Jessani, N. & Cravatt, B. F. The development and application of methods for activity-based protein profiling. *Curr. Opin. Chem. Biol.* **8**, 54–59 (2004).

81. Barglow, K. T. & Cravatt, B. F. Activity-based protein profiling for the functional annotation of enzymes. *Nat. Methods* **4**, 822–827 (2007).

82. Sadaghiani, A. M., Verhelst, S. H. & Bogyo, M. Tagging and detection strategies for activity-based proteomics. *Curr. Opin. Chem. Biol.* **11**, 20–28 (2007).

83. Nodwell, M. B. & Sieber, S. A. ABPP methodology: introduction and overview. *Top. Curr. Chem.* **324**, 1–41 (2012).

84. Shi, H., Liu, K., Xu, A. & Yao, S. Q. Small molecule microarray-facilitated screening of affinity-based probes (AfBPs) for gamma-secretase. *Chem. Commun. (Camb).* 5030–2 (2009). doi:10.1039/b910611a

85. Shi, H., Zhang, C.-J., Chen, G. Y. J. & Yao, S. Q. Cell-based proteome profiling of potential dasatinib targets by use of affinity-based probes. *J. Am. Chem. Soc.* **134**, 3001–14 (2012).

86. Kolb, H. C. & Sharpless, K. B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **8**, 1128–37 (2003).

87. Speers, A. E. & Cravatt, B. F. Profiling Enzyme Activities In Vivo Using Click Chemistry Methods. *Chem. Biol.* **11**, 535–546 (2004).

88. Cravatt, B. F., Wright, A. T. & Kozarich, J. W. Activity-Based Protein Profiling: From Enzyme Chemistry to Proteomic Chemistry. *Annu. Rev. Biochem.* **77**, 383–414 (2008).

89. Rix, U. & Superti-Furga, G. Target profiling of small molecules by chemical proteomics. *Nat. Chem. Biol.* **5**, 616–624 (2009).

90. Ito, T. *et al.* Identification of a Primary Target of Thalidomide Teratogenicity. *Science (80-.).* **327**, 1345–1350 (2010).

91. Margarucci, L., Monti, M. C., Tosco, A., Riccio, R. & Casapullo, A. Chemical

Proteomics Discloses Petrosapongiolide M, an Antiinflammatory Marine Sesterterpene, as a Proteasome Inhibitor. *Angew. Chemie Int. Ed.* **49**, 3960–3963 (2010).

92. Sun, L. *et al.* Mixed Lineage Kinase Domain-like Protein Mediates Necrosis Signaling Downstream of RIP3 Kinase. *Cell* **148**, 213–227 (2012).

93. Huber, K. V. M. *et al.* Stereospecific targeting of MTH1 by (S)-crizotinib as an anticancer strategy. *Nature* **508**, 222–7 (2014).

94. Capolupo, A. *et al.* Determination of Gymnemic Acid I as a Protein Biosynthesis Inhibitor Using Chemical Proteomics. *J. Nat. Prod.* **80**, 909–915 (2017).

95. Capolupo, A. *et al.* Proteasome as a New Target for Bio-Inspired Benzo[*k* , *l*]xanthene Lignans. *Chem. - A Eur. J.* **23**, 8371–8374 (2017).

96. Shiyama, T., Furuya, M., Yamazaki, A., Terada, T. & Tanaka, A. Design and synthesis of novel hydrophilic spacers for the reduction of nonspecific binding proteins on affinity resins. *Bioorg. Med. Chem.* **12**, 2831–2841 (2004).

97. Speers, A. E., Adam, G. C. & Cravatt, B. F. Activity-Based Protein Profiling in Vivo Using a Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *J. Am. Chem. Soc.* **125**, 4686–4687 (2003).

98. Hong, V., Presolski, S. I., Ma, C. & Finn, M. G. Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed. Engl.* **48**, 9879–83 (2009).

99. Kaida, D. *et al.* Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. *Nat. Chem. Biol.* **3**, 576–83 (2007).

100. Wang, G., Shang, L., Burgett, A. W. G., Harran, P. G. & Wang, X. Diazonamide toxins reveal an unexpected function for ornithine delta-amino transferase in mitotic cell division. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 2068–73 (2007).

101. Speers, A. E. & Cravatt, B. F. A tandem orthogonal proteolysis strategy for high-content chemical proteomics. *J. Am. Chem. Soc.* **127**, 10018–9 (2005).

102. van der Veken, P. *et al.* Development of a novel chemical probe for the selective enrichment of phosphorylated serine- and threonine-containing peptides. *Chembiochem* **6**, 2271–80 (2005).

103. Verhelst, S. H. L., Fonović, M. & Bogyo, M. A Mild Chemically Cleavable Linker System for Functional Proteomic Applications. *Angew. Chemie Int. Ed.* **46**, 1284–1286 (2007).

104. Raida, M. Drug target deconvolution by chemical proteomics. *Curr. Opin. Chem. Biol.* **15**, 570–575 (2011).

105. Yamamoto, K., Yamazaki, A., Takeuchi, M. & Tanaka, A. A versatile method of identifying specific binding proteins on affinity resins. *Anal. Biochem.* **352**, 15–23 (2006).

106. Terstappen, G. C., Schlüpen, C., Raggiaschi, R. & Gaviraghi, G. Target deconvolution strategies in drug discovery. *Nat. Rev. Drug Discov.* **6**, 891–903 (2007).

