Synthesis of Aspartic Proteases Probes and their Application for Interaction Identification and Binding Hotspots Mapping

Zur Erlangung des akademischen Grades eines

Dr.rer.nat

von der Fakultät Bio- und Chemieingenieurwesen
der Technischen Universität Dortmund
genehmigte Dissertation

vorgelegt von

M.Sc. Suyuan Chen

aus

Mianyang, Sichuan, China

Tag der mündlichen Prüfung: 29.11.2023

1. Gutachter/-in: Prof. Dr. Albert Sickmann
2. Gutachter/-in: Prof. Dr. Markus Kaiser

Dortmund 2023

Angefertigt im Zeitraum von September 2017 bis Juni 2023

am Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V.

in der Otto-Hahn-Straße 6b, 44227 Dortmund

Table of Content

Achievements during PhD

1 Peer-reviewed publications

Publications related with thesis

- **Chen, S.**; 1 Liang, C.; Li, H.; Yu, W.; Kopczynski, D.; Loroch, S.; Fransen, M.; Verhelst, S., Pepstatin-based probes for photoaffinity labeling of aspartic proteases and application in target identification, ACS chemical biology, 2023, 18 (4), 686-692.

- Korovesis, D.; Beard, H.; **Chen, S.;** Verhelst, S. H., Cleavable linkers and their application in MS-based target identification. Mol. Omics, 2021, 17(2), 197-209.

Publications from external collaborations

- Sun, J.; Ru, J.; Ramos-Mucci, L.; Qi, F.; Chen, Z.; **Chen, S.;** Cribbs, A.; Deng, L.; Wang, X., DeepsmirUD: the Prediction of Regulatory Effects on microRNA Expression Mediated by Small Molecules Using Deep Learning. Int. J. Mol. Sci., 2023 (3), 1878.

- Yang, L.; **Chen, S.;** Yi, D.; Chen, Q.; Zhang, J.; Xie, Y.; Sun, H., Synthesis and fluorescence properties of red-to-near-infrared-emitting push–pull dyes based on benzodioxazole scaffolds. Journal of Materials Chemistry B, 2021, 9(40), 8512-8517. (Co-first author)

Publications related with the work during master

- Yang, L.; **Chen, S.;** Yi, D.; Fu, Q.; Liang, W.; Zhang, Z.; Du, F.; Ji, J.; Zeng, Q., Study on the synthesis of sulfenylindoles with Dess—Martin periodinane—promoted. (IN CHINESE), West China Journal of Pharmaceutical Sciences 2018, 33 (1), 8-12.

2 Oral presentation

- The endless frontier: Five things that learned from the winning of Merck Innovation Cup. Chinese-German Chemical Association (CGCA) 33rd Annual Conference: Chemistry in pandemic prevention and green energy, Dec 19th, 2020, Online, host in Germany.

- Size Matters - CARS monitoring of small molecules (Joint talk with Dr. Daniel Krahn), ISAS Postdoc Pitch Day, Oct 18th, 2021

- One more step towards target deconvolution of bio-active molecule: Sulfoxide diazirine (SODA) enables unbiased binding hotspots mapping via LC-MS^3, Leibniz Alumni Online Network, Jan 30th, 2023

- MS1 filtering of crosslinked peptides using the small molecule interface. Skyline Course, March 20th- 23rd, 2023, International course hosted at ISAS (Dortmund).

3 Poster presentation

- Suyuan Chen, Steven H. L. Verhelst, A cleavable photo-cross-linking strategy for "on" and "off" targets profiling. The 12th International Activity Based Protein Profiling Meeting (Leuven, Belgium), Mar 27th - 29th, 2019.

- Suyuan Chen, Steven H. L. Verhelst, A cleavable photo-cross-linking strategy for "on" and "off" targets profiling. Gesellschaft Chinesischer Chemiker und Chemieingenieure in der Bundesrepublik Deutschland (GCCCD) 31st Annual Conference: Strategic Development in Chemistry and Chemical Engineering (Mainz, Germany), Dec 7th, 2019.

- Suyuan Chen, Stefan Loroch, Sem Tamara, Riccardo Z. Chiozzi, Albert J. R. Heck, Albert Sickmann, Steven H. L. Verhelst, Sulfoxide diazirine (SODA) enables unbiased binding hotspots mapping via LC-MS^3. 2022 International Symposium on Chemical Biology (Geneva, Switzerland), Nov 8th -10th, 2022.

4 Stipends and Grants

- CSC scholarship (201704910881), 2017-2021 (48 months), cover living cost and health insurance in Germany.

- Scientific challenge – BioSolveIT GmbH, Fragment-based discovery of novel covalent inhibitors of rhomboid intramembrane proteases, Jun 2019 - Jun 2020, one-year software license.

- EPIC-XS (Collaboration Grant, #266), A MS-cleavable photoreactive group for identification of target proteins binding to short linear peptide motifs with top-down proteomics approach, Dec 2020 - Present, cover 1-2 weeks travel and accommodation costs at collaborating institute (Utrecht University, Utrecht, The Netherlands) plus experiment cost at collaborating institute.

- Travel Grant, 2022 International Symposium on Chemical Biology (NCCR-Chemical Biology), Geneva, Switzerland, Nov 8th -10th, 2022, 200 CHF.

5 Awards

- Merck Innovation Cup 2021, Winner (Targeted protein degradation: Molecular Glue).

Erklärung vorab veröffentlichter Inhalte

Die Inhalte dieser Forschungsarbeit sind am ISAS im Rahmen des Forschungszweiges „Krankheitsmechanismen und Targets “ im Projekte „Chemische Sonden“ erarbeitet worden.

Teile dieser Arbeit sind bereits vom Autor veröffentlicht und präsentiert worden oder basieren auf Messdaten, die im Rahmen betreuter studentischer Arbeiten am ISAS entstanden sind. Insbesondere die Synthese der verwendeten chemischen Sonden und die biologischen Evaluationen dieser sind explizit genannten Beiträgen:

Kapitel 1	1.6 in Teilen modifiziert aus	[A]
Kapitel 3	3.1 in Teilen modifiziert aus	[B]
Kapitel 4	4.1 in Teilen modifiziert aus	[B]
Kapitel 5	5.1 in Teilen modifiziert aus	[B]

Publikationen

[A] Korovesis, D.; Beard, H.; **Chen, S.;** Verhelst, S. H., Cleavable linkers and their application in MS-based target identification. Mol. Omics, 2021, 17(2), 197-209.

[B] **Chen, S.**; 1 Liang, C.; Li, H.; Yu, W.; Kopczynski, D.; Loroch, S.; Fransen, M.; Verhelst, S., Pepstatin-based probes for photoaffinity labeling of aspartic proteases and application in target identification, ACS chemical biology, 2023, 18 (4), 686-692.

Abstract

Aspartic proteases play a crucial role in human physiology and pathologyincluding as biomarkers for breast cancer and Alzheimer's disease, and as potential drug targets for infectious diseases. However, chemical probes for photoaffinity labeling (PAL) of these proteases are underdeveloped.

We develop a full on-resin synthesis of clickable PAL probes based on the natural product inhibitor pepstatin, incorporating a minimal diazirine photo-reactive group. The positioning of this group in the inhibitor determines the labeling efficiency. Effective probes sensitively detect cathepsin D, a biomarker for breast cancer, in cell lysates. Through chemical proteomics experiments and deep learning algorithms, we also identify sequestosome-1 as a direct interaction partner and substrate of cathepsin D.

PAL combined with tandem mass spectrometry (MS^n) can reveal interactions between small molecule drugs and protein in biological environments. However, the direct detection of the ‘hotspots’ of the photo-crosslinking sites by MS^n is challenging because of the unexpected fragmentation of small molecule drugs, especially when these are small peptides. We synthesize and introduce sulfoxide diazirine (SODA) building blocks to peptide-like probes for PAL. Those MS-cleavable probes enable a MS^2 cleavage event that generates a probe-derived reporter ion and a minimal fragment on the modified peptide. Following a subsequent MS^3 fragmentation event, we show that this strategy allows for unbiased identification of modification sites and mapping of binding hotspots of peptide-like bio-active molecules.

Overall, our study presents the synthesis of aspartic proteases probes and their application for interaction identification and binding hotspots mapping. However, further improvement is required for this study to achieve broader application. We propose a number of possible follow-up experiments and discuss future prospects in **chapter 6**.

Zusammenfassung

Aspartatproteasen spielen eine entscheidende Rolle in der menschlichen Physiologie und Pathologie, unter anderem als Biomarker für Brustkrebs und Alzheimer, sowie als potenzielle Wirkstoffziele für Infektionskrankheiten. Allerdings ist die Toolbox chemischer Sonden für die Photoaffinitätsmarkierung (PAL) dieser Proteasen nur unzureichend entwickelt.

Wir entwickeln eine vollständige On-Resin-Synthese für klickbare PAL-Sonden auf der Basis des Naturstoff-Inhibitors Pepstatin, die eine minimale Diazirin-Reaktivgruppe tragen. Es wurde gezeigt, dass die Platzierung der Gruppe im Inhibitor die Markierungseffizienz bestimmt und, dass diese Sonden den sensitiven Nachweis von Cathepsin D, einem Biomarker für Brustkrebs, in Zelllysaten erlauben. Dafür wurde eine Serie von Experimenten in Verbindung mit Deep-Learning-Algorithmen-gestützter Datenauswertung durchhgeführt, wobei Sequestrosom-1 als direkter Interaktionspartner und Substrat von Cathepsin D identifiziert wurde.

