UNIVERSITÀ DEGLI STUDI DI MILANO

Doctoral School in Pharmaceutical Sciences – XXXV cycle

DIPARTIMENTO DI SCIENZE FARMACEUTICHE

Quantitative proteomic approaches to study drug mechanism of action

CHIM/08

SILVIA CASTELLI
R12583

Tutor: Prof. Giancarlo ALDINI; Dott.ssa Barbara VALSASINA

Coordinator: Prof. Giancarlo ALDINI

Academic year 2021/2022

To my beloved family

Index

1 Abstract

In recent years an increased number of covalent protein kinase inhibitors has been approved for cancer therapy and many more are undertaking clinical trials. Covalent binding is usually obtained by introducing in these drugs an electrophilic warhead able to bind specific nucleophilic sites in the protein target.

Covalent inhibition of oncogenic protein kinases allows to obtain a stronger and prolonged therapeutic effect compared to reversible inhibition; however, the choice of the dose to be administered to patients and the evaluation of the selectivity among the kinase family are mandatory to reach pharmacological results reducing possible side effects. From the DMPK perspective for covalent inhibitors, *in vitro* and *in vivo* data extrapolation, to obtain human pharmacokinetic projection, can be challenging and numerous efforts have to be undertaken in developing methods to accurately and quantitively determine inhibitor target engagement in preclinical and more importantly in clinical studies. Chemoproteomic approaches, taking advantage of the use of chemical probes, provide powerful tools to analyze binding characteristics between small molecules and proteins and validate the mode of action of these drug candidates. Click chemistry is a cycloaddition reaction between an azide and alkyne group to generate a 1,4-disubstituted 1,2,3-triazole ring. In this way, it is possible to functionalize the inhibitor under development with an alkyne moiety that could be labelled with a fluorescent-azide for target engagement detection and quantitation in native environments and a desthiobiotin-azide molecule for identification of the inhibitor off-target binders and proteome-wide selectivity. The functionalization of the inhibitor with a small group such as the alkyne group (the molecule obtained is called probe) allows the incubation of the compound directly *in situ*, with minimal alteration of inhibitor features (e.g. potency and permeability).

Thus, the aim of this project is the optimization and application of click chemistry reaction in order to study covalent mechanism of action of ibrutinib, a covalent inhibitor of the tyrosine protein kinase BTK, approved by FDA in 2013 for the treatment of B cell malignancies. We used ibrutinib as case study, but the developed protocols can be applied to the study of other covalent inhibitors.

Abstract

Target engagement conditions were initially optimized on the recombinant BTK protein using an ibrutinib derivative, bearing an alkyne group, and then transferred to cell extracts.
Competition experiments were set up on extracts and then an i*n cell* target engagement experiment was conducted, treating cells with 5 μM ibrutinib. The significant decreasing of the signal intensity of fluorescent probe labeled BTK, after pre-incubation of cells with ibrutinib, suggested a full target occupancy of BTK.
Nevertheless, with the aim to precisely calculate this target occupancy, a chemoproteomic workflow coupled to quantitative mass spectrometry analysis has been optimized and set.
In this case, click chemistry reaction was performed coupling to ibrutinib probe an azide functionalized with a desthiobiotin moiety in order to capture desthiobiotinylated peptides with a streptavidin resin, taking advantage of high affinity between streptavidin and desthiobiotin.
The developed protocols have been validated in cells treated with ibrutinib and, in addition to BTK, two additional protein kinases JAK3 and BLK have been identified modified on cysteine 909 and 319, respectively. These cysteines are located in the ATP-binding site of the two kinases in a position corresponding to the cysteine 481 in BTK. Interaction with ibrutinib for JAK3 and BLK was confirmed by additional analyses on recombinant protein and cell lysates.
Total proteome analysis, both in label free quantitation and TMT mode, has been undertaken to additionally characterize ibrutinib treated cells.
The work allowed to better understand the preclinical profile of an oncological target therapy drug in term of potency and selectivity in living cells. The process is transferable to other covalent drugs and applied not only in preclinical models but also in clinical trials, helping in the definition of the optimal dose for patients to obtain the best efficacy, limiting side effects.

2 ABBREVIATIONS

ABPP	Activity-based proteome profiling
ACN	Acetonitrile
AGC	Automatic gain control
Ambic	Ammonium Bicarbonate
BTK	Bruton's tyrosine kinase
BSA	Bovine serum albumin
CuAAC	Copper-catalyzed alkyne azide cycloaddition
DDA	Data dependent acquisition
DMSO	Dimethyl sulfoxide
DTT	DL-Dithiothreitol
ELISA	Enzyme-linked immunosorbent assay
ESI	Electrospray ionization
FA	Formic acid
FDR	False discovery rate
HCD	Higher-Energy Collision Dissociation
HEPES	N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid)
HPLC	High-performance liquid chromatography
IC50	Half maximal inhibitory concentration
LC-MS	Liquid chromatography – Mass spectrometry

LFQ	Label-free quantification
MALDI	Matrix Assisted Laser Desorption Ionization
MBR	Match-between run
MOPS	3-(N-morpholino)propanesulfonic acid
NCE	Normalized collision energy
PBS	Phosphate-buffered saline
PK-PD	Pharmacokinetic-pharmacodynamic
PRM	Parallel reaction monitoring
PSM	Peptide-spectrum match
PVDF	Polyvinylidene difluoride
SDS	Sodium Dodecyl Sulphate
SDS-PAGE	Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis
TBS	Tris-buffered saline
TCEP	Tris(2-carboxyethyl)phosphine
TEAB	Triethylammonium bicarbonate buffer
TFA	Trifluoroacetic acid
THPTA	Tris-hydroxypropyltriazolylmethylamine
TMT	Tandem mass tag
TOF	Time of flight
Tris-HCl	Tris(hydroxymethyl)aminomethane hydrochloride

3 Introduction

3.1 Proteomics

Proteome can be defined as the entire set of proteins expressed in a cell, tissue, or individual. The term was first coined by Wilkins in 1996 to describe "proteins as a complement of genomic data". (*1*)

Proteomics denotes the characterization of proteome, including expression, structure, functions, interactions and post translational modifications. Proteins are the effectors of biological functions and protein levels depend not only on the levels of corresponding mRNA but also on host translational control and regulation. (*2*)

Thus, proteomics would be considered as the most relevant data set to characterize a biological system. The analysis of proteins however is challenging due to their high dynamic range and the diversity of molecular size, hydrophobicity, hydrophilicity and the presence of a number of post translational modifications.

With the purpose of studying proteome, different techniques, from the mid-1970s, were used (*3*) starting from two-dimensional gel electrophoresis (2D) (*4*). This approach is based on separating proteins according to their isoelectric point (IsoElectroFocusing) and subsequently by their molecular weight through SDS-PAGE. Protein detection is usually performed by staining the gel with Blue Coomassie or fluorescent dyes while the identification is carried out by specific antibodies through Western blotting technique or later on by SDS-PAGE band trypsin digestion and mass spectrometry analysis. Two-dimensional gel electrophoresis, unfortunately, presents some negative aspects that have limited its use in the analysis of the whole proteome: a limited capacity of separation of very complex mixtures, the low reproducibility and the difficulty in detection of proteins with very different levels of expression.

Advances in mass spectrometry-based techniques have contributed to a better interrogation of proteome as well as drug-protein interaction. Biological mass spectrometry (MS) was introduced with the development of electrospray and MALDI ionization techniques. This advance made biological molecules readily analyzable by mass spectrometry and entailed the Chemistry Nobel Prize in 2002.

Proteomic analysis can detect a wide range of processes such as protein expression profiling, protein modifications, protein-protein interactions, protein structure, and protein function and the results

obtained, can be used to understand disease processes, provide diagnosis and prognosis of diseases in drug development phase.

To this purpose, proteomic approaches have become popular in cancer studies; proteomics-based technologies have enabled the identification of protein expression patterns that can be used to assess tumor prognosis, prediction, tumor classification and to identify potential biomarkers for specific therapies.
In addition, proteomics techniques have been applied to understand the basic biology of cancer in order to dissect how the signalling pathways in tumor cells are altered, increasing the possibility of identifying specific protein targets to be inhibited with selective drug molecules. (*5-8*)

3.2 Anticancer drug mechanism of action

3.2.1 Cancer

Cancer is a multifactorial disease, and it is one of the leading causes of death worldwide.
According to estimates from the World Health Organization (WHO) in 2019, cancer is the first or second leading cause of death before the age of 70 years in 112 of 183 countries and ranks third or fourth in a further 23 countries (Figure 1). (*9*)

There were an estimated 19.3 million new cases and 10 million cancer deaths worldwide in 2020. Regarding the distribution of all-

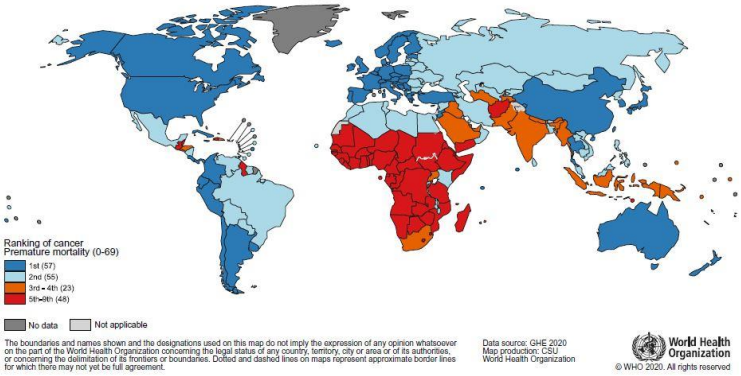


Figure 1 - National ranking of cancer as a cause of death at ages <70 in 2019. The number of countries in each ranking are included in the legend.

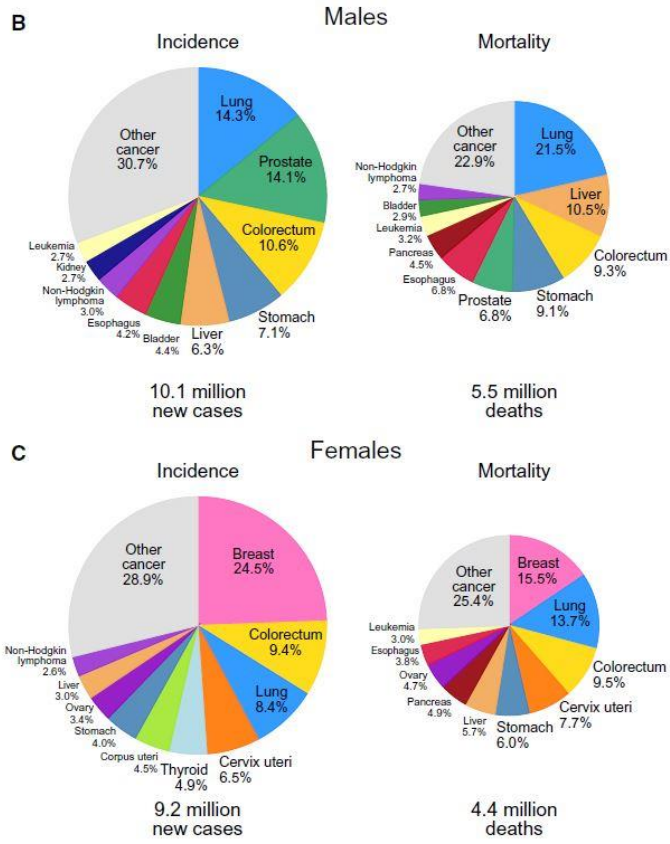


Figure 2 - Distribution of Cases and Deaths for the Top 10 Most Common Cancers in 2020 for Men and Women. Nonmelanoma skin cancers (excluding basal cell carcinoma for incidence) are included in the "other" category. Source: GLOBOCAN 2020.

cancer incidence and mortality according to world region, for both sexes combined, one-half of all cases and 58.3% of cancer deaths are estimated to occur in Asia in 2020, where 59.5% of the global population resides. Europe accounts for 22.8% of the total cancer cases and 19.6% of the cancer deaths, although it represents 9.7% of the global population, followed by the Americas' 20.9% of incidence and 14.2% of mortality worldwide.

Figure 2 shows the top 10 cancer types for estimated cases and deaths worldwide for men and women separately. Female breast cancer is the most commonly diagnosed cancer (11.7% of total cases),

closely followed by lung (11.4%), colorectal 10.0%), prostate (7.3%), and stomach (5.6%) cancers. Lung cancer is the leading cause of cancer death (18.0% of the total cancer deaths), followed by colorectal (9.4%), liver (8.3%), stomach (7.7%), and female breast (6.9%) cancers. Lung cancer is the most frequently occurring cancer and the leading cause of cancer death in men, followed by prostate and colorectal cancer for incidence and liver and colorectal cancer for mortality. In women, breast cancer is the most commonly diagnosed cancer and the leading cause of cancer death, followed by colorectal and lung cancer for incidence.

Cancer is a multistep process characterized by a sustained proliferative signaling, growth suppressors evasion, cell death resistance, immortal replication, angiogenesis induction and invasion and metastasis activation (Figure 3-A).

These hallmarks of cancer were coined by Douglas Hanahan and Robert Weinberg in their paper *The Hallmarks of Cancer* published January 2000 in Cell (*10*). They were originally composed by six biological capabilities acquired during the multistep development of tumors and have been updated in 2011 to eight capabilities (*11*), including abnormal metabolic pathways and evasion of the immune system and two enabling characteristic: genome instability and mutation and tumor-promoting inflammation (Figure 3-B).

Further studies identified different cancer subtypes driven by specific targets activating pathways recapitulated in the hallmark of cancers and changed the mindset of pathologists and clinicians going from a histological and organ related tumor classification to a molecular classification where the same target can drive tumors in different organs. (*12, 13*)

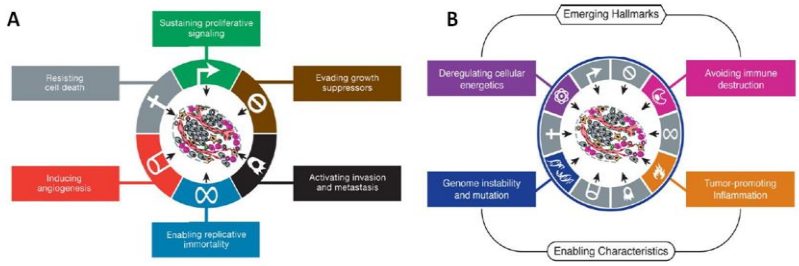


Figure 3 - Hallmark capabilities of cancer; A: Hallmarks originally proposed by D. Hanahan and R. Weinberg in 2000; B: Emerging hallmarks and enabling characteristics proposed by D. Hanahan and R. Weinberg in 2011

These advances in knowledge about mechanisms led to the development of specific therapies that interfere with each of the acquired capabilities necessary for tumor growth and progression.
Drug treatment together with surgical operation, radiotherapy and biotherapy constitute the main approaches to cancer treatment. For a long time, chemotherapy, (*14*) which is a method of killing tumor cells and inhibiting the growth and proliferation of tumor cells by chemical drugs (i.e. alkylating agents, antimetabolites, mitotic and topoisomerase inhibitors), was the only approach to cancer drug therapy. The biggest limitation of chemotherapy is the inability to distinguish between cancer cells and normal cells, resulting in systemic toxicity in normal rapidly proliferating tissues and consequent dose reduction or treatment discontinuation (*15*).

More recently new molecules, specifically inhibiting targets responsible for driving tumor growth, have been explored. The concept was introduced for the first time by Ehrlich in 1906 (*16*) that proposed the idea of the drug as "magic bullet" that would selectively find its target and only destroy it without affecting the organism; more recently, this concept was (*17*) encouraged by the FDA approval in 2001 of the first small-molecule tyrosine kinase inhibitor (TKI) Imatinib mesylate (GlivecR) that is able to inhibit the oncogenic fusion protein BCR-ABL, generated by a chromosomal rearrangement (Philadelphia chromosome- Ph) in patients with chronic myelogenous leukemia (CML) and acute lymphocytic leukemia (ALL) that are -positive (Ph+) (*18*). Another successful example of precision medicine comes from Piccart-Gebhart MJ., (*19*) with the development of the monoclonal antibody Trastuzumab able to target the single gene mutation in the human epidermal growth factor receptor 2 (HER2) gene in breast cancer cells; giving birth to the era of targeted therapy.

3.2.2 Oncological molecular targeted therapy

Oncological targeted therapy works by targeting cancer-specific genes, proteins or tissue environment that contribute to cancer growth and survival (*20*). In this way it is possible to kill tumor cells as the inhibitor interacts with specific targets present on malignant cells, without threatening the normal cells, thus reducing toxicity.
Agents used in molecular targeted therapy are classified into small molecules, monoclonal antibodies, immunotherapeutics and gene therapies and they can block signals that favor the promotion of cancer

cell growth, interfere with the regulation of cell cycle or induce cell death to kill cancer cells (*21*).

Antibodies (*22*) are typically characterized by high selectivity; however, their targets are often restricted to the cell surface and they require intravenous or subcutaneous dosing because of their large molecular weight. Small molecule inhibitors, instead, vary in selectivity, and can potentially bind a wider range of extracellular and intracellular targets. There are a total of 89 anti-cancer small molecules approved in the United States and China. Figure 4 summarizes the small-molecule anti-cancer drugs approved by the US FDA and National Medical Products Administration (NMPA) of China since 2001. The targets of these drugs include kinases, epigenetic regulatory proteins, DNA damage repair enzymes, and proteasomes (*23*).

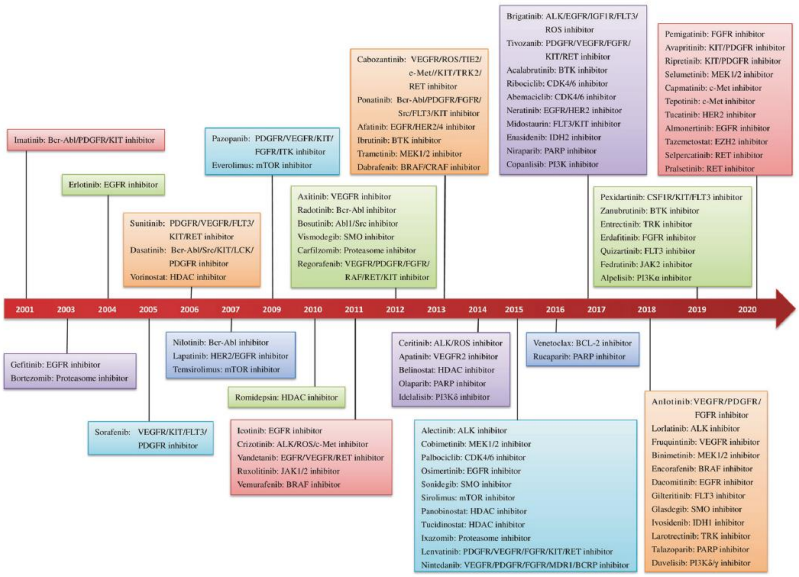


Figure 4 - Timeline for the approval of small molecule targeted anti-cancer drugs

3.2.3 Small molecule protein kinase inhibitors

Small molecule drugs are defined as compounds with relatively low molecular weight (<900 Da) that are able to penetrate into cells to target specific protein within cells (*24*).

Different small molecule kinases inhibitors were developed in order to interrupt the signaling pathways which are dysregulated in tumors: currently, there are multiple examples of small molecule kinase inhibitors that have produced clinical benefit. Among the first developed oncogenic protein kinases inhibitors it is worthwhile to mention, imatinib (*25*) utilized to inhibit BCR-ABL in chronic myelogenous leukemia (CML) and acute lymphoblastic leukemia (ALL) bearing Philadelphia chromosome; crizotinib (*26*) and other ALK kinase inhibitors for cancers driven by *ALK* fusions; lapatinib (*27*) for *ERBB2/ HER2*-amplified tumors; gefitinib (*28*) and erlotinib for *EGFR* mutated tumors and vemurafenib (*29*) for *BRAF* mutant tumors. In each of these cases, superior clinical benefit was observed for the targeted agent in molecularly selected patients compared with prior standard-of-care regimens. Second generation molecules against these targets have been also developed more recently to overcome tumor drug resistance mechanisms (*30*) based on kinase mutations reducing drug affinity for the target of interest.

Protein kinases are indeed a class of enzymes that catalyze the transfer of the γ-phosphate group of ATP to the hydroxyl group of substrate proteins; thus, altering the activity and/or structure of the substrate. Kinases play important roles in signal transduction and regulate a variety of cellular processes including metabolism, transcription, cell cycle progression, cytoskeletal rearrangement, apoptosis and differentiation.

According to the substrate residues, protein kinases can be classified as tyrosine kinases (including both receptor and nonreceptor tyrosine kinases), serine/threonine kinases, and tyrosine kinase-like enzymes.

Because of their overexpression and genetic alterations such as mutations and translocations, kinase signaling pathway dysregulation is associated with a variety of conditions involved in processes leading to tumor cell proliferation and survival. For instance, kinase signaling pathways have been shown to drive many of the hallmark phenotypes of tumor biology, including proliferation, survival, motility, metabolism, angiogenesis, and evasion of antitumor immune

responses. Thus, kinases represent promising targets for therapeutic approaches (*31*).

Mutated or translocated kinases often have transforming capacity and are therefore considered to be oncogenic**.** They become constitutively activated and indispensable for cancer cell survival, making them essential for survival and/or proliferation of the cancer cell. This process, called "oncogene addiction" makes cancers susceptible to appropriate kinase inhibition. For example, the V600e mutation, located in the activation loop of BRAF, has been implicated in development of carcinomas of the skin, ovary, thyroid, colon and pancreas (*32, 33*). Moreover, a number of altered receptor tyrosine kinases are driving tumors such as FLT3 in acute myeloid leukemia (AML), RET in Thyroid and a small number of lung tumors, ALK in neuroblastoma. Another important kinase target is correlated to kinases not oncogenic or mutated but required for the survival and /or proliferation of cancer cells and located in key signaling pathways. For example, MEK1 and MEK2 (MAP2K1 and MAP2K2) that are located in critical MAPK pathway (*34*). Other examples of targetable kinases include cyclin-dependent kinases (CDKs), which regulate cell cycle transitions; the Aurora kinases, which are essential for chromosomes division and the Polo-like kinases, which are important during both mitosis and cytokinesis.

The human genome contains about 500 protein kinase genes, representing approximately 2% of the whole human proteome and it is well recapitulated by Manning *et al.,* in a paper published in 2002 where the kinase tree has been represented for the first time (*35*). Protein kinases typically share conserved secondary structure elements that are arranged into 12 subdomains folding into a bilobal catalytic core structure with the site of phosphate transfer located between these two lobes with ATP occupying a narrow hydrophobic cleft located between the lobes. The adenine ring forms hydrogen bonds with the backbone of a peptide linker, referred as the kinase "hinge" (the segment that connects the amino- and carboxy-terminal kinase domains). The ribose and triphosphate groups of ATP bind in a hydrophilic channel extending to the substrate binding site. All kinases have a conserved activation loop, which is important in regulating kinase activity and is marked by conserved DFG (Aspartic acid – Phenylalanine - Glycine) and APE (Alanine – Proline - Glutamic acid) motifs. The active sites of protein kinases are highly dynamic and the

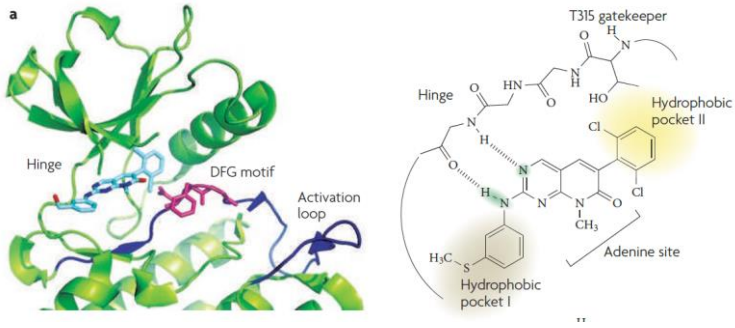


Figure 5 - Kinase inhibitor binding models. ABL1 in complex with the type 1 ATP-competitive inhibitor PD166326

movement of these motifs, allows kinases to interconvert between catalytically active and inactive forms.

Although all protein kinases share the common pocket allocating ATP, differences in aminoacid composition inside the pocket can be exploited in order to obtain selectivity towards the desired kinase target among the other components of the family (*36*). ATP pocket indeed has been largely studied to discover and develop small molecules able to compete with ATP and generate potent and selective inhibitors of oncogenic protein kinases.

Protein kinases inhibitors are classified as (*37*):

Type I

They constitute the majority of ATP-competitive inhibitors and recognize the active conformation of the kinase by miming the purine ring of the adenine moiety of ATP. These inhibitors (Figure 5) usually contain a heterocyclic ring that occupies the purine binding site, where it serves as a scaffold for side chains that occupy adjacent hydrophobic regions. The hydrophilic regions of the enzyme, occupied by the ribose moiety of ATP, may be used to exploit the solubility of the drugs. Numerous type I kinase inhibitors for the treatment of cancer have been approved by the FDA (i.e. bosutinib, crizotinib, dasatinib, erlotinib and gefitinib) (*31*).

Type II

These inhibitors recognize the inactive conformation of the kinase which is characterized by the movement of the DFG motif. This

displacement results in the Asp of the DFG-motif to be unable to coordinate a magnesium ion that is important for phosphate transfer and prevents the Phe side chain from participating in a regulatory network of hydrophobic residues (*38*). These types of inhibitors are designed to occupy the pocket created by the movement of the Phe side chain in the DFG-motif and in general, are sterically incompatible with the active forms of protein kinases. They have a high degree of selectivity towards unwanted kinases determining an increase in the safety profile of type II kinase inhibitors. Examples of type II inhibitors, FDA approved, include the ABL1, KIT and PDGFR inhibitors Imatinib and Nilotinib and the multikinase inhibitor Sorafenib.

Type III

This class of compounds binds outside the ATP-binding site and modulate kinase activity in an allosteric manner. These types of inhibitors tend to exhibit the highest degree of kinase selectivity because of their binding sites and regulatory mechanism that are unique to a particular kinase. These drugs are noncompetitive inhibitors because ATP cannot prevent their interaction with the target kinase. One of the earliest allosteric inhibitors is CI-1040, which inhibits MEK1 and MEK2 by occupying a pocket adjacent the ATP binding site.

Type IV

The fourth class of kinase inhibitors forms an irreversible, covalent bond to the kinase active site, most frequently by reacting with a nucleophilic cysteine residue located inside or nearby the ATP pocket (e.g. BTK inhibitors ibrutinib and acalabrutinib). These inhibitors are named target covalent kinase inhibitors (TCKIs) and the main features are described below.

3.2.4 Target covalent kinase inhibitors

TCKIs are designed to covalently bind and inhibit target kinase proteins. They bind to their target in two steps: the compound must first bind non-covalently to the target protein, these assures that a sufficient degree of selectivity is achieved. Then, a covalent bond formation takes place, permanently disabling enzymatic activity (Figure 6). Importantly, only the first of these two steps is ATP-competitive.

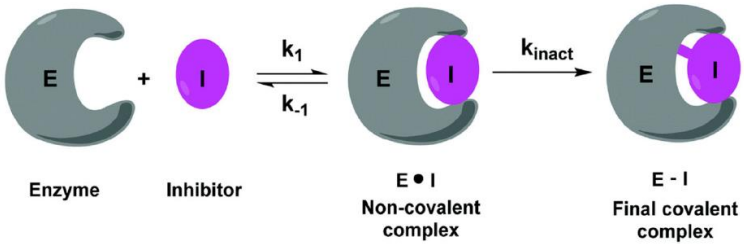


Figure 6 - General mechanism of covalent interaction between a small molecule and its target

In order to obtain these two steps, TCKIs are characterized by a moiety that recognizes a specific binding site of the protein target (i.e. the ATP pocket) and an electrophilic warhead able to covalently bind specific nucleophilic sites (cysteine, serine, tyrosine, lysine, threonine) in or close to the protein target binding site (*39, 40*).

Irreversible covalent inhibition is a non-equilibrium process: irreversible inhibitors interact with their targets in a time-dependent manner, and the reaction proceeds to completion rather than to equilibrium. The potency and selectivity of covalent inhibitors is governed by two parameters: the affinity of initial non-covalent binding, K_i (koff/kon), and the rate of the subsequent bond-forming reaction, k_{inact} *(41, 42)*.
The potency and selectivity of conventional reversible inhibitors are typically defined in terms of the equilibrium binding affinity for the target, or the concentration of the compound that is required to achieve 50% inhibition in a biochemical or cellular assay (IC_{50}). However, the potency of irreversible covalent inhibitors must be considered differently. If the reaction is allowed to proceed for a sufficient period of time, any inhibitor concentration would be expected to result in essentially complete inhibition. Therefore, consideration of the time-dependence of inhibition is essential to any assessment of the activity of covalent inhibitors. Importantly, the potency and selectivity of an irreversible inhibitor can be optimized by altering the structure of the compound to modulate either its non-covalent binding to the target (K_i), or the rate at which it reacts with the target nucleophile after it is bound (k_2). (*41, 43*)

For high selectivity, the non-covalent affinity (K_i) of the inhibitor must be high enough to ensure that the compound binds selectively to the desired target and achieves a residence time that is sufficient for a

covalent reaction. In parallel, the reaction rate of the bound inhibitor (k_{inact}) must be high enough to give a high probability that the reaction will occur within the lifetime of the non-covalent complex that is formed in the initial step of the reaction. However, because highly reactive electrophiles must be avoided, this reaction rate must be achieved by optimal positioning of the electrophile relative to the nucleophile on the target.