107. Sinz, A. Chemical cross-linking and mass spectrometry for mapping three-dimensional structures of proteins and protein complexes. *J. Mass Spectrom.* **38**, 1225–1237 (2003).

108. Huang, B. X., Kim, H.-Y. & Dass, C. Probing three-dimensional structure of bovine serum albumin by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **15**, 1237–47 (2004).

109. Huang, B. X. & Kim, H.-Y. Probing Akt-inhibitor interaction by chemical cross-linking and mass spectrometry. *J. Am. Soc. Mass Spectrom.* **20**, 1504–13 (2009).

110. Boelt, S. G. *et al.* Mapping the Ca(2+) induced structural change in calreticulin. *J. Proteomics* **142**, 138–48 (2016).

111. Chen, Z. A. & Rappsilber, J. Quantitative cross-linking/mass spectrometry to elucidate structural changes in proteins and their complexes. *Nat. Protoc.* **14**, 171–201 (2019).

112. West, G. M. *et al.* Quantitative proteomics approach for identifying protein-drug interactions in complex mixtures using protein stability measurements. *Proc. Natl. Acad.*

Sci. U. S. A. **107**, 9078–82 (2010).

113. Tran, D. T., Adhikari, J. & Fitzgerald, M. C. Stable Isotope Labeling with Amino Acids in Cell Culture (SILAC)-Based Strategy for Proteome-Wide Thermodynamic Analysis of Protein-Ligand Binding Interactions. *Mol. Cell. Proteomics* **13**, 1800–1813 (2014).

114. Park, C. & Marqusee, S. Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. *Nat. Methods* **2**, 207–212 (2005).

115. Park, C. & Marqusee, S. Probing the High Energy States in Proteins by Proteolysis. *J. Mol. Biol.* **343**, 1467–1476 (2004).

116. Pace, C. N. & McGrath, T. Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. *J. Biol. Chem.* **255**, 3862–5 (1980).

117. Liu, P.-F., Kihara, D. & Park, C. Energetics-Based Discovery of Protein–Ligand Interactions on a Proteomic Scale. *J. Mol. Biol.* **408**, 147–162 (2011).

118. Chang, Y., Schlebach, J. P., VerHeul, R. A. & Park, C. Simplified proteomics approach to discover protein-ligand interactions. *Protein Sci.* **21**, 1280–7 (2012).

119. Adhikari, J. & Fitzgerald, M. C. SILAC-Pulse Proteolysis: A Mass Spectrometry-Based Method for Discovery and Cross-Validation in Proteome-Wide Studies of Ligand Binding. *J. Am. Soc. Mass Spectrom.* **25**, 2073–2083 (2014).

120. Zeng, L. *et al.* Discovery of Nicotinamide Adenine Dinucleotide Binding Proteins in the Escherichia coli Proteome Using a Combined Energetic- and Structural-Bioinformatics-Based Approach. *J. Proteome Res.* **16**, 470–480 (2017).

121. Pantoliano, M. W. *et al.* High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J. Biomol. Screen.* **6**, 429–40 (2001).

122. Vedadi, M. *et al.* Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 15835–40 (2006).

123. Niesen, F. H., Berglund, H. & Vedadi, M. The use of differential scanning fluorimetry

to detect ligand interactions that promote protein stability. *Nat. Protoc.* **2**, 2212–2221 (2007).

124. Fedorov, O. *et al.* A systematic interaction map of validated kinase inhibitors with Ser/Thr kinases. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 20523–8 (2007).

125. Jafari, R. *et al.* The cellular thermal shift assay for evaluating drug target interactions in cells. *Nat. Protoc.* **9**, 2100–2122 (2014).

126. Franken, H. *et al.* Thermal proteome profiling for unbiased identification of direct and indirect drug targets using multiplexed quantitative mass spectrometry. *Nat. Protoc.* **10**, 1567–93 (2015).

127. Werner, T. *et al.* Ion coalescence of neutron encoded TMT 10-plex reporter ions. *Anal. Chem.* **86**, 3594–601 (2014).

128. Lomenick, B. *et al.* Target identification using drug affinity responsive target stability (DARTS). *Proc. Natl. Acad. Sci.* **106**, 21984–21989 (2009).

129. Englander, S. W. & Rolfe, A. Hydrogen exchange studies of respiratory proteins. 3. Structural and free energy changes in hemoglobin by use of a difference method. *J. Biol. Chem.* **248**, 4852–61 (1973).

130. Henzler-Wildman, K. & Kern, D. Dynamic personalities of proteins. *Nature* **450**, 964–972 (2007).

131. Lomenick, B., Jung, G., Wohlschlegel, J. A. & Huang, J. Target Identification Using Drug Affinity Responsive Target Stability (DARTS). in *Current Protocols in Chemical Biology* **3**, 163–180 (John Wiley & Sons, Inc., 2011).

132. Lomenick, B., Olsen, R. W. & Huang, J. Identification of Direct Protein Targets of Small Molecules. *ACS Chem. Biol.* **6**, 34–46 (2011).

133. Kim, D. *et al.* FK506, an Immunosuppressive Drug, Induces Autophagy by Binding to the V-ATPase Catalytic Subunit A in Neuronal Cells. *J. Proteome Res.* **16**, 55–64 (2017).