Somit wurde gezeigt, dass PAL in Kombination mit Tandem-Massenspektrometrie (MS^n) eingesetzt werden kann, um Interaktionen zwischen kleinen Moleküldrogen und Proteinen in biologischen Umgebungen aufzudecken. Die direkte Detektion der "Hotspots" der Photo-Crosslink-Stellen durch MS^n ist jedoch aufgrund der komplexen und nicht vorhersehbaren Fragmentierung von kleinen Moleküldrogen, insbesondere bei kleinen Peptiden, weiterhin eine Herausforderung.

Wir synthetisieren und führen Sulfoxid-Diazirin (SODA) Bausteine in Peptid-ähnliche Sonden für PAL ein. Dabei entsteht eine MS-labile Gruppe, die im MS^2 zu einem ein Reporterion und ein modifiziertes Peptidion führt. Durch eine anschließende MS^3 Fragmentierung zeigen wir, dass diese Strategie eine Identifizierung von Modifikationsstellen und die Kartierung von Bindungshotspots von Peptid-ähnlichen bioaktiven Molekülen ermöglicht.

Zusammenfassend, zeigt die Arbeit, die erfolgreiche Synthese von Aspartatprotease-Sonden und deren Anwendung zur Identifikation von Interaktionspartnern einschließlich der exakten Interaktionsstelle. Allerdings sind weitere Optimierungen erforderlich, um eine breitere Anwendung zu ermöglichen. Wir schlagen eine Reihe möglicher Folgeexperimente vor und diskutieren zukünftige Perspektiven in **Kapitel 6**.

List of abbreviations

ABPs (ABP)	Activity-based probes
ABPP	Activity-based protein profiling
ACN	Acetonitrile
AcOH	Acetic acid
AfBPP	Affinity-based protein profiling
AfBPs	affinity-based probes
AGC	Automatic gain control
AI	Artificial intelligent
asRNA	antisense ribonucleic acid
BACE1/2	Beta-site APP cleaving enzyme 1/2
BAL	Backbone amide linker
BCA protein assay	Bicinchoninic acid protein assay
Boc	tert-Butyloxycarbonyl protecting group
CADD	Computer-aided drug design
$CDCl_3$	Deuterated chloroform
cDNA	complementary deoxyribonucleic acid
CID	Collision-induced dissociation
cryo-EM	Cryogenic electron microscopy
CuAAC	Cu(I)-catalyzed azide-alkyne cycloaddition
CYM	Chymosin
DABCO	1,4-diazabicyclo[2,2,2]octane
DARTS	Drug Affinity Responsive Target Stability
DCM	Dichloromethane
DDA	Data dependent acquisition mode
DDA^2	Data dependent acquisition square
DIEA	*N,N*-Diisopropylethylamine
DIQ	(S, S, S)-decahydro-isoquinoline-3-carbonyl
DMEM	Dulbecco's Modified Eagle Medium
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DPBS	Dulbecco's phosphate-buffered saline
DTT	Dithiothreitol
EBSS	Earle's Balanced Salt Solution
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assay
EM	Electron microscopy
ER	Astrogen receptor
ESI-source	Electrospray ionization source
ETD	Electron transfer dissociation

EThcD	Electron-transfer/higher-energy collision dissociation
FA	Formic acid
FAIMS	High-field asymmetric waveform ion mobility spectrometry
FBS	Fetal Bovine Serum
FDA	The United States Food and Drug Administration
FDR	False discovery rate
FKBP12	Tacrolimus (FK506) binding protein 12
Fmoc	Fluorenylmethoxycarbonyl protecting group
Fmoc-Ala-OH	Fmoc-L-alanine
Fmoc-photoLeu-OH	(*S*)-2-(Fmoc-amino)-3-(3*H*-diazirin-3-yl)butanoic acid
Fmoc-Sta-OH	N-Fmoc-L-statine
Fmoc-Val-OH	Fmoc-L-valine
FRB domain	FKBP-rapamycin binding (FRB) domain
GluC	Glutamyl endopeptidase
HBTU	*N,N,N′,N′*-Tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate
HDX-MS	Hydrogen-Deuterium Exchange-Mass Spectrometry
HIV-1	Human immunodeficiency virus 1
HPLC	High-performance liquid chromatography
HRMS	High-resolution mass
HRPF	Hydroxyl radical protein footprinting
IAA	Iodoacetamide
idotp	Isotope dot product
IGF1	Insulin-like growth factor 1
iTRAQ	Isobaric Tag for Relative and Absolute Quantitation
LC	Liquid chromatography
LC-MS	Liquid chromatography–mass spectrometry
LC-MS/MS	Liquid Chromatography with tandem mass spectrometry (MS^1 + MS^2)
LC-MS^3	Liquid Chromatography with tandem mass spectrometry (MS^1 + MS^2 + MS^3)
LFQ	Label-free quantification
LiP-MS	Limited proteolysis-coupled mass spectrometry
MeOH	Methanol
MPs	Metalloproteases
MS	Mass spectrometry, mass spectrometer or MS survey scan
MS^1	First stage Ions survey scan
MS^2	Second stage survey scan of fragement products after the fragmentation of selective ions from MS^1
MS^3	Third stage survey scan of fragement products after the fragmentation of selective ions from MS^2
MS^n	Tandem mass spectrometry with Nth stage survey scan
mTOR	Mammalian target of rapamycin
nCE	Normalized collision energy
NHS	*N*-hydroxysuccinimide

NMP	N-Methylpyrrolidone
NMR	Nuclear magnetic resonance
NRIP3	Nuclear receptor interacting protein
PAL	Photo-affinity labeling
PCA	Principal component analysis
PDB	Protein Data Bank
pH	Potential of hydrogen
PPIs	Protein-protein interactions
PRM^2	Parallel reaction monitoring square
PSMs	Peptide spectrum matches
ROC curve	Receiver operating characteristic curve
RT	Room temperature
SBDD	Structure-based drug design
SDS-PAGE	sodium dodecyl sulfate–polyacrylamide gel electrophoresis
SILAC	Stable isotope labeling by amino acids in cell culture
SIM-PAL	Small molecule interactome mapping by photo-Affinity Labeling
SLiMs	Short linear motifs
SODA	Sulfoxide diazirine
SP3	Solid phase-enhanced sample-preparation
SPP	signal peptide peptidase
SPPL2a	Signal Peptide Peptidase Like 2a
SPPL2b	Signal Peptide Peptidase Like 2b
SPPL2c	Signal Peptide Peptidase Like 2c
SPPL3	Signal Peptide Peptidase Like 3
SPPS	Solid-phase peptide synthesis
SQSTM1	Sequestosome-1
TAMRA	Carboxytetramethylrhodamine
TAMRA-azide	Carboxytetramethylrhodamine-azide
TBST buffer	Tris-buffered saline with 0.1% Tween 20 detergent
TEA	Triethylamin
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TIPS	Triisopropyl silane
TLC	Thin-layer chromatography
TMT	Tandem Mass Tag
TNBC	Triple-negative breast cancer
UV	Ultraviolet
UVPD	Ultraviolet photodissociation
Xcorr	Cross-correlation
XL-MS	Cross-linking mass spectrometry

1 Introduction

1.1 Aspartic proteases

Peptidases – commonly referred to as proteases – are one of the largest enzyme families. To date, more than 4100 proteases have been identified (MEROPS database; http://merops.sanger.ac.uk).[1] They catalyze the cleavage of peptide bonds in proteins and polypeptides, which leads to protein activation and maturation, protein catabolism, protein transport, and the regulation of numerous physiological and pathological processes, including embryonic development, cell proliferation and tissue remodeling, blood coagulation, blood pressure control, inflammation, infection, and cancer.[2]

Based on the difference in their catalytic mechanism, proteases have been classified into seven types (aspartic proteases, Cysteine proteases, Glutamic proteases, Metalloproteases, Asparagine peptide lyases, Serine proteases, Threonine proteases). Aspartic proteases are a relatively small group of proteases with approximately 20 members in the human genome, divided into two clans (AA and AD) each with a common evolutionary origin (Figure 1A).[1] Clan AA has three families, family A1 including the classical aspartic proteases (pepsin A, pepsin C, beta-site APP cleaving enzyme 1/2 (BACE1/2), renin, cathepsin D, cathepsin E, and napsin A), family A2 containing integrated proteases of retroviruses like HIV protease, and family A28 including nuclear receptor interacting protein (NRIP3) and ubiquitin-dependent protease (DDI1 and DDI2), which hydrolyze substrates only when they are modified by long ubiquitin chains. Clan AD contains intramembrane proteases: presenilin-1, presenilin-2 and signal peptide peptidase (SPP) and related proteases Signal Peptide Peptidase Like 2a (SPPL2a), Signal Peptide Peptidase Like 2b (SPPL2b), Signal Peptide Peptidase Like 2c (SPPL2c) and Signal Peptide Peptidase Like 3 (SPPL3).