Covalent inhibitors have many desirable features, including increased biochemical efficiency of target inhibition, being not subjected to classical equilibrium kinetics and not limited by the competition with high endogenous substrate concentrations (ATP); moreover, the mechanism of action is prolonged as the target activity can only be recovered by new protein synthesis. Since they are highly potent, with low IC50 values, and long binding duration, the necessary dose and intake frequency are lower compared to normal drugs; the prolonged duration of drug action on the target separates the pharmacodynamics of the drug from the pharmacokinetics of exposure, as target inhibition persists after the drug has been cleared. This property enables less frequent dosing and the potential for lower drug doses. Uncoupling PK from PD could also minimize off-targets inhibition, improving safety. Indeed, it has been shown that the compliance of patients increases significantly with lower dosing frequency, especially with once-daily dosing, which can be achieved with a covalent inhibitor (*44*).

Zhao et al. (*45*) hand-curated the distribution of covalent kinase inhibitors (CKIs) from published reviews and databases. A total of 202 CKIs were described, distributed across 55 kinases with EGFR having the highest number of released CKIs. For every kinase, the aminoacid residues involved in the covalent interaction are shown in Figure 7-B. These residues comprise cysteine, lysine and aspartic acid and were binned into 12 regions based on their spatial locations in the tertiary structure of the kinase domain. The three locations with the most CKIs are Front Pocket (128 CKIs), P-loop (35 CSKIs) and DFG-1 (18 CSKIs).
As reported above, CKIs are characterized by a warhead that improves the binding affinity and selectivity by forming a covalent interaction with the kinase reactive residue.

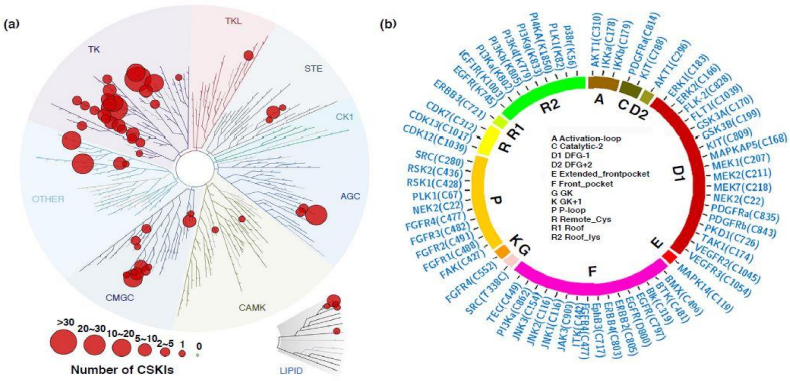


Figure 7 - Distribution of kinases, amino-acid and locations for recognized CKIs

Zhao et al summarized 14 types of warheads classified for the frequency and targeted residues. Warhead type 1-11 target cysteines distributed at different locations on the kinase domains; mostly among the front pocket (Figure 8). Warhead 1, acrylamide, is the most popular reactive group with 94 CSKIs. Warhead 2 is related to warhead 1 and have 31 CSKIs developed. The spatial size of the electrophile requires a larger space to be accommodated; thus, the targeted residues are distributed at the front pocket and remote cysteine regions, where larger reaction group can be tolerated.

Based on the active conformation of the kinase catalytic domain, Knapp *et al.* (*46*) identified a total of 18 positions harboring cysteines (Figure 9), that can be targetable by electrophilic inhibitors, within or in close proximity to the ATP binding site clustered in the glycine rich loop (P1–P4), the DFG-1 (D1), roof (R1), and gatekeeper (G1) positions,

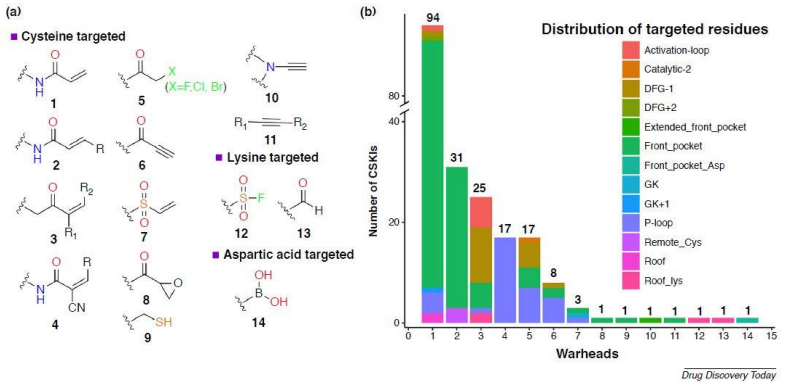


Figure 8 - Type and distribution of warheads

the hinge region (H1–H3), the solvent exposed front region (F1-F4), a distant side common to DK12 and CDK13 (S1) and three diverse sites including the activation segment (O1–O3). These cysteines have an accessible side chain thiol that can potentially be covalently linked to ATP-mimetic inhibitors.
As a consequence of the identification of targetable cysteines that are unique to the desired target, the clinical development of the selective JAK3 covalent inhibitor Ritlecitinib (*47*) has been successful since it targets selectively cysteine 909 of JAK3 that is absent in the other members of the family JAK1, JAK2 and TYK2.

Currently, 8 covalent kinase inhibitors (Table 1) have been approved by the FDA (*48*) for treating various types of cancer in clinic. Afatinib, a covalent inhibitor of EGFR was approved by the FDA for the treatment of EGFR-driven non-small cell lung carcinoma (NSCLC) in 2013; dacomitinib, a structural analog of afatinib, developed by Pfizer, is used in metastatic NSCLC with exon 19 deletions. Osimertinib is a third-generation mutant-selective EGFR inhibitor approved for the treatment of metastatic NSCLC with EGFR T790M resistance mutation. Ibrutinib, a covalent inhibitor of BTK, was firstly approved by the FDA in 2013 for the treatment of mantle cell lymphoma (MCL) and later approved for the treatment of chronic lymphocytic leukemia (CLL).
Neratinib which is a quinoline-derived pan-HER-inhibitor for treatment of HER2-positive breast cancer and ibrutinib, acalabrutinib and zanubrutinib for covalent inhibition of BTK.

All of these kinase inhibitors bind to the ATP binding site of the tyrosine kinase family and target a cysteine thiol.
Despite the advantages of covalent kinase inhibitors described above, the reactivity of covalent inhibitors can lead to non-specific binding to other proteins, increasing the risk of developing off-targets effects.

Target	Cys	Stage	Indication
EGFR	C797	Afatinib, dacomitinib, osimertinib	Lung cancer
HER2	C805	Neratinib	Breast cancer
BTK	C481	Ibrutinib, acalabrutinib, zanubrutinib	Lymphoma
KRAS	G12C	Sotorasib	Lung cancer

Table 1 - List of covalent kinase inhibitors FDA approved

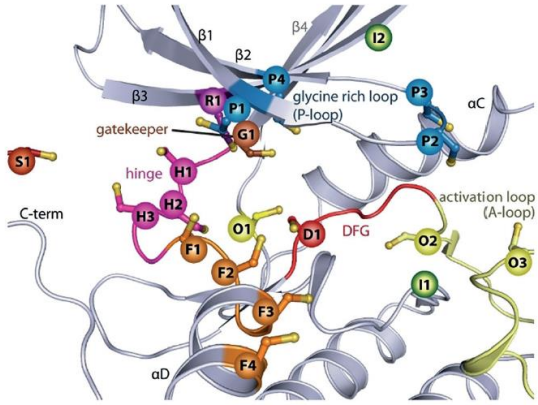


Figure 9 - Positions of cysteines with accessible side chain thiols that can potentially be covalently linked to ATP-mimetic inhibitors are highlighted as spheres

It is essential to determine the identity of these off-targets in order to improve the selectivity and efficacy of covalent inhibitors. Indeed, it is important to confirm that a covalent inhibitor interacts with its protein target in a living system; this type of parameter is referred to as target engagement.

3.3 Target engagement studies

One of the challenges in the development of covalent kinase inhibitors is to identify compounds that are significantly more selective for the target in order to minimize unwanted side effects. This has been facilitated by technology platforms that allow for the profiling of inhibitors in every stage of the drug discovery in order to reduce attrition rates and development costs (*49*).

Only between one in 10 and one in 20 drugs that enter Phase I trials will eventually become approved for clinical use; moreover, Phase II (Figure 10) is the key driver of attrition, blocking two-thirds of compounds that enter it. In addition, the preclinical stage itself only has a 51% success rate for discovery research and 69% for preclinical development (*50*).

A key reason for this derives from an incomplete picture of the pharmacokinetic-pharmacodynamic (PK-PD) relationships of the drug candidate and its target under investigation: it is important to confirm that the drug engages the nominated target in the relevant

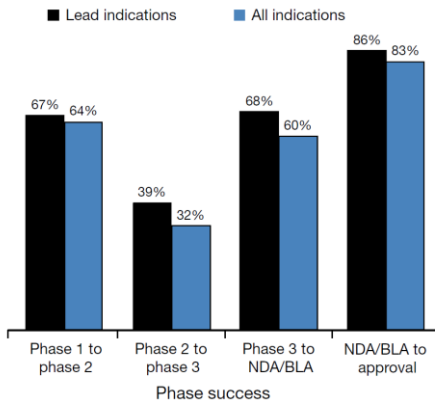


Figure 10 - Phase success rate for lead and all indications. The rate represents the probability that a drug will successfully advance to the next phase

compartment but also to quantify the relationship between drug-target occupancy and its functional consequences.

Furthermore, a recent analysis demonstrates that a high proportion of cancer drugs achieve their cytotoxicity through off-target effects, further increasing the need to validate potential compound targets before clinical development (*51*).

An analysis of 44 phase II programs at Pfizer (*52*) identified "lack of efficacy" as the most common cause of attrition in their discovery programs. They suggested three key aspects for increasing the chance of clinical efficacy and reducing attrition in late drug development stages, termed the 'three pillars of survival': the unbound ligand has to reach the location of its pharmacological target (e.g. cellular compartment) and it has to reside there for a sufficient period of time; the molecule has to engage the target and binding to the target has to results in a pharmacological effect.

AstraZeneca scientists have published a review (*53*) in which they identified five critical technical aspects of project success. They coined the 'five R': the right target, the right patient, the right tissue, the right safety and the right commercial potential. To assess the right tissue, it is necessary to evaluate PK properties and target engagement to assess PK/PD correlation.

Detection of protein-ligand interaction using purified proteins does not always translate into ligand-protein binding in living systems due to changes in small molecule substrate or cofactors or complexes between target protein and biomolecules in cells; also, indirect

approaches do not reveal off-targets of the small molecule and if no effect is observed it remains unclear if the molecule did not access the target or if binding does not result in the desired effect (*54, 55*).

On the contrary, studying native proteins in cell extracts, tissues or *in vivo,* allows to preserve protein integrity, folding, post-translational modification and interaction with other proteins; additionally, it is essential to attribute the pharmacological effects to perturbation of the protein of interest and the potential drug interactions with off-target proteins in order to define the drug selectivity (*56*).

An indirect way to measure target engagement is the analysis of a functional effect downstream of the predicted target protein upon treatment with a small molecule (e.g. visualizing calcium mobilization using fluorescent calcium indicators after ibrutinib treatment (*57*)).
The advantage of this assay is the direct evaluation of a desired phenotype upon compound treatment. Moreover, a true cellular effect not biased by sample processing can be observed. This indirect approach, however, has limitations regarding target engagement determination: it does not reveal off-targets of the investigated small molecule, it can achieve misleading results in case an effect on a modulated pathway is compensated, and if no effect is observed, it remains unclear if the molecule did not access the target or if binding does not result in the desired effect. Further, setting up functional assays for a certain target can be challenging and for some effects a suitable functional readout may not be available.

In recent years, methodologies (*49*) have emerged that enable detection of target engagement in cells, tissues and *in vivo*. An ideal assay would measure the extent of target engagement which can help to determine drug doses that produce efficacy, limiting side effects and the potential for drug interactions with off-target proteins.
The first class of approaches requires the modification of both small molecule and target protein with fluorescence or bioluminescence resonance energy transfer measurements (FRET and BRET). In FRET approach, the inhibitor is linked to a fluorophore (acceptor) at a site that does not affect binding and the target protein is expressed in cells fused to a fluorophore protein (donor). When acceptor and donor are located in close proximity, the donor transfers non-radiative energy to the acceptor, generating fluorescence. FRET technique coupled with FLIM (fluorescence lifetime imaging microscopy), allows monitoring the target engagement in cells providing information on the spatial and

temporal distribution of the protein-inhibitor complex. BRET is the process of energy transfer between a luminescent donor (i.e. target protein expressed in cells fused with luciferase) and a fluorescent acceptor (i.e. inhibitor tagged with a fluorophore). When the inhibitor is in close proximity of the target protein, resonance energy is transferred to the acceptor which, upon excitation, emits light. However, functionalization of a target protein might alter its endogenously expression and activity; hence, a second class of approaches, that requires modification only at the level of the small molecule, was developed, based on chemical proteomic techniques.

3.4 Chemical Proteomics to measure drug-target engagement

Chemoproteomic approaches provide powerful tools to analyze binding characteristics between small molecules and proteins and validate the mode of action of drug candidates. With this technique, it is possible to attribute the pharmacological effects to perturbation of the protein of interest and the potential drug interactions with off-target proteins in order to define the drug selectivity.

Target engagement and occupancy studies in cells confirm direct binding of a ligand to its intended target protein and provide the binding affinity. Combined with biomarkers to measure the functional consequences of target engagement, these experiments can increase confidence in the relationship between *in vitro* pharmacology and observed biological effects.

3.4.1 Activity and affinity-based proteome profiling

Different chemical proteomic approaches have been developed for evaluating drug-target engagement.

Activity-based protein profiling (ABPP) is a chemical proteomic technique that uses small-molecule probes to directly understand the functional state of enzymes in biological systems. Developed by Benjamin F. Cravatt and Matthew Bogyo in the late 1990s, ABPP (*58-60*) employs activity-based probes (ABPs) to specifically label proteins in complex samples (Figure 11).

These ABPs are designed to covalently modify the active site of the target protein; for example, considering protein kinases, an acyl phosphate ATP derivative has been generated in order to covalently

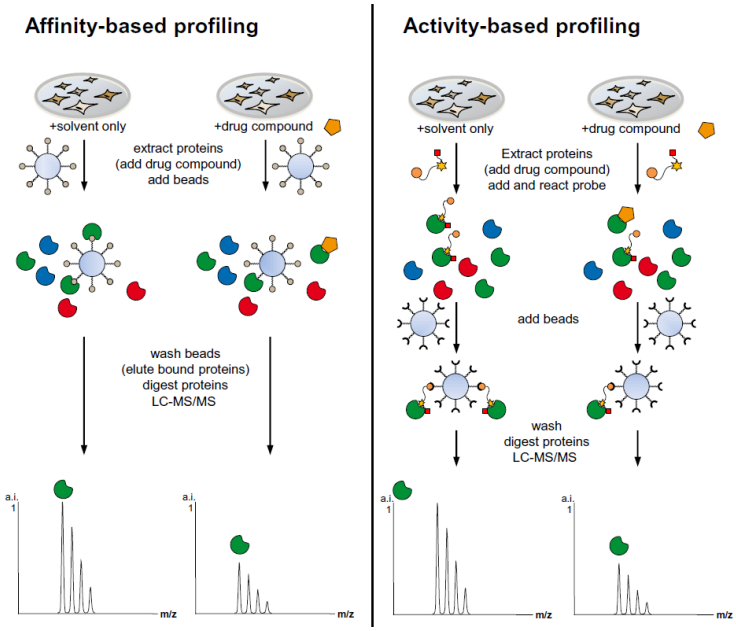


Figure 11 - Direct Affinity- and Activity-Based Chemoproteomics

derivatize catalytic lysine in active protein kinases (*61*). A critical step is the design and synthesis of the chemical probes. They derived from covalent inhibitors of the target protein and they are characterized by a reactive group (the warhead that covalently labeled the aminoacid of the protein target), reporter group (fluorescent dyes and biotin that allow labeled proteins to be visualized and enriched for subsequent studies) and the linker that can be a hydrophilic chain, a lipophilic chain or a peptide, in which its basic function is to provide adequate space between the reactive and the reporter groups (*59*).

For proteins that lack covalent inhibitors, a complementary chemoproteomic platform named affinity-based proteome profiling has also been introduced. These assays involved treating cells with inhibitors followed by cell lysis and broad profiling of kinase activities in native proteomes. Bantscheff *et al.* (*62*) incubated proteomes with bead-immobilized broad-spectrum kinase inhibitors (kinobeads) and then bound kinases were analyzed and quantified by LC-MS.

Accordingly, using these approaches, the development of chemical probes of covalent inhibitors has enabled selectivity profiling and target engagement studies using both in gel fluorescence and mass

spectrometry. Chemical probes can be added to untreated cells to identify reactive proteins or used in competitive mode to identify proteins whose signals are blocked by a pre-treatment of cells with unlabeled chemical probe. However, bulky reporter groups can also impair cell permeability and protein interactions. Therefore, significant advances were achieved by functionalizing the molecule with alkyne or azide groups, which impose minimal steric alteration of the parent inhibitor and that can be coupled to reporter tags through biologically inert (biorthogonal) reactions like copper(I)-catalyzed azide-alkyne cycloaddition (CuAAC).

3.4.2 Chemoproteomic studies based on click chemistry reaction

Click chemistry has emerged as a powerful approach to characterize the selectivity of covalent inhibitors. It is a chemical strategy that utilizes covalent probe to profile the functional state of enzymes in complex proteomes. This type of reaction was introduced by Sharpless in 2001 (*63*) and it is characterized by these features: *The reaction must be modular, wide in scope, give very high yields, generate only inoffensive byproducts that can be removed by nonchromatographic methods, and be stereospecific (but not necessarily enantioselective). The required process characteristics include simple reaction conditions (ideally, the process should be insensitive to oxygen and water), readily available starting materials and reagents, the use of no solvent or a solvent that is benign (such as water) or easily removed, and simple product isolation*.

The most widely used click chemistry reaction is the copper (I)-catalyzed cycloaddition reaction (CuAAC) between an alkyne and an azide group to generate a 1,4-disubstituted 1,2,3-triazole ring (Figure 12).

Replacing bulky reporter tags (biotin or fluorophore) with alkyne and azide groups enables the probe-labeling step to occur in vivo within live cells and organism. Moreover, the alkyne is incorporated

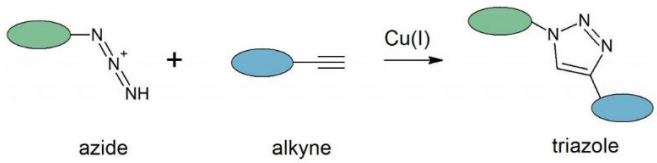


Figure 12 - CuAAC reaction scheme

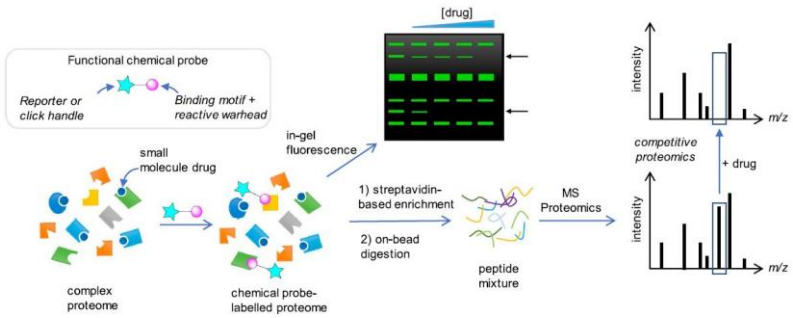


Figure 13 - General chemical probe-based labeling strategy to validate target engagement in complex proteomes

into the covalent inhibitor with minimal disruption to the cell permeability. The lysates can, then, be subjected to click chemistry conjugation with the desired reporter tag for the identification of the targeted proteins.

The detection of the probe–protein complex can be carried out using different reactive–reporter partners, as reported in Figure 13: imaging reporters (fluorescent reporter tag) and affinity reporters (biotin/desthiobiotin reporter tag). For the imaging detection, probe-treated proteomes are labeled through click chemistry with an azide/alkyne functionalized with a fluorophore or a biotin moiety. Thus, the probe–protein complex can be visualized by in-gel fluorescence analysis (for fluorescent probes) or streptavidin blotting (for biotinylated probes). Fluorescent detection is useful to rapidly verify and visualize the inhibitor-target engagement and to rapid screening proteomes by 1D sodium dodecyl sulfate (SDS)-PAGE in order to visualize if other proteins are labelled by inhibitor. The identification of these off-targets is possible with chemoproteomic approaches using affinity reporters. In this case, probe-treated proteomes are labeled through click chemistry with an azide/alkyne functionalized with a biotin/desthiobiotin moiety. Biotin binds to streptavidin with an extremely high affinity, allowing enrichment and isolation of the probe-protein complexes by pulldown experiments. The probe–protein complex can be then detected with mass spectrometry analysis. The experiment can be conducted in a competitive format to evaluate the potency and selectivity of target inhibitors in native biological samples. Inhibitors compete with probe for enzyme targets and this competition is detected by loss of MS signals. With the help of mass spectrometry analysis identification of

on- and off-targets of covalent probes is obtained in a physiological, cellular context without prior knowledge of protein function, localization, or activation state.

3.4.3 Chemoproteomic applications to target engagement studies

Considerable effort has been placed on developing target engagement assays that are compatible with human clinical studies and applied in a dynamic way to define PK/PD models.
To better define target engagement profile in clinical settings, the assay has to be applied to samples obtained in a minimally invasive way and possibly withdrawn more than once during the clinical trial; restricting usually the field to blood, plasma or blood cells.

A competitive ABPP assay was used to evaluate the off-targets profiles of bortezomib and carfilzomib (two proteasome inhibitors used for the treatment of multiple myeloma) (*64*) directly on human peripheral blood mononuclear cells (PBMCs) taken from human subjects treated with these inhibitors. Different serine proteases were identified as selectively inactivated by bortezomib but not carfilzomib. These chemoproteomic findings suggest that cross-reactivity with serine proteases could contribute to the peripheral neuropathy identified as principal side-effect of bortezomib.

Selinexor, a covalent XPO1 inhibitor, is approved in USA in combination with dexamethasone for the treatment of adult patients with relapsed or refractory multiple myeloma. Selinexor covalently binds to XPO1 active site cysteine 528; a clickable probe based on selinexor (*65*) was developed to label XPO1 in living cells and assess target engagement and selectivity.

Bogyo and collegues (*66*) have introduced quenched near-infrared fluorescent activity-based probes to image cysteine protease activities in tumor xenograft *in vivo* and *ex vivo*. These probes emit a fluorescent signal only after covalently modifying a specific protease target.

Evans and collegues (*67*) developed a sensitive and quantitative assay to measure CC-292-BTK engagement. Starting from the covalent inhibitor, it was developed a probe, chemically linked to a biotin moiety.
The BTK occupancy was detected with an ELISA assay using a monoclonal antibody that selectively recognizes BTK immobilized on the streptavidin substrate by the covalent probe (Figure 14). By

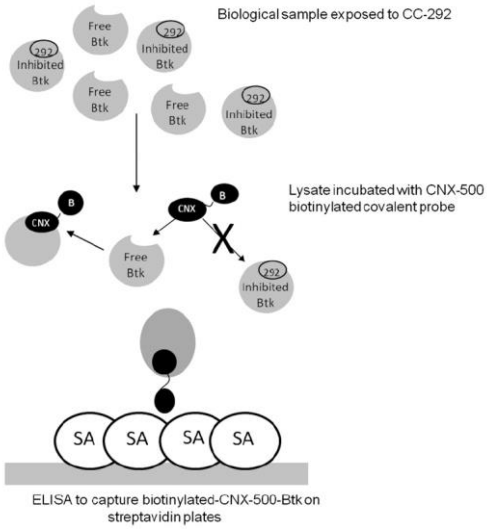


Figure 14 - Workflow to measure CC-292-BTK engagement

building a standard curve with known amounts of recombinant protein bound to the probe, the amount of BTK in any sample has been quantified. Furthermore, used in a competition assay, this probe detected only free BTK that was excluded from interaction with BTK previously labeled by the CC-292: samples, derived from cell culture, animal tissues or clinical studies, were treated with CC-292, lysed and then incubated with biotin probe. Uninhibited BTK was derivatized by the probe and captured by streptavidin-coated Elisa plate. Normalization to untreated control sample allowed the determination of the percentage of BTK occupancy; moreover, the degree of BTK covalently bounded by CC-292, referred to BTK occupancy, was correlated with inhibition of BTK kinase activity and efficacy in B cell functional assays.

Furthermore, chemoproteomic methods can determine the full spectrum of proteins that interact with the inhibitor, helping to elucidate the mechanism of action for inhibitor that can interact with multiple protein targets. Here commonly used chemical proteomics strategies are summarized in order to identify drug off-targets.

Schulke and his collegues (*68*) have developed a clickable photoaffinity probe to measure target engagement of MP-10 (an inhibitor of PDE10A) and characterize the chemoproteomic profile of the clinical candidate. The biotinylated affinity probes allowed the identification of drug-

interaction partners in rodent and human tissue and quantitative mass spectrometry analysis revealed highly specific binding of MP-10 to PDE10A with no off-target binding.

A key limitation of these methods is that sometimes the identification of these off-targets, is performed at the probe-modified proteins level but did not localize sites of probe-modification. Indeed, identification of cysteine modification sites definitively proves that the modification occurs.
Furthermore, efforts to improve quantitative measurement of target engagement and its relation to the pharmacological effects of chemical probes in both preclinical and clinical settings are in place in different research centers and indeed recently quantitative mass spectrometry methods have been incorporated into chemical proteomics by enabling relative comparison between different samples or against negative controls.

3.5 From chemoproteomics to quantitative chemoproteomic approaches

The continuous advances in mass spectrometry-based techniques have contributed to a better interrogation of proteome, enabling to identify low-abundant proteins in complex mixtures, characterize post-translational modifications, identify protein complexes as well as analyze drug-protein interactions in cell extracts and living cells. Moreover, the development of quantitative proteomic mass spectrometry approaches, allows nowadays differential display analyses between tumor and normal tissue samples, among different tumor cell lines or between drug treated and control samples.

3.5.1 Proteomic and chemoproteomic mass spectrometry technologies

Mass spectrometric measurements are carried out in the gas phase on ionized analytes and the mass-to-charge ratio (m/z) of these gas-phase ions is measured. Mass spectrometers consist of an ion source that converts analyte molecules into gas-phase ions, a mass analyzer that separates ions based on m/z ratio, and a detector that records the number of ions at each m/z value. The two used ionization methods for biological samples are nowadays electrospray ionization (ESI) and matrix assisted laser desorption ionization (MALDI) and. These

techniques do not cause protein fragmentation and allow for the determination of intact protein/peptide molecular weight.
Using the ESI ion source, the sample flows into a needle to which a high voltage is applied. In case of positive ion mode, a high positive voltage is applied and by electrostatic repulsion, a fine Taylor cone is formed from which droplets, with an excess of positive charges on the surface, are accelerated towards the ion optics. Solvent evaporation reduces the droplet size until surface tension and electrostatic repulsion of the positive charges on the droplet surface are balanced (Rayleigh limit). After exceeding this limit, droplets fragment into smaller drops. This process continues until generating a current of positively charged ions.
The development of electrospray ionization (ESI) (*69*) and matrix-assisted laser desorption/ionization (MALDI) (*70*), the two soft ionization techniques capable of ionizing peptides or proteins (Figure 15), revolutionized protein analysis using MS. ESI ionizes the analytes out of a solution and is readily coupled to liquid-based (for example, chromatographic and electrophoretic) separation tools.
MALDI source instead sublimates and ionizes the samples out of a dry, crystalline matrix via laser pulses. MALDI-MS is normally used to analyse simple peptide mixtures, whereas integrated liquid-

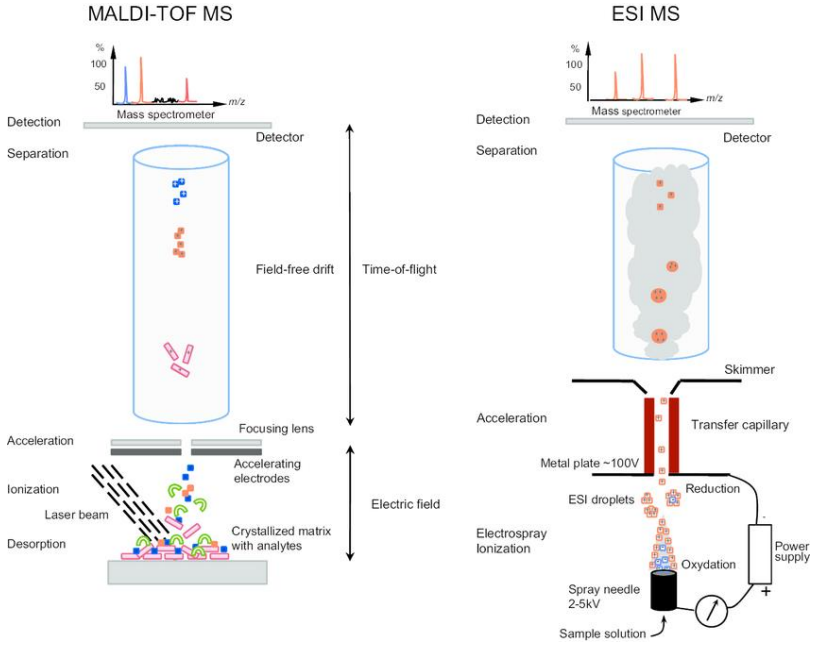


Figure 15 - Ionization soft techniques for proteome analysis

chromatography ESI-MS systems (LC-MS) are preferred for the analysis of complex samples (*13*).
The mass analyzer is central to MS technology. Its key parameters are sensitivity, resolution, mass accuracy and the ability to generate ion mass spectra from peptide fragments (MS/MS spectra). For proteomics research, four types of mass analyzers are commonly used: quadrupole (Q), ion trap, time-of-flight (TOF), and Fourier-transform ion cyclotron resonance (FTICR). They vary in their physical principles and analytical performance; they can be stand alone or put together to take advantage of the strengths of each.
An example of combined mass analyzer is constituted by Thermo Exploris 240.
After electrospray ionization indeed in the Exploris 240 instrument, the ions are accelerated by a radio-frequency quadrupole (RF lens) and transferred through to a mass filtering quadrupole (Figure 16). A quadrupole consists of four cylindrical rods electrodes (hexapole or octapoles) in which voltage is applied in order to stabilize ions within a certain mass-to-charge (*m/z*) range in the quadrupole. Thus, these ions might pass the quadrupole on stable trajectories without hitting the rods.
To this extent, it is possible to guide and select only ions with a defined *m/z* ratio through this ion filter. Then ions are trapped in the C-trap in which the modulation of the voltage allows for the storage and the selective release of ions in the Orbitrap mass analyzer. In the Orbitrap, the ions are stabilized on their trajectories around the central electrode and oscillate in axial direction.