134. Stankunas, K. *et al.* Conditional protein alleles using knockin mice and a chemical

inducer of dimerization. *Mol. Cell* **12**, 1615–24 (2003).

135. Banaszynski, L. A., Chen, L.-C., Maynard-Smith, L. A., Ooi, A. G. L. & Wandless, T. J. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* **126**, 995–1004 (2006).

136. Cohen, A. A. *et al.* Dynamic Proteomics of Individual Cancer Cells in Response to a Drug. *Science (80-.).* **322**, 1511–1516 (2008).

137. Wijayaratne, A. L. & McDonnell, D. P. The human estrogen receptor-alpha is a ubiquitinated protein whose stability is affected differentially by agonists, antagonists, and selective estrogen receptor modulators. *J. Biol. Chem.* **276**, 35684–92 (2001).

138. Feng, Y. *et al.* Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044 (2014).

139. Fontana, A. *et al.* Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* **51**, 299–321 (2004).

140. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics. *Cell* **138**, 795–806 (2009).

141. Yost, R. A. & Enke, C. G. Triple quadrupole mass spectrometry for direct mixture analysis and structure elucidation. *Anal. Chem.* **51**, 1251–64 (1979).

142. Kuster, B., Schirle, M., Mallick, P. & Aebersold, R. Scoring proteomes with proteotypic peptide probes. *Nat. Rev. Mol. Cell Biol.* **6**, 577–83 (2005).

143. Mallick, P. *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **25**, 125–131 (2007).

144. Deutsch, E. W., Lam, H. & Aebersold, R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* **9**, 429–34 (2008).

145. Keller, A., Eng, J., Zhang, N., Li, X. & Aebersold, R. A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol. Syst. Biol.* **1**, 2005.0017 (2005).

146. Craig, R. & Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **20**, 1466–1467 (2004).

147. Hubbard, T. J. P. *et al.* Ensembl 2007. *Nucleic Acids Res.* **35**, D610–D617 (2007).

148. Piazza, I. *et al.* A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **172**, 358-372.e23 (2018).

149. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).

150. Moulder, R., Goo, Y. A. & Goodlett, D. R. Label-Free Quantitation for Clinical Proteomics. in *Methods in molecular biology (Clifton, N.J.)* **1410**, 65–76 (2016).

151. Margarucci, L., Monti, M. C., Tosco, A., Riccio, R. & Casapullo, A. Chemical proteomics discloses petrosapongiolide M, an antiinflammatory marine sesterterpene, as a proteasome inhibitor. *Angew. Chem. Int. Ed. Engl.* **49**, 3960–3 (2010).

152. Capolupo, A. *et al.* Determination of Gymnemic Acid I as a Protein Biosynthesis Inhibitor Using Chemical Proteomics. *J. Nat. Prod.* **80**, 909–915 (2017).

153. Margarucci, L., Monti, M. C., Fontanella, B., Riccio, R. & Casapullo, A. Chemical proteomics reveals bolinaquinone as a clathrin-mediated endocytosis inhibitor. *Mol. BioSyst.* **7**, 480–485 (2011).

154. Cassiano, C. *et al.* Chemical proteomics reveals heat shock protein 60 to be the main cellular target of the marine bioactive sesterterpene suvanine. *Chembiochem* **13**, 1953–8 (2012).

155. Margarucci, L. *et al.* Chemical proteomics-driven discovery of oleocanthal as an Hsp90 inhibitor. *Chem. Commun. (Camb).* **49**, 5844–6 (2013).

156. Cassiano, C. *et al.* Heteronemin, a marine sponge terpenoid, targets TDP-43, a key factor in several neurodegenerative disorders. *Chem. Commun. (Camb).* **50**, 406–8 (2014).

157. Margarucci, L. *et al.* N-Formyl-7-amino-11-cycloamphilectene, a marine sponge

metabolite, binds to tubulin and modulates microtubule depolymerization. *Mol. Biosyst.* **10**, 862–7 (2014).

158. Del Gaudio, F. *et al.* Biomolecular proteomics discloses ATP synthase as the main target of the natural glycoside deglucoruscin. *Mol. Biosyst.* **12**, 3132–8 (2016).

159. Lomenick, B. *et al.* Target identification using drug affinity responsive target stability (DARTS). *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21984–9 (2009).

160. Roe, S. M. *et al.* Structural basis for inhibition of the Hsp90 molecular chaperone by the antitumor antibiotics radicicol and geldanamycin. *J. Med. Chem.* **42**, 260–6 (1999).

161. Schulte, T. W. *et al.* Antibiotic radicicol binds to the N-terminal domain of Hsp90 and shares important biologic activities with geldanamycin. *Cell Stress Chaperones* **3**, 100–108 (1998).

162. Khandelwal, A. *et al.* Structure-guided design of an Hsp90β N-terminal isoform-selective inhibitor. *Nat. Commun.* **9**, 425 (2018).

163. Morretta, E. *et al.* Discovering the Biological Target of 5-epi-Sinuleptolide Using a Combination of Proteomic Approaches. *Mar. Drugs* **15**, 312 (2017).

164. Lakshmi, V. & Kumar, R. Metabolites from Sinularia species. *Natural Product Research* **23**, 801–850 (2009).

165. Blunt, J. W., Copp, B. R., Keyzers, R. A., Munro, M. H. G. & Prinsep, M. R. Marine natural products. *Natural Product Reports* **33**, 382–431 (2016).