The crystal structures of most A1 human aspartic proteases have been determined. [3, 4, 5] In general, they follow a typical fold with three topologically unique regions: an N-terminal domain, a C-terminal domain, and an interdomain consisting of six-stranded antiparallel sheets connecting the other two domains (Figure 1B). Both the N-terminal domain and the C-terminal domain each contribute one catalytic aspartic acid residue to the active site, resulting in a total of two catalytic aspartic acid residues. The majority of clan AA aspartic proteases feature a flap that closes down on top of the substrate or inhibitor, protecting the active site from solvent and generating an active site with binding pockets on both sides of the catalytic residues.

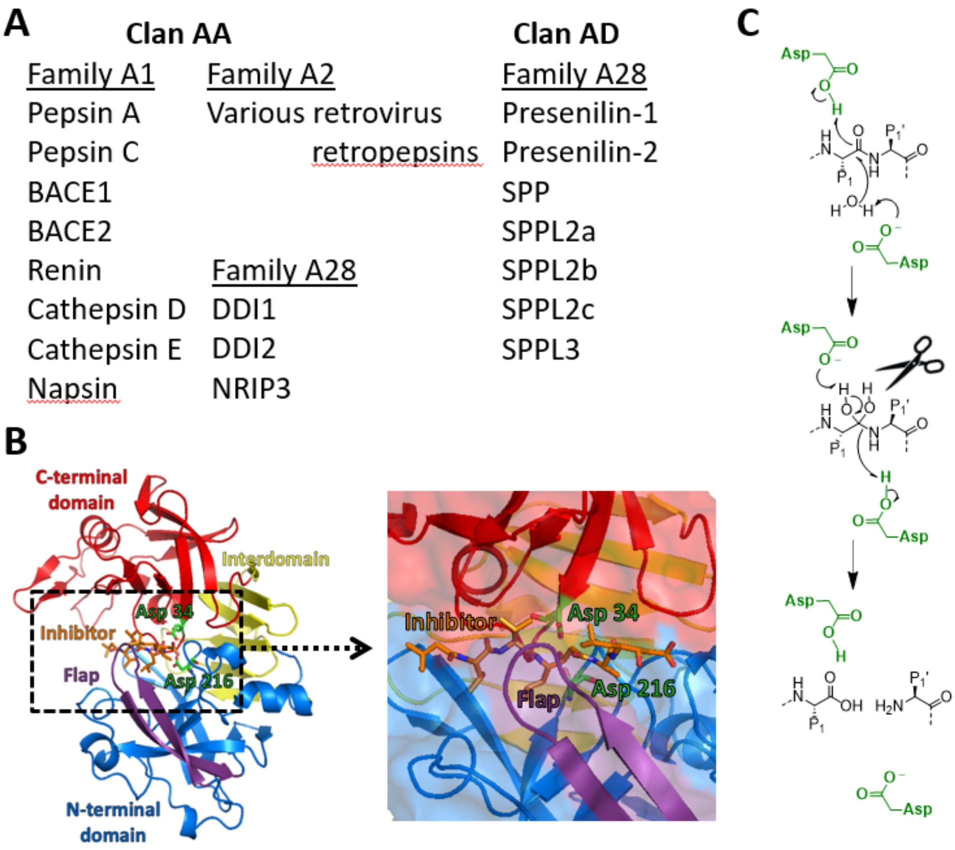


Figure 1. Aspartic proteases and the general aspartic protease catalytic mechanism. (A) Overview of aspartic proteases, which comprises two evolutionary separate clans (AA and AD). **(B)** Crystal structure of Bovine Chymosin in complex with Pepstatin A (PDB code: 4AUC). Protein depicted in cartoon mode with N-terminal domain in blue, C-terminal domain in red, interdomain in yellow and a loop in purple covering the active site (Asp 34 and Asp 216). Pepstatin is depicted in stick model (orange). Picture rendered with PyMol.[6] **(C)** Catalytic mechanism of amide bond hydrolysis. A protease substrate upon attack of a water molecule at the carbonyl of the scissile bond (indicated with scissors). Amino acid residues at the N-terminal side are named P1, P2 etc., whereas the residues at the C-terminal side are denoted with an apostrophe (P1' etc).

The cleavage of peptide bonds is accomplished by a general acid-base catalytic mechanism (Figure 1C).[7] [8] In the enzyme-substrate complex, one of the two catalytic aspartic residues undergoes protonation. The other aspartic residue functions as a general base, activating a water molecule, which then attacks the carbonyl carbon of the scissile amide bond, culminating in the production of a tetrahedral geminal diol intermediate. In the end, the proton of hydroxyl group is accepted by one of the catalytic aspartates, while the leaving amine is activated at the same time by the other protonated aspartic residue, resulting in the breakage of the peptide bond at the end of the reaction.

Aspartic proteases, despite the fact that they constitute a minor proportion of the protease population, play a crucial role in human physiology and pathological processes. These include cathepsin D, associated with breast cancer[9] as a poor prognosis biomarker; presenilin, a component of the γ-secretase complex which has an important role in the Notch signaling pathway and has been considered as a target in Alzheimer's disease[10] as well as in cancer therapy.[11] Aside from that, aspartic proteases of infectious pathogens, such as the HIV protease[12] and plasmepsins from the malaria-causing parasite Plasmodium falciparum, are also potential drug targets. [13]

1.2 Cathepsin D

1.2.1 Maturation of cathepsin D

Aspartic endo-protease cathepsin D is widely distributed in lysosomes. Initially, the primary role of cathepsin D was thought to be the degradation of proteins in lysosomes at an acidic pH [14]. Over the last three decades, it has been shown that cathepsin D, besides its activity as a main protein-degrading enzyme in lysosome, may also process and activate certain proteins in specialized cells [15]. For instance, essential growth factors are provided by cathepsin D to particular epithelial cells to facilitate tissue remodeling and regeneration.[16]

The pre-pro-enzyme form of cathepsin D is produced in the rough endoplasmic reticulum (RER) and undergoes numerous proteolytic cleavages during biosynthesis to create the mature form.[17] [18] [19] As soon as the signal peptide is removed from pro-cathepsin D, two *N*-linked glycosylation sites are connected and the enzyme is delivered to the Golgi. As a result of binding to mannose-6-phosphate (M6P) receptors, an intermediate 48 kDa single chain of cathepsin D is relocated to the lysosome. [20] [21]

The processing and activation of aspartic proteases generally fall into one of three categories. (1) Porcine pepsinogen may fully activate on its own. [22] (2) Pro-renin is completely activated with the assistance of a cofactor. (3) Cathepsin D undergoes a combination of partial auto-activation and enzyme-associated activation that results in the mature enzyme. [17] [23] [24] Double-knockout experiments of cathepsin B and L in the brains of mice indicated that the absence of cathepsin B and L increased the amounts of intermediate and mature cathepsin D. [25] Besides, the use of two cysteine protease inhibitors, CLICK-148 and CA-074-Me, led to a hypothesis that cathepsin L and cathepsin B are involved in the processing of intermediate 48 kDa cathepsin D to mature 34 kDa cathepsin D. [26]

1.2.2 Role of cathepsin D in cancer progression and metastasis

Cathepsin D has been shown to promote the growth of cancer cells, according to a number of studies. The isolated pro-cathepsin-D of MCF-7 breast cancer cells enhances MCF-7 cell proliferation in vitro. [27] Additionally, increased mitogenesis caused by pro-cathepsin D was also observed in prostate cancer cells. [28] [29] [30] Even more remarkable, when transfected with cathepsin D complementary deoxyribonucleic acid (cDNA), the growth of the 3Y1-Ad12 rat cancer cells was boosted in vitro at low and high cell densities, and the metastatic potential of the cells in in vivo experiments was also increased. [31] [32] [33]

Cathepsin D does not only stimulate the proliferation of cancer cells, but it also enhances angiogenesis. [34] Cathepsin D expression was significantly associated with increased vascular counts in 102 invasive breast carcinomas in a reported clinical investigation. [35] The function of cathepsin D in angiogenesis has not yet been completely understood. The release of extracellular matrix (ECM)-bound basic fibroblast growth factor (bFGF) by cathepsin D in breast cancer cells was first hypothesized to enhance angiogenesis in a preliminary investigation. [36] Pro-cathepsin D released by prostate cancer cells may also be responsible for the synthesis of angiostatin, which is a particular inhibitor of angiogenesis in vivo. [37] Moreover, cathepsin D also cleaves human prolactin, resulting in numerous 16K-like N-terminal prolactin fragments with antiangiogenic effects. [38]

Apart from cancer progression and angiogenesis, cathepsin D is involved in cancer metastasis. Experiments on rat tumor cells have shown that transfection-induced overexpression of cathepsin D has a direct effect on the ability of these tumor cells to metastasize. [31] [32] Instead of increasing invasiveness in this rat tumor model, it seems that cathepsin D had a favorable impact on cell proliferation and encouraged the formation of micro-metastases. [31] [32] [33] [34] [39] Particularly, in MDA-MB-231 cancer cells, an antisense ribonucleic acid (asRNA) approach revealed that cathepsin D influences growth, tumorigenicity, and lung colonization. [40]