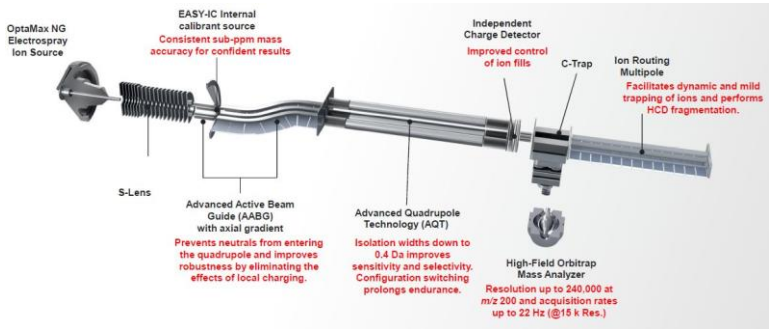


Figure 16 - Orbitrap Exploris 240 MS instrument overview

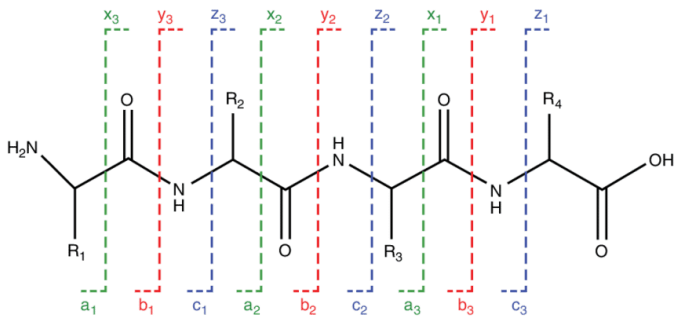


Figure 17 - Peptide fragmentation nomenclature

These oscillations along the z-axis are depending on the *m/z* ratio of the ions and induce a current in the outer electrodes which can be converted into a mass spectrum by Fourier transformation.
Concerning the fragmentation, after precursor ion selection by the quadrupole, the precursor ion is transferred via the C-trap into the HCD (high energy collision dissociation) cell. In the HCD cell, precursor ions are excited to high energy and are collided with an inert gas such as helium or nitrogen. These collisions result in the breakage of the amide bonds between two amino acids into product ions. Depending on the position of the retaining charge, these ions are called b-ions (charge retained on N-terminus) or y-ions (charge retained on C- terminus). Ideally, b- and y-ions allow for the determination of the amino acid sequence of the fragmented peptide.
The fragmented ions, once concentrated in the C-trap, are released to the Orbitrap mass analyzer for the detection.
Peptide ions are measured at high resolution in a data-dependent mode: after each full MS scan, the most intense peptide ions are fragmented to generate MS/MS spectra. After obtaining the MS1 spectrum, single precursor ions are selected, isolated and subjected to a collision cell where peptides are fragmented. Then, fragments are again transferred to the mass analyzer, and their m/z ratio is determined (MS/MS or MS2). Resulting fragments containing the N-terminus of the peptide are called a, b or c, depending on the position of bond breakage. Respective C-terminal fragments are referred to as x, y or z-ions (Figure 17).
The mass of each fragmented peptide together with its fragment ion pattern is searched against databases for peptide identification and bottom-up protein assembly.

3.5.2 Top-down and bottom-up analysis

In principle, two major workflows are used in high-throughput proteomics: top-down and bottom-up proteomic approaches. In top-down proteomics (Figure 18-A), intact proteins are analyzed: intact protein molecular ions, generated by ESI, are introduced into the mass analyzer and subjected to gas-phase fragmentation. The main advantages of these technique are a higher sequence coverage of target proteins; the ability to detect sequence variants, post-translational modifications (PTMs) and a more straightforward sample preparation than bottom-up strategy. However just a limited number of proteins can be identified, and the information obtained with this kind of approach on total proteomic samples remain very limited.

The alternative approach used, is called 'bottom-up' or 'shotgun' proteomics (Figure 18-B). With this technique, protein mixtures, are cleaved into peptides prior to MS analysis. For the generation of peptides, proteins are digested with proteolytic enzymes, being trypsin the most used one, with the aim to generate peptides cleaved in specific sites specifically in the case of trypsin after lysine and arginine residues. The mixture of peptides is then subjected to the LC-MS/MS

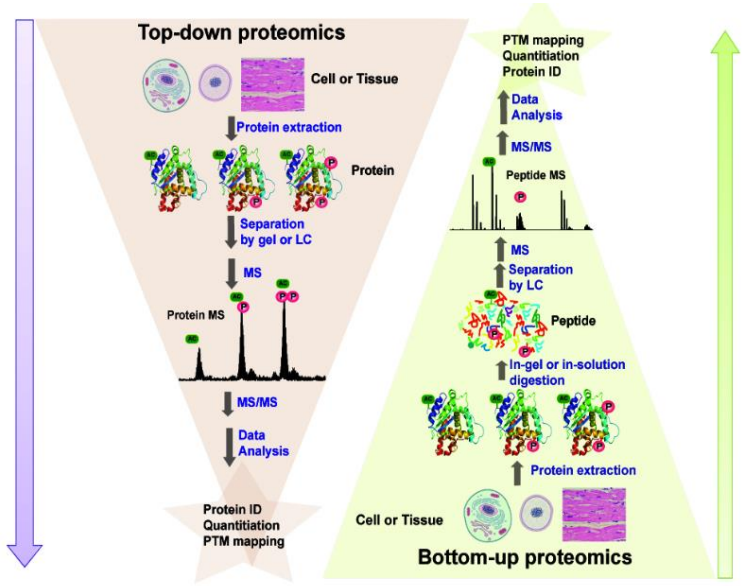


Figure 18 - Top-down and bottom-up proteomic approacches

analysis where peptides are separated, ionized, and analyzed in the mass spectrometer determining not only their mass-to-charge (*m/z*) ratio (precursor ions) but also upon fragmentation obtaining information about the amino acid sequence of the peptide. Then, data achieved during the MS analysis, including the *m/z* value of the precursor and the MS/MS spectra of the fragments associated, are matched against an *in silico* database to identify the sample's proteins.

Bottom-up strategy presents limitations in the analysis of protein modifications, mainly because of the only partial sequence obtained. In addition, a variable fraction of the digested peptides is recovered, and there is a high probability to find redundant peptide sequences.

3.5.3 Bottom-up mass spectrometry applied to chemoproteomic

Beside mass spectrometry analysis, two additional steps need attention in a shotgun mass spectrometry-based chemoproteomic experiment: the preparation of the sample and the separation of peptides (e.g., liquid chromatography).

Sample preparation

Proteins are first extracted from their cellular/tissue context by non-denaturing lysis conditions. The choice of lysis buffer and addition of detergents influence the lysis efficiency as well as the state of proteins. Chemical probe can be added directly in cells in order to evaluate target engagement in physiological environment and on lysates; in this case, subsequent treatment step of lysates with chemical probe, requires functional proteins; therefore, cells should be lysed under physiological conditions conserving the native three-dimensional protein structure, as well as post translational modifications and stable protein complexes. After the addition of the chemical probe, proteins are enriched by Streptavidin-bound resin and digested with trypsin. It is possible to separate the sample in SDS-PAGE and then perform the in-gel digestion or to directly digest the whole sample in solution.

Liquid chromatography-mass spectrometry

The resulting peptide mixture is separated by an on-line reversed-phase liquid chromatography. The separation power of this type of chromatography is generated by interactions of non-polar side chains of peptides with the non-polar stationary phase consisting in octadecyl carbon chain-coated silica (C18) and mixtures of water and organic

solvents (e.g., acetonitrile) as a mobile phase. Acidic supplements like formic or trifluoroacetic acid as proton donors help to generate the net positively charged species.
Upon elution with increasing concentrations of acetonitrile, peptides are separated according to hydrophobicity, influenced by the size and polarity of peptides. Chromatographic peptide separation prior to mass spectrometric analysis reduces sample complexity and, thus, helps to improve proteome coverage. The peptides are then ionized via electrospray ionization and subjected to MS/MS analysis.
The development of 'soft' ionization techniques made the routine mass spectrometric analysis of large polar organic molecules like proteins and peptides possible. For studying very complex proteome, electrospray is the ionization method of choice because is the most suitable if coupled to liquid chromatography.

3.5.4 Quantitative chemoproteomic approaches

Quantitative chemoproteomic experiments aim to identify and determine the amount of target and off-targets proteins sensitive toward covalent modifications by a selected chemical probe. With a competitive experiment, it is possible to evaluate the potency and selectivity of target inhibitors in native biological samples: a loss of probe labeling, that is determined by a loss of probe-protein intensity quantification in MS analysis, is indicative of inhibitor binding.
The most used quantitative chemical proteomics methods, as reported in Figure 19, are label-free quantification, stable isotope tagging with metabolic tag (SILAC) and stable isotope tagging with chemical tag (tandem mass tag).

Label-free approaches

In an article written by Rooden and his group, (*71*) it is reported a label-free quantitative method in order to identify the in vivo selectivity profile of covalent inhibitors. Using label-free quantitative proteomics protocol, they confirmed covalent inhibitor-target binding and they identified several off targets proteins.
Label-free methods aim to compare two or more experiments by comparing the direct mass spectrometric signal intensity for any given peptide, quantifying the integrated peptide peak areas observed in extracted ion chromatograms (XICs), derived from the same protein (*72*); the protein abundance ratio is measured by comparing the XIC

intensities of same peptides of proteins from different experiments (*73*).
Alternatively, using the number of MS/MS sequencing events that lead to peptide identification originating from the same protein during an LC-MS/MS run. This type of quantification is named spectral counting and its assumption is that the abundance of the proteins correlates with MS/MS spectrum matches (*74, 75*).
Label-free methods can be applied to samples that cannot be metabolically labeled and the number of samples that can be compared to each other is unlimited.
However, label-free quantitation leads to an increase in measuring time due to parallelized measurements, the requirement for simple, robust sample handling and is more prone to potential inaccuracy (*72*).
Advances in label-free quantitation were made through improvements in instrumentation and quantitation algorithms. With fast acquisition, high mass resolution and accuracy MS instruments, more data points can be collected.

Stable isotope labeling by amino acid in cell culture (SILAC)

Cravatt and co-workers (*76*) applied an activity-based protein profiling coupled with a quantitative mass spectrometry approach (SILAC) to globally map the targets, both specific and non-specific, of covalent kinase inhibitors in human cells. With this approach many proteins were identified, and it was possible to define windows of selectivity for covalent kinase inhibitors.
This approach is based on stable isotope labelling theory which states that a stable isotope-labelled peptide is chemically identical to its native counterpart and therefore the two peptides also behave identically during chromatographic and mass spectrometric analysis. Given that a mass spectrometer can recognize the mass difference between the isotopically labeled and unlabeled forms of a peptide, quantification is achieved by comparing their respective signal intensities.
SILAC approach was introduced by Mann and co-workers in 2002 and it is a powerful method to identify and quantify relative differential changes in complex protein samples. The SILAC method uses metabolic incorporation of stable isotope into proteins during cell growth and division. It requires growing mammalian cells in specialized media supplemented with light or heavy forms of essential amino acids $^{12}C_6$ and $^{13}C_6$ L-Lysine. The heavy or the light amino acids

are incorporated into proteins through natural cellular protein synthesis. The proteins from both cell populations (incubated with heavy and light amino acids) can be combined and analysed by mass spectrometry. Chemically identical peptides labelled with different stable-isotope composition can be differentiated by mass spectrometry owing to their mass difference. The ratio of peak intensities in the mass spectrum for such peptide pairs reflects the abundance ratio for the two proteins in the two different population. SILAC approach is limited to cell culture experiments and a maximum of three conditions can be mixed and analysed together.

Tandem mass tag approaches (TMT)

Weerapana and his group (*77*) described instead a chemoproteomic method, termed isoTOP-ABPP, to profile quantitatively the reactivity of cysteine residues directly in native biological systems with quantitative mass spectrometry. They developed an azide molecule functionalized with a TEV tag, a biotin group for streptavidin enrichment and an isotopically labelled valine for quantitative mass spectrometry measurements. After cells treatment with a probe characterized by the presence of alkyne group, they applied click chemistry reaction with the azide molecule. After streptavidin enrichment and on-bead proteolytic digestions with trypsin, probe labelled peptides were eluted with TEV protease and analysed by mass spectrometry to identify off-target proteins and quantify their extent of labelling. They identified and quantified hyper-reactive cysteines in several proteins of uncharacterized functions, using isobaric mass tag. Among isobaric mass tags, TMT is the most used one. Schirle and his group (*78*) described a competition-based experimental design, using three broad specific kinase inhibitors, in combination with quantitative mass spectrometry TMT in order to rank-order interactions of inhibitors to kinase by binding affinity.

The tags contain four regions: a mass reporter region, a cleavable linker region, a mass normalizer region and a protein reactive group. TMT labels react with primary amine, groups of tryptic peptides. The same peptides labelled with different isobaric tags have the same mass and co-elute in LC separations. When they are analysed by mass spectrometry, they fragment into reporter ions of different masses in the MS/MS scan. The intensities of the different reporter ions are then

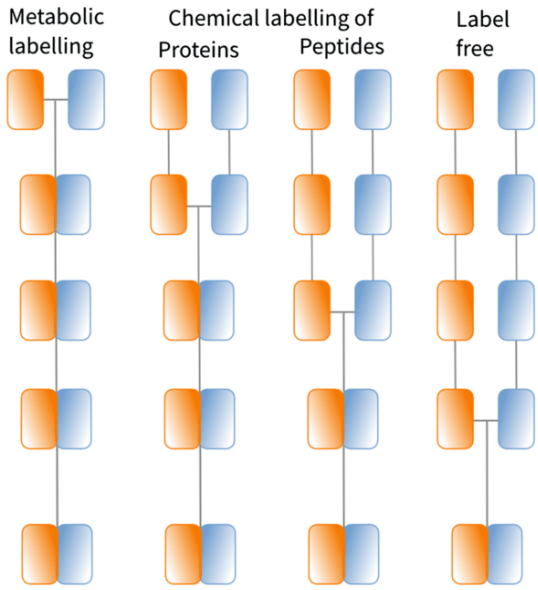


Figure 19 - Overview on different mass-spectrometry based protein quantification methods.

used to determine the relative abundance of the corresponding peptides and proteins in different sample.
Quantitative proteomics via tandem mass tag (TMT) can be more reliable due to its sample multiplexing capability, high precision and throughput: TMT approaches enable parallel quantitation by monitoring the reporter ions generated from the isobaric precursor ions of multiplexed samples. Moreover, multiplexing samples provides increased sensitivity for precursor and product ions and gives the possibility to mix up to 16 samples, reducing the starting material quantity for each condition.

3.5.4.1 Parallel Reaction Monitoring quantification (PRM)

The technique was introduced by Peterson in 2012 (*79*), following the introduction of the Q Exactive bench-top quadrupole-Orbitrap MS (Thermo Fisher). The instrument was characterized by a high resolution and precision; due to a 12 Hz scan rate at a resolution of 17,500 for both MS and MS/MS scanning, 0.2 Th of quadrupole mass filter isolation windows and 1 ppm of mass measurement errors with internal calibration and 5 ppm with external calibration. With these

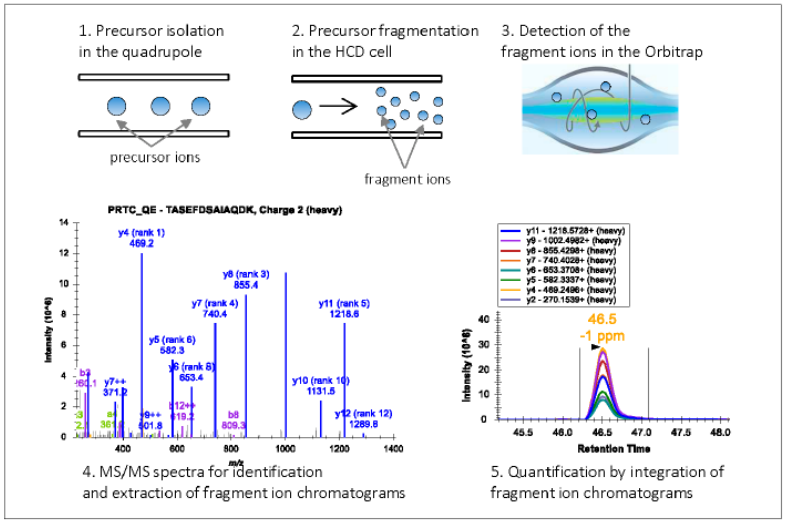


Figure 20 - PRM workflow

performance characteristics, they developed a targeted proteomics method where all products of a target peptide were simultaneously monitored.

As it is shown in Figure 20, in PRM, a target precursor ion, previously defined, is isolated in the quadrupole analyzer and fragmented in the HCD cell; the fragment ions are then detected in the Orbitrap mass analyzer. PRM generates high resolution, full MS/MS data. During data processing, the MS/MS spectrum is used for the confident identification of the peptide; subsets of fragment ions with highest intensities in the MS/MS spectrum are used for quantification.

The approach is identical to the selected reaction monitoring (SRM) scan (*80*), largely used with triple quadrupole MS even for small molecules, except that all transitions are co-detected and distinguished from one another, and from background, by the final mass analysis. The PRM technique has several potential advantages over the SRM approach. First, PRM spectra are highly specific because all potential product ions of a peptide are available to confirm the identity of the peptide; in SRM only 3–5 transitions are followed. Second, PRM could provide a higher tolerance for co-isolated background peptides because of the availability of numerous ions for identification and quantitation and because of the separation of the interfering ions from the product ions thanks to the high resolution of the instrument.

Moreover, monitoring all transitions does not require a prior knowledge of target transitions before analysis.

3.5.5 Software for proteomic data analysis

After data acquisition, peptide m/z-ratios and fragments need to be identified and attributed to a protein. With the advent of high-throughput proteomic experiments, manual spectra interpretation became outdated and unfeasible. Hence, data produced from large-scale proteomic analyses require computational power, sophisticated search engines and highly complex data analysis algorithms to perform peptide and protein identification, covalent inhibitor modification localization and quantitation with high confidence (*81, 82*).

In data dependent acquisition (DDA), spectra originated by peptides fragmentation are used for proteins identification. First, some search parameters need to be specified, and the spectra are searched against protein databases based on the sample organism. Samples usually are digested with protease like trypsin, and expected peptide modifications, coming from post-translational modifications or from covalent inhibitor modification, need to be specified. The modifications can be fixed, which leads to the addition of this mass to each peptide mass, and variable so that each peptide is searched in modified and unmodified version.

A predefined database is digested in silico to generate theoretical tryptic peptides which are expected to be in the sample. The comparison between theoretical spectra and acquired data provides protein identification. For the comparison, search algorithms like Andromeda (*83*), Sequest or Mascot can be used, returning a probability-based ranking of matching peptide sequences. Some peptides can occur in several proteins which makes protein identification more difficult; a peptide occurring in only one protein is called unique; instead, a peptide preferentially occurring in one protein group are called razor.

The identification is based on a score value which reflects the quality of a peptide-spectrum-match (PSM) between a theoretical and an experimental spectrum (*84*). To increase the confidence level, identifications are filtered according to their false discovery rate (FDR). The FDR (*85*) is a statistical approach to estimate and control the number of false positive identified proteins. The FDR is estimated by matching the experimental data against a target-decoy database,

containing the forward and the reversed protein sequences derived from the forward database. If fragmentation spectra match one of these decoy hits, it has to be a false hit by definition. Generally, an FDR of 1% is an accepted value for large datasets.

Some search engines are stand-alone software such as Mascot, whereas the search engine Andromeda is integrated in the proteomic software platform MaxQuant (*83, 86*) and the search engine SEQUEST is integrated in Proteome Discoverer software developed by Thermo Fisher.

MS-based quantifications is usually complicated by the different peptide species because there is no proportionality between peptide amounts and the signal intensities they generate in the mass spectrometer; this is due to the diverse physicochemical properties of peptides such as hydrophobicity or the presence of basic residues, resulting in varying ionization efficiencies. Nevertheless, the behaviour of the same peptide should be identical in different mass-spectrometric runs allowing for relative quantification of peptide abundance.

A simple method available to determine the relative abundance of peptides and proteins is label-free quantification, where all processing and measurement steps are performed separately. A basic version of label-free quantification is called spectral counting (*87, 88*) and compares the numbers of times a peptide has been fragmented. Since there is a tendency to fragment more abundant peptides more often, this approach is useless in quantification of low abundant proteins. Another more accurate and quantitative method is based on precursor ion intensity (*89*). Here, intensities are derived from integration of the peak area of precursor peptide, providing the comparison of intensities or areas under the curve directly between samples. It correlates well with the concentration of a peptide and covers a high dynamic range. The MaxQuant platform employs a sophisticated algorithm for label-free quantification termed MaxLFQ; here, LFQ intensity is represented by a normalized intensity profile that is generated according to the algorithms described in Cox *et al.* article (*90*).

Label-free technique and data-dependent acquisition is susceptible to stochasticity: as the samples are analysed separately, label-free can result in inaccuracy and missing data; to overcome this problem, several chemical and computational methods have been applied (*91-*

95). The most popular method is the match between run algorithm (MBR), included in Maxquant software (*72, 83*).

Briefly, MBR algorithm assesses each identified peak in a MS1 spectrum from one MS run and compares its retention time to unidentified peak in another. The identification is transferred if an unidentified peak, with the same m/z ration and charge state, is found within a specified retention time window, after an alignment of chromatograms. Gygi *et al (96)* evaluated the error associated with these transfers comparing a human containing 10% yeast spike-in sample with a human-only sample. They first analysed human samples and then human + yeast samples; they found that 44% of yeast proteins were incorrectly transferred at least once to the human only sample; however, these transfers were assigned zero or near zero quantification values, preventing incorrect quantification; moreover, they found that in identical samples MBR increased the number of peptides identified by an average of 43%, attenuating the missing-values problems.

3.6 Case study: BTK inhibitors for the treatment of B-Cell Leukemia/lymphoma

3.6.1 BTK and B cell receptor

BTK (Bruton's tyrosine kinase) is a nonreceptor tyrosine kinase member of the TEC kinase family (*97*). It is involved in B cell receptor (BCR) signaling and governs B-lymphocyte development, differentiation, signaling and survival (*98-100*) (Figure 21).

Antigen binding to the BCR, results in B cell receptor oligomerization, SYK and LYN kinase activation, followed by BTK activation (*101*). Once activated, BTK forms a signaling complex with BLNK, LYN and SYK and phosphorylates phospholipase C (PLCγ2). This leads to downstream release of intracellular Ca^{2+} stores and propagation of the BCR signaling pathway through extracellular signal-regulated kinase and nuclear factor-kB signaling, resulting in transcriptional changes to B cell survival, proliferation and differentiation.

Deregulation of BTK is observed in B-cell-derived malignancies including acute lymphoblastic leukemia (ALL), chronic lymphocytic leukemia (CLL), non-Hodgkin's lymphoma (NHL), mantle cell lymphoma (MCL), Waldenstrom's macroglobunemia (WM), and

multiple myeloma (MM) (*102-104*). One therapeutic approach for the

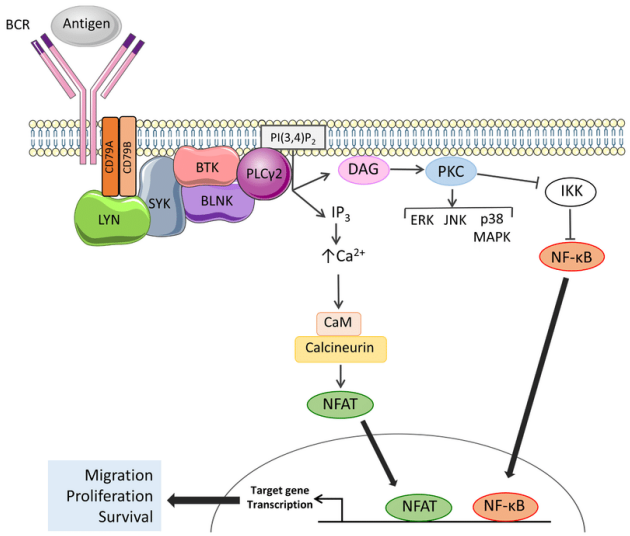


Figure 21 - BTK and B cell receptor signaling

treatment of B-cell derived malignancies is the inhibition of BTK using small molecule inhibitors (*105*).
Covalent inhibition of BTK, targeting Cys 481 in the ATP pocket of the kinase, resulted in a prolonged inhibition of kinase activity (*106*). In contrast to reversible inhibition with the BTK inhibitor dasatinib (*107*), for which kinase activity had almost completely returned 6 hours after drug removal, recovery of BTK activity after 1-hour exposure to CC-292 (a covalent inhibitor of BTK) continued to be suppresses for 8 hours in drug-free media. This prolonged period of BTK inhibition correlated with BTK protein turnover assayed in the presence of the protein synthesis inhibitor cycloheximide (*67*).

3.6.2 Ibrutinib and off-targets effects

The first-in-class BTK inhibitor Ibrutinib (Imbruvica®, PCI-32765) (*108, 109*) has been approved by the US FDA for the treatment of mantle cell lymphoma MCL in 2013 and for the treatment of chronic lymphocytic leukemia CLL in 2014. It irreversibly blocks BTK activity through covalent modification of Cys481 (Figure 22) within the enzyme ATP-binding pocket following Michael addition of the cysteine thiol to the acrylamide moiety. This event prevents the phosphorylation of Tyr223, an essential step for BTK activation. Ibrutinib inhibits BTK

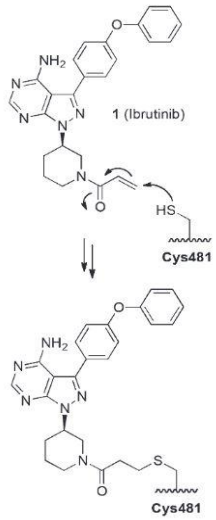
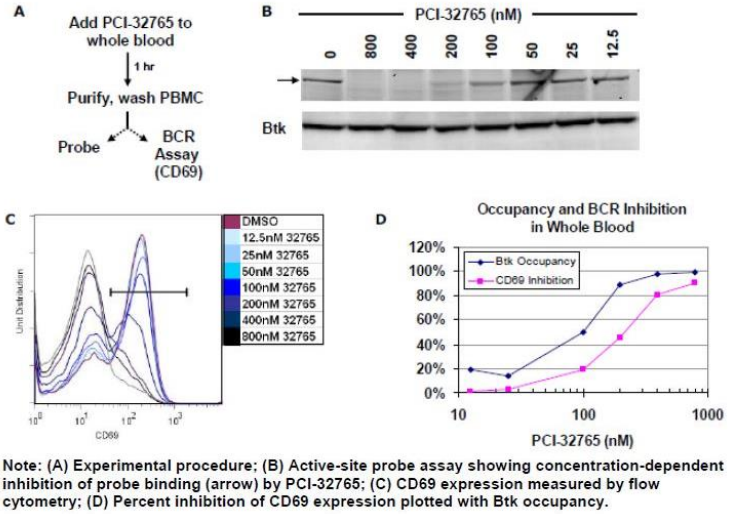


Figure 22 - Mechanism-based inactivation of Bruton's tyrosine kinase (BTK) by Ibrutinib

autophosphorylation on Tyr223 and phosphorylation of the PLCγ2 substrate on Tyr1217 as well as activation of the downstream kinase extracellular signal-regulated kinase.
In cells, Ibrutinib promotes apoptosis, inhibits proliferation, and blocks the response to survival stimuli from their environment.
Ibrutinib is generally well tolerated with rapid and durable response (*110*); however, ibrutinib-associated adverse events (AEs), including atrial fibrillation and hemorrhage (*111, 112*), have led to treatment interruption and long-term discontinuation. Some treatment associated toxicities may be explained by ibrutinib off-target effects due to the inhibition of other kinases: cysteine 481 is poorly conserved among the kinome and can be found in only eleven kinases comprising the TEC-kinase family members BMX, BTK, ITK, TEC, and TXK/RLK as well as the kinases BLK, EGFR, ErbB2, ErbB4, JAK3, and MAP2K7.