166. Radhika, P., Rao, P. R., Archana, J. & Rao, N. K. Anti-inflammatory activity of a new sphingosine derivative and cembrenoid diterpene (lobohedleolide) isolated from marine soft corals of Sinularia crassa Tixier-Durivault and Lobophytum species of the andaman and nicobar islands. *Biol. Pharm. Bull.* **28**, 1311–1313 (2005).

167. Liang, C. H. *et al.* Apoptosis effect of Sinularia leptoclados, S. depressan and S. inflate extracts in human oral squamous cell carcinomas. *J. Taiwan Inst. Chem. Eng.* **41**, 86–91 (2010).

168. Cheng, S.-Y. *et al.* Bioactive norditerpenoids from the soft coral Sinularia gyrosa. *Bioorg. Med. Chem.* **18**, 3379–86 (2010).

169. Sheu, J.-H., Ahmed, A. F., Shiue, R.-T., Dai, C.-F. & Kuo, Y.-H. Scabrolides A-D, four new norditerpenoids isolated from the soft coral Sinularia scabra. *J. Nat. Prod.* **65**, 1904–8 (2002).

170. Takaki, H., Koganemaru, R., Iwakawa, Y., Higuchi, R. & Miyamoto, T. Inhibitory effect of norditerpenes on LPS-induced TNF-alpha production from the Okinawan soft coral, Sinularia sp. *Biol. Pharm. Bull.* **26**, 380–2 (2003).

171. Chang, Y.-T. *et al.* Sinuleptolide inhibits proliferation of oral cancer Ca9-22 cells involving apoptosis, oxidative stress, and DNA damage. *Arch. Oral Biol.* **66**, 147–54 (2016).

172. Tseng, Y.-J., Wang, S.-K. & Duh, C.-Y. Secosteroids and norcembranoids from the soft coral Sinularia nanolobata. *Mar. Drugs* **11**, 3288–96 (2013).

173. Liang, C.-H. *et al.* 5-epi-Sinuleptolide induces cell cycle arrest and apoptosis through tumor necrosis factor/mitochondria-mediated caspase signaling pathway in human skin cancer cells. *Biochim. Biophys. Acta* **1820**, 1149–57 (2012).

174. Fujii, T., Iwane, A. H., Yanagida, T. & Namba, K. Direct visualization of secondary structures of F-actin by electron cryomicroscopy. *Nature* **467**, 724–728 (2010).

175. Oda, T., Iwasa, M., Aihara, T., Maéda, Y. & Narita, A. The nature of the globular- to fibrous-actin transition. *Nature* **457**, 441–5 (2009).

176. Dominguez, R. & Holmes, K. C. Actin Structure and Function. *Annu. Rev. Biophys.* **40**, 169–186 (2011).

177. Wegner, A. & Isenberg, G. 12-fold difference between the critical monomer concentrations of the two ends of actin filaments in physiological salt conditions. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 4922–4925 (1983).

178. Graceffa, P. & Dominguez, R. Crystal structure of monomeric actin in the ATP state.

Structural basis of nucleotide-dependent actin dynamics. *J. Biol. Chem.* **278**, 34172–80 (2003).

179. Wakatsuki, T., Schwab, B., Thompson, N. C. & Elson, E. L. Effects of cytochalasin D and latrunculin B on mechanical properties of cells. *J. Cell Sci.* **114**, 1025–1036 (2001).

180. Suresh, S. Biomechanics and biophysics of cancer cells. *Acta Biomater.* **3**, 413–438 (2007).

181. Weissmann, C. & Brandt, R. Mechanisms of neurodegenerative diseases: Insights from live cell imaging. *Journal of Neuroscience Research* **86**, 504–511 (2008).

182. Lázaro-Diéguez, F. *et al.* Dynamics of an F-actin aggresome generated by the actin-stabilizing toxin jasplakinolide. *J. Cell Sci.* **121**, 1415–25 (2008).

183. Morretta, E. *et al.* Crellastatin A, a PARP-1 Inhibitor Discovered by Complementary Proteomic Approaches. *ChemMedChem* **15**, 317–323 (2020).

184. Mehbub, M. F., Lei, J., Franco, C. & Zhang, W. Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. *Mar. Drugs* **12**, 4539–77 (2014).

185. Essack, M., Bajic, V. B. & Archer, J. A. C. Recently confirmed apoptosis-inducing lead compounds isolated from marine sponge of potential relevance in cancer treatment. *Mar. Drugs* **9**, 1580–606 (2011).

186. Bergquist, P. R., Karuso, P., Cambie, R. C. & Smith, D. J. Sterol composition and classification of the Porifera. *Biochem. Syst. Ecol.* **19**, 17–24 (1991).

187. Okada, Y., Matsunaga, S., Van Soest, R. W. M. & Fusetani, N. Sokodosides, steroid glycosides with an isopropyl side chain, from the marine sponge Erylus placenta. *J. Org. Chem.* **71**, 4884–4888 (2006).

188. John, V. *et al.* Biosynthetic studies of marine lipids. 20. Sequence of double-bond introduction in the sponge sterol 24.beta.-methylcholesta-5,7,22,25-tetraen-3.beta.-ol. *J. Org. Chem.* **54**, 1642–1647 (1989).