1.2.3 Cathepsin D as a prognostic factor in breast cancer

As early as 1980, researchers discovered that the estrogen receptor (ER) positive human breast cancer cell line MCF-7 secretes a 52 kDa glycoprotein in response to estradiol treatment.[41] Later, it was determined that the protein was pro-cathepsin D, which is also highly expressed in triple-negative breast cancer (TNBC) cell lines such as BT-20 and MDA-MB231. [42] [43] Cathepsin D overexpression does not seem to be a result of gene amplification or significant chromosomal rearrangements in estrogen receptor positive-breast cancer cell lines.[44] Instead, estrogens and certain growth hormones (e.g., Insulin-like growth factor 1 (IGF1), Epidermal growth factor (EGF)) have a strong influence on cathepsin D production.[45] [46] Estrogen hormones boost the transcription of cathepsin D gene in ER-positive breast cancer cell lines,[47] [48] which subsequently leads to an increase of cathepsin D expression.[46]

In human breast cancer, cathepsin D was suggested as a tumor marker a long time ago. [49] Since then, many investigations aimed to link cathepsin D protein levels or activities to the clinical outcome of breast cancer patients. Several independent clinical studies have shown that the cathepsin D level in primary breast cancer cytosol is an independent prognostic parameter correlated with the incidence of clinical metastasis and shorter survival times.[50] [51] A meta-analysis of studies on node-negative breast cancer, [52] as well as a complete study on 2810 patients have confirmed the value of high expression level of cathepsin D as a marker of aggressiveness.[53] Furthermore, the expression level of cathepsin D was suggested to predict prognosis independent of any other clinical prognostic factor. Other studies clarified that cathepsin D levels in the tumor specimen, but not its proteolytic activity in patient serum, has prognostic value. [54]

1.3 Activity-based protein profiling (ABPP) for proteases

The activity-based protein profiling (ABPP) technique, developed by Benjamin F. Cravatt and Matthew Bogyo in the late 1990s[55] [56], is a chemical proteomic approach that uses activity-based probes (ABPs) to specifically label active proteins in biological samples and provide a direct readout of the functional state of proteins in biological systems (Figure 2B).[57] [58]

Because of the unique enzymatic reaction (substrate cleavage) offered by proteases, selective protease ABPs can be developed for use in various applications including proteome-wide profiling and imaging protease activity. ABPs for proteases normally have three components[59]: (1) an electrophilic group (also known as a warhead) that covalently reacts with the nucleophilic active site amino acid residue (serine, cysteine, threonine) in the catalytic pocket of certain proteases; (2) a tag that can be utilized to identify the covalent enzyme-probe complex using a variety of methods such as fluorescence detection, biotin blot, mass spectrometry measurements; (3) a peptide linker can be utilized to fuse the two previous parts together, and create selectivity for a specific class of proteases (Figure 2A). Due to their mode of action, these activity-based probes are limited to specific proteases. This is because other types of proteases, such as metalloproteases and aspartic proteases, employ an activated water molecule in their hydrolysis action, which would result in hydrolysis of the warhead. For such cases, the photoaffinity labeling technique (PAL) was developed to overcome the limitation of electrophilic warheads. The group of Benjamin Cravatt has developed a photoaffinity labeling approach for profiling of Metalloproteases (MPs), which promoted selective binding and labeling of MPs by coupling a benzophenone photocrosslinker to a zinc-chelating hydroxamate (Figure 3C).[60] [61] Also, the photoaffinity labeling of γ-secretase (one of the aspartic proteases) was developed by installing a benzophenone photocrosslinker on a hydroxylethylene-based inhibitor (Figure 3D).[62]

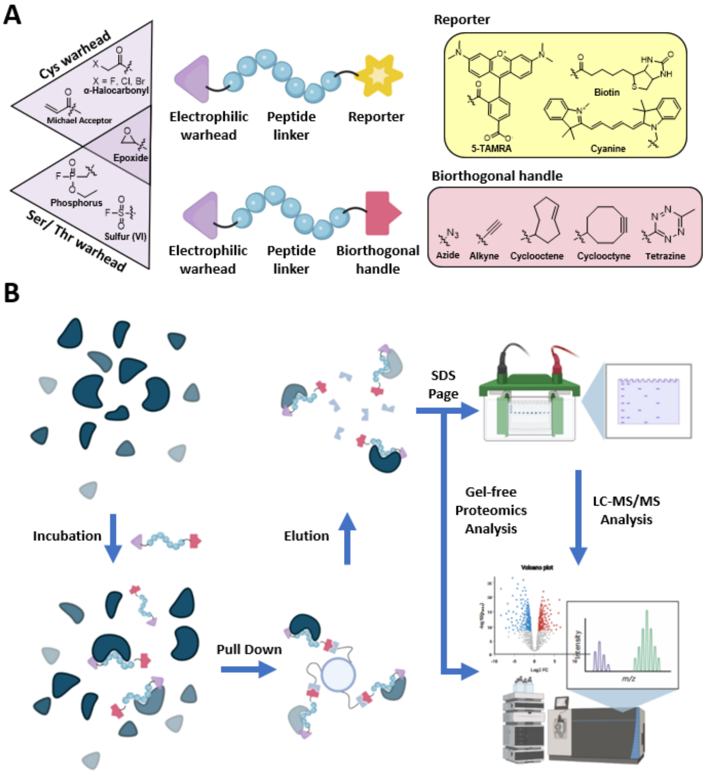


Figure 2. Activity-based probes and activity-based protein profiling (ABPP). (A) Overview of

activity-based probes. Representative cysteine (Cys) warhead: α-Halocarbonyl, Michael Acceptor and Epoxide. Representative Serine (Ser) / Threonine (Thr) warhead: Phosphorus, Sulfur(VI) Fluoride (Sulfur(VI)) and Epoxide. Reporter: 5-TAMRA (5-Carboxytetramethylrhodamine), Biotin and Cyanine. Biorthogonal handle: Azide, Alkyne, Cyclooctene, Cyclooctyne and Tetrazine **(B)** General workflow of ABPP. ABPs are applied to incubate with proteome of interest. ABP labeled proteins are enriched by streptavidin bead. Active proteins are eluted from the bead after washing. Active proteins are analyzed by SDS-page and LC-MS/MS or gel-free proteomics.

1.4 Photoaffinity labeling technique: affinity-based probes (AfBPs)

The PAL-based probes are often referred to as AfBPs because they rely on the binding affinity of the scaffold to the target protein.[63] Since its development by Westheimer et al. in 1962,[64] PAL has provided an approach for investigating ligand-receptor interactions. Upon UV irradiation, the photo-crosslinking group forms a highly reactive intermediate that reacts with its adjacent molecule, ultimately leading to the creation of a covalent connection between the probe and the target protein (Figure 3A).

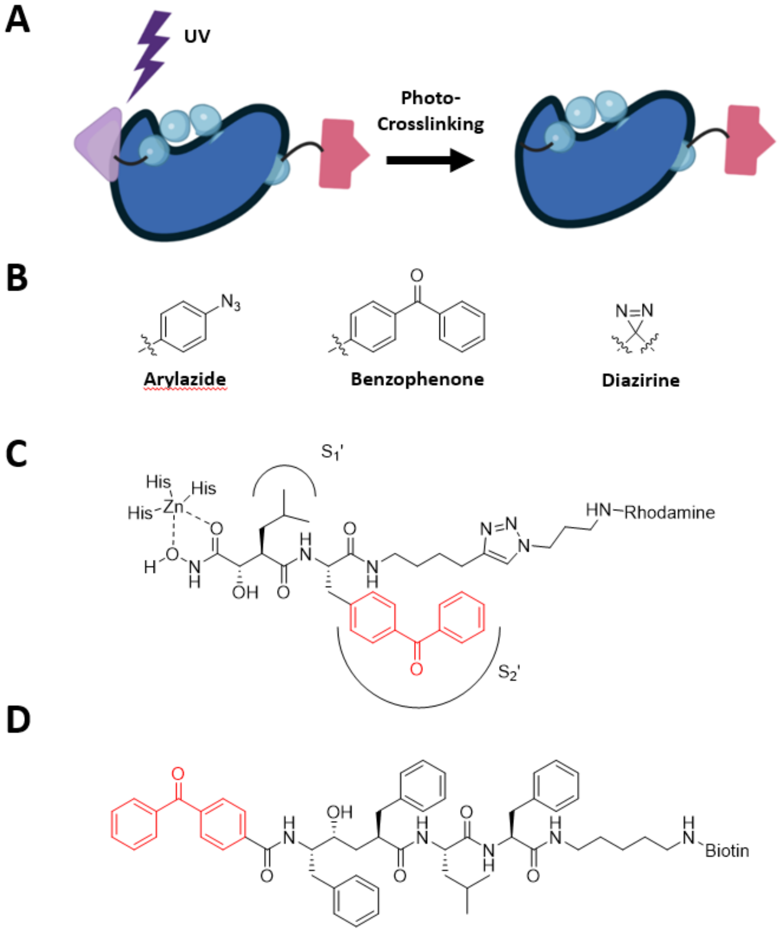


Figure 3. Affinity-based probes (AfBPs). (A) Overview of AfBPs. Probe is covalently crosslinked on proteins of interest. **(B)** Chemical structure of photo-crosslinking group: Arylazide, Benzophenone and Diazirine. **(C)** Chemical structure of metalloprotease AfBP. **(D)** Chemical structure of γ-secretase AfBP.