Second-generation BTK inhibitors with improved selectivity have developed in order to be more selective than Ibrutinib, resulting in less adverse effects like diarrhea and rash. Acalabrutinib (*113, 114*) was approved for the treatment of mantle cell lymphoma in 2017 and is currently undergoing phase III clinical trials for chronic lymphocytic leukemia and phase II clinical trials for several types of cancer.



Note: (A) Experimental procedure; (B) Active-site probe assay showing concentration-dependent inhibition of probe binding (arrow) by PCI-32765; (C) CD69 expression measured by flow cytometry; (D) Percent inhibition of CD69 expression plotted with Btk occupancy.

Figure 23 - Target occupancy determination of Ibrutinib reported in pharmacological review submitted to FDA

Another TCI of BTK is CC-292. This compound follows the same mechanism as Ibrutinib and is currently undergoing phase I clinical trial for the treatment of several lymphoma types and completed phase II clinical trial for the treatment of rheumatoid arthritis.

Chemical probes, starting from ibrutinib inhibitor, were developed to identify proteome-wide spectrum of interaction of ibrutinib. In the Pharmacology Review presented by Pharmacyclics for the NDA submission of Ibrutinib to FDA in 2013 (*115*), the correlation between Ibrutinib (PCI-32765) binding to BTK active site and B-cell inhibition was confirmed using an active site-directed fluorescent probe: the fluorescent signals were blocked by the presence of covalently bound Ibrutinib in cell-lysates and in isolated human PBMCs. BTK is required for the inhibition of anti-IgM induced B cell activation, as measured by CD69 upregulation.
Covalently labeling of Ibrutinib to BTK active site reduces the level of active BTK. Thus, the fluorescent probe was used to determine the concentration of Ibrutinib in whole blood required to occupy the active site of BTK in PMBCs. The complete occupancy of BTK in PBMCs was indeed achieved at 200nM Ibrutinib as well as approximately 50% inhibition of CD69. (see Figure 23)

In an article reported by Lanning *et al* in 2014 *(116),* a clickable probe, based on ibrutinib molecule and functionalized with an alkyne moiety, was developed in order to map ibrutinib-protein interactions in cells; nevertheless, this technique is based on the identification and quantification of modified proteins but did not identify and localize the sites of covalent inhibitor modification. Therefore, once identified these potential off-targets, biochemical assays are required to determine the exact site of covalent engagement. Indeed, identification of protein modification sites definitively prove that the modification occurs.

4 AIM OF THE WORK

The recent clinical success and increasing number of targeted covalent kinase inhibitors developed, renewed the interest in irreversible binding mechanisms. One of the challenges in the development of these inhibitors is to identify compounds that are significantly more selective for the target in order to minimize unwanted side effects; moreover, many candidate drugs fails during clinical studies because of an incomplete picture of the PK-PD relationships of the drug candidate: in case of lack of efficacy in humans and lack of target engagement assays, it has been difficult to conclude whether poor clinical activity was due to the drug (target is not engaged by the compound) or the target (drug-target interaction did not produce the desired effect). This has been facilitated by technology platforms that allow for the profiling of inhibitors in every stage of the drug discovery in order to reduce attrition rates and development costs; indeed, important efforts have to be done in developing assays that can be transferable to the clinical phase. Rather than lysates and recombinant protein-based assays, cellular chemoproteomics assays can assess engagement of endogenously expressed targets under physiologically relevant conditions.

Thus, the aim of my research is to optimize and apply quantitative chemoproteomic approaches, that take advantages of click chemistry reaction, to study target engagement of covalent molecules. We used, as case study, the BTK inhibitor ibrutinib with the purpose of characterizing its mechanistic profile in cell-systems.

In particular, the workflow of my Ph.D. program will be the visualization of the drug-target engagement *in cell system* by in gel-fluorescence analysis after optimization steps on recombinant and lysate samples. Then, once verified the target engagement, the next step will be the detection of the proteome-wide spectrum of interaction of ibrutinib in order to identify specifically interaction sites of covalent inhibitor, after optimization of the workflow.

During the project, an in-depth analysis of two quantitative mass-spectrometry methods (LFQ and TMT) and related available software for quantitative proteomic analysis will be performed and specific features of each program considered with the aim to improve the whole process.

5 MATERIALS AND METHODS

5.1 Click chemistry reaction

Ibrutinib probe (PF-06658607, Sigma-Aldrich # PZ0242.) (Figure 24) was used as case study to validate approaches suitable to study covalent drug mechanism of action.
This molecule was developed by Pfizer® by functionalizing ibrutinib with an alkyne moiety in a position not perturbing the intrinsic reactivity of the Michael acceptor (α-β-unsaturated amide electrophile) and not interfering with target kinase interactions (BTK).
The probe (5 mg, 464.52 Da) is reconstituted by adding 1.09 ml of DMSO to make a 10 mM stock solution, stored at -20°C.
$CuSO_4$ (Sigma, C1297-100G) and Ascorbate (Sigma, A7631-100G) are freshly prepared at a concentration of 50 mM for CuSO4 and 200 mM for Ascorbate. Cy-5.5-azide (Jena Bioscience, CLK-1059-1) and dethiobiotin-PEG3-azide (Jena Bioscience, CLK-1107-10) are diluted in DMSO and stored at -80°C. THPTA (Jena Bioscience, CLK-1010-100) is diluted in H_2O and stored at -20°C.
The active CU(I) catalyst is generated from $CuSO_4$ using sodium ascorbate as reducing agent in at least five equivalent excess. THPTA serves as dual purpose of protecting proteins from hydrolysis by CU(II) and intercepting the radicals or peroxides that oxidize histidine and other residues; also, it chelates and stabilizes Cu(I) that is instable in solution.
Click chemistry reaction conditions reported below have been optimized during the master-degree thesis project.

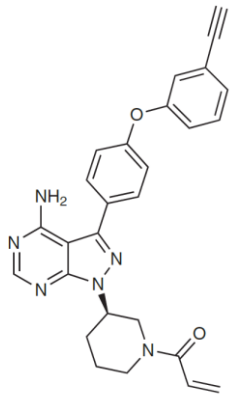


Figure 24 - Structure of ibrutinib probe

5.1.1 Cy-5.5-Azide

The molecule (Figure 25) is constituted by an azide moiety linked to the fluorophore Cy5.5, emitting fluorescence at 694 nm if excited at a wavelength of 678 nm. The detection can be performed with the near infrared imager Odyssey system (LI-COR Biosciences).
The molecule (1 mg) is reconstituted by adding 100 µl of DMSO to make a 10 mM stock solution and stored at -80°C. The solution in sensible to light and has to be stored in the dark.
Click chemistry reaction conditions used are 1 µl of 250 µM Cy-5.5-azide, 1 µl of 100 mM THPTA, 1 µl of 50 mM $CuSO_4$ and 1 µl of 200 mM Ascorbate. The reaction is incubated for 1 hour at room temperature in the dark.

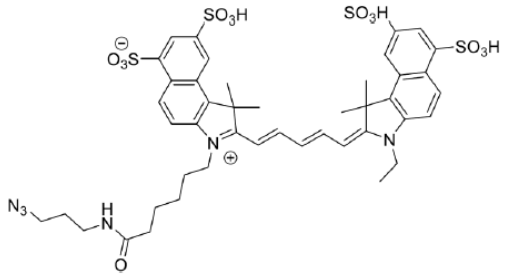


Figure 25 - Cy-5.5-Azide

5.1.2 Biotin-peg3-Azide and desthiobiotin-peg3-Azide

These molecules are composed by the azide moiety linked to biotin and desthiobiotin (Figure 26) for the affinity enrichment on streptavidin resin. Unlike Biotin, desthiobiotin binding to Streptavidin is easily reversible under acidic elution conditions and this allows high recovery of labelled proteins and peptides; due to the above-mentioned properties, desthiobiotin has been chosen to perform all the experiments reported in this dissertation.

Desthiobiotin-PEG3-azide (25 mg, 414.50 Da) was reconstituted in DMSO to make a 100 mM stock solution and stored at -80°C. Biotin-PEG3-azide (Jena Bioscience, CLK-AZ104P4-25) was reconstituted in DMSO to make a 10 mM stock solution and stored at -80°C.

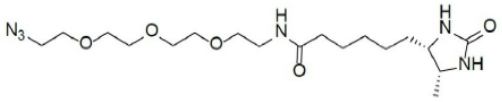


Figure 26 - Desthiobiotin-peg3-azide

Click chemistry reaction conditions used are 300 µM desthiobiotin/biotin-azide (final concentrations), 6 mM THPTA, 2 mM $CuSO_4$ and 14 mM Ascorbate. The reaction is allowed to proceed at room temperature for 2 hours.

5.2 Recombinant protein derivatization and detection

The kinase domain of recombinant BTK produced in insect cells was used to validate ibrutinib probe. Reaction was carried out in the following conditions: 2 µM BTK KD (382 aa – 659 aa) was incubated with 4 µM Ibrutinib or Ibrutinib probe in PBS as reaction buffer for 1 hour at room temperature. The reaction was stopped with 1 µl of 10% TFA and the protein derivatization was confirmed with LC-MS analysis.

The validation of JAK3 as ibrutinib off-target was conducted on recombinant protein produced in insect cells: 2 µM GST-JAK3 was derivatized with 4 µM or 20 µM Ibrutinib probe in PBS as reaction buffer for 2 hours at room temperature. The protein derivatization was confirmed with LC-MS analysis.

Then 2 µM GST-JAK3 was derivatized with 20 µM of ibrutinib for 2 hours at room temperature. The complex was analysed by LC-MS.

5.3 Cell lysis and protein extraction

When performing the CuAAC reaction, the choice of buffer used to generate cell lysates can affect the efficiency of the reaction: in general, buffers containing primary amines, like Tris (tris(hydroxymethyl) aminomethane), should be avoided, as they slow down CuAAC reactions due to the Cu binding to Tris molecule. Amine free buffers such ad Phosphate (PBS), acetate, HEPES are the most widely used buffers for CuAAC reactions. High concentration of strong, ionic detergents like SDS, should be avoided due to a negative impact on reaction efficiency.

Ramos cell line (Burkitt lymphoma) was used due to the high expression levels of BTK in this cell line. Cells were washed in PBS, centrifuged, and stored as pellets at -80°C. Then, they were lysed with 3 cycles of freeze and thawing by adding a lysis buffer compatible with the click chemistry reaction, containing protease inhibitors (Complete™ EDTA free, Sigma-Aldrich, 11836170001), 0.1% v/v NP-40 and 1x PBS. Lysate was then centrifuged for 30 minutes at 4°C at 13000g and supernatant (cytosolic fraction) was separated from the pellets. Bradford assay was used to measure the protein concentration.

5.3.1 Bradford protein assay

Bradford protein assay is used to measure protein concentration in a sample. It is based on the shift in absorbance maximum of Coomassie Brilliant Blue G-250 dye from 465 to 595 nm after binding to denatured proteins in solution. The reagent contains 15% Coomassie Brilliant Blue G-250 in methanol and 55% phosphoric acid. In these acidic conditions, the brown form of the dye is converted into its blue form, when bound to the protein being assayed. This method relies on the presence of the basic amino acid residues, arginine, lysine and histidine, which contributes to the formation of the protein-dye complex. The protein concentration is calculated detecting the absorbance of the solution at 595 nm and comparing the value with a standard curve obtained with known concentration of BSA (Sigma Aldrich, A7030-100G)

Brief, 1 to 10 µl of sample were added to 1 ml of Coomassie brilliant blue solution; samples particularly rich in protein can be diluted before analysis in order to have a concentration within the limits of the standard curve in use. The solution absorbance was read at 595 nm in a Ultraspec 3000 spectrophotometer (Amersham Pharmaceutics). The protein concentration was determined by correlating the obtained absorbance values with a calibration curve calculated using known concentrations of BSA.

5.4 Sample processing for gel-based analysis

5.4.1 *In-vitro* lysates derivatization and competition experiment

The ibrutinib probe was validated on Ramos cell lysates, incubating the lysate with increasing concentration of probe. Briefly, lysate (1 mg/ml protein concentration in PBS) was treated with 1 nM, 10 nM, 100 nM, 1 µM, 10 µM of probe for 2 hours at room temperature. Click chemistry, with fluorescent-azide, was then applied, using reagent concentrations reported above (5.1.1). Samples were separated by SDS-PAGE and analyzed by fluorescent gel scanning using a Licor Odyssey instrument. Gel was then transferred to a Hybond ECL nitrocellulose membrane (GE Healthcare, RPN303D). Membrane was scanned on a Li-Cor Odyssey scanner using fluorescence detection.

For inhibitor competitive experiment, lysates were treated with 1 nM, 10 nM, 50 nM, 0.1 µM, 0.2 µM, 0.5 µM, 1 µM and 5 µM of ibrutinib for 1 hour at room temperature. The pre-incubation step was followed by incubation with 1 µM ibrutinib probe for 2 hours at 25°C. Then click chemistry reaction was applied, proteins were separated by SDS-PAGE and the fluorescence detected.

5.4.2 *In situ* cell treatment

Ramos cells were treated with 5 µM ibrutinib for 24 hours. Following incubation, the medium was removed, and the cells were washed with DPBS. Cells were then harvested and stored as pellets at -80°C. Then, cell pellets were freeze and thawed and after centrifugation, lysates were treated with 1 µM ibrutinib probe for 2 hours at room temperature. Click chemistry was set up with the same reagents and concentration reported above (5.1). 4X sample buffer was added to samples and 40 µg of each sample loaded on SDS-PAGE gel. The gel was imaged as describe below (5.4.3).

5.4.3 SDS-PAGE

SDS-PAGE is an analytical method for the electrophoretic migration of proteins in a polyacrylamide gel matrix that separates proteins on the basis of their molecular weight. Protein samples are treated with SDS which consists in a polar sulphate moiety and twelve carbon chain.

SDS is a detergent with a strong protein-denaturing effect and binds to the protein backbone at a constant molar ratio. In the presence of SDS and a reducing agent that reduces protein disulfide bonds, proteins unfold into linear chains with negative charge proportional to the polypeptide chain length due to SDS binding. The proteins coated with SDS are loaded in wells on the top of acrylamide gel. When electric tension is applied, proteins migrate through the gel towards the positive pole with a speed depending on their mass; the smaller proteins running faster than the bigger ones. The pore size on the gel can vary and depends on the percentage of acrylamide used to cast them. In our case the following commercial pre-cast gels were used where acrylamide formed a 4-12% gradient in buffer bis-Tris: CRITERION™ XT (Biorad®) and SurePage™ BisTris (GenScript). MOPS 1X (NuPAGE®) was used as running buffer.
Samples were prepared by adding Laemmli Sample Buffer 4X (0.25 m TRIS-HCl pH 8, 40% glycerol, 8% SDS, 8% β-mercaptoethanol, 0.005% bromophenol blue), heating at 95°C for 5 minutes. In every gel, one lane is dedicated to the molecular weight marker. The markers used in this dissertation are Chameleon® Duo Pre-stained Protein Ladder (Figure 27) (LI-COR, 928-60000) that provide two-color near-infrared detection (700 nM and 800 nM). The separation was performed by applying constant 60 mA until the end of the run.
Gel was imaged and fluorescence analysed using the Odyssey (LI-COR Biosciences) fluorescence scanner, with detection at 700 nM.

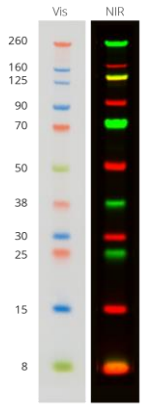


Figure 27 - Chameleon duo pre-stained protein ladder

5.4.4 Transfer of proteins to a nitrocellulose membrane

This technique is based on transferring proteins from gel to a membrane and it is named Western blot. The transfer blotting buffer used is composed by 1x Tris-glycine (Bio-Rad), 20% v/v methanol. After the transfer, the membrane is first colored with Ponceau S staining solution (Sigma, P7170-1L) in order to stain and detect proteins on nitrocellulose and PVDF membranes for evaluating the transfer efficiency of a western blot and verify the total protein normalization between samples. The staining protocol is simple and results in red bands that are easily imaged. The stain is reversible and can be removed with a short incubation in TBST (Tris-buffered saline with 0.1% Tween® 20 detergent).
Next, the membrane is saturated with a solution containing high amount of proteins (e.g. non-fat milk or BSA solution) in order to prevent unwanted membrane-protein interaction. To visualize the protein of interest the membrane is first probed using a primary protein-specific antibody followed by a labeled secondary antibody used for the detection.

In this dissertation, membrane (Hybond ECL nitrocellulose membrane, GE Healthcare, RPN303D) was blocked with 5% non-fat dry milk in TBS-Tween buffer (20 mM Tris-HCl, 500 mM NaCl pH 7.5, 0.1% v/v Tween 20) for 1 hour at room temperature and then incubated with α-BTK primary antibody (EPITOMICS, S2100) diluted 1:1000 in 5% milk overnight at 4°C. Membrane was washed in TBST and incubated with IRDye® 800CW Goat anti-Rabbit Secondary Antibody (LI-COR, 926-32211) diluted 1:5000 in 5% milk. Membrane was scanned on a Li-Cor Odyssey scanner and fluorescence detection performed.
The membrane can be scanned with both λ 700 nm and 800 nm at the same time (Figure 28) in order to overlap the signal corresponding the fluorescent azide (red) and the signal corresponding the primary antibody α-BTK (green); in this case the colour of the corresponding lane will be yellow.

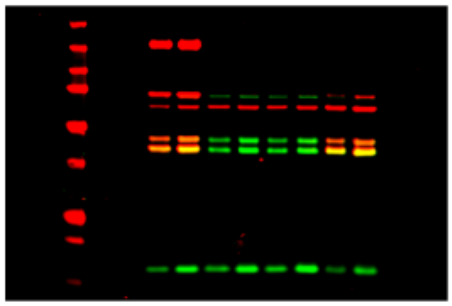


Figure 28 - Example of membrane scanned with Odyssey at both λ 700 nM and 800 nM

5.5 Sample processing for LC-MS/MS analysis

5.5.1 S-Trap™ Proteins digestion

S-Trap™ are spin columns developed by Protifi useful for bottom-up proteomic sample preparation that combine sample concentration, clean up and digestion.
There are different sizes of columns available:

- µS-Trap™ suitable to digest up to 100 µg proteins
- mini S-Trap™ suitable to digest from 100 µg to 300 µg of proteins
- midi S-Trap™ suitable to digest more than 1 mg of proteins

For the digestion of recombinant proteins and lysates for total proteome analysis, µS-Trap™ (Protifi, C02-micro-80) can be used; however, for chemoproteomic approaches that require higher quantity of starting materials, midi S-Trap™ (Protifi, C02-midi-40) were used.
The digestion protocol for micro and midi included the same steps, but different volumes of buffers are used in the workflow. Herein, the protocol for µS-Trap is reported; regarding the application to midi columns, all the volumes have to be multiplied by 20.

First proteins are solubilized in 5% SDS, reduced with 5 mM TCEP for 10 minutes at 95°C and alkylated with 20 mM iodoacetamide for 1 hour in the dark. After these steps (Figure 29), samples are acidified with 10% v/v of 12% phosphoric acid in order to denature proteins. Proteins are then loaded on the S-Trap™ resin after dilution in S-Trap™ binding buffer constituted by 90% methanol and 100 mM Tris-HCl

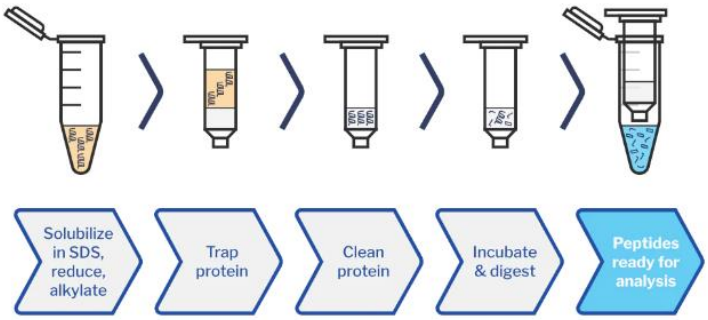


Figure 29 - S-Trap™ workflow

pH 7.1. The column is centrifuged at 4000g until all the buffer passes through; in this way the formed protein suspension is trapped in the column. The matrix is washed for three times with 150 µl S-Trap binding buffer and then digestion is performed adding 25 µl trypsin solution (Sigma-Aldrich, T8658) (1:20 wt/wt ratio in 50 mM Ambic) at 47°C for 2 hours. At the end of digestion step, peptides are eluted first with 10% ACN, 0.2% formic acid and then with 60% ACN 0.2% formic acid. Eluates are pulled and dried in a vacuum centrifuge.

5.5.2 Application of click chemistry on S-Trap

The following protocol was applied to midi S-Trap™ (Figure 30). After SDS solubilization and acidification with 12% phosphoric acid, 2 mg proteins were added to midi S-Trap™ column. After washing step with 3 ml S-Trap™ binding buffer, click chemistry reaction was applied on column by adding 1 µl of 100 mM desthiobiotin-azide, 35 µl of 100 mM THPTA, 18 µl of 50 mM CuSO4 and 37 µl of 200 mM ascorbate in PBS for 2 hours at room temperature. Then, washing step with S-Trap™ binding buffer was performed and 1:20 wt:wt trypsin in 350 µl of 50 mM Ambic was added to the top of the protein trap and incubated for 2 hours at 47°C. Then, peptides were eluted with 800 µl of 10% ACN,

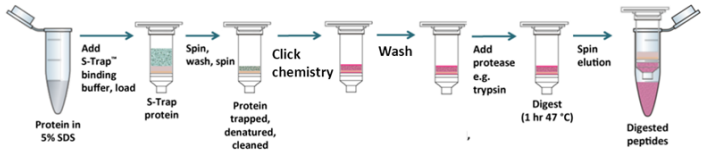


Figure 30 - Application of click chemistry on S-Trap™

0.2% formic acid and with 800 µl of 60% ACN, 0.2% formic acid. The eluates were dried in a speed-vacuum.

5.5.3 Desthiobiotin-modified peptides enrichment

This method allows for the enrichment of probe-labelled peptides. Lysates, treated with 100 µM ibrutinib probe for 2 hours, were normalized to 2 mg/ml. 2 mg of protein extract was solubilized in 5% SDS, reduced by adding TCEP and alkylated with iodoacetamide. Then, samples were acidified with 12% v/v phosphoric acid and loaded on midi S-Trap™ for click chemistry and digestion as reported above (5.5.2). After elution step, peptides were dried with speed-vacuum and resuspended in 400 µl of PBS. Then, enrichment step was performed incubating samples with 80 µl of streptavidin resin slurry (Biozen™ magbeads streptavidin coated, Phenomenex®, KS0-9532) overnight at 4°C on wheel. The resin was pre-washed with 400 µl of 1x PBS for 4 times. After incubation, resin was washed 3 times with 400 µl of 2 M Urea and 10 mM Tris-HCl pH 8, 3 times with 400 µl 1x PBS and 2 times with 400 µl of dH_2O.
Then enriched peptides were eluted with 125 µl of 50% ACN, 0.1% TFA at 65°C for 3 minutes. This elution step was repeated twice. Eluted peptides were filtered on 0.45 µm centrifugal filter Durapore®-PVDF (Millipore, UFC30HV00) to remove possible resin residues.
After filter step, eluates were dried with SpeedVac and analyzed with LC-MS/MS analysis in data-dependent mode and with PRM.

5.5.4 Full proteome label-free sample preparation

Cells were pelleted and lysed. The Bradford assay was used to quantify the proteins for each sample. The digestion of samples was performed with µS-Trap™ (ProtiFi) following the protocol reported in paragraph 5.5.1: briefly 100 µg of lysates were diluted in SDS buffer (5% SDS, 100 mM Tris-HCl) and, after reduction with 5 mM DTT and alkylation with 20 mM iodoacetamide, digested with 1:20 wt/wt trypsin. Peptides were then eluted with 40 µl of 10% ACN, 0.2% FA and 35 µl of 60% ACN, 0.2% FA. Digested samples were desalted and purified with ZipTip® (5.5.7) and resuspended in 2% ACN, 0.1% FA for LC-MS/MS analysis.

5.5.5 Full proteome TMT-based sample preparation

Cells were pelleted and lysed. The Bradford assay was used to quantify the proteins in each sample. The digestion of samples was performed with µS-Trap™ (ProtiFi) following the protocol reported in paragraph 5.5.1: briefly 100 µg of lysates were diluted in SDS buffer (5% SDS, 100 mM Tris-HCl) and, after reduction with 5 mM DTT and alkylation with 20 mM iodoacetamide, digested with 1:20 wt/wt trypsin. Peptides were then eluted with 40 µl of 10% ACN, 0.2% FA and 35 µl of 60% ACN, 0.2% FA. Peptide concentrations were determined by Pierce™ quantitative colorimetric peptide assay kit (Thermo Fisher Scientific, 23275) and concentration between samples were normalized with ZipTip Pipette Tips™ (Millipore, ZTC18S096) before TMT labeling step. 10 µg of tryptic peptides were reconstituted in 100 mM TEAB and labelled with 40 µg of TMT reagent (TMT duplex™, Thermo Fisher Scientific, 90060) (TMT : protein ratio 4:1) at 25°C for 1 hour. 2 µg of each sample was quenched with 0.7% hydroxylamine for 15 minutes, pooled and analysed by LC-MS/MS for labelling efficiency test; the remaining sample was put at -80°C without quenching the reaction. After estimated the labelling efficiency, the reaction of the remaining part was stopped with 0.7% hydroxylamine for 15 minutes. Labelled samples were then combined and fractionated using Thermo Scientific™ Pierce™ High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher Scientific, 84868).

5.5.6 High pH reversed-phase peptide fractionation

Fractionation of peptides was performed with Pierce™ high pH reversed-phase fractionation kit (Thermo fisher scientific, 84868). After reconstituted in 300 µl H_2O, 0.1% TFA solution, peptides are loaded into an equilibrated, high-pH, reversed-phase fractionation spin column (Figure 31). Peptides are bound to the hydrophobic resin under aqueous conditions and desalted by washing the column with water. A step gradient of increasing acetonitrile concentrations in high pH elution solution is then applied to the columns to elute bound peptides intro eight fractions collected by centrifugation:

1. 5% ACN, 0.1% triethylamine (to remove unreacted TMT reagent)
2. 10% ACN, 0.1% triethylamine

3. 12.5% ACN, 0.1% triethylamine
4. 15% ACN, 0.1% triethylamine
5. 17.5% ACN, 0.1% triethylamine
6. 20% ACN, 0.1% triethylamine
7. 22.5% ACN, 0.1% triethylamine
8. 25% ACN, 0.1% triethylamine
9. 50% ACN, 0.1% triethylamine

Elutions were dried with speed-vacuum and re-suspend in 0.1% FA for LC-MS/MS analysis.

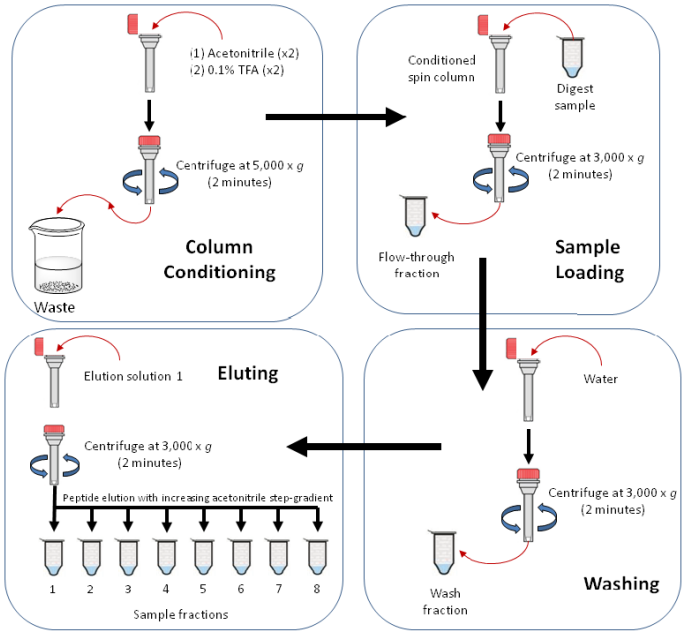


Figure 31 - Pierce™ high pH reversed-phase sample fractionation workflow

5.5.7 ZipTip® for sample purification

The ZipTip®(Millipore, ZTC18S096) are pipette tips used for concentrating and purifying peptide or protein samples. They are 10 µl pipette tips containing C-18 reversed-phase resin and the binding maximum capacity is typically 5 µg when used with saturating amounts of analyte.
The solutions used are:

- Acetonitrile
- Water, 0.1% TFA (loading and washing solution)
- 60% ACN, 0.1% TFA (elution buffer)

The workflow was automated using DigestPro™ (Figure 32) (CEM), a protein digestion and sample preparation system for MS based proteomics. It completely automates all of protein digestion protocols and works with both in-solution and in-gel based methods.