189. Fukuzawa, S., Matsunaga, S. & Fusetani, N. Isolation of 13 New Ritterazines from the Tunicate Ritterella tokioka and Chemical Transformation of Ritterazine B(1). *J. Org. Chem.* **62**, 4484–4491 (1997).

190. Pettit, G. R. *et al.* Isolation and structure of the hemichordate cell growth inhibitors cephalostatins 2, 3, and 4. *J. Chem. Soc. Chem. Commun.* 865–867 (1988). doi:10.1039/C39880000865

191. Moser, B. R. Review of cytotoxic cephalostatins and ritterazines: isolation and synthesis. *J. Nat. Prod.* **71**, 487–91 (2008).

192. D'Auria, M. V. *et al.* Crellastatin A: A Cytotoxic Bis-Steroid Sulfate from the Vanuatu Marine Sponge Crella sp. *J. Org. Chem.* **63**, 7382–7388 (1998).

193. Thomas, C. *et al.* Hit and run versus long-term activation of PARP-1 by its different domains fine-tunes nuclear processes. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 9941–9946 (2019).

194. Ray Chaudhuri, A. & Nussenzweig, A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nature Reviews Molecular Cell Biology* **18**, 610–621 (2017).

195. Krietsch, J. *et al.* Reprogramming cellular events by poly(ADP-ribose)-binding proteins. *Molecular Aspects of Medicine* **34**, 1066–1087 (2013).

196. Malanga, M., Pleschke, J. M., Kleczkowska, H. E. & Althaus, F. R. Poly(ADP-ribose) binds to specific domains of p53 and alters its DNA binding functions. *J. Biol. Chem.* **273**, 11839–11843 (1998).

197. Althaus, F. R. *et al.* Poly ADP-ribosylation: a DNA break signal mechanism. *Mol. Cell. Biochem.* **193**, 5–11 (1999).

198. Pleschke, J. M., Kleczkowska, H. E., Strohm, M. & Althaus, F. R. Poly(ADP-ribose) binds to specific domains in DNA damage checkpoint proteins. *J. Biol. Chem.* **275**, 40974–40980 (2000).

199. Gagné, J.-P. *et al.* Proteome-wide identification of poly(ADP-ribose) binding proteins

and poly(ADP-ribose)-associated protein complexes. *Nucleic Acids Res.* **36**, 6959–76 (2008).

200. Krupitza, G. & Cerutti, P. Poly(ADP-ribosylation) of histones in intact human keratinocytes. *Biochemistry* **28**, 4054–60 (1989).

201. Althaus, F. R. *et al.* Interactions of poly(ADP-ribose) with nuclear proteins. *Biochimie* **77**, 423–432 (1995).

202. Kraus, W. L. & Hottiger, M. O. PARP-1 and gene regulation: Progress and puzzles. *Molecular Aspects of Medicine* **34**, 1109–1123 (2013).

203. Kameshita, I., Matsuda, M., Nishikimi, M., Ushiro, H. & Shizuta, Y. Reconstitution and poly(ADP-ribosyl)ation of proteolytically fragmented poly(ADP-ribose) synthetase. *J. Biol. Chem.* **261**, 3863–8 (1986).

204. D'Amours, D., Desnoyers, S., D'Silva, I. & Poirier, G. G. Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions. *Biochemical Journal* **342**, 249–268 (1999).

205. D'Silva, I. *et al.* Relative affinities of poly(ADP-ribose) polymerase and DNA-dependent protein kinase for DNA strand interruptions. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1430**, 119–126 (1999).

206. Pion, E. *et al.* Poly(ADP-ribose) polymerase-1 dimerizes at a 5' recessed DNA end in vitro: a fluorescence study. *Biochemistry* **42**, 12409–17 (2003).

207. Lonskaya, I. *et al.* Regulation of poly(ADP-ribose) polymerase-1 by DNA structure-specific binding. *J. Biol. Chem.* **280**, 17076–83 (2005).

208. Eustermann, S. *et al.* The DNA-binding domain of human PARP-1 interacts with DNA single-strand breaks as a monomer through its second zinc finger. *J. Mol. Biol.* **407**, 149–70 (2011).

209. Langelier, M.-F., Planck, J. L., Roy, S. & Pascal, J. M. Crystal structures of poly(ADP-ribose) polymerase-1 (PARP-1) zinc fingers bound to DNA: structural and functional

insights into DNA-dependent PARP-1 activity. *J. Biol. Chem.* **286**, 10690–701 (2011).

210. Langelier, M.-F., Servent, K. M., Rogers, E. E. & Pascal, J. M. A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation. *J. Biol. Chem.* **283**, 4105–14 (2008).

211. Langelier, M.-F., Ruhl, D. D., Planck, J. L., Kraus, W. L. & Pascal, J. M. The Zn3 domain of human poly(ADP-ribose) polymerase-1 (PARP-1) functions in both DNA-dependent poly(ADP-ribose) synthesis activity and chromatin compaction. *J. Biol. Chem.* **285**, 18877–87 (2010).

212. Langelier, M.-F., Planck, J. L., Roy, S. & Pascal, J. M. Structural basis for DNA damage-dependent poly(ADP-ribosyl)ation by human PARP-1. *Science* **336**, 728–32 (2012).