PAL frequently employs three different kinds of photoactivatable groups: arylazides, benzophenones, and diazirines (including aryl diazirines and aliphatic diazirines). (Figure 3B)

1.4.1 Aryl azides

The usage of aryl azide in PAL was first described in 1969 by J. R. Knowles and co-workers.[65] Aryl azide produces a singlet nitrene after UV irradiation, which primarily undergoes ring expansion to form 1,2-didehydroazepine and subsequently reacts with nucleophiles such as an amino group on the target protein (Figure 4A).[66] Furthermore, aryl azides are moderately stable in the dark and strongly reactive when exposed to ultraviolet light or ultraviolet radiation.

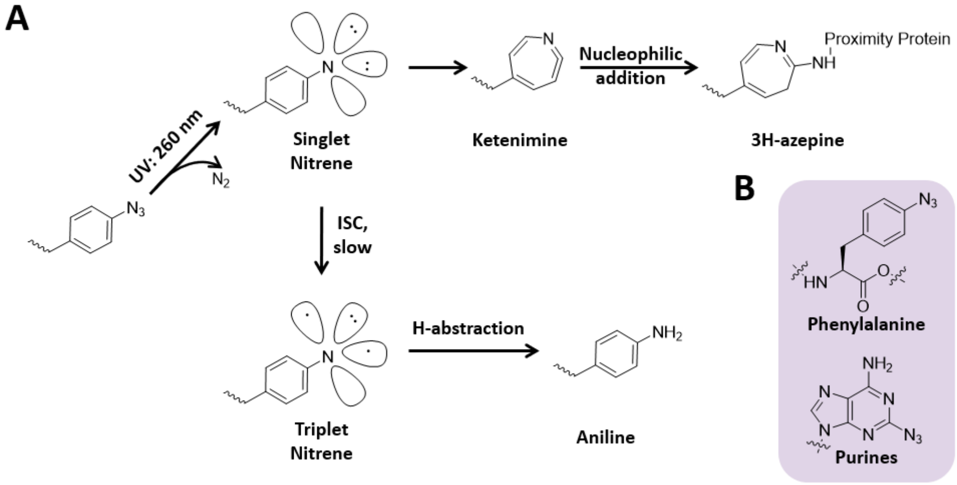


Figure 4. Aryl Azides (A) Photo-crosslinking mechanism of aryl azides. **(B)** Representative scaffold of aryl azides: phenylalanine and purines.

Because of the short activation wavelength of aryl azides (the main absorption peak is at approximately 260 nm) and their low crosslinking yields generated by side reactions, aryl azides have been used in just a few applications[67, 68]. They are, nevertheless, simple to synthesize and implement into AfBPs. The scaffold is also rather small, which makes it particularly effective in the modification of aromatic moieties that are widely present in bioactive chemicals, such as phenylalanine[69] and purines.[70] (Figure 4A)

1.4.2 Benzophenones

The use of benzophenones in PAL has grown in popularity since its initial report in 1973.[71] When exposed to the appropriate wavelength of light (ranging from 320 to 360 nm), benzophenones form a triplet ketyl biradical that can react with protein amino acid residues. In this process, the ketyl oxygen of triplet ketyl biradical abstracts the hydrogen from an amino acid residue as a consequence of the highly reactive n-π* state of the triplet ketyl.[72] The resulting radical on the residue then couples with the ketyl radical, leading to a covalent crosslink with the protein.[66] (Figure 5)

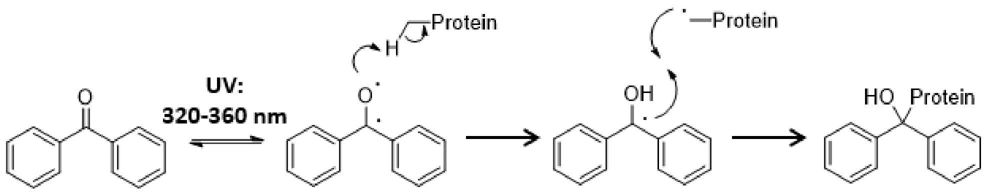


Figure 5. Benzophenone. Photo-crosslinking mechanism of benzophenone.

Benzophenones are particularly well-suited for biological applications as their produced reactive triplet state is almost inert to water.[71] They are triggered by longer (320 - 360 nm) wavelengths of light compared with aryl azides (approximately 260 nm). These wavelengths are less damaging to the aromatic residues of nucleic acids and proteins. Additionally, it should be emphasized that the excitation of benzophenones is a reversible process: if hydrogen abstraction does not occur during the lifetime of the excited state, benzophenones revert to their ground-state configuration. The ground state is then ready for re-excitation.[73] These two advantages (the resistance to water quenching and the length of the irradiation process) can result in significant non-specific labeling.

1.4.3 Diazirine

A diazirine is a two-nitrogen and one-carbon three-membered ring system. Its chemical synthesis was first reported in 1959,[74] [75] [76] while the first use of diazirines as PAL chemical tools occurred in 1973.[77] Depending on whether the diazirine ring is connected directly to an aromatic ring or to an aliphatic carbon atom, it is categorized into aromatic diazirine or aliphatic diazirine. (Figure 6B) A singlet carbene and molecular nitrogen are formed when an appropriate wavelength of UV is applied on diazirines. Along with forming carbenes, diazirines can also convert to linear diazo compounds via isomerization under UV irradiation, which can in turn produce carbenes or carbocations. Carbenes are highly reactive and rapidly develop covalent connections with X–H (X = C, N, O, or S) through insertion.[78] (Figure 6A)

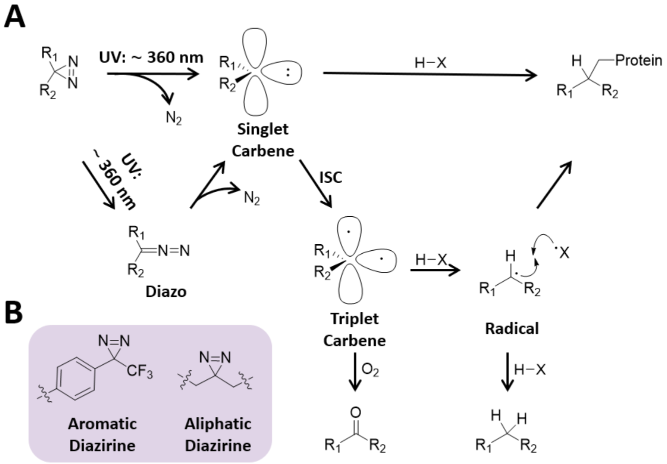


Figure 6. Diazirine (A) Photo-crosslinking mechanism of diazirine. **(B)** Representative scaffold of aryl azides: aromatic diazirine and aliphatic diazirine.

Nevertheless, through intersystem crossing (ISC), the singlet carbenes can convert into triplet carbenes in which the electrons with parallel spins occupy two distinct orbitals. When the triplet carbene interacts with an X–H bond, it subsequently becomes a radical intermediate that either conducts an insertion process like the singlet carbene or extracts a second hydrogen atom from a separate C–H bond and ends in reduction. Alternatively, molecular oxygen may oxidize the triplet carbene and produce the corresponding ketone. [67, 68] (Figure 6A) Overall, the reactive species in diazirine photolabeling are predominantly comprised of the singlet carbene and the diazo species, the proportion of which varies depending on the chemical structure of the diazirines.[79]

Aromatic diazirines are chemically stable in a broad range of circumstances, such as highly acidic, strongly basic, oxidizing and reducing conditions.[80] The phenyl group stabilizes the carbene and prevents rearrangements.[78] Aromatic diazirines allows the less amount of harm to biological systems, because it can be activated under less harmful wavelength at 365 nm. Furthermore, the active carbenes are easily quenched by water and result in reduced non-specific labeling, which avoids the inducing of unexpected reaction in the biological systems. [81] However, the bioactivity of modified compounds might be affected by the fact that aromatic diazirines are relatively bulky. Moreover, cleavage of the photo-labeling may happen via hydrogen fluoride (HF) elimination and enamine hydrolysis, while the most popular trifluoromethylphenyl diazirine is inserted between N–H bonds.[82]

Aliphatic diazirines are becoming more and more popular, despite their moderate crosslinking efficiency resulting from the lower yield of carbenes and the higher tendency for the rearrangement process. The small size of aliphatic diazirines makes them ideal for functionalizing small-molecule drugs, because the drug–target interactions may be disturbed by large photo-crosslinking groups. Therefore, the development of "minimalist" linkers represents a chemically appropriate method for affinity-based protein profiling in vitro and in vivo.[83] [84]

1.5 Target identification

Following AfBP labeling, the target proteins can be identified by different techniques. One option is the use of protein microarrays.[85] [86] Alternatively, proteins labeled by AfBPs can be enriched through the tag (mostly by the use of biotin or a clickable handle, to which a biotin or a bead introduced). The target proteins can be identified using electrophoretic gel-based proteomic analysis[87] [88] and gel-free quantitative proteomics analysis. [89] [90] [91]