Figure 32 - DigestPro™ system

5.6 MS-based analysis

5.6.1 LC-MS analysis for recombinant protein

Recombinant proteins characterization was performed by liquid chromatography and mass spectrometry (LC-MS) analysis. The instrument employed was an Agilent 6230 ESI-TOF mass spectrometer coupled with an Agilent 1260 Infinity II HPLC.

In the reverse phase chromatography, the components of the mixture are separated using the different interactions with the hydrophobic stationary phase. Generally, recombinant proteins are partially retained by the hydrophobic resin, whereas salts and other buffer components that would interfere with the MS analysis are washed away by the hydrophilic mobile phase. Proteins are then eluted from the column by the increasing concentration of organic solvent (ACN) in the mobile phase, which competes for the hydrophobic binding. The eluate from the chromatographic column is sent to the ion source, that in the ESI-TOF instrument is an electrospray.

ESI ionization produces soft ionization and involves the formation of multicharged ions (z >1). The spectra obtained consist, therefore, of multi-charged peaks.

The mass spectra are than analysed by Agilent Mass Hunter Qualitative Analysis Software where raw MS spectra are deconvoluted with different algorithms in order to obtain the protein molecular weight. The mass values obtained are then compared to the theoretical ones in order to confirm the correct sequence and to identify specific modifications such as phosphorylation or covalent adduct obtained after the reaction with the probe.

The injected protein amount ranges from 2 to 10 µg. The liquid chromatography was performed using a FortisBio C4 column 100x2.1 mm – 1.7 µm (CPS analitica, BIO304-020501). The quantity of recombinant protein injected was 2 µg.
Mobile phase used:

- Mobile phase A: H_20, 0.05% TFA
- Mobile phase B: Acetonitrile, 0.05% TFA

Proteins were separated using a linear gradient from 5% to 95% of phase B in 24 minutes at 0.2 mL/min constant flow rate.
The molecular weight was obtained through deconvolution of mass spectra using Mass Hunter (Agilent) software and pMod algorithm. The

mass value was compared with the theoretical one based on the amino acidic sequence acquired from Paws® software and reported below.

- BTK 382 – 659 aa mass: 32384.2 Da
- GST-JAK3 781-end mass: 66791.6 Da
- Ibrutinib probe mass: 464.52 Da
- Ibrutinib mass: 440.5 Da

The settings for the mass detector were the following:

ESI SOURCE		**MS-TOF**	
Gas temperature	300°C	Fragmentor	175 V
Drying gas	10 l/min	Skimmer	65 V
Nebulizer	45 psi	Octupole radio frequency voltage	750 V
Capillary voltage	3500 V		

5.6.2 MALDI-TOF mass spectrometry

The MALDI-TOF/TOF mass spectrometry (Matrix assisted Laser Desorption/ Ionization – Time of Flight) allows to separate a mixture of ions according to their mass/charge ratio (m/z). The instrument consists of a MALDI source and a TOF analyzer. The samples are mixed with an organic compound that acts as a matrix. The matrix is able to absorb the light emitted by the laser source. The laser beam, which hits the sample, allows desorption and ionization of the sample. The created ions undergo a strong acceleration thanks to the electric field in the ionization chamber and then enter the flight tube where they are analyzed depending on the time necessary to cover the analyzer-detector distance. All ions enter the analyzer with the same kinetic energy (KE). Since the speed at which the ion moves and therefore the time taken to cover the source-analyzer distance, depends on its mass-to-charge ratio (m/z), the ions that possess lower m/z will reach the detector before those having a greater m/z.

MALDI ionization produces soft ionization and involves the formation of monocharged ions (z = 1). The spectra obtained consist, therefore, of single-charged peaks. The matrix used is α-cyano- 4 hydroxycinnamic in a concentration of 10 mg/ml; it is suitable to identify proteins and peptides below 10 kDa.

The experimental procedure is:

1. 0.4 µl sample are loaded into a well of the plate
2. 0.4 µl of the previously prepared matrix solution is added to the sample

3. The plate is dried so that the matrix crystallizes together with the sample
4. The plate is loaded into the instrument
5. When the instrument is ready and in high vacuum, the plate is displayed, and the laser can be directed to the well where the crystallized sample is located.
6. The laser is activated, and data are collected by the instrument. The software elaborates them and provide a spectrum reporting the m/z ratios of the ionized peptides in X-axis and the relative intensity in Y-axis.

5.6.3 LC-MS/MS proteome analysis

DDA and PRM analysis were carried out on an Orbitrap Exploris 240 mass spectrometer (Thermo Fisher Scientific) coupled to an Ultimate 3000 RLSCnano system (Thermo Fisher Scientific). Peptides were trapped on a PepMap™ 0.3 x 5 mm – 5 µM precolumn cartridge packed with 100 Å C18 particles (Thermo Fisher Scientific, 160454). Subsequent peptide separation was on a 75 µm x 150 mm long EASY-spray™ column packed with 100 Å C18 stationary phase (Thermo Fisher Scientific, ES800A) for chemoproteomic approaches; however, a 75 µm x 500 mm long EASY-spray™ column packed with 100 Å C18 stationary phase (Thermo Fisher Scientific, ES903) was used for total proteome analysis.

Mobile phase used are:

- Nano pump phase A: H_2O, 0.1% FA
- Nano pump phase B: 80% ACN, 0.1% FA
- Loading pump solution: H_2O, 0.1% FA

Regarding ion source parameters, a positive spray voltage of 1500 V and temperature of ion transfer tube of 275°C were used.

Chemoproteomic experiment

Dried eluates (5.5.3) were redissolved in 8 µl of 2% ACN, 0.1% FA and 1 µl injected in the instrument. Peptides were loaded on the precolumn at 10 µL/min in water with 0.1% (v/v) FA. Then, they were separated on a 75 µm x 150 mm long EASY-spray™ column packed with 100 Å C18 stationary phase. Linear gradient elution, flowing at 300 nl/min, was performed from 4% to 28% phase B over 108 min, from 28 to 40% in 10 minutes and from 40 to 95% in 1 minute. At the beginning of the gradient, 20 minutes of desalting step on cartridge

were introduced in order to better purify sample before separation on the column. The total gradient time is 164 min (Figure 33-A).

Data were collected in data-dependent acquisition method with 3 seconds as cycle time. MS1 spectra were collected using the following parameters: scan range 350 – 1200 m/z, resolving power 120 K, normalized AGC target 300% and maximum injection time of 20ms. Intensity threshold was set to $1x10^5$ and 60 sec of dynamic exclusion. MS2 spectra were collected using the following parameters: HCD normalized collision energy 30%, isolation window 1.4 m/z, normalized AGC target $1x10^4$ and maximum injection time of 22 ms. Precursors with an assigned monoisotopic m/z and a charge state of 2–5 were interrogated. Samples were analyzed in three technical replicates.

Total proteome experiment with TMT

Dried fractionated eluates (5.5.6) were redissolved in 12 µl of 2% ACN, 0.1% FA and 5 µl injected in the instrument. Peptides were loaded on the precolumn at 10 µL/min in water with 0.1% (v/v) FA. Then, they were separated on a 75 µm x 500 mm long EASY-spray™ column packed with 100 Å C18 stationary phase (Thermo Fisher Scientific). Linear gradient elution, flowing at 300 nl/min, was performed from 5% to 28% phase B over 143 min, from 28 to 45% in 18 minutes and from 45 to 95% in 1 minute. The total gradient time is 200 min (Figure 33-B).

Data were collected with a data-dependent acquisition method with 3 seconds as cycle time. MS1 spectra were collected using the following parameters: scan range 350 – 1200 m/z, resolving power 120 K, normalized AGC target 300% and maximum injection time of 20ms. Intensity threshold was set to $1x10^5$ and 45 sec of dynamic exclusion. MS2 spectra were collected using the following parameters: HCD normalized collision energy 38%, isolation window 0.7 m/z, normalized AGC target $1x10^4$ and maximum injection time of 22 ms. Precursors with an assigned monoisotopic m/z and a charge state of 2–5 were interrogated. Fractions were analyzed in technical duplicate.

Total proteome experiment with label free quantitation

Dried eluates (5.5.4) were redissolved in 25 µl of 2% ACN, 0.1% FA and 2 µl injected in the instrument. Peptides were loaded on the precolumn at 10 µL/min in water with 0.1% (v/v) FA. Then, they were separated on a 75 µm x 500 mm long EASY-spray™ column packed with 100 Å C18 stationary phase (Thermo Fisher Scientific). Linear gradient elution, flowing at 300 nl/min, was performed from 4% to 28% phase B over 100 min, from 28 to 40% in 10 minutes and from 40 to 95% in 1 minute. The total gradient time is 144 min (Figure 33-C).

Data were collected in data-dependent acquisition method with 3 seconds as cycle time. MS1 spectra were collected using the following parameters: scan range 350 – 1200 m/z, resolving power 120 K, normalized AGC target 300% and maximum injection time of 20ms. Intensity threshold was set to $1x10^5$ and 60 sec of dynamic exclusion. MS2 spectra were collected using the following parameters: HCD normalized collision energy 30%, isolation window 1.4 m/z, normalized AGC target 75% and maximum injection time of 22 ms. Precursors with an assigned monoisotopic m/z and a charge state of 2–5 were interrogated.

Samples were analyzed in five replicates.

PRM analysis of chemoproteomic samples

Regarding the PRM analysis, chemoproteomic samples were run using the same gradient used for data-dependent acquisition. PRM data are acquired at 30000 resolution at m/z 200 with an AGC target of 100% and 100 ms of ion injection time. Isolation list (Table 2), containing 4 m/z targets from BTK and BLK, was imported. The isolation window was set to 1.6 m/z. NCE was set to 30% and the fixed first mass was set to 110 m/z. Maximum number of multiplexed ions was set to 3.

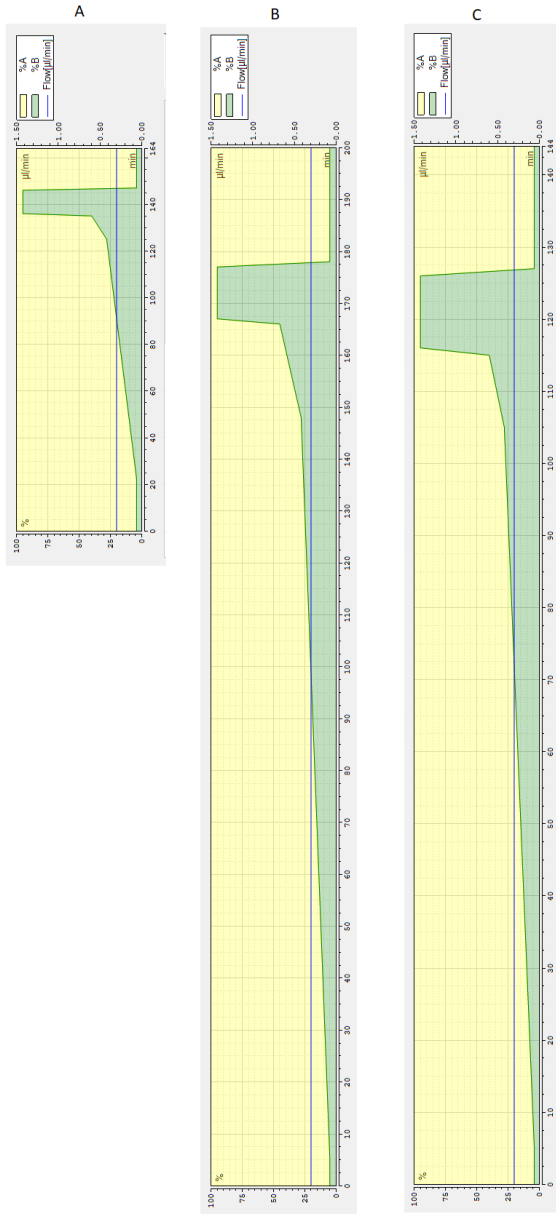


Figure 33 – A. Linear gradient elution used for chemoproteomic samples; B. Linear gradient elution used for total proteome analysis with TMT quantification method; C. Linear gradient elution used for total proteome analysis with LFQ quantification method

Compound	Gene symbol	m/z	z	t start (min)	t stop (min)
QRPIFITEYMANG**C[878]**LLNYLR	BTK-Cys 481	852.4408	4	135	145
G**C[319]**LLDFLK	BLK-Cys 319	596.3154	3	120	150
G**C[319]**LLDFLKTDEGSR	BLK-Cys 319	811.4041	3	130	140
DLLGLC[878]EQKR	XPO1-Cys 528	685.0315	3	125	145

Table 2 - Inclusion list of target m/z selected for PRM analysis

5.7 Data analysis

5.7.1 Chemoproteomic

Raw MS files were analyzed with MaxQuant (version 1.6.10.43) using default settings. The data were searched with Andromeda search engine against Uniprot human proteome database (UP000005640_9606). Search criteria included carbamidomethylation of cysteine, chemical probe (Figure 34), N-term acetylation and methionine oxidation as variable modifications.
Diagnostic peaks were added in chemical probe modification (Table 3) and Trypsin/P with two missed cleavages was selected as enzyme. Mass tolerance was set at 4.5 ppm for precursor ions and 20 ppm for fragment ions. Peptide spectrum match and protein FDR were set at 0.01. Minimum score for modified peptides was decreased to 15.
The label-free quantification of proteins was obtained by MaxLFQ algorithm in MaxQuant. Match between the runs option was enabled and remaining default parameters were permitted.

Name	Composition	$M+H^+$
m/z 240	H(21) C(12) N(3) O(2)	240.170653
m/z 913	H(60) O(7) C(45) N(12) S	913.450139

Table 3 - Table of diagnostic peaks added in Maxquant modification

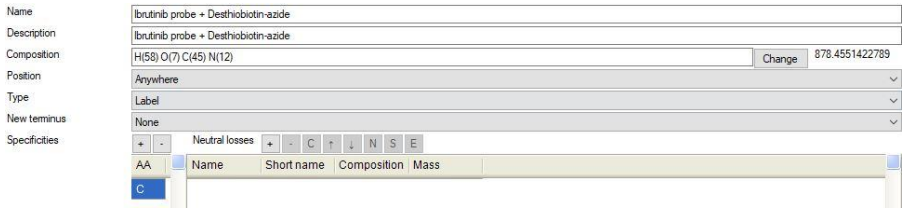


Figure 34 - Ibrutinib probe + desthiobiotin-azide complex added as chemical modification in Maxquant

5.7.2 PRM

Data-dependent acquisition was used to construct spectral library. DDA analysis was conducted with Proteome Discoverer, as described in chapter 0 adding chemical probe modification on cysteine as variable modification. Skyline software (version 21.1.0.278) was used for PRM assay development of trypsin digested samples. Skyline peptide settings included 7-40 amino acids. Four transitions for peptide were selected based on the spectral library rank and dotp value greater than 0.95.

5.7.3 Total proteome analysis

LFQ analysis

Raw MS files for total protein data sets were uploaded in Proteome Discoverer (version 2.5) (Thermo Fisher Scientific) using SequestHT and Mascot search engine (Matrix science) and in Maxquant software using Andromeda search engine. All analyses were performed searching against Uniprot human proteome database (UP000005640_9606). Carbamidomethylation of cysteine was set as fixed modification and N-term acetylation, methionine oxidation and phosphorylation (STY) as variable modifications. The minimum peptide length was 7 amino acids and digestion mode was set to specific with trypsin as digestion enzyme and two missed cleavages were allowed. Mass tolerances were set to 10 ppm for precursor ions and 0.02 Da as fragment mass. MSPepSearch node was used to increase the number of identifications and the spectral library added is NIST Human Orbitrap HCD. Precursor detector was used in order to identify others additional precursors co-eluted and co-isolated. The consensus workflow further included false discovery rates of 0.05 for both protein and peptide identifications.

The quantification was performed with Minora feature detector node for the processing part and feature mapper and precursor ion identifier for the consensus part. Unique and razor peptides were used for the quantification and minimum % of replicate features were set to 30; normalization was performed on total peptide amount. Pairwise ratio calculation approach was used to calculate protein ratio and p-value was calculated with t-test background based; imputation mode was set to none.

Regarding Maxquant analysis, protein identification and quantification was performed using MaxQuant (1.6.10.43) software.

Extracted peak lists were searched against a human UniProt database. Carbamidomethyl (57.02 Da) was specified as fixed modification of cysteine; N-term acetylation, oxidation of methionine (15.99 Da) and phosphorylation (STY) were set as variable modifications. Trypsin/P was specified as the proteolytic enzyme and up to two missed cleavage sites were allowed as the labelling would prevent cleavage at the labelled lysine. The mass tolerance of the precursor ion was set to 20 ppm, second search peptide tolerance 4.5 ppm. The false discovery rate (FDR) was set at 1% for both peptide and proteins. For peptide quantification, match between runs was performed with a matching time window of 0.7 min and an alignment window of 20 min; for protein quantification intensity-based label free quantification was performed using MaxQuant LFQ algorithm: LFQ intensities were calculated from razor and unique peptides with a minimum ratio count of two peptides across samples. A minimum length of seven amino acids was needed for the identification. Normalization was performed via the summation of the total intensity of all the identified human peptides.
LFQ intensities were used for the subsequent quantitative analysis in Perseus. Protein were filtered for contaminants and the LFQ intensities were log2 transformed. Proteins found in at least 30% of all replicates, for one group, were retained for analysis and missing values were replaced by simulating values of low abundant proteins. Using this data, volcano plot was generated for comparison between samples using standard setting of FDR cut-off at 0.05. Statistically significant differences in protein abundance between control and treated samples were used to generate volcano plot.

TMT analysis

Raw MS files for total protein data sets were uploaded in Proteome Discoverer (version 2.5) using SequestHT and Mascot search engine. All analyses were performed searching against Uniprot human proteome database (UP000005640_9606). For the calculation of labeling efficiency, TMT modification for K and peptide N-terminus was set as variable. When confirmed that labelling efficiency is greater than 98% for all channels, TMT modification was set as fixed.
TMT new quantification method was created editing the table with TMT corrections factors reported in the certificate of analysis; as reported in Figure 35.

For processing workflow, carbamidomethylation of cysteine and TMT modification were set as fixed modification and methionine oxidation and phosphorylation (STY) as variable modifications. The digestion mode was set to specific with trypsin as the digestion enzyme and two missed cleavages were allowed. Mass tolerances were set to 10 ppm for precursor ions and 0.02 Da for-fragment ions.
Spectra were selected with a maximum precursor mass of 5000 Da. For PSM validation, Percolator, a target-decoy search strategy, was selected and for TMT quantification reporter ions quantifier node was used. Regarding consensus workflow part, in reporter ion quantifier node, the search criteria further included co-isolation threshold of 50% and 10 for average reporter S/N threshold. Normalization was performed on total peptide amount; pairwise ratio based was used for protein ratio calculation and p-value was calculated with t-test background based; imputation mode was set to none. False discovery rates of 0.05 for both protein and peptide identifications.

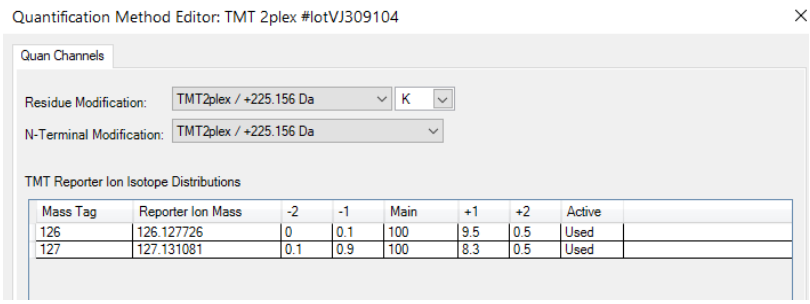


Figure 35 - Study-specific TMT2plex (#lotVJ309104) quantification method used in this dissertation

6 RESULTS AND DISCUSSION

6.1 Recombinant protein characterization

6.1.1 Intact MS protein analysis

Covalent interaction between BTK kinase domain (382-659aa) and the alkyne probe of the ibrutinib inhibitor or ibrutinib inhibitor per se, has been demonstrated with the help of LC/MS analysis. The complete derivatization of recombinant BTK (32384.2 Da) with the inhibitor (440.5 Da) was indeed confirmed (Figure 36), incubating recombinant BTK protein with a 2-fold excess of compound or in a solution containing the same percentage of organic solvent (DMSO) included in the reaction with the compound, for 1 hour at room temperature and analyzing the mixtures by LC/ESI-MS. The peak shown in Figure 36-A corresponds to the deconvolute mass spectrum of the control protein incubated in DMSO, while the one reported in Figure 36-B corresponds to the protein incubated with the compound and presents a mass shift of 440.5 Da indicative of a complete covalent addition of one single molecule of ibrutinib to BTK.

The validation of the probe (Figure 37) was performed in the same way, by treating recombinant BTK with ibrutinib probe at a 2-fold excess for 1 hour at room temperature. The peak in Figure 37-B corresponds to the protein with a mass shift of 464.52 Da; confirming again that the protein was completely modified by a single molecule of probe.

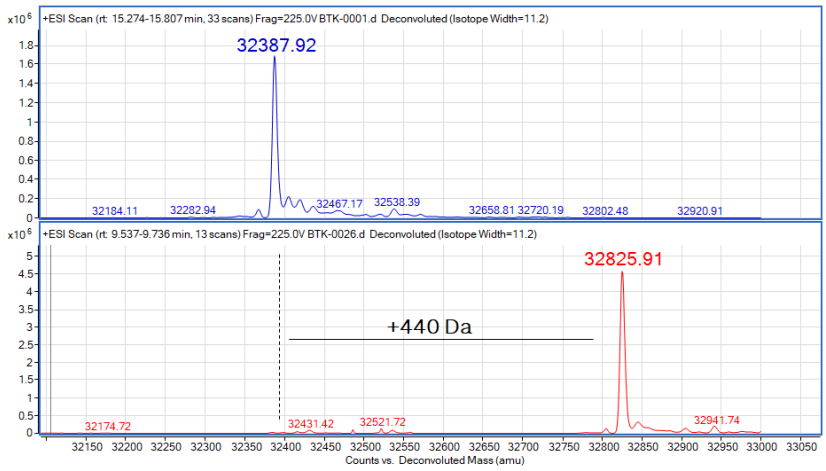


Figure 36 - LC-MS analysis of recombinant BTK kd derivatized with ibrutinib; A. recombinant BTK kd control; B. recombinant BTK kd labelled with ibrutinib

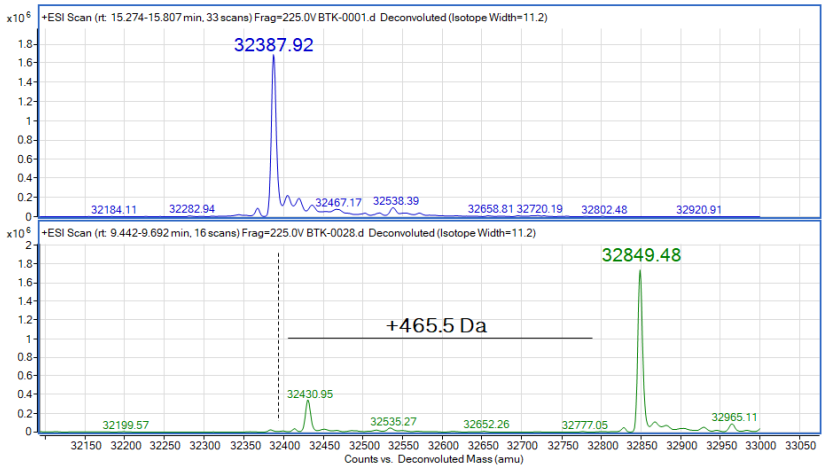


Figure 37 - LC-MS analysis of recombinant BTK kd derivatized with ibrutinib probe; A. rec BTK kd control; B. rec BTK kd labelled with ibrutinib probe

6.1.2 Confirmation of covalent mechanism of action by mass spectrometry

Then, the protein complex with ibrutinib and ibrutinib probe was subjected to trypsin digestion with µSTrap followed by LC-MS/MS analysis on an Orbitrap Exploris 240 mass spectrometer (Thermo Fisher Scientific) coupled to an Ultimate 3000 RLSCnano system (Thermo Fisher Scientific) analysis. Elaboration of obtained data using Proteome Discoverer 2.5 version software (Thermo Fisher Scientific), allowed to detect the peptide containing Cys 481 (QRPIFIITEYMANG**C**LLNYLR) containing a mass addition of 464.52 Da (consistent with the mass of Ibrutinib Probe) and 440.5 Da (corresponding to the mass of Ibrutinib). Strong confidence in the results is given by the attribution of both b and y ions in the spectrum (Figure 38 and Figure 39) confirming the identity of the peptide and probe targeted amino acid.

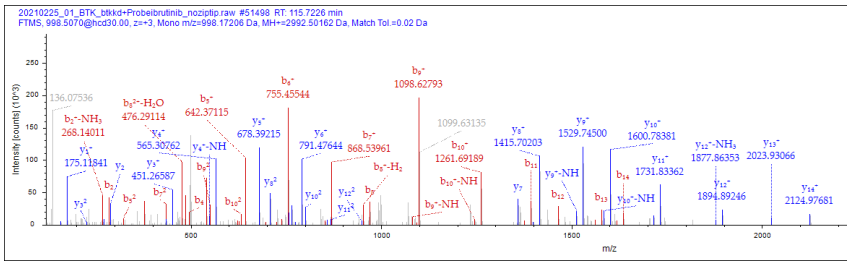


Figure 38 - ms/ms spectrum of BTK peptide containing cys481 covalently labelled by probe ibrutinib

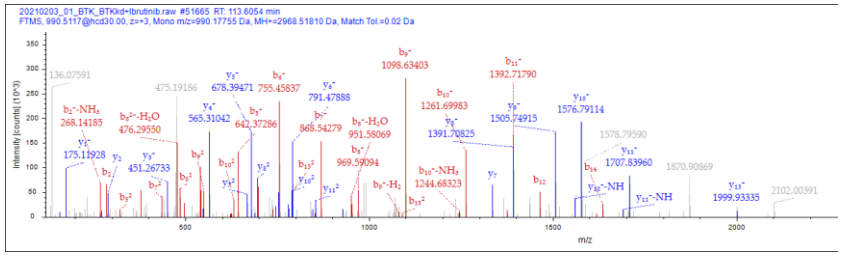


Figure 39 - ms/ms spectrum of BTK peptide containing cys481 covalently labelled by Ibrutinib

6.2 Target engagement studies

To perform target engagement studies, an azide functionalized with a fluorescent moiety (5.1.1) and SDS-PAGE coupled to gel fluorescence detection were used.
Regarding the strategy (Figure 40), target engagement experiment has to be conducted in a competitive format: in cells treated only with ibrutinib probe, probe will label BTK and other possible proteins; however, in cells pre-treated with ibrutinib, the inhibitor competes with probe for BTK labeling, and this competition corresponds to a lower labeling between BTK, which is already labeled by ibrutinib, and probe. This is detected by a loss of in gel probe-labeled fluorescent signal.

Cell lysis and fluorescent click chemistry reaction conditions previously identified have been applied. Briefly, in order to obtain

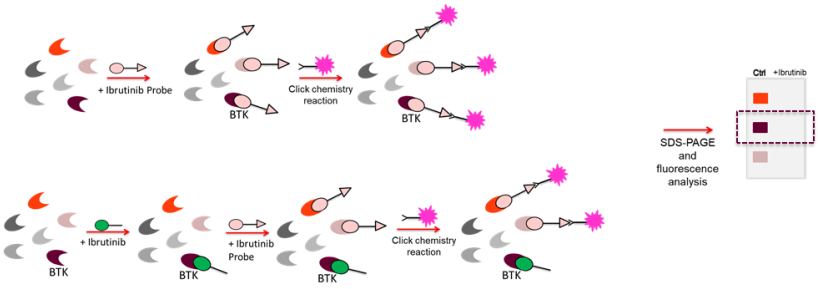


Figure 40 – Scheme of fluorescent click chemistry experiment for the visualization of ibrutinib-BTK target engagement in cells

optimal results, buffer containing primary amines should be avoided, as they slow CuAAC reaction; moreover, high concentration of strong, ionic detergents like SDS, reduces reaction efficiency, however, if needed, 0.1% NP40 can be added without negative impact.

6.2.1 Click chemistry validation on lysates

First, we evaluated probe ibrutinib ability to label BTK in cell lysates and determine the optimal concentration for probe labeling. We used Ramos cells, a B lymphocyte lineage tumor cell line with high expression level of BTK.

Ramos lysates were treated with increasing concentrations of ibrutinib probe (1 nM, 10 nM, 100 nM, 1 µM and 10 µM) for 1 hour at room temperature. In parallel, as a negative control of the reaction, an untreated sample was prepared, in which lysates were treated only with click chemistry reagents (fluorescent azide, THPTA, $CuSO_4$, Ascorbate); in this way it was possible to verify possible azide background effects.