213. Chapman, J. D., Gagné, J. P., Poirier, G. G. & Goodlett, D. R. Mapping PARP-1 auto-ADP-ribosylation sites by liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* **12**, 1868–1880 (2013).

214. Gagné, J.-P. *et al.* Quantitative site-specific ADP-ribosylation profiling of DNA-dependent PARPs. *DNA Repair (Amst).* **30**, 68–79 (2015).

215. Poirier, G. G., de Murcia, G., Jongstra-Bilen, J., Niedergang, C. & Mandel, P. Poly(ADP-ribosyl)ation of polynucleosomes causes relaxation of chromatin structure. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 3423–7 (1982).

216. Kim, M. Y., Mauro, S., Gévry, N., Lis, J. T. & Kraus, W. L. NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. *Cell* **119**, 803–14 (2004).

217. Muthurajan, U. M. *et al.* Automodification switches PARP-1 function from chromatin architectural protein to histone chaperone. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12752–7 (2014).

218. Dawicki-McKenna, J. M. *et al.* PARP-1 Activation Requires Local Unfolding of an Autoinhibitory Domain. *Mol. Cell* **60**, 755–768 (2015).

219. Steffen, J. D. *et al.* Targeting PARP-1 allosteric regulation offers therapeutic potential against cancer. *Cancer Res.* **74**, 31–37 (2014).

220. Hottiger, M. O., Hassa, P. O., Lüscher, B., Schüler, H. & Koch-Nolte, F. Toward a unified nomenclature for mammalian ADP-ribosyltransferases. *Trends Biochem. Sci.* **35**, 208–19 (2010).

221. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).

222. Mozzicafreddo, M. *et al.* Environmental pollutants directly affect the liver X receptor alpha activity: Kinetic and thermodynamic characterization of binding. *J. Steroid Biochem. Mol. Biol.* **152**, 1–7 (2015).

223. Ba, X. & Garg, N. J. Signaling mechanism of poly(ADP-ribose) polymerase-1 (PARP-1) in inflammatory diseases. *Am. J. Pathol.* **178**, 946–55 (2011).

224. Chiarugi, A. & Moskowitz, M. A. Poly(ADP-ribose) polymerase-1 activity promotes NF-κB-driven transcription and microglial activation: Implication for neurodegenerative disorders. *J. Neurochem.* **85**, 306–317 (2003).

225. Martire, S., Mosca, L. & d'Erme, M. PARP-1 involvement in neurodegeneration: A focus on Alzheimer's and Parkinson's diseases. *Mech. Ageing Dev.* **146**–**148**, 53–64 (2015).

226. Soriano, F. G., Virág, L. & Szabó, C. Diabetic endothelial dysfunction: role of reactive oxygen and nitrogen species production and poly(ADP-ribose) polymerase activation. *J. Mol. Med. (Berl).* **79**, 437–48 (2001).

227. Helleday, T., Bryant, H. E. & Schultz, N. Poly(ADP-ribose) polymerase (PARP-1) in homologous recombination and as a target for cancer therapy. *Cell Cycle* **4**, 1176–1178 (2005).

228. Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917–921 (2005).

229. Gaymes, T. J. *et al.* Inhibitors of poly ADP-ribose polymerase (PARP) induce apoptosis

of myeloid leukemic cells: potential for therapy of myeloid leukemia and myelodysplastic syndromes. *Haematologica* **94**, 638–46 (2009).

230. Habener, L. J., Hooper, J. N. A. & Carroll, A. R. Chemical and Biological Aspects of Marine Sponges from the Family Mycalidae. *Planta Med.* **82**, 816–31 (2016).

231. Giordano, F., Mayol, L., Notaro, G., Piccialli, V. & Sica, D. Structure and absolute configuration of two new polybrominated C 15 acetogenins from the sponge Mycale rotalis. *J. Chem. Soc. Chem. Commun.* 1559–1561 (1990). doi:10.1039/C39900001559

232. Wang, B.-G., Gloer, J. B., Ji, N.-Y. & Zhao, J.-C. Halogenated organic molecules of Rhodomelaceae origin: chemistry and biology. *Chem. Rev.* **113**, 3632–85 (2013).

233. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).

234. Anckar, J. & Sistonen, L. Regulation of HSF1 function in the heat stress response: implications in aging and disease. *Annu. Rev. Biochem.* **80**, 1089–115 (2011).

235. Beere, H. M. & Green, D. R. Stress management - heat shock protein-70 and the regulation of apoptosis. *Trends Cell Biol.* **11**, 6–10 (2001).

236. Rosenzweig, R., Nillegoda, N. B., Mayer, M. P. & Bukau, B. The Hsp70 chaperone network. *Nature Reviews Molecular Cell Biology* **20**, 665–680 (2019).

237. Kaul, S. C., Deocaris, C. C. & Wadhwa, R. Three faces of mortalin: a housekeeper, guardian and killer. *Exp. Gerontol.* **42**, 263–74 (2007).

238. Schilke, B. *et al.* Evolution of mitochondrial chaperones utilized in Fe-S cluster biogenesis. *Curr. Biol.* **16**, 1660–5 (2006).

239. Liu, Q., D'Silva, P., Walter, W., Marszalek, J. & Craig, E. A. Regulated cycling of mitochondrial Hsp70 at the protein import channel. *Science (80-.).* **300**, 139–141 (2003).