1.5.1 SDS electrophoretic gel-based identification

SDS-PAGE and two-dimensional electrophoresis can be used to separate the target proteins after target enrichment. This was mainly done in the past when tandem MS was not as powerful as it is today. To visualize the gel-separated proteins, Coomassie brilliant blue staining or silver staining is generally used. Gel cutting and in-gel digestion are then performed. Tandem mass spectrometry is subsequently used to identify the peptides of enriched proteins. (Figure 7) The targets are eventually identified via database searching of the results from probe treated samples and controls. For instance, the biotin-conjugated ABP DCG-04 was developed to profile the target proteins of the natural product E-64. In experiments using rat kidney lysate, following a 2D electrophoretic gel separation and LC-MS/MS analysis cathepsin H was proven to be the one of target proteins.[56]

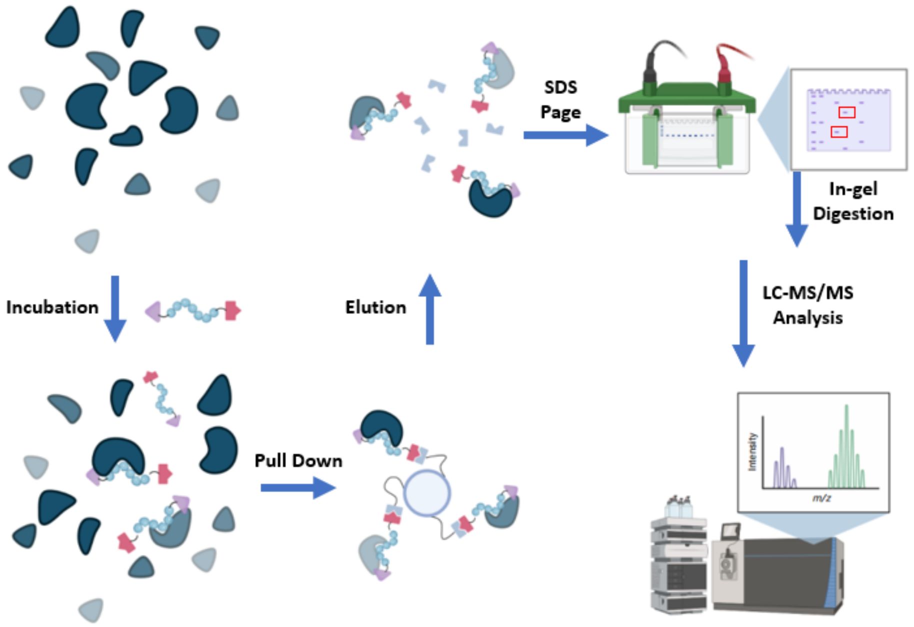


Figure 7. General workflow of SDS electrophoretic gels-based targets identification. ABPs are incubate with the proteome of interest. ABP labeled proteins are enriched by streptavidin bead. Active proteins are eluted from the bead after washing. Active proteins are separated by SDS-page and LC-MS/MS. Separated proteins are cut from the SDS-page and digested within the gel, and eventually analyzed by LC-MS/MS.

Despite the simple protocol, the application of gel-based target identification is limited by the sensitivity of gel staining. Moreover, non-specifically enriched proteins may overlap with the desire bands in the gel and result in false positive identification.

1.5.2 Gel-free quantitative proteomics analysis

Benefiting from the sensitivity of the current mass spectrometers and powerful statistical analysis, gel-free quantitative proteomics analysis overcomes the limitations of less sensitive identification and reduces false positive identification from gel-based identification.[92] [93] [94] [95] Briefly, the enriched proteome is enzymatically digested and subsequently analyzed by LC-MS/MS. The relative abundance of each identified proteins across various samples against controls is then quantified, resulting in the identification of statistically significant enriched targets. Usually, cut-offs are used for samples versus controls of a Log2 abundance ratio $\geqslant$ 1 and a p-value $\leqslant$ 0.05, although more stringent criteria may be applied. Quantitative proteomics analysis includes label-free quantification (LFQ),[92] stable isotope labeling by amino acids in cell culture (SILAC),[90] and peptide N-terminal chemical labeling by

Isobaric Tag for Relative and Absolute Quantitation (iTRAQ) [96] or Tandem Mass Tag (TMT).[97]

In the case of label-free quantification, protein abundance is determined using peptide intensity from the corresponding precursor at the MS^1 level. (Figure 8A) Using label-free quantification allows for a more straightforward experimental design, which saves money and time by circumventing the need for isotope labeling steps.[98] [99] Since there is no limit to the number of samples that are compared, the experimental design is easily adapted to the specific application. In addition, label-free approaches also allow for greater proteome coverage since samples do not need to be mixed. However, label-free quantification has a few drawbacks, including the requirement for very stable LC separation and spray conditions, as well as the need for technical replicates. Furthermore, the requirement to align the runs increases the amount of time required for data processing.[92]

Starting from 2002, SILAC has emerged to be one of the most widely used quantitative proteomics method for the identification of targets.[100] In short, the same type of cells are grown in separate mediums (one containing regular (“light”) amino acids and the other containing isotopically tagged "heavy" amino acids (e.g., labeled with ^{13}C, or ^{15}N)) to produce proteins with differing molecular weights after a few passages of cell culture (Figure 8B). The “light” and “heavy” cells are treated separately with control (e.g. DMSO, competition cocktail) or a chemical probe and mixed in a 1:1 ratio afterwards.[101] After enrichment and on-bead digestion of the samples, the mass-to-charge shift is used to identify peptides from the treated sample or control. By comparing the relative protein abundances of the two groups, the particular target proteins can be easily identified. One of the most important benefits of SILAC is preventing technical error through pooling treated “light” and “heavy” proteome before the enrichment process, which prevents variation during sample preparation and results in unbiased target identification.[100] SILAC has a limitation in that the incorporation of heavy amino acids into cells may induce a slight perturbation in the cellular biochemistry. Because this includes metabolic conversion of arginine to proline in eukaryotes, which generates multiple satellite peaks for all tryptic peptides containing proline, the perturbation is compromising the accuracy of SILAC. An internal correction can be achieved by heavy proline using [$^{15}N_4$]-arginine in combination with normal lysine in the light condition and [$^{13}C_6$,$^{15}N_4$]-arginine in combination with [$^{13}C_6$,$^{15}N_2$]-lysine in the heavy condition. [102]