Click chemistry was then applied, samples separated by SDS-PAGE analysis and probe addition detected by fluorescent gel scanning. The presence and the amount of BTK protein in the samples was then detected by Western blot, transferring the gel to a nitrocellulose membrane and incubating it with the antibodies of interest (α-BTK and secondary antibody for the detection). The membrane was then analyzed by fluorescence scanning on selected wavelengths for the protein and the probe.

As reported in Figure 41-A, higher probe incubation time corresponds to an increasing of the fluorescent intensity of the bands,

corresponding to an increasing of proteins-probe labeling; it also increases the background signal that can be attributed to a non-specific binding of the probe to off-target proteins and not to click chemistry reaction thanks to the presence of negative control.
In Figure 41-C, obtained by analyzing the membrane at the λ of α-BTK primary antibody (nm excitation/emission, green) and λ of fluorescent azide (nm excitation/emission, red) simultaneously, it is possible to confirm that BTK, recognized by the antibody, is also derivatized by the probe. In fact, the green color corresponds to the signal of BTK, the red color to the fluorescent azide; when these two signals overlap, a yellow signal is obtained indicating that the target protein is derivatized by the probe. The signal intensity of fluorescent BTK significantly increased after incubation of lysates with increasing concentration of ibrutinib probe and a maximum signal, corresponding of a full target occupancy of BTK, was obtained after the incubation of lysates with 1 μM of probe.

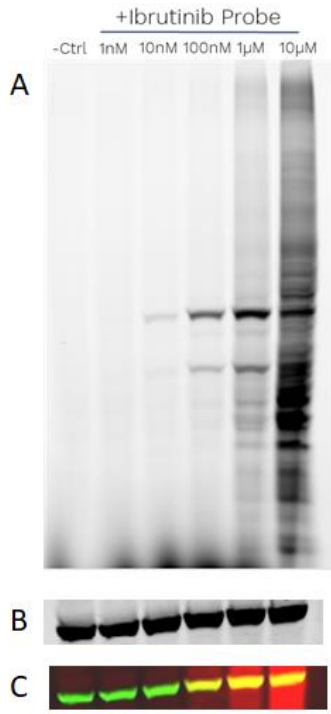


Figure 41 - Probe labeling of Ramos lysates followed by Cy-5.5-azide mediated click chemistry; A. In gel fluorescence detection; B. Western blotting with α-BTK; C. Dual analysis of membrane scanned at both λ 700 nM (detection of fluorescence azide) and 800 nM (detection of α-BTK)

This concentration was selected as the optimal labeling concentration for competition experiment, resulting in sufficient labeling of BTK, without saturating the signal; treating lysates with 10 μM of ibrutinib probe, strongly increased the background without further increase in the band corresponding to BTK signal.

Results obtained indicated that structural modification of ibrutinib with the alkyne handle did not substantially alter inhibitor potency of ibrutinib for BTK.

6.2.2 *In vitro* target engagement

Next, a competition experiment was performed in order to determine at which concentration, a complete target engagement of Ibrutinib is obtained: lysates were incubated with increasing concentrations of ibrutinib and then treated with a fixed concentration of ibrutinib probe. Labeled lysates were then "clicked" with the

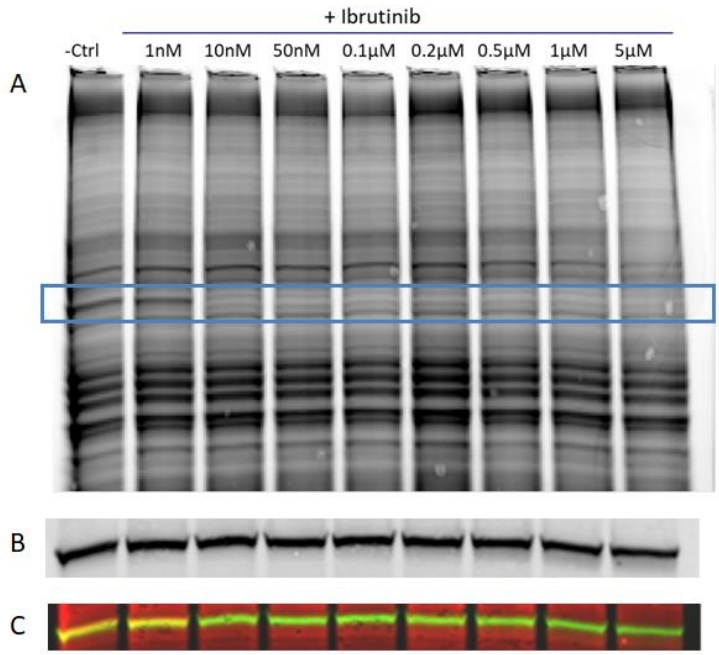


Figure 42 - Ibrutinib target engagement in Ramos lysates; A. Fluorescence detection of BTK labeled to Cy-5.5-azide; B. Detection of α-BTK; C. Fluorescence detection of BTK labeled to fluorescent-azide (red) and α-BTK primary antibody (green)

fluorescent probe and probe modified proteins visualized by in gel fluorescence.

As reported in Figure 42-A, the prominent fluorescent band at 70 kDa was competed by ibrutinib in a concentration-dependent manner: in Figure 42-C, obtained by analyzing the membrane at the λ of α-BTK primary antibody (green) and λ of fluorescent azide (red) simultaneously, the signal intensity of fluorescent BTK significantly decreased after pre-incubation of lysates with ibrutinib, suggesting a full target occupancy of BTK with an incubation between 1 and 10 nM of ibrutinib. This is in agreement with the apparent IC50 reported for the same compound in literature using substrate assay (*108*).

Also, the concentration-dependent labeling of several additional proteins, that was also detected in Ramos lysates incubated with Ibrutinib Probe (Figure 41), were not competed by ibrutinib pre-treatment, confirming the selectivity of ibrutinib for its protein target.

6.2.3 *In cell* target engagement

Results obtained with lysates do not reflect the real drug-protein interaction that occurs under physiological conditions: proteins in cells are included in cellular sub-compartments, form complexes. Furthermore, the affinity of a ligand for the drug target can be modulated by changing the activation state of the target protein, for example, by phosphorylation or binding of regulatory proteins. Consequently, it is important to directly monitor target engagement inside cells (Figure 40). Therefore, to verify ibrutinib interaction and derivatization of BTK directly *in situ*, Ramos cells were treated with 5 μM of ibrutinib for 24 hours and in parallel, as control, with 2% DMSO, the solvent in which the inhibitor is dissolved. Samples were then processed as reported above and the visualization of the probe-bound protein *in cells* was fluorescently detected. Obtained results (Figure 43) showed a prominent fluorescent band at 70 kDa that was competed by *in cells* ibrutinib pre-treatment; in Figure 43-C, obtained by analyzing the membrane at the λ of α-BTK primary antibody (green) and λ of fluorescent azide (red) simultaneously, this prominent band correspond to BTK and it significantly decreased after pre-incubation of cells with ibrutinib, suggesting a target occupancy of BTK by ibrutinib. The others fluorescent bands present in the control sample, are not competed by ibrutinib pre-treatment; thus, corresponding to ibrutinib probe off-targets, whose identification is

possible using higher resolution techniques such as mass spectrometry, as reported in paragraph below.

As a result of click chemistry reaction, the functionalization of the probe with a small group, such as the alkyne group, rather than bulky groups such as fluorescent groups or biotin moiety, maintains ibrutinib cell permeability and thus makes it possible to use it in cell cultures. Furthermore, this approach demonstrates the utility of these clickable probe to profile the cellular activity of BTK inhibitor and the ability to apply this technique for the verification of ibrutinib-BTK target engagement in physiological conditions as cells. The approach can also be extended to animal tissue or even to patient specimens to evaluate the effect of a covalent compound during clinical phases. However, to establish a relationship between drug-target occupancy and its functional effects, especially in human clinical studies, a more quantitative approach has to be considered.

To achieve this purpose, a quantitative chemoproteomic approach has been set and optimized and reported in the next chapters.

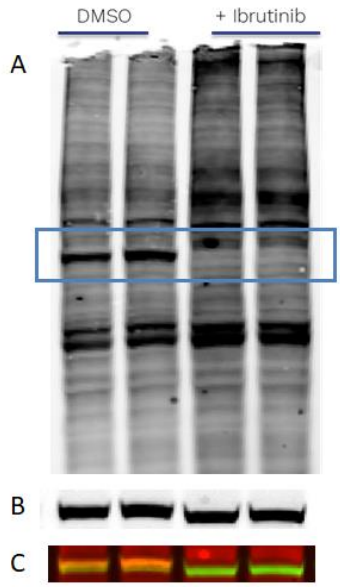


Figure 43 - Ibrutinib in cell target engagement; A. Fluorescence detection of BTK labeled to Cy-5.5-azide; B. Detection of α-BTK; C. Fluorescence detection of BTK labeled to fluorescent-azide (red) and α-BTK primary antibody (green)

Furthermore, it was possible to characterize the ibrutinib profile of interaction with others possible off-targets in order to define a suitable concentration window, across which ibrutinib could achieve complete inhibition of BTK without proteome wide-reactivity.

6.3 Ibrutinib proteome-wide spectrum of interaction determination with quantitative chemoproteomics

Versatility of click chemistry reaction, allowing the binding of different tags to the same compound of interest, fits the different purposes. Addition of an azide group, functionalized with a desthiobiotin moiety, to ibrutinib probe allowed to capture and enrich probe labelled targets with a streptavidin resin, taking advantage of high affinity between streptavidin and desthiobiotin (5.1.2).
Quantitative chemoprotemic methods reported in literature are based on identification at probe-modified proteins level but did not localize sites of probe-modification. Indeed, identification of cysteine modification sites definitively proves that the modification occurs and can inform on stoichiometry and specificity of the potential covalent interaction. Therefore, a new protocol was set, requiring probe-modified peptides enrichment on streptavidin resin: the experiments have to be conducted in competitive format to evaluate the potency and selectivity of ibrutinib in native biological samples (Figure 44). Ibrutinib competes with probe for enzyme targets and this competition is read out by loss of MS signals: cells are treated with ibrutinib probe,

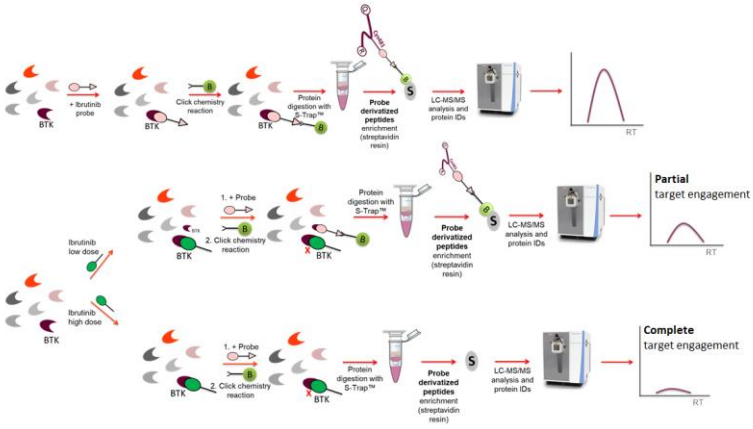


Figure 44 - Strategy for chemoproteomic sample preparation

probe will label BTK and, after digestion step, probe-modified peptides will be captured on streptavidin resin and detected with MS analysis; however, pre-treating cells with an high or low dose of ibrutinib, correspond to a fewer engagement of probe-BTK, which is already labeled by ibrutinib, that correspond to a decrease of probe-modified peptide capture on streptavidin resin and consequently, a decrease of its MS signal. Treating cells with different doses of ibrutinib can help in identifying only specific proteins that are covalently labeled by ibrutinib in a dose-dependent manner and in identifying only the one labeled by ibrutinib at an effective therapeutic dose.

The initial steps were dedicated to sample preparation optimization and mass spectrometry analysis in order to get as much information as possible about target engagement between probe-on target and probe-off-target hits.

6.3.1 Probe ibrutinib and desthiobiotin-azide characterization

We started with the validation and characterization of ibrutinib probe and desthiobiotin-azide interaction by LC-ESI-MS analysis. In Figure 45, chromatograms, reporting the total ion counts (TIC) for desthiobiotin-azide, probe-ibrutinib and complex after click chemistry reaction, are reported. As we can see in figure, LC-ESI MS can clearly discriminate among the probe (Figure 45-C), desthiobiotin (Figure 45-A) and the product of reaction (Figure 45-B); thus, confirming that this method can be applied for the determination of the reaction complex. The hydrophilic nature of desthiobiotin-azide, make the complex

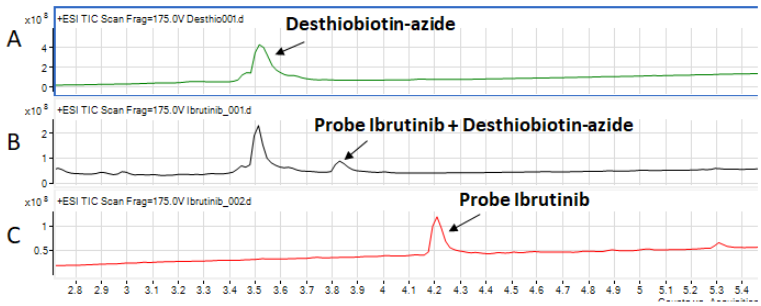


Figure 45 - LC-MS characterization of probe ibrutinib, desthiobiotin-azide and the complex after click chemistry reaction; A. Desthiobiotin-azide; B. probe ibrutinib – desthiobiotin-azide complex; C. probe ibrutinib

probe ibrutinib – desthiobiotin-azide more hydrophilic (the RT of the peak is 3.8 min) then the probe ibrutinib only (RT: 4.2 min).

6.3.2 Desthiobiotin-azide click chemistry optimization

The optimization of the pulldown experiment is performed on the recombinant protein target of the inhibitor of interest. The standard protocol consists on protein derivatization, click chemistry reaction for dethiobiotin-azide binding to alkyne group, protein digestion, elimination of desthiobiotin-azide excess through dialysis and the enrichment of modified peptides on streptavidin resin. After elution step, peptides are purified and concentrated by ZipTip™ Pipette Tips and then analyzed by nLC-MS/MS (Orbitrap Exploris 240, Thermo Fisher Scientific).

In order to increase the efficiency of the reaction, desthiobiotin-azide must be used in high excess and then removed before streptavidin enrichment in order to not saturate the resin. Dialysis or other purification methods, such as methanol/chloroform protein precipitation, are reported in literature, however, with low recovery yield (<50%). With the aim to improve recovery, we optimized click chemistry protocol applying the reaction on S-Trap™. S-Trap™ are spin columns useful for bottom-up proteomic sample preparation that combine sample concentration, clean up and digestion: the method is based on SDS-mediated protein solubilization (SDS ≤ 10%) and protein capture in the submicron pores of the S-Trap™. In this way, proteins present an extremely high surface area to volume, and can be easily cleaned of detergent (e.g. SDS) and contaminants and digested with trypsin, giving the possibility to use lysis buffer that could not be compatible with MS techniques.

S-Trap™ sample processing column has been adapted to chemoproteomic sample preparation as reported in 5.5.2. In our previous experience, we compared click chemistry on S-Trap with click chemistry in solution; we observed a peak corresponding to the peptide derivatized only by the probe (50% of the protein not derivatized) still present after application of in solution click chemistry; however, where click chemistry was conducted on S-Trap™, the peptide derivatized by the desthiobiotin-azide was present, while the corresponding one, derivatized only by the probe, was not observed; indicating a complete click chemistry reaction on S-Trap™.

These results suggest that proteins on this column present a high surface area and this increases the probability of the desthiobiotin-azide to bind efficiently the proteins target; furthermore, with S-Trap™, the elimination of the excess of dethiobiotin-azide is simpler and more reproducible than other classic methods (chloroform-methanol precipitation, dialysis).

Thus, we focused on the analysis methods in order to increase and maximize the level of identification of peptides modified by the probe: as with any chemical modification of peptides, understanding the fragmentation pattern that results from probe modification is essential to improve its detection by LC-MS/MS and increase the confidence of the identification.
First, recombinant BTK kinase domain (382-659aa) was derivatized with Ibrutinib Probe and the complete derivatization of the protein was confirmed. Then click chemistry was applied on S-Trap™ (as described above), coupling desthiobiotin-azide to the probe-modified peptides. Click chemistry reaction was conducted using desthiobiotin-azide because it binds to streptavidin with equal specificity but less affinity than biotin (Kd = 10^{11} vs. 10^{15} M, respectively). Consequently, desthiobiotinylated peptides can be eluted specifically from streptavidin resin using milder conditions (e.g. acidic conditions) than the ones required for elution of biotin molecules.
Digested peptides were analyzed by LC-MS/MS analysis on an Orbitrap Exploris 240 mass spectrometer (Thermo Fisher Scientific) coupled to an Ultimate 3000 RLSCnano system (Thermo Fisher Scientific) analysis. Using Proteome Discoverer and Maxquant software, precursor ion 852.4408 m/z, corresponding to the molecular weight of

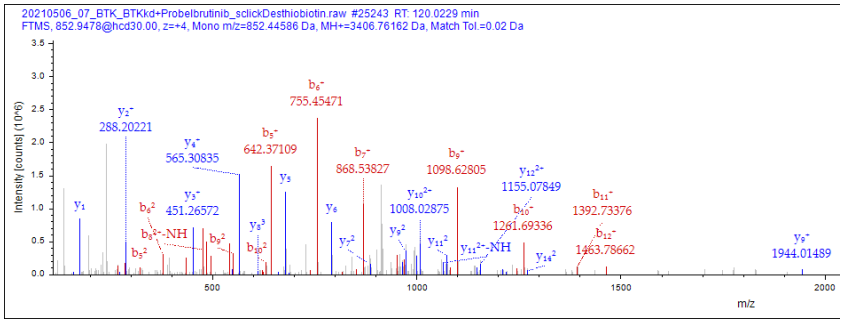


Figure 46 - ms/ms spectrum of BTK peptide containing cysteine 481 covalently labelled by probe ibrutinib and desthiobiotin-azide

quadruple charged peptide (QRPIFIITEYMANG**C**LLNYLR) containing cystein 481, was identified as modified by the probe and desthiobiotin-azide. Both b and y ions are labeled in the above spectrum (Figure 46) confirming the identity of the peptide and probe targeted amino acid. As it is shown in the ms/ms spectrum in Figure 47, two intense peaks were not assigned neither by Maxquant software nor by Proteome Discoverer. These peaks correspond to inhibitor fragments generated during the Higher-energy Collisional Dissociation fragmentation of the compound bound to the peptide in the mass spectrometer. The exact chemical formula of these peaks, called diagnostic, is reported in Figure 48, shown below, and can be included in the analysis during the identification of desthiobiotin-probe-modified peptides. We compared the Maxquant analysis performed using diagnostic peaks in ibrutinib probe chemical modification and the analysis performed without these peaks. The Andromeda score, indicative of the statistical significance of the analysis, increased from 191.07 not considering these peaks (Figure 49) to 215.75 considering diagnostic peaks for the modified peptide containing cys 481. Accordingly, understanding the exact chemical formula of these peaks could be helpful in the identification of peptides modified by the probe during chemoproteomic experiments, increasing the level of confidence in the identification of endogenously modified peptides. Proteome Discoverer software (version 2.5) doesn't consider these peaks during data analysis; thus, we have used only Maxquant software for the analysis of chemoproteomic samples.

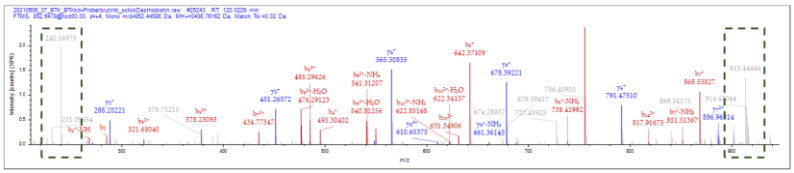
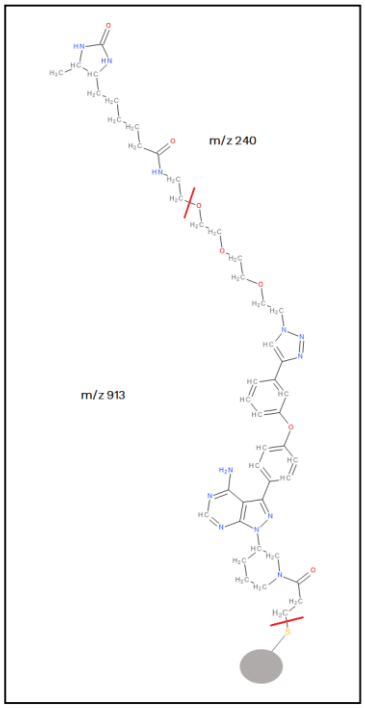


Figure 47 - Focus on diagnostic peaks in ms/ms spectrum of BTK peptide containing cysteine 481 covalently labelled by probe ibrutinib and desthiobiotin-azide



m/z	Chemical Formula	Neutral mass
240	H21 C12 N3 O2	239.16338
913	H60 O7 C45 N12 S	912.44286

Figure 48 - Chemical formula andstructure of diagnostic peaks of probe ibrutinib –desthiobiotinAzide complex addedin Maxquant software

The same workflow was applied using biotin-peg3-azide.
As reported above, desthiobiotin has been chosen to perform all the experiments reported in this dissertation because, unlike biotin, its binding to streptavidin is easily reversible under acidic elution conditions and this allows high recovery of labelled proteins and peptides; in any case, we apply click chemistry reaction on probe-derivatized recombinant BTK, using biotin-azide in order to consolidate the results. Digested peptides were analyzed with nano LC-MS/MS and data analysis performed with Proteome Discoverer. Precursor 688.1453 m/z, corresponding to the peptide containing cysteine 481, fifth charged, was identified as modified by the probe and biotin-azide. Both b and y ions resulted labeled in the spectrum (Figure 50-A), confirming the identity of the peptide and Probe targeted amino acid.

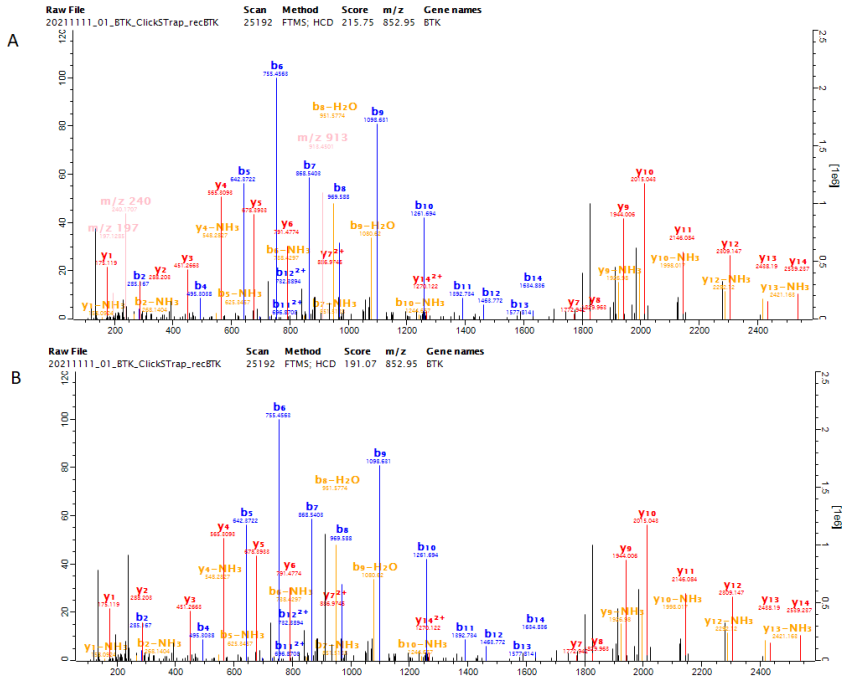


Figure 49 - ms/ms spectrum of BTK peptide containing cys481 labelled with probe ibrutinib and desthiobiotin-Azide analyzed with Maxquant software; A. analysis performed including probe ibrutinib diagnostic peaks; B. analysis performed not including probe ibrutinib diagnostic peaks

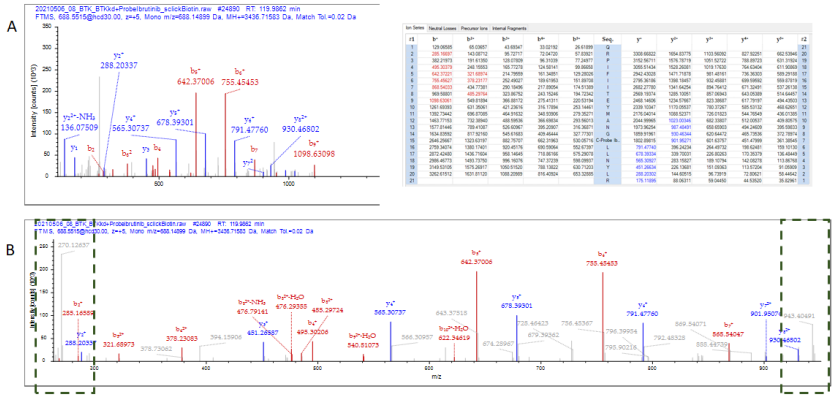
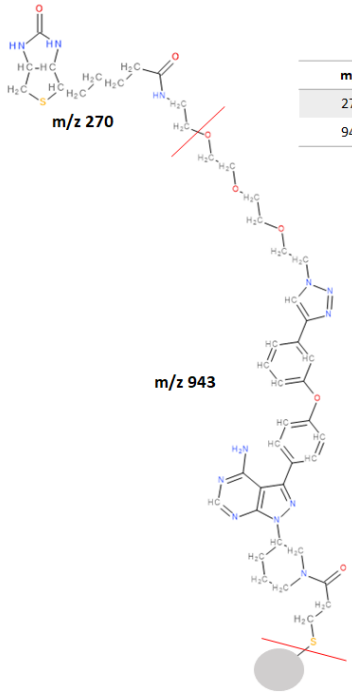


Figure 50 - ms/ms spectrum of peptide containing cys 481 labelled with ibrutinib probe and biotin-azide; A. total ms/ms spectrum; B. focus on inhibitor diagnostic peaks



m/z	Chemical Formula	Neutral mass
270	H19 C12 N3 O2 S	269.11979
943	H58 O7 C45 N12 S2	942.39928

Figure 51 - Chemical formula and structure of diagnostic peaks of probe ibrutinib – biotinAzide complex added in Maxquant software

However, also in this case and as it is shown in the ms/ms spectrum in Figure 50-B, two intense peaks, m/z 270.1273 and m/z 943.4049 were not assigned by Proteome Discoverer and Maxquant software used for the data analysis. These peaks correspond to inhibitor fragments generated during the Higher-energy Collisional Dissociation fragmentation of the compound bound to the peptide in the mass spectrometer. The chemical formula and structure of these fragments is reported in Figure 51 shown above.

The detection and more important the quantification of endogenously probe-modified peptides, more specifically considering BTK peptide containing cysteine 481, can be difficult due to the high dynamic range of proteins, even considering enriched chemoproteomic samples. Therefore, we developed a method for the detection and quantitation of selected probe-modified peptides, which employs PRM analysis using a high resolution orbitrap instrument.

First, a spectral library was constructed on Skyline software using the DDA file obtained by the analysis with Proteome Discoverer (Thermo Fisher Scientific) of digested recombinant BTK, previously derivatized with ibrutinib probe and desthiobiotin-azide. The spectral library is essential for PRM analysis in order to match MS/MS spectrum of the peptide acquired by PRM method.

Then, a PRM test was performed on recombinant protein, in order to target identify peptide containing cys 481 labelled with ibrutinib probe and desthiobiotin-azide. As shown in Figure 52, the peak at 125.26 min corresponds to co-eluted fragments ions of BTK precursor 852.4408 m/z, the quadruple charged probe ibrutinib-modified peptide on cysteine 481 (QRPIFIITEYMANG**C**LLNYLR).

Fragments peak area intensities are compared with the ones presented in the spectral library, confirming the assignment of the peptide. Dot product (dotp) value indicates the degree of the match between spectral library MS/MS and the extracted ion chromatograms of the corresponding transitions; high dotp indicates the absence of interfering signals. The total integrated fragment ion signal for the peptide is plotted as a bar graph; contribution from each individual fragment ion is displayed as a different color in the bars.

In order to better purify peptides before LC column separation, a 20 min desalting step on cartridge pre-column was introduced. The new gradient was tested on recombinant BTK (Figure 53) in order to determine the new retention time (i.e. 140 min) of BTK peptide containing cysteine 481 used for the next PRM analysis.

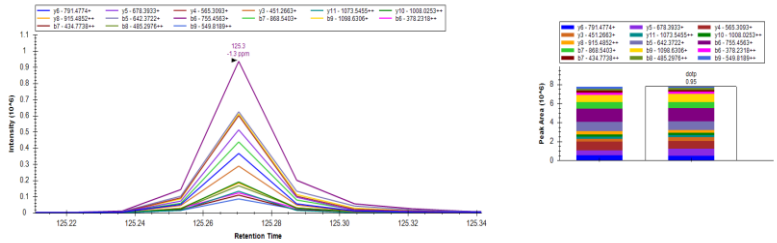


Figure 52 - Chromatograms of fragment ions extracted from the peptide of recombinant BTK containing cys 481. Mass measurement error and retention time of the most intense transition are annotated above the peak.