240. Liu, Y., Liu, W., Song, X.-D. & Zuo, J. Effect of GRP75/mthsp70/PBP74/mortalin overexpression on intracellular ATP level, mitochondrial membrane potential and ROS accumulation following glucose deprivation in PC12 cells. *Mol. Cell. Biochem.* **268**, 45–

51 (2005).

241. Wadhwa, R. *et al.* Hsp70 family member, mot-2/mthsp70/GRP75, binds to the cytoplasmic sequestration domain of the p53 protein. *Exp. Cell Res.* **274**, 246–53 (2002).

242. Ma, Z. *et al.* Mortalin controls centrosome duplication via modulating centrosomal localization of p53. *Oncogene* **25**, 5377–90 (2006).

243. Burbulla, L. F. *et al.* Dissecting the role of the mitochondrial chaperone mortalin in Parkinson's disease: functional impact of disease-related variants on mitochondrial homeostasis. *Hum. Mol. Genet.* **19**, 4437–52 (2010).

244. Qu, M. *et al.* Mortalin overexpression attenuates beta-amyloid-induced neurotoxicity in SH-SY5Y cells. *Brain Res.* **1368**, 336–45 (2011).

245. Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* **346**, 623–8 (1990).

246. Vogel, M., Mayer, M. P. & Bukau, B. Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker. *J. Biol. Chem.* **281**, 38705–11 (2006).

247. Zhuravleva, A. & Gierasch, L. M. Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 6987–6992 (2011).

248. Morshauser, R. C., Wang, H., Flynn, G. C. & Zuiderweg, E. R. The peptide-binding domain of the chaperone protein Hsc70 has an unusual secondary structure topology. *Biochemistry* **34**, 6261–6 (1995).

249. Kityk, R., Vogel, M., Schlecht, R., Bukau, B. & Mayer, M. P. Pathways of allosteric regulation in Hsp70 chaperones. *Nat. Commun.* **6**, (2015).

250. Mayer, M. P. *et al.* Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat. Struct. Biol.* **7**, 586–93 (2000).

251. Shida, M. *et al.* Direct inter-subdomain interactions switch between the closed and open forms of the Hsp70 nucleotide-binding domain in the nucleotide-free state. *Acta*

Crystallogr. D. Biol. Crystallogr. **66**, 223–32 (2010).

252. Amick, J. *et al.* Crystal structure of the nucleotide-binding domain of mortalin, the mitochondrial Hsp70 chaperone. *Protein Sci.* **23**, 833–42 (2014).

253. Kaul, S. C., Reddel, R. R., Mitsui, Y. & Wadhwa, R. An N-terminal region of mot-2 binds to p53 in vitro. *Neoplasia* **3**, 110–4 (2001).

254. Gabizon, R. *et al.* Specific recognition of p53 tetramers by peptides derived from p53 interacting proteins. *PLoS One* **7**, e38060 (2012).

255. Ahn, B. Y. *et al.* Tid1 is a new regulator of p53 mitochondrial translocation and apoptosis in cancer. *Oncogene* **29**, 1155–66 (2010).

256. Ahmad, A. *et al.* Heat shock protein 70 kDa chaperone/DnaJ cochaperone complex employs an unusual dynamic interface. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 18966–71 (2011).

257. Wadhwa, R. *et al.* Selective toxicity of MKT-077 to cancer cells is mediated by its binding to the hsp70 family protein mot-2 and reactivation of p53 function. *Cancer Res.* **60**, 6818–21 (2000).

258. Kityk, R., Kopp, J., Sinning, I. & Mayer, M. P. Structure and Dynamics of the ATP-Bound Open Conformation of Hsp70 Chaperones. *Mol. Cell* **48**, 863–874 (2012).

259. Qi, R. *et al.* Allosteric opening of the polypeptide-binding site when an Hsp70 binds ATP. *Nat. Struct. Mol. Biol.* **20**, 900–907 (2013).

260. Deocaris, C. C., Kaul, S. C. & Wadhwa, R. Mortalin's Machinery. in *Mortalin Biology: Life, Stress and Death* 21–30 (Springer Netherlands, 2014). doi:10.1007/978-94-007-3027-4_2

261. Jones, G., Willett, P., Glen, R. C., Leach, A. R. & Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748 (1997).

262. Woodhead, A. J. *et al.* Discovery of (2,4-Dihydroxy-5-Isopropylphenyl)-[5-(4-Methylpiperazin-1-Ylmethyl)-1,3-Dihydroisoindol-2-Yl]Methanone (at13387), a Novel

Inhibitor of the Molecular Chaperone Hsp90 by Fragment Based Drug Design. *J.Med.Chem.* **53**, 5956 (2010).

263. Wright, L. *et al.* Structure-Activity Relationships in Purine-Based Inhibitor Binding to Hsp90 Isoforms. *Chem.Biol.* **11**, 775 (2004).

264. Bowden, B. F., Coll, J. C., Mitchell, S. J., Mulder, J. & Stokie, G. J. Studies of australian soft corals. ix: A novel nor-diterpene from the soft coral sinularia leptoclados. *Aust. J. Chem.* **31**, 2049–2056 (1978).