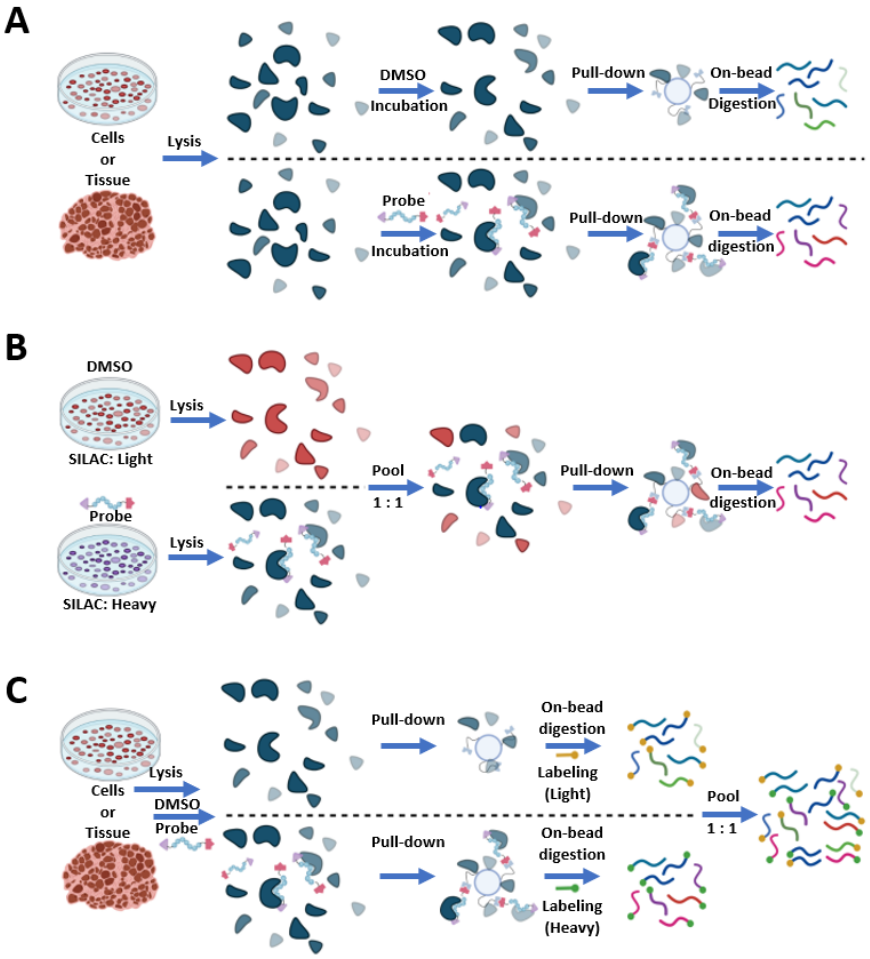


Figure 8. Workflows of gel-free quantitative proteomics analysis. Note: ABPs can be applied either before or after cell lysis. **(A)** Workflow of LFQ. DMSO or ABPs are separately incubated with cells or tissue lysates. ABP-labeled proteins are subsequently enriched utilizing streptavidin beads. Enriched proteins are then digested on-bead after washing. Digested proteins are analyzed by LC-MS/MS. Protein abundance is determined using peptide intensity from the corresponding precursor at the MS^1 level. The relative abundance of each identified protein across various samples against controls is then quantified, resulting in the identification of statistically significant enriched targets. **(B)** Workflow of SILAC. Cells are separately cultured in light media ([^{12}C, ^{14}N]-L-Lysine, [^{12}C, ^{14}N]-L-Arginine) or heavy media (([^{13}C, ^{15}N]-L-Lysine, [^{13}C, ^{15}N]-L-Arginine). DMSO or ABP is separately incubated with cells. After mixing “light” and “heavy” samples 1:1, ABP-labeled proteins are subsequently enriched by streptavidin beads. Enriched proteins are digested on the bead after washing. Digested proteins are analyzed by LC-MS/MS. Protein abundance is determined using peptide intensity from the corresponding precursor at the MS^1 level. The relative abundance of each identified protein across various samples against controls is then quantified, resulting in the identification of statistically significant enriched targets. (**C)** Workflow of N-terminal chemical labeling. DMSO or ABP is separately incubated with cells or tissues lysates. ABP-labeled proteins are subsequently enriched on streptavidin beads. Enriched proteins are digested on the bead after washing. N-termini of DMSO or ABP-treated samples are separately labeled by “light” or “heavy” iTRAQ or TMT reagents. After mixing “light” and “heavy” samples, ABP-labeled proteins are analyzed by LC-MS/MS. The protein abundance ratio is determined using corresponding iTRAQ or TMT reporter ions at the MS^2 level. The relative abundance ratio of each identified protein across various samples against controls is then quantified, resulting in the identification of statistically significant enriched targets.

A disadvantage of SILAC is the limited number of channels to do multiplex labeling. In addition, nondividing cells and human samples cannot be labeled by SILAC, and it is extremely expensive to use on mammalian models.[103] To overcome these shortcomings, chemical labeling with isobaric tandem mass tags (iTRAQ[96] and TMT[97]) was developed. In general, N-termini and lysine side chains of peptides from the digested proteome are labeled by a series of N-hydroxysuccinimide reagents and coupled with a bipartite adduct that contains a mass balance and mass reporter.[96] Bipartites of N-hydroxysuccinimide have the same chemical structure and contain different isotopic mass reporters, while the molecular weight of each is equalized by the isotopic mass balance. The same chemical structure and equal mass of bipartites enable isobaric labeling of peptides. As a result, the same peptides with different bipartites elute at same retention times from the LC and are detected with the same mass from the full MS^1 scan. The subsequent MS^2 event will dissociate the bipartite as well as the peptide bonds during the peptide fragmentation process. The different m/z values of mass reporters can be detected in MS^2 spectra and can be determine the relative abundance of corresponding proteins (Figure 8C).[104] Currently, there are two types of isobaric labeling reagents available on the market, which are iTRAQ® for 2 plex, 4 plex and 8 plex labeling and TMT ™ for 2 plex, 6 plex, 8 plex, 10 plex, 11 plex, 16 plex and 18 plex. When integrated with a semi-automated proteomic sample preparation approach, isobaric labeling has the potential to boost the throughput capabilities of classical ABPP workflow.[105]

1.5.3 protein microarrays

In combination with AfBPs, protein microarrays have also been used to discover the targets of bioactive compounds. [106] In short, AfBP-labeled proteins of interest are immobilized on a high-density array of anti-body and then the interaction can be detected by means of the tag on the AfBPs (biotin, fluorophore, or radioactive isotope). (Figure 9) High-throughput microarrays can be used to identify target and off-target proteins in the entire proteome at the same time. [107] [108] Nevertheless, reporting tags must be added to AfBPs, which might interfere the compound's original activities..

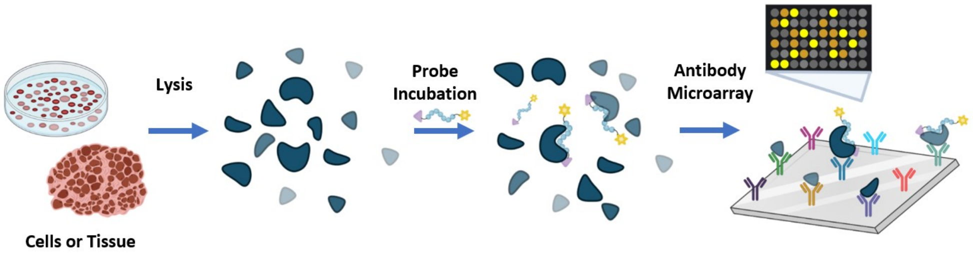


Figure 9. Workflows of protein microarrays. ABPs are incubated with cells or tissue lysates. Subsequent incubation on an antibody microarray immobilizes proteins of interest and results in the identification of targets.

1.6 Cleavable linkers

To identify ABPs or AfBPs targets using tandem-MS, one important step is to enrich the target proteins to distinguish between labeled and unlabeled proteins. The biotin-streptavidin or biotin-avidin interaction is used in the majority of target enrichment strategies because of its extremely high binding affinity (Kd: ~ 10^{-15} mol/L), which permits enrichment of even extremely diluted proteins.[109] To break the biotin-streptavidin binding, severe conditions such as denaturation must be used to release the biotinylated target proteins. As a result, not only the real probe targets, but also nonspecifically bound proteins, endogenously biotinylated proteins, and large amount of streptavidin will be released at the same time. And that will generate lots of background noise in the tandem-MS analysis. Furthermore, most chemical proteomics experiments offer an indirect inference of probe binding (drug targets) through the quantification of constituent peptides from the enriched proteome. The direct detection of probe-modified peptides may give a clear conclusion of probe binding (drug targets). However, the probe-modified peptides are too complex to be detected in mass spectrometry due to multiple reason such as being hard to recover from beads, too bulky to be separated by HPLC, unable to be ionized, or resulting in complex spectra during MS^2 fragmentation .[110]

Those difficulties of releasing and identification have been addressed by introducing cleavable linkers that can be cleaved by chemical agents (acid, base, oxidation or deduction), UV-irradiation or enzymes during the Affinity-based protein profiling (AfBPP) workflow[111] [112]. It allows for selective elution of targets over nonspecifically bound proteins from the enrichment and increasing recovery rates of target proteins by the mild conditions. Moreover, cleavable linkers can also circumvent the poor ionization or unpredictable fragmentation of probe modified (especially biotin-modified) peptides during tandem MS measurements by eliminating the biotin and/or a portion of the AfBPs. Accordingly, it increases the possibility of the direct identification of probe-modified peptides and provide a direct assignment to probe targets, as well as the probe binding site.

A few reviews have comprehensively summarized varieties of cleavable linkers.[111] [112] One of the main goals of my PhD study is the integration of MS cleavable linkers into the AfBPP workflow. Therefore, this section is focused on MS cleavable linkers.

MS-cleavable linkers allow for unambiguous identification of cross-linked peptides in MS^n analysis, as well as for simplified database searches in the following steps. Several cross-linking MS (XL-MS) applications, including the clarification of protein conformations and the mapping of protein–protein interactions, have been developed using these linkers. XL-MS has recently been reviewed in depth elsewhere,[113] [114] [115], including the usage of MS-cleavable linkers. In order to dismantle crosslinks and fragment both peptides, an MS-cleavable linker should be cleaved at a lower fragmentation energy than the peptide backbone. Collisional (CID/HCD) and electron transfer (ETD/ECD) fragmentation have been used to accomplish these fragmentations with different kinds of linkers.[112] CID-cleavable linkers are the most common.

Collision-induced dissociation (CID) is a process in which the selected precursor ions collide with an inert gas such as helium and subsequently produce fragment ions. Upon reaching the fragmentation threshold, the weakest bonds are broken. For regular peptide ions, this results in the formation of characteristic y- and b-type fragment ions.[116]

For peptide bond cleavage, a certain threshold of collision-induced dissociation (CID) energy is necessary. Most experiments use a standard 35% normalized CID collision energy. Some chemical bonds can be broken at lower CID energy, and these represent ideal cleavable linkers. For example, the carbonyl sulfoxide-type linker is a typical CID-cleavable linker because it undergoes a McLafferty-type rearrangement at a lower fragmentation energy than the peptide backbone. The rearrangement proceeds through a five-membered ring transition state (Figure 10A), in which the sulfoxide oxygen abstracts a beta-hydrogen, resulting in the formation of an R-SOH and a cis-1,2 eliminated olefin as fragment ions.[117] [118] There are several different sulfoxide-type cleavable linkers available[112] that have been used in crosslinking MS to get insight into protein structure, conformation, and protein–protein interactions.