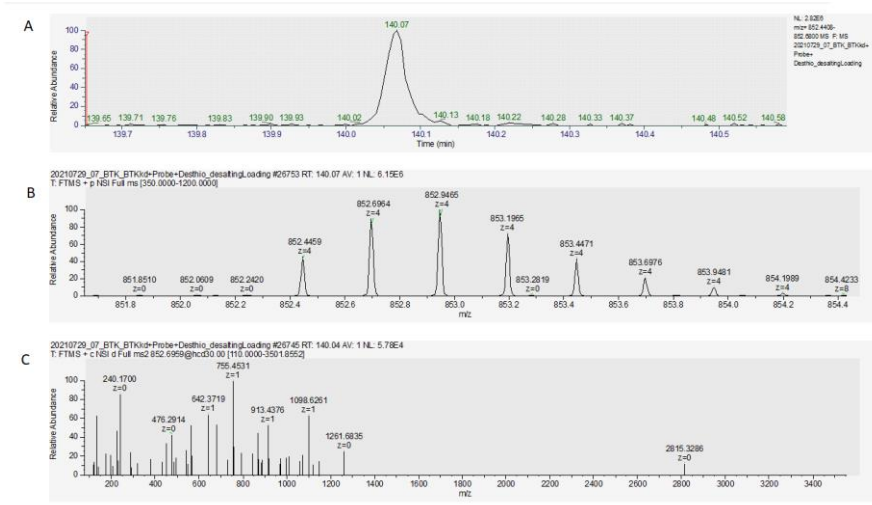


Figure 53 - Test of the new gradient on recombinant BTK and focus on peptide containing cys481 labelled with ibrutinib probe and desthiobiotin-azide; A. Eluted chromatogram precursor peak at 140 min; B. ms spectrum of precursor 852.4459 (z = 4); C. ms/ms spectrum of the corresponding precursor. Pictures obtained from Freestyle software

6.3.3 In cell target engagement with chemoproteomic

The PRM method reported above was performed in order to quantify ibrutinib-BTK target occupancy after *in cell* inhibitor treatment.

Briefly, RAMOS cells were pretreated with DMSO and 5 µM Ibrutinib for 24 hours and samples processed as described in Figure 44: lysates were treated with ibrutinib probe and "clicked" to desthiobiotin-azide. After digestion, probe-modified peptides were enriched on streptavidin resin and eluates analyzed with PRM analysis on Orbitrap exploris 240, target detecting peptide containing cysteine 481 covalently labeled by probe ibrutinib and desthiobiotin-azide. The same samples were used, in order to target detect other ibrutinib probe off-targets (6.3.5). In Figure 54-A, the chromatograms of fragment ions extracted from the peptide QRPIFIITEYMANG**Cys481**LLNYLR (m/z 852.440, 4+) labeled with Ibrutinib Probe and desthiobiotin-azide, are presented in control sample (i.e. cells treated with DMSO). The same chromatogram can't be identified in sample pre-treated with ibrutinib (Figure 54-B), confirming that the target engagement of ibrutinib is complete, and there is no free cysteine that could be labeled by ibrutinib probe. The peak indicated in Figure 54-B with a black arrow have a dotp 0.45, revealing that it doesn't correlate with reference ms/ms spectrum.

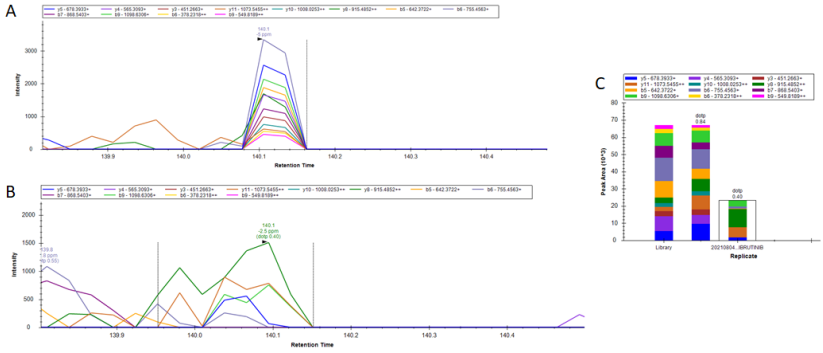


Figure 54 - Chromatograms of fragment ions extracted from the peptide of BTK containing cys 481. Mass measurement error and retention time of the most intense transition is annotated above the peak; A. sample corresponding to cells in DMSO; B. sample corresponding to cells treated with Ibrutinib

With targeted mass spectrometry-based chemoproteomic workflow, it was possible to calculate the target occupancy of ibrutinib for BTK that resulted to be complete after the treatment of cells with 5 µM of ibrutinib for 24 hours. Next steps were focused on the determination of cellular selectivity of ibrutinib.

6.3.4 Identification of probe ibrutinib off-targets

We started with the validation of chemoproteomic workflow that was further optimized with the aim to fully characterize ibrutinib profile of interaction and identify its off-targets. The experiment was performed on Ramos lysates, by treating lysates with high concentration of ibrutinib probe with the aim to identify a high number of probe off-targets; then, probe-modified proteins were "clicked" with desthiobiotin-azide and after digestion, peptides were enriched on streptavidin resin and eluates analyzed by LC-MS/MS. Raw data were analyzed with Maxquant software. At this probe concentration, 562 peptides were identified as modified by the probe, corresponding to 358 proteins.
As reported in Figure 55, we have compared probe-modified proteins identified in this experiment (selected and blue colored in figure) with their total abundances (reported on Y-axis), determined after the

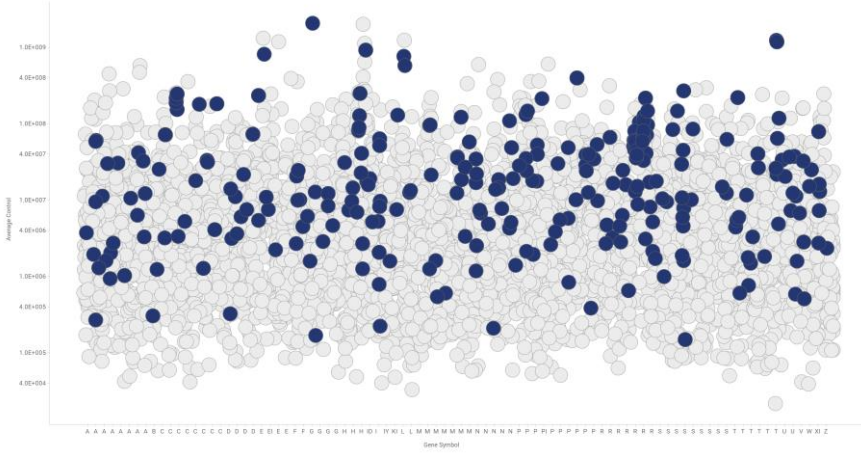


Figure 55 - Comparison of proteins identified modified by ibrutinib probe with their total abundances in RAMOS cell line; on x-axis are reported gene name and on y-axis total abundances log transformed. The protein colored and labeled are identified with probe modification. The others lightly colored are identified only in total proteome

analysis of Ramos cell line total proteome, in order to understand if derivatization occurs in most cases on abundant proteins or even low abundant reactive proteins can be enriched by this approach. Interestingly applying this technique, we have identified also low abundant proteins bringing the covalent modification; confirming that the enrichment could help, reducing the complexity of the sample, in the identification of probe modified peptides.

We matched these results with other works published in these years in which chemoproteomic approaches were used in order to detect reactive cysteines. The majority of these proteins, containing cysteines labelled by ibrutinib probe, were identified in these works reporting reactive cysteine that could be labelled by electrophilic warhead used with this high concentration (*117*). We identified RTN4 protein modified on cysteine 1101; this cysteine was identified in a paper published by Cravatt group as reactive with iodoacetamide-desthiobiotin (*118*).

Moreover, XPO1 was identified modified by ibrutinib probe on the same cysteine that is targeted by selinexor, a first-in-class selective inhibitor of nuclear export (*65*).

Focusing on XPO1 and peptide containing cysteine 528, target of selinexor, the score of the peptide probe-modified, without inclusion of diagnostic peaks, is 32.7; however, the score of modified-peptide including diagnostic peaks is 50.4. This data confirms the utility of these peaks in increasing the score of identification of probe-labeled peptides; especially, considering endogenously and low abundant proteins.

In order to clarify what of these targets could be ibrutinib off-targets and quantify the target occupancy, we performed two strategies: one targeted strategy, based on the selection of two interesting proteins and the validation of them and another strategy based on an unbiased proteome-wide competitive and quantitative chemoproteomic enrichment.

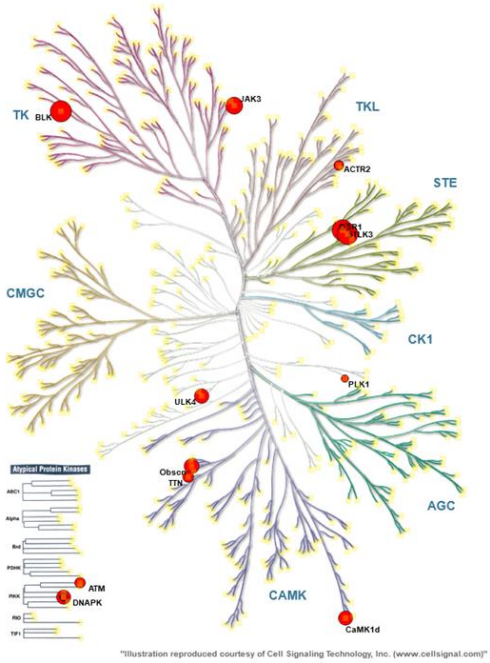


Figure 56 – Probe labelled protein kinases identified and reported in kinome tree; the size of the red circle corresponds to the identification score

Regarding the first strategy, we focused on probe labelled protein kinases identified and reported on kinome tree with red circle and shown in Figure 56.
From these, JAK3 and BLK were identified modified on cysteine 909 and 319, respectively. Their sequences and identification score are reported in the Table 4 below.

Peptide sequence	Gene symbol	Score
GC[878]LLDFLK	BLK-Cys 319	66
LVMEYLPSGC[878]LR	JAK3 – Cys 909	48

Table 4 - Peptide sequence of JAK3 and BLK covalently modified by ibrutinib probe

BTK	T	E	Y	M	A	N	G	C	L	L	N
ITK	F	E	F	M	E	H	G	C	L	S	D
TEC	T	E	F	M	E	R	G	C	L	L	N
BMX	T	E	Y	I	S	N	G	C	L	L	N
RLK/TXK	T	E	F	M	E	N	G	C	L	L	N
BLK	T	E	Y	M	A	R	G	C	L	L	D
JAK3	M	E	Y	L	P	S	G	C	L	R	D
EGFR	T	Q	L	M	P	F	G	C	L	L	D
ErbB2/HER2	T	Q	L	M	P	Y	G	C	L	L	D
ErbB4/HER4	T	Q	L	M	P	H	G	C	L	L	E

Figure 57 - Alignment of kinases having a cysteine residue in the ATP-binding site corresponding to cys 481 in BTK

These proteins, together with other 7 kinases, have a corresponding cysteine residue in the ATP binding site (Figure 57).

These cysteines are located in the ATP-binding site in a position corresponding to the one of cysteine 481 in BTK, so they could be considered for further characterization as possible ibrutinib off-targets. Thus, we started with the validation of these kinases as ibrutinib off-targets.

6.3.5 Jak3 and BLK validation as ibrutinib off-targets

In order to validate Jak3 as ibrutinib off-target, recombinant Jak3 was tested as it was already available in house; thus, GST-JAK3 (781-end) was treated with Ibrutinib and Ibrutinib Probe at a 10-fold excess for 1 hour at room temperature. A control was prepared incubating the protein with a solution containing the same percentage of organic solvent (DMSO) included in the reaction with the two compounds. The complete derivatization of the protein was confirmed by LC-MS (Figure 59): after 1 hour of incubation, there is a peak corresponding to the protein (66791.6 Da), whose deconvolute control mass spectrum was reported in Figure 59-A and C, with a mass shift of 440.5 Da for ibrutinib molecule (Figure 59-B) and 464.52 Da for probe ibrutinib (Figure 59-D); confirming that the protein was completely modified by a single molecule of probe and its corresponding parent inhibitor.

We digested then the complexes in Figure 59-B and D in order to verify that the modification occurs at cysteine 909 level.

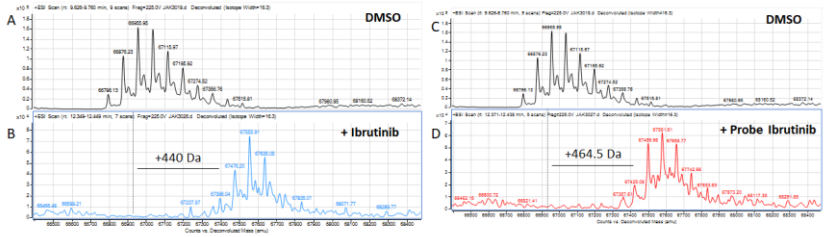


Figure 59 - LC-MS analysis of recombinant GST- JAK3 (781-end) derivatized with ibrutinib (440.5 Da) and ibrutinib probe (464.5 Da); A and C. recombinant JAK3 control; B. recombinant JAK3 labelled with ibrutinib; D. recombinant JAK3 labelled with probe ibrutinib

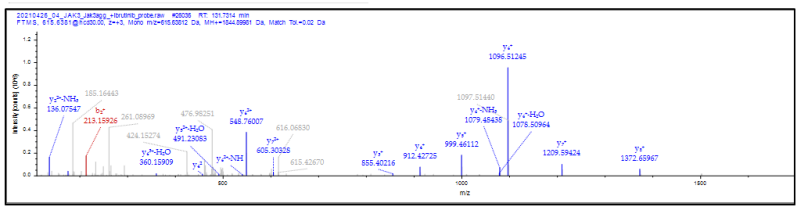


Figure 58 - ms/ms spectrum of JAK3 peptide containing cys909 covalently labelled by probe Ibrutinib

As reported in Figure 58, using Proteome Discoverer software (Thermo Fisher Scientific), the peptide containing cysteine 909 (LVMEYLPSG**C**LR) was identified containing a mass addition of 464.52 Da (consistent with the mass of Ibrutinib Probe) and 440.5 Da (corresponding to the mass of Ibrutinib). Both b and y ions are labeled in the above spectrum (Figure 58) confirming the identity of the peptide and ibrutinib targeted amino acid.

We focused then on the validation of BLK as ibrutinib off-targets. In this case, the recombinant protein was not present in house; thus, we performed a PRM analysis in order to detect probe ibrutinib modified BLK after competition experiment in cells: Ramos cells were treated with 5 µM Ibrutinib or DMSO for 24 hours. Lysates were then treated with Ibrutinib probe and "clicked" to desthiobiotin-azide. After digestion, probe-modified peptides were enriched on streptavidin resin and eluates analyzed with PRM analysis on Orbitrap exploris 240, detecting BLK peptide containing cysteine 319 covalently labeled by probe ibrutinib and desthiobiotin-azide. A DDA (data dependent

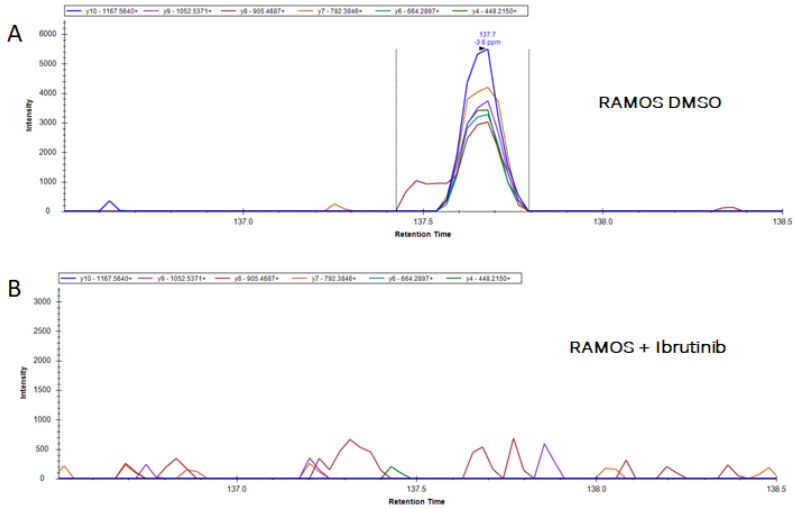


Figure 60 - Chromatograms of fragment ions extracted from the peptide of BLK containing cys 319. Mass measurement error and retention time of the most intense transition is annotated above the peak; A. sample corresponding to cells in DMSO; B. sample corresponding to cells treated with Ibrutinib

acquisition) analysis was previously performed in order to construct a spectral library for target identification of BLK peptide and identify the retention time window (i.e. 137 min) that must be added during the PRM analysis. DDA data were analyzed using Proteome Discoverer; Skyline software was used for PRM analysis.

In Figure 60-A, the chromatogram of fragment ions extracted from the peptide G**Cys319**LLDFLKTDEGSR (m/z 811.4041, 3+) labeled with Ibrutinib Probe and desthiobiotin-azide in control sample (i.e. cells treated with DMSO) is shown. The same chromatogram is not present in sample pre-treated with ibrutinib (Figure 60-B), confirming that the target engagement of ibrutinib for BLK is complete, and there is no free cysteine that could be labeled by ibrutinib probe.

We compared these results with data reported in literature: among kinases most potently engaged by Ibrutinib, BLK was identified. In an article reported by Cravatt and co-workers (*118*), the authors reported a chemoproteomic technique used to profile the ligandable proteome of spleen tissue extracted from mice treated intraperitoneally with 10 mg/kg or 20 mg/kg ibrutinib in biological duplicate. They identified BTK cysteine 481 as one of the most significantly ibrutinib target; in addition, they found cysteine 313 on BLK, that correspond to human

cysteine 319, as a target of ibrutinib with a competition ratio (% reduction of probe modified peptide signal intensity) similar to the one calculated for BTK.
Moreover, even if the potency of irreversible covalent inhibitors should be described by the K_i parameter (non-covalent binding to the target) and K_{inact} (the rate at which it reacts with the target nucleophile) rather than IC50, the measurement of IC50 at a specified time (IC50(t)) can be done in order to compare and describe the potency of covalent inhibitors. In the Pharmacology Review presented by Pharmacyclics for the NDA submission of Ibrutinib to FDA in 2013, the IC50 of ibrutinib for BTK, BLK and JAK3 were reported to be respectively 0.46 nM, 0.52 nM and 16 nM. As we can observe, ibrutinib possess a lower potency for JAK3 than BLK and BTK, attesting that JAK3 can be ibrutinib off-target only using a higher dose of inhibitor (*119*).

We then focus on the second strategy; an unbiased proteome-wide quantitative chemoproteomic experiment in order to detect and quantify others possible ibrutinib off-targets. The strategy is reported in Figure 44. An important issue is related to the quantification of ibrutinib-off target occupancy because the methods provided for modified-peptides enrichment and the differential abundances are calculated at peptide level and not at protein level, thus affecting the accuracy of the final ratio. So, the next part of the project was focused on the optimization and comparison of two quantitative mass spectrometry approaches, label free quantification and tandem mass tag, in order to quantify targets that specifically interact with ibrutinib after competition experiments. The optimized workflow will be then conducted after treatment of cells with an increasing dose of ibrutinib, in order to identify and quantify only the targets that specifically interacts with the inhibitor in dose-dependent manner.

6.4 Quantitative mass spectrometry approaches optimization

We initially optimized the analysis applying the two methods on the study of Ramos total proteome after treatment of cells with ibrutinib or DMSO.
A two-sample t-test was employed to define proteins differentially regulated in ibrutinib treated and control groups respectively. Protein log2fold changes were evaluated by comparing the average of LFQ intensities and TMT intensities between treated and control groups.

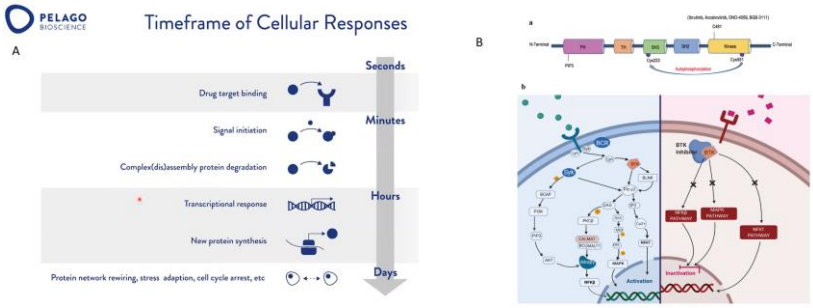


Figure 61 - A. Timeframe of cellular responses after inhibitor treatment; B. B cell receptor signaling cascade after ibrutinib treatment

Statistically significant (p value <0.05) differentially expressed proteins were selected. Regarding these proteins, many of them present a high percentage of ratio variability and the changes in term of differential expression were not so high. We have to consider that ibrutinib covalently label BTK and the first inhibition effects will be at B cell receptor signaling cascade level, as it shown in Figure 61, and in order to observe effects at protein expression level, longer treatment should be performed.

Thus, analysis of enriched proteome is very useful for the study of inhibitor mechanism of action, taking account of immediately functional and activity effects after ibrutinib treatment. Enrichment of phosphorylated peptides and the study of in-depth characterization of signal transducing events can be performed with phosphoproteomic analysis and the activity effects can be studied with chemoproteomic analysis.

6.4.1 Tandem Mass Tag

Regarding TMT technique, RAMOS cells in DMSO and treated with ibrutinib were processed as described in 5.4.2. The raw files were analyzed with Proteome Discoverer software. 108512 peptides group with high confidence, corresponding to 7702 proteins with at least medium confidence, and 6933 high confidence proteins with a minimum of 2 peptides, were identified by the Sequest and Mascot Distiller search engine in Proteome Discoverer™ using Uniprot_Homo sapiens database.

We started to evaluate the efficiency of sample preparation steps and in particular considering the main critical parts: normalization of

peptides in each sample before TMT labeling, the efficiency of TMT labeling and the fractionation of peptides before mass spectrometry analysis.
The normalization of peptides prior to TMT labeling is of high importance because relative quantification relies on equal amounts of labelled peptides being mixed. Different techniques are used: pooling of a little amount of TMT-labeled sample and analysis of the mixture by nano LC-MS/MS in order to calculate the relative concentration using the average intensity of each TMT reporter ion or using quantitative colorimetric peptide assay in order to calculate peptide concentration in each sample. When having a high number of samples, the automation of this step could help in reducing systematic errors and increase the reproducibility of sample preparation; thus, we optimized a protocol of peptides purification with DigestPro™, (5.5.7) relying on the idea that, when performing a double purification step, it is possible to saturate and elute 10 µg, or at least the same amount, of peptides per each sample, using 10 µL ZipTips® C18 with 5 µg capacity. We check the not normalized median log10 abundances across each sample group, after nanoLC-MS/MS analysis of peptide mixture, and it is very similar, confirming that this procedure of normalization is useful to label and mix the same amounts of peptides, increasing the reproducibility of sample preparation.

We then focus on TMT labeling. The standard protocol of TMT relies on labelling 25-100 µg peptides with 800 µg of TMT reagent; however, the cost for labeling reagents represents a major contribution to the overall experiment cost, limiting the use of these kits. Storage conditions of TMT reagents have to be carefully checked because TMT are very sensitive to hydrolysis caused by absorbed water from ambient air; so, cost-benefit use of TMT labelling has been evaluated in detail.

Zecha et al. (*120*) presented an optimized TMT labeling approach, reducing the quantity of required labeling reagent, considering that 100 µg peptides correspond to 116 nmol primary amines and 800 µg TMT correspond 2.3 µmol. They suggested that the standard protocol uses at least a 20-fold molar excess of the labeling reagent. Therefore, we decreased TMT and peptide quantities, reducing the reaction volume to maintain the standard protocol concentrations that are

	Initial conc.	Final conc.	Amount in nmol
Peptides (10 µg)	1 mg/ml	0.7 mg/ml	11.6 nmol
TMT Reagent (40 µg)	10 mg/ml	2.9 mg/ml	118 nmol
% ACN finale		32	

Table 5 - Concentration of peptides and TMT used for the labeling, using the down-scaled TMT labeling

reported in Table 5: 10 µg of peptides were labelled with 40 µg of TMT reagent and we calculated the TMT efficiency.
In order to verify the efficiency of TMT labeling with the new protocol, we performed a new data analysis, setting up, in the software, TMT as variable modification: 10% of peptides resulted to be not modified by TMT reagent confirming that the quantity of TMT used for the labeling is enough to label 10 µg of peptides.

We investigated if the TMT labelling proceed more efficiently on lysine or N-terminal of peptides to better define quantitation in the presence of different N-terminus amino acid of peptides and insert the correct parameters into the analysis software: the number of peptides containing lysine are 129982 and the one with TMT modification are 126422; so the percentage of lysine peptides modification is 97%; however, the N-terminal peptides without N-acetyl modification are 161529 and the number of these peptides with TMT modification is 133923, corresponding to a 83%. We can observe that the TMT modification is more efficient on lysine than peptide N-terminal; as a result, when obtaining % of TMT labeling fewer than 90%, TMT modification on lysine can be added as static and TMT modification on peptide N-terminus can be added as variable, in order to reduce the complexity of the search space.

Regarding the fractionation performed with Pierce™ high pH reversed-phase fractionation kit (5.5.6), the distribution of the peptide groups identified in every fraction is reported in Figure 62 and it is possible to observe that the majority of peptides elute using a percentage of acetonitrile greater than 12.5%. I reported the chromatogram of each fraction in Figure 63 and we can notice that increasing the concentration of acetonitrile in elution buffer, correspond to a shift of eluted peptides during chromatographic separation, corresponding to a gradual elution of hydrophobic peptides.

We analyzed even the labeled peptide mixture before fractionation in order to test the importance of the fractionation step: the number of proteins identified in fractionated samples was increased of 50% compared to the unfractionated one, confirming that fractionation is an important step to reduce the complexity of the sample and increase protein coverage.

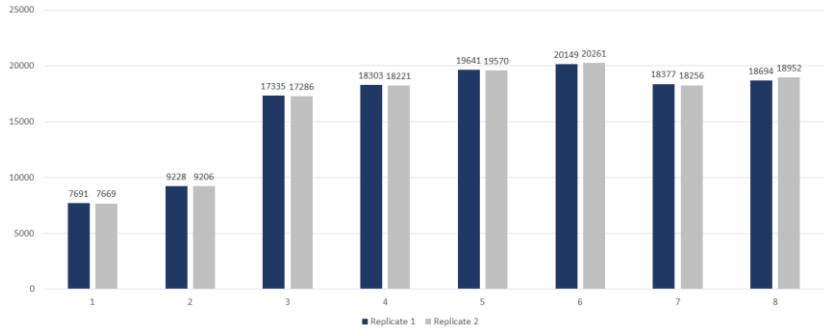


Figure 62 - Number of peptide groups identified in single fractions and in each replicate

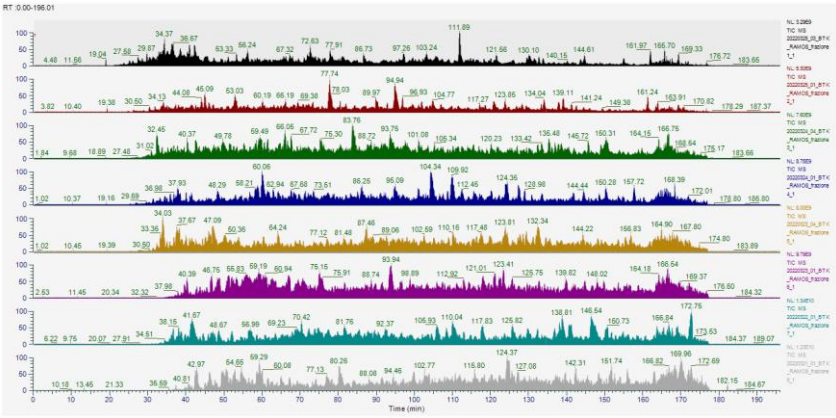


Figure 63 - Chromatograms of the 8 fractions of Ramos total proteome samples

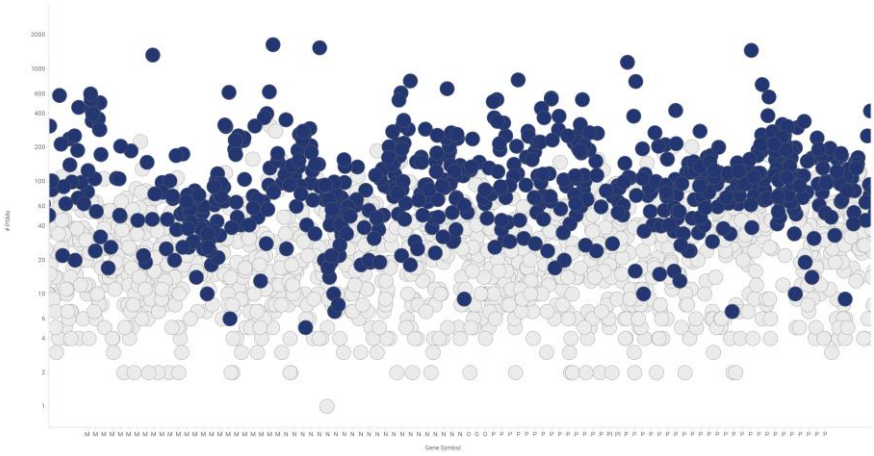


Figure 64 - Number of PSM identified for each protein in fractionated and not fractionated samples; on y-axis the number of PSM of each protein in fractionated sample is reported, PSM of proteins in not fractionated sample are blue colored; on x-axis gene name of proteins identified

In the Figure 64 reported above, I compared the number of peptides spectral matches (PSM), reported on y-axis, identified for each protein of fractionated sample (blue colored) with the number of PSM of proteins identified in sample before fractionation (grey colored). We can observe that with fractionation, we have identified proteins with low number of PSM that were not identified in the sample not fractionated (grey colored). Thus, considering that the number of PSM can correlate with the protein abundances, with fractionation we have reduced the complexity of sample, identifying even low abundant proteins.