265. Turner, K. E., Bowden, B. F., Stokie, G. J. & Howard, C. J. (4R*,8S*,11R*,13S*,14R*)-8,11-Epoxy-14-hydroxy-11-methyl-4-(1-methylvinyl)-6,9-dioxocyclotetradec-1-ene-1,3-carbolactone. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **35**, 1283–1284 (1979).

266. Chen, X., Ni, F., Kondrashkina, E., Ma, J. & Wang, Q. Mechanisms of leiomodin 2-mediated regulation of actin filament in muscle cells. *Proc.Natl.Acad.Sci.USA* **112**, 12687–12692 (2015).

267. Langelier, M.-F., Planck, J. L., Roy, S. & Pascal, J. M. Structural Basis for DNA Damage-Dependent Poly(ADP-ribosyl)ation by Human PARP-1. *Science (80-.).* **336**, 728–732 (2012).

268. Hanwell, M. D. *et al.* Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, 17 (2012).

269. Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, NA-NA (2009).

270. Durrant, J. D. & McCammon, J. A. NNScore: A Neural-Network-Based Scoring Function for the Characterization of Protein−Ligand Complexes. *J. Chem. Inf. Model.* **50**, 1865–1871 (2010).

271. Notaro, G., Piccialli, V. & Sica, D. A further C15 nonterpenoid polybromoether from

the encrusting sponge Mycale rotalis | IRIS Università degli Studi di Napoli Federico II. *J. Nat. Prod.* **55**, 626–632 (1992).

272. Gong, W. B., Yang, J., Zhang, H. & Perrett, S. Structure of glutathionylated hHsp70 SBD (385-641). *TO BE Publ.* doi:10.2210/PDB5GJJ/PDB

273. Wisniewska, M. *et al.* Substrate binding domain of the human Heat Shock 70kDa protein 9 (mortalin). *TO BE Publ.* doi:10.2210/PDB3N8E/PDB

274. Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W. & Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins Struct. Funct. Genet.* **52**, 609–623 (2003).

275. Stierand, K., Maass, P. C. & Rarey, M. Molecular complexes at a glance: automated generation of two-dimensional complex diagrams. *Bioinformatics* **22**, 1710–6 (2006).

276. Leth-Larsen, R. *et al.* Metastasis-related plasma membrane proteins of human breast cancer cells identified by comparative quantitative mass spectrometry. *Mol. Cell. Proteomics* **8**, 1436–1449 (2009).

277. Lund, R., Leth-Larsen, R., Jensen, O. N. & Ditzel, H. J. Efficient Isolation and Quantitative Proteomic Analysis of Cancer Cell Plasma Membrane Proteins for Identification of Metastasis-Associated Cell Surface Markers. *J. Proteome Res.* **8**, 3078–3090 (2009).

278. Lu, P., Weaver, V. M. & Werb, Z. The extracellular matrix: A dynamic niche in cancer progression. *Journal of Cell Biology* **196**, 395–406 (2012).

279. Fang, M., Yuan, J., Peng, C. & Li, Y. Collagen as a double-edged sword in tumor progression. *Tumor Biology* **35**, 2871–2882 (2014).

280. Hamidi, H. & Ivaska, J. Every step of the way: integrins in cancer progression and metastasis. *Nat. Rev. Cancer* 1–16 (2018). doi:10.1038/s41568-018-0038-z

281. Zaidel-Bar, R., Itzkovitz, S., Ma'ayan, A., Iyengar, R. & Geiger, B. Functional atlas of the integrin adhesome. *Nat. Cell Biol.* **9**, 858–867 (2007).

282. Horton, E. R. *et al.* Definition of a consensus integrin adhesome and its dynamics during

adhesion complex assembly and disassembly. *Nat. Cell Biol.* **17**, 1577–1587 (2015).

283. Horton, E. R. *et al.* The integrin adhesome network at a glance. *J. Cell Sci.* **129**, 4159–4163 (2016).

284. Winograd-Katz, S. E., Fässler, R., Geiger, B. & Legate, K. R. The integrin adhesome: From genes and proteins to human disease. *Nat. Rev. Mol. Cell Biol.* **15**, 273–288 (2014).

285. Schindler, J. & Nothwang, H. G. Aqueous polymer two-phase systems: effective tools for plasma membrane proteomics. *Proteomics* **6**, 5409–17 (2006).

286. Johansson, G. & Walter, H. Partitioning and concentrating biomaterials in aqueous phase systems. *Int. Rev. Cytol.* **192**, 33–60 (2000).

287. Barinaga-Rementeria Ramírez, I., Ekblad, L. & Jergil, B. Affinity partitioning of biotinylated mixed liposomes: Effect of charge on biotin-NeutrAvidin interaction. *J. Chromatogr. B Biomed. Sci. Appl.* **743**, 389–396 (2000).

288. Gierow, P., Sommarin, M., Larsson, C. & Jergil, B. Fractionation of rat liver plasma-membrane regions by two-phase partitioning. *Biochem. J.* **235**, 685–91 (1986).

289. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–40 (2016).

290. Dennis, G. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, (2003).

291. Huang, D. W. *et al.* DAVID Bioinformatics Resources: Expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* **35**, (2007).

292. Rontogianni, S. *et al.* Proteomic profiling of extracellular vesicles allows for human breast cancer subtyping. *Commun. Biol.* **2**, (2019).

293. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

294. Szklarczyk, D. *et al.* The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368 (2017).