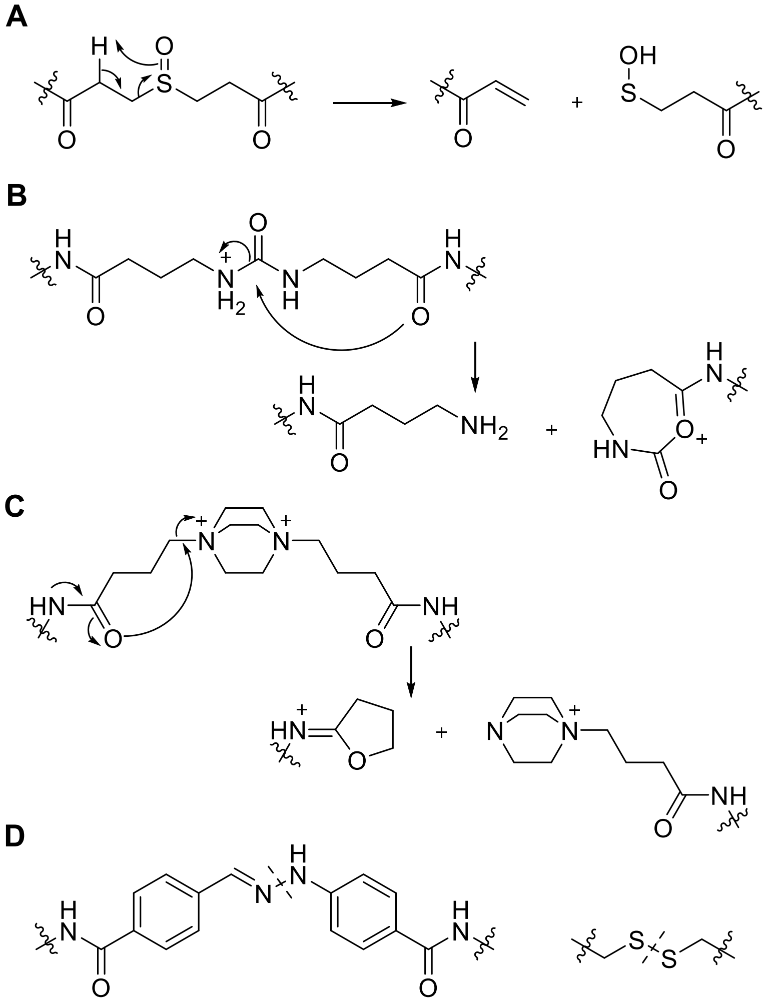


Figure 10. Examples of MS-cleavable linkers. (A) Cleavable mechanism of sulfoxide linkers. **(B)** Cleavage mechanism of urea-based linkers. **(C)** Cleavage mechanism of the DABCO-based

quaternary ammoniumbased linker. **(D)** ETD-cleavable bisaryl hydrazone and disulfide linkages (cleavable bond indicated with a dashed line).

Beside sulfoxide-type linkers, Sinz and colleagues have also developed a urea-based MS-cleavable linker. In this case, one of the amide carbonyl helps fragmentation by generating a seven-membered structure in the linker (Figure 10B).[119] A similar mechanism was proposed for quarternary ammonium linkers containing a central DABCO moiety (1,4-diazabicyclo[2,2,2]octane) (Figure 10C).[120]

Apart from linkers that are MS-cleavable by CID, alternative fragmentation techniques, such as electron transfer dissociation (ETD), have been used. Interestingly, ETD may also be used to cleave various chemically cleavable linkers. Bisaryl hydrazones[121] [122] and disulfides (Figure 10D) are examples of these compounds, although cleavage of the latter has mostly been employed to map cysteine–cysteine connections in proteins.[123] [124]

1.7 The importance of ligand binding sites in structure-based drug design

Resolving the atomic structure of a ligand-protein complex provides detailed structural information on the location of the binding site and the precise nature of the interactions between the functional groups of the drug and the amino acids in the protein that are important for binding. Insight into the binding site of a small molecule or peptide-like drug to its target gives important clues about the mechanism of action. Moreover, it provides possibilities to improve potency or selectivity.

In the 1950s and 1960s, when the first x-ray crystal structures were resolved, these insights were instantly related to physiology and the development of new drugs. For example, after discovering the structure of haemoglobin, Max Perutz and John Kendrew were capable to understand sickle cell disease. [125] [126] Another important milestone was reached by Dorothy Hodgkin, who determined the structure of insulin and began studying insulin redesign, leading to synthetic forms of insulin that were used for treatment of diabetes. [127]

Since 1990s, when the first crystal structures of ligand-protein complexes were submitted to the Protein Data Bank (PDB), [128] lots of ligand-protein complex structures have been resolved using X-ray crystallography, nuclear magnetic resonance (NMR) and – more recently - cryo-electron microscopy (cryo-EM). The development of retroviral protease inhibotors for the human immunodeficiency virus (HIV)-1, approved by The United States Food and Drug Administration (FDA) in 1995, represents one of the most successful cases. [129] The experimental crystal structures of HIV-1 retroviral protease has been resolved after the successful expression and purification of the recombinant protein. [130] [131] [132] [133] These studies confirmed the hypothesis of genomic sequence comparison, which observed a signature sequence (Asp-Thr-Gly) and suggested that the HIV-1 retroviral protease was a pepsin-like aspartic proteases. [134] [135] Nevertheless, the crystal structure disclosed that HIV-1 retroviral protease is a homodimer made up of four short strands, rather than the six long strands present in the pepsins (Figure 11A). [136]

Hydrogen bonds stabilize the active site triad (Asp25-Thr26-Gly27), which is located in a loop identical to the one seen in eukaryotic enzymes. Two Asp25 carboxylate groups from both chains are approximately coplanar and demonstrate tight contact with each other. Hydrogen bonds from the Thr26 main-chain NHs of the opposite loop make the network highly stable, making it very hard to break (Figure 11B). [137] In the enzyme substrate complex, one of the two catalytic aspartic residues undergoes protonation. The other aspartic residue functions as a general base, activating a water molecule, which then attacks the carbonyl carbon of the scissile amide bond, culminating in the production of a tetrahedral geminal diol intermediate (Chapter 1, Figure 11C).

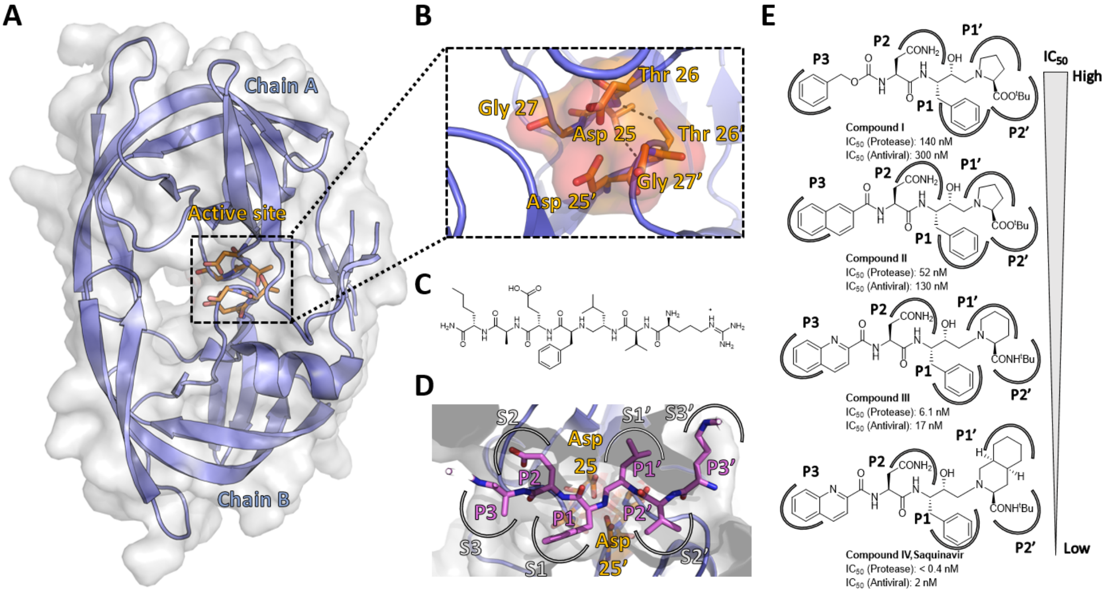


Figure 11. **(A)** Crystal structure of human immunodeficiency virus 1 (HIV-1) protease. PDB ID: 6o48. **(B)** Active site of HIV-1 protease. Asp(25)-Thr(26)-Gly(27), Asp(25')-Thr(26')-Gly(27'). **(C)** Chemical structure of substrate analog CA-p2 (H-Arg-Val-Unk-Phe-Glu-Ala-Nle-NH_2). **(D)** Substrate analog CA-p2 in the active site. S1–S3 and S1′–S3′: three subsites of binding pocket. P1–P3 and P1′–P3′: Side chains of substrates or inhibitors. **(E)** Discovery of Saquinavir (Ro 31-8959): Structure-Activity Relationship (SAR). [138]

Most HIV-1 retroviral protease inhibitors have a hydroxyl group in their structure as a mimic of the tetrahedral intermediate of a substrate. This hydroxyl is in close enough proximity to form hydrogen bonds with at least one of the carboxylate oxygens of each aspartate and inhibit the hydrolytic activity of the enzyme. [139] Within the structure of the complex of HIV retroviral protease and its substrate analog CA-p2 (H-Arg-Val-Unk-Phe-Glu-Ala-Nle-NH2), it is possible to resolve a number of different subsites that are able to accept side chains of the inhibitors. (Figure 11A-11C) There are three subsites (S1–S3 and S1′–S3′) on either side of the non-scissile link. In addition to the aspartates found in the active site, the protease side chains that make up the pockets S1 and S1' are predominantly composed of hydrophobic residues (Figure