The efficiency of fractionation step was confirmed by calculating the percentage of unique peptides identified only in a single fraction that correspond to 78% (Figure 65), indicating a great fractional resolution with spin-columns.

Interestingly, we calculated the % of TMT modified peptides eluted in each fraction, and there is an enrichment of these peptides at high concentration of acetonitrile, confirming that the TMT modification impair to the peptide an higher hydrophobicity and this is the reason

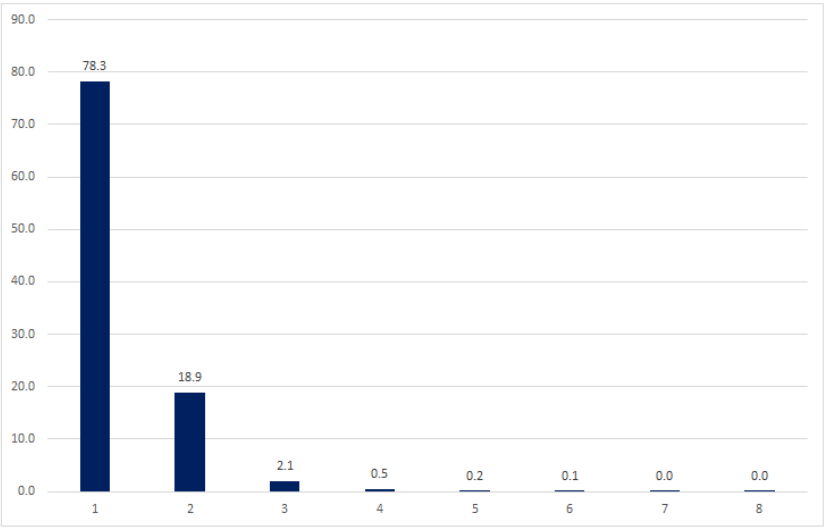


Figure 65 - Percentage of fraction efficiency in single, double fractions

why there is the an additional fraction elution at 22.5% acetonitrile in the standard protocol (5.5.6).

6.4.2 Label-free quantification

We next focused on label-free quantification method. The raw files obtained by label-free analysis of RAMOS in DMSO and treated with Ibrutinib for 24 hours were analyzed with Proteome Discoverer software. 70544 peptides group with high confidence, corresponding to 6028 proteins with at least medium confidence, and 5126 high confidence proteins, with a minimum of 2 peptides, were identified by the Sequest and Mascot Distiller search engine in Proteome Discoverer™ using Uniprot_Homo sapiens database.

With label free data, MS PepSearch node was used in order to increase the number of proteins identified as reported in the Venn diagram shown below (Figure 66).

The high dynamic range of proteins make the detection of low abundant proteins extremely difficult with label free with consequently decreased reproducibility; thus, high attention has to be put when considering protein ratio of low abundant proteins that can be affected due to the low number of MS/MS spectra obtained for low abundance proteins in comparison with high abundant ones. This

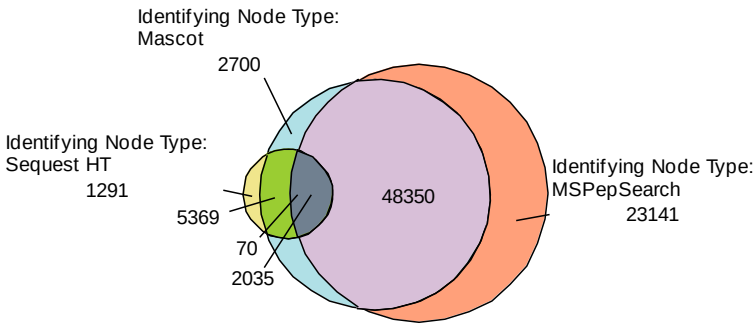


Figure 66 - Venn diagram of proteins identified in the three different search nodes: SequestHT, Mascot and MSPepSearch

issue is partially compensated by an analysis software parameter named match between runs.

MBR is included in MaxQuant and Proteome Discoverer software and it compares retention time of identified peak in an MS1 spectrum from an LC-MS/MS run to unidentified peaks in another samples. An identification is transferred if an unidentified peak with the same properties (m/z, charge state) is found within a specified retention time window. This algorithm was developed in order to overcome the problem of missing values and inaccuracies that occur when label free quantification is performed; however, in the absence of tandem MS information, the risk of false transfers from MBR algorithm can occur; especially considering less abundant proteins in a proteome wide label free analysis. Maxquant is more stringent than Proteome Discoverer in applying this algorithm, because it has a “safety” mechanism, which ensures that in a sample group, including a number of replicates for the same sample, there should be at least one peptide identified by MS/MS to include the abundance of the peptide in the calculated abundance of the total protein. Furthermore, MBR can be switched off when using Maxquant software while, in the Proteome Discoverer software, is included in the analysis without the possibility to switching it off.

We compared the number of proteins and peptides identified and quantified between the two methods of quantification: considering the number of proteins identified and quantified, the number is similar; however, if we consider the number of peptides identified and quantified, the difference is bigger. Thus, confirming that LFQ can

result in inaccuracies and missing data that can alter the final ratio of a protein between two different conditions.

In conclusion, with label free quantification, it is possible to compare an unlimited number of samples and the sample preparation is easier and faster than TMT because LFQ doesn't require the labelling step; however, the sensitivity can be affected due to missing values and the risk of false transfer, especially considering less abundant proteins. Moreover, samples are separately processed and, in particular in case of inclusion in the method of enrichment steps, this can increase the risk to introduce systematic errors, decreasing the reproducibility of the runs. Many technical and biological replicates have to be used, thus, increasing the instrument analysis time. Quantitative proteomics via TMT can be more reliable due to its sample multiplexing capability, high precision and throughput. Multiplexing provides increased sensitivity for precursor and product ions; giving the possibility to mix up to 16 samples, reducing the starting material quantity for each condition and saving instrument time. For label free approaches, we processed 1-2 mg of starting material for each sample; however, for TMT approaches, it is possible to process up to 100-200 µg per sample.

Furthermore, with TMT, there is the possibility to perform fractionation to reduce the complexity of the sample and increase protein coverage. However, the technique requires the use of labeling kit that increases sample preparations costs. We have observed that especially with complex spectra, coeluting peptides within the isolation window affect the accuracy of quantification, underestimating the real protein/peptide abundance differences: this is known as ratio compression phenomenon.

Of course, analysis of less complicated mixture of peptides (e.g. phosphoproteomic or chemoproteomic sample), reducing isolation window and increasing gradient time during mass spectrometry analysis, can result in a better protein ratio quantification.

Additional analysis is needed to apply these two quantitative methods to chemoproteomic workflow, reported in the previously chapters, after ibrutinib dose-dependent treatments, in order to in-depth identify and quantify other off-targets that are covalently labeled by ibrutinib at an effective and therapeutic dose.

7 CONCLUSIONS

The determination of target engagement in living system (cells, animal models, human) is fundamental to determine the mechanism of action of a small molecule covalent inhibitor during drug development

The generation of a chemical probe, starting from the covalent inhibitor, is a very useful method to better characterize the extent of target engagement under physiological conditions and to better identify possible off-targets that can then lead to toxicity in drug development program.

In recent years, many papers described methods based on the use of chemical probes to determine target engagement level in physiological conditions by the use of fluorescent or mass spectrometry-based methods; however, standardized tools for mass spectrometry data analysis in this field, are still in the early days.

The study of ibrutinib mechanism of action was used as case study in order to test and optimize the workflow, define the critical steps and standardize the approach with the possibility to transfer the protocol to additional developed covalent inhibitors.

Click chemistry reaction, thanks to its simplicity and robustness, allowed the evaluation of ibrutinib target engagement and selectivity profile directly *in cell*. The functionalization of the molecule with a small group such as the alkyne group ensured probe cellular permeability and the maintenance of an activity profile similar to ibrutinib parent inhibitor.

The versatility of click chemistry probe allowed not only to verify target engagement of BTK in physiological conditions functionalizing the azide with a fluorescent tag and in-gel fluorescence detection but also functionalizing the azide with a desthiobiotin moiety. High-resolution quantitative mass spectrometry analysis enabled the quantification of target occupancy for BTK and moreover the identification and quantification of other possible off-targets.

In addition, the development of a chemoprotemic method coupled to PRM analysis allowed the quantification of the ibrutinib-BTK target engagement and the verification of site (cysteine 481) covalently modified by ibrutinib. Different peptides were identified covalently modified by the chemical probe and from them, two interesting proteins were selected: BLK and JAK3. Both proteins were confirmed to be ibrutinib true off-targets modified at cysteine 319 and 909 respectively.

Quantitative chemoprotemic methods reported in literature are based on identification at probe-modified proteins level but did not localize sites of probe-modification or compare with the total protein levels in the cells. Indeed, identification of cysteine modification sites definitively proves that the modification occurs. We developed new competitive chemoproteomic protocol based on probe-modified peptides enrichment. Thus, we compared and optimized two quantitative mass spectrometry approaches with the aim of defining the selectivity profile of ibrutinib, applying the methods to the study of a total proteome.

Further analysis is therefore needed to apply these quantitative methods to chemoproteomic workflow after ibrutinib dose-dependent treatments in order to identify and quantify off-targets that are covalently labeled by ibrutinib at an effective and therapeutic dose.

8 REFERENCES

1. M. R. Wilkins, K. L. Williams, Appel Ron D., D. F. Hochstrasser. (Springer Berlin, Heidelberg, 1997), pp. XVIII, 245.
2. J. Cox, M. Mann, Is proteomics the new genomics? *Cell* **130**, 395-398 (2007).
3. K. Chandramouli, P. Y. Qian, Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genomics Proteomics* **2009**, (2009).
4. S. Magdeldin *et al.*, Basics and recent advances of two dimensional- polyacrylamide gel electrophoresis. *Clin Proteomics* **11**, 16 (2014).
5. Y. W. Kwon *et al.*, Application of Proteomics in Cancer: Recent Trends and Approaches for Biomarkers Discovery. *Front Med (Lausanne)* **8**, 747333 (2021).
6. Y. Zhou *et al.*, Proteomic signatures of 16 major types of human cancer reveal universal and cancer-type-specific proteins for the identification of potential therapeutic targets. *J Hematol Oncol* **13**, 170 (2020).
7. J. Parsons, C. Francavilla, 'Omics Approaches to Explore the Breast Cancer Landscape. *Front Cell Dev Biol* **7**, 395 (2019).
8. S. Saleem *et al.*, Proteomics analysis of colon cancer progression. *Clin Proteomics* **16**, 44 (2019).
9. W. Health, O. (WHO). (Who.int/ data/gho/data/theme s/morta lity-andgloba l-healt
   h-estim ates/ghe-leadi ngcause s-of-death, Accessed December 11,2020.)
10. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
11. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
12. Q. Jiang, M. Li, H. Li, L. Chen, Entrectinib, a new multi-target inhibitor for cancer therapy. *Biomed Pharmacother* **150**, 112974 (2022).
13. R. Aebersold, M. Mann, Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).

14. J. Verweij, M. J. de Jonge, Achievements and future of chemotherapy. *Eur J Cancer* **36**, 1479-1487 (2000).
15. V. V. Padma, An overview of targeted cancer therapy. *Biomedicine (Taipei)* **5**, 19 (2015).
16. K. Strebhardt, A. Ullrich, Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat Rev Cancer* **8**, 473-480 (2008).
17. T. A. Baudino, Targeted Cancer Therapy: The Next Generation of Cancer Treatment. *Curr Drug Discov Technol* **12**, 3-20 (2015).
18. D. G. Savage, K. H. Antman, Imatinib mesylate--a new oral targeted therapy. *N Engl J Med* **346**, 683-693 (2002).
19. M. J. Piccart-Gebhart *et al.*, Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N Engl J Med* **353**, 1659-1672 (2005).
20. B. K. Mishra, P. M. Parikh, Targeted Therapy in Oncology. *Med J Armed Forces India* **62**, 169-173 (2006).
21. Y. T. Lee, Y. J. Tan, C. E. Oon, Molecular targeted therapy: Treating cancer with specificity. *Eur J Pharmacol* **834**, 188-196 (2018).
22. L. M. Weiner, R. Surana, S. Wang, Monoclonal antibodies: versatile platforms for cancer immunotherapy. *Nat Rev Immunol* **10**, 317-327 (2010).
23. L. Zhong *et al.*, Small molecules in targeted cancer therapy: advances, challenges, and future perspectives. *Signal Transduct Target Ther* **6**, 201 (2021).
24. W. D. Joo, I. Visintin, G. Mor, Targeted cancer therapy--are the days of systemic chemotherapy numbered? *Maturitas* **76**, 308-314 (2013).
25. N. Iqbal, Imatinib: a breakthrough of targeted therapy in cancer. *Chemother Res Pract* **2014**, 357027 (2014).
26. J. C. Chuang, J. W. Neal, Crizotinib as first line therapy for advanced ALK-positive non-small cell lung cancers. *Transl Lung Cancer Res* **4**, 639-641 (2015).
27. F. L. Opdam, H. J. Guchelaar, J. H. Beijnen, J. H. Schellens, Lapatinib for advanced or metastatic breast cancer. *Oncologist* **17**, 536-542 (2012).
28. F. Nurwidya, F. Takahashi, K. Takahashi, Gefitinib in the treatment of nonsmall cell lung cancer with activating epidermal growth factor receptor mutation. *J Nat Sci Biol Med* **7**, 119-123 (2016).
29. R. Fisher, J. Larkin, Vemurafenib: a new treatment for BRAF-V600 mutated advanced melanoma. *Cancer Manag Res* **4**, 243-252 (2012).

30. X. Wang, H. Zhang, X. Chen, Drug resistance and combating drug resistance in cancer. *Cancer Drug Resist* **2**, 141-160 (2019).
31. J. Zhang, P. L. Yang, N. S. Gray, Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* **9**, 28-39 (2009).
32. H. Davies *et al.*, Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954 (2002).
33. M. J. Garnett, R. Marais, Guilty as charged: B-RAF is a human oncogene. *Cancer Cell* **6**, 313-319 (2004).
34. J. Y. Wang, K. M. Wilcoxen, K. Nomoto, S. Wu, Recent advances of MEK inhibitors and their clinical progress. *Curr Top Med Chem* **7**, 1364-1378 (2007).
35. G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The protein kinase complement of the human genome. *Science* **298**, 1912-1934 (2002).
36. C. Arter, L. Trask, S. Ward, S. Yeoh, R. Bayliss, Structural features of the protein kinase domain and targeted binding by small-molecule inhibitors. *J Biol Chem* **298**, 102247 (2022).
37. R. Roskoski, Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol Res* **103**, 26-48 (2016).
38. S. B. Hari, E. A. Merritt, D. J. Maly, Sequence determinants of a specific inactive protein kinase conformation. *Chem Biol* **20**, 806-815 (2013).
39. K. M. Backus, Applications of Reactive Cysteine Profiling. *Curr Top Microbiol Immunol* **420**, 375-417 (2019).
40. Q. Liu *et al.*, Developing irreversible inhibitors of the protein kinase cysteinome. *Chem Biol* **20**, 146-159 (2013).
41. J. M. Strelow, A Perspective on the Kinetics of Covalent and Irreversible Inhibition. *SLAS Discov* **22**, 3-20 (2017).
42. E. De Vita, 10 years into the resurgence of covalent drugs. *Future Med Chem* **13**, 193-210 (2021).
43. J. Singh, R. C. Petter, T. A. Baillie, A. Whitty, The resurgence of covalent drugs. *Nat Rev Drug Discov* **10**, 307-317 (2011).
44. D. S. Johnson, E. Weerapana, B. F. Cravatt, Strategies for discovering and derisking covalent, irreversible enzyme inhibitors. *Future Med Chem* **2**, 949-964 (2010).
45. Z. Zhao, P. E. Bourne, Progress with covalent small-molecule kinase inhibitors. *Drug Discov Today* **23**, 727-735 (2018).
46. A. Chaikuad, P. Koch, S. A. Laufer, S. Knapp, The Cysteinome of Protein Kinases as a Target in Drug Development. *Angew Chem Int Ed Engl* **57**, 4372-4385 (2018).

47. M. F. Robinson *et al.*, Efficacy and Safety of PF-06651600 (Ritlecitinib), a Novel JAK3/TEC Inhibitor, in Patients With Moderate-to-Severe Rheumatoid Arthritis and an Inadequate Response to Methotrexate. *Arthritis Rheumatol* **72**, 1621-1631 (2020).
48. P. Y. Lee, Y. Yeoh, T. Y. Low, A recent update on small-molecule kinase inhibitors for targeted cancer therapy and their therapeutic insights from mass spectrometry-based proteomic analysis. *FEBS J*, (2022).
49. M. Schürmann, P. Janning, S. Ziegler, H. Waldmann, Small-Molecule Target Engagement in Cells. *Cell Chem Biol* **23**, 435-441 (2016).
50. M. Hay, D. W. Thomas, J. L. Craighead, C. Economides, J. Rosenthal, Clinical development success rates for investigational drugs. *Nat Biotechnol* **32**, 40-51 (2014).
51. A. Lin *et al.*, Off-target toxicity is a common mechanism of action of cancer drugs undergoing clinical trials. *Sci Transl Med* **11**, (2019).
52. P. Morgan *et al.*, Can the flow of medicines be improved? Fundamental pharmacokinetic and pharmacological principles toward improving Phase II survival. *Drug Discov Today* **17**, 419-424 (2012).
53. D. Cook *et al.*, Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov* **13**, 419-431 (2014).
54. K. V. M. Huber, G. Superti-Furga, Profiling of Small Molecules by Chemical Proteomics. *Methods Mol Biol* **1394**, 211-218 (2016).
55. M. M. Hann, G. L. Simpson, Intracellular drug concentration and disposition--the missing link? *Methods* **68**, 283-285 (2014).
56. G. M. Simon, M. J. Niphakis, B. F. Cravatt, Determining target engagement in living systems. *Nat Chem Biol* **9**, 200-205 (2013).
57. E. Schepers, G. Glorieux, A. Dhondt, L. Leybaert, R. Vanholder, Flow cytometric calcium flux assay: evaluation of cytoplasmic calcium kinetics in whole blood leukocytes. *J Immunol Methods* **348**, 74-82 (2009).
58. Y. Liu, M. P. Patricelli, B. F. Cravatt, Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci U S A* **96**, 14694-14699 (1999).
59. S. Wang, Y. Tian, M. Wang, G. B. Sun, X. B. Sun, Advanced Activity-Based Protein Profiling Application Strategies for Drug Development. *Front Pharmacol* **9**, 353 (2018).

60. H. Fang *et al.*, Recent advances in activity-based probes (ABPs) and affinity-based probes (AfBPs) for profiling of enzymes. *Chem Sci* **12**, 8288-8310 (2021).
61. C. E. Franks, K. L. Hsu, Activity-Based Kinome Profiling Using Chemical Proteomics and ATP Acyl Phosphates. *Curr Protoc Chem Biol* **11**, e72 (2019).
62. M. Bantscheff *et al.*, Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat Biotechnol* **25**, 1035-1044 (2007).
63. H. C. Kolb, M. G. Finn, K. B. Sharpless, Click Chemistry: Diverse Chemical Function from a Few Good Reactions. *Angew Chem Int Ed Engl* **40**, 2004-2021 (2001).
64. S. Arastu-Kapur *et al.*, Nonproteasomal targets of the proteasome inhibitors bortezomib and carfilzomib: a link to clinical adverse events. *Clin Cancer Res* **17**, 2734-2743 (2011).
65. J. G. Martin *et al.*, Chemoproteomic Profiling of Covalent XPO1 Inhibitors to Assess Target Engagement and Selectivity. *Chembiochem* **22**, 2116-2123 (2021).
66. G. Blum, G. von Degenfeld, M. J. Merchant, H. M. Blau, M. Bogyo, Noninvasive optical imaging of cysteine protease activity using fluorescently quenched activity-based probes. *Nat Chem Biol* **3**, 668-677 (2007).
67. E. K. Evans *et al.*, Inhibition of Btk with CC-292 provides early pharmacodynamic assessment of activity in mice and humans. *J Pharmacol Exp Ther* **346**, 219-228 (2013).
68. J. P. Schülke *et al.*, Chemoproteomics demonstrates target engagement and exquisite selectivity of the clinical phosphodiesterase 10A inhibitor MP-10 in its native environment. *ACS Chem Biol* **9**, 2823-2832 (2014).
69. J. B. Fenn, M. Mann, C. K. Meng, S. F. Wong, C. M. Whitehouse, Electrospray ionization for mass spectrometry of large biomolecules. *Science* **246**, 64-71 (1989).
70. M. Karas, F. Hillenkamp, Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Anal Chem* **60**, 2299-2301 (1988).
71. E. J. van Rooden *et al.*, Mapping in vivo target interaction profiles of covalent inhibitors using chemical proteomics with label-free quantification. *Nat Protoc* **13**, 752-767 (2018).
72. J. Cox *et al.*, Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol Cell Proteomics* **13**, 2513-2526 (2014).

73. P. V. Bondarenko, D. Chelius, T. A. Shaler, Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal Chem* **74**, 4741-4749 (2002).
74. H. Liu, R. G. Sadygov, J. R. Yates, A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* **76**, 4193-4201 (2004).
75. M. P. Washburn, D. Wolters, J. R. Yates, Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotechnol* **19**, 242-247 (2001).
76. S. Niessen *et al.*, Proteome-wide Map of Targets of T790M-EGFR-Directed Covalent Inhibitors. *Cell Chem Biol* **24**, 1388-1400.e1387 (2017).
77. E. Weerapana *et al.*, Quantitative reactivity profiling predicts functional cysteines in proteomes. *Nature* **468**, 790 (2010).
78. M. Schirle *et al.*, Kinase inhibitor profiling using chemoproteomics. *Methods Mol Biol* **795**, 161-177 (2012).
79. A. C. Peterson, J. D. Russell, D. J. Bailey, M. S. Westphall, J. J. Coon, Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* **11**, 1475-1488 (2012).
80. P. Picotti, R. Aebersold, Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat Methods* **9**, 555-566 (2012).
81. M. Bantscheff, B. Kuster, Quantitative mass spectrometry in proteomics. *Anal Bioanal Chem* **404**, 937-938 (2012).
82. P. Mallick, B. Kuster, Proteomics: a pragmatic perspective. *Nat Biotechnol* **28**, 695-709 (2010).
83. J. Cox *et al.*, Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **10**, 1794-1805 (2011).
84. C. L. Xiao *et al.*, Dispec: a novel peptide scoring algorithm based on peptide matching discriminability. *PLoS One* **8**, e62724 (2013).
85. J. E. Elias, S. P. Gygi, Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4**, 207-214 (2007).
86. J. Cox, M. Mann, MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).

87. W. Zhu, J. W. Smith, C. M. Huang, Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol* **2010**, 840518 (2010).
88. L. Arike, L. Peil, Spectral counting label-free proteomics. *Methods Mol Biol* **1156**, 213-222 (2014).
89. S. Mehta *et al.*, Precursor Intensity-Based Label-Free Quantification Software Tools for Proteomic and Multi-Omic Analysis within the Galaxy Platform. *Proteomes* **8**, (2020).
90. S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**, 2301-2319 (2016).
91. J. D. O'Connell, J. A. Paulo, J. J. O'Brien, S. P. Gygi, Proteome-Wide Evaluation of Two Common Protein Quantification Methods. *J Proteome Res* **17**, 1934-1942 (2018).
92. C. Lazar, L. Gatto, M. Ferro, C. Bruley, T. Burger, Accounting for the Multiple Natures of Missing Values in Label-Free Quantitative Proteomics Data Sets to Compare Imputation Strategies. *J Proteome Res* **15**, 1116-1125 (2016).
93. J. J. O'Brien *et al.*, The effects of nonignorable missing data on label-free mass spectrometry proteomics experiments. *Ann Appl Stat* **12**, 2075-2095 (2018).
94. M. Bantscheff, S. Lemeer, M. M. Savitski, B. Kuster, Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **404**, 939-965 (2012).
95. Y. V. Karpievitch, A. R. Dabney, R. D. Smith, Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* **13 Suppl 16**, S5 (2012).
96. M. Y. Lim, J. A. Paulo, S. P. Gygi, Evaluating False Transfer Rates from the Match-between-Runs Algorithm with a Two-Proteome Model. *J Proteome Res* **18**, 4020-4026 (2019).
97. C. M. Lewis, C. Broussard, M. J. Czar, P. L. Schwartzberg, Tec kinases: modulators of lymphocyte signaling and development. *Curr Opin Immunol* **13**, 317-325 (2001).
98. J. M. Lindvall *et al.*, Bruton's tyrosine kinase: cell biology, sequence conservation, mutation spectrum, siRNA modifications, and expression profiling. *Immunol Rev* **203**, 200-215 (2005).
99. J. A. Woyach, A. J. Johnson, J. C. Byrd, The B-cell receptor signaling pathway as a therapeutic target in CLL. *Blood* **120**, 1175-1184 (2012).

100. A. B. Satterthwaite, O. N. Witte, The role of Bruton's tyrosine kinase in B-cell development and function: a genetic perspective. *Immunol Rev* **175**, 120-127 (2000).
101. Y. Baba *et al.*, BLNK mediates Syk-dependent Btk activation. *Proc Natl Acad Sci U S A* **98**, 2582-2586 (2001).
102. M. Cinar *et al.*, Bruton tyrosine kinase is commonly overexpressed in mantle cell lymphoma and its attenuation by Ibrutinib induces apoptosis. *Leuk Res* **37**, 1271-1277 (2013).
103. M. F. de Rooij *et al.*, The clinically active BTK inhibitor PCI-32765 targets B-cell receptor- and chemokine-controlled adhesion and migration in chronic lymphocytic leukemia. *Blood* **119**, 2590-2594 (2012).
104. R. E. Davis *et al.*, Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. *Nature* **463**, 88-92 (2010).
105. J. Liu, C. Chen, D. Wang, J. Zhang, T. Zhang, Emerging small-molecule inhibitors of the Bruton's tyrosine kinase (BTK): Current development. *Eur J Med Chem* **217**, 113329 (2021).
106. Z. Pan *et al.*, Discovery of selective irreversible inhibitors for Bruton's tyrosine kinase. *ChemMedChem* **2**, 58-61 (2007).
107. O. Hantschel *et al.*, The Btk tyrosine kinase is a major target of the Bcr-Abl inhibitor dasatinib. *Proc Natl Acad Sci U S A* **104**, 13283-13288 (2007).
108. L. A. Honigberg *et al.*, The Bruton tyrosine kinase inhibitor PCI-32765 blocks B-cell activation and is efficacious in models of autoimmune disease and B-cell malignancy. *Proc Natl Acad Sci U S A* **107**, 13075-13080 (2010).
109. S. E. Herman *et al.*, Bruton tyrosine kinase represents a promising therapeutic target for treatment of chronic lymphocytic leukemia and is effectively targeted by PCI-32765. *Blood* **117**, 6287-6296 (2011).
110. J. C. Byrd *et al.*, Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *N Engl J Med* **369**, 32-42 (2013).
111. F. Caron, D. P. Leong, C. Hillis, G. Fraser, D. Siegal, Current understanding of bleeding with ibrutinib use: a systematic review and meta-analysis. *Blood Adv* **1**, 772-778 (2017).
112. S. Ganatra *et al.*, Ibrutinib-Associated Atrial Fibrillation. *JACC Clin Electrophysiol* **4**, 1491-1500 (2018).
113. P. Ghia, M. Dlugosz-Danecka, L. Scarfò, W. Jurczak, Acalabrutinib: a highly selective, potent Bruton tyrosine kinase inhibitor for the treatment of chronic lymphocytic leukemia. *Leuk Lymphoma* **62**, 1066-1076 (2021).

114. T. Barf *et al.,* Acalabrutinib (ACP-196): A Covalent Bruton Tyrosine Kinase Inhibitor with a Differentiated Selectivity and In Vivo Potency Profile. *J Pharmacol Exp Ther* **363**, 240-252 (2017).
115. Pharmacyclics NDA (2013), https://www.accessdata.fda.gov/drugsatfda_docs/nda/2013/205552Orig1s000PharmR.pdf
116. B. R. Lanning *et al.,* A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nat Chem Biol* **10**, 760-767 (2014).
117. K. M. Backus *et al.,* Proteome-wide covalent ligand discovery in native biological systems. *Nature* **534**, 570-574 (2016).
118. M. Kuljanin *et al.,* Reimagining high-throughput profiling of reactive cysteines for cell-based screening of large electrophile libraries. *Nat Biotechnol* **39**, 630-641 (2021).
119. A. Berglöf *et al.,* Targets for Ibrutinib Beyond B Cell Malignancies. *Scand J Immunol* **82**, 208-217 (2015).
120. J. Zecha *et al.,* TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol Cell Proteomics* **18**, 1468-1478 (2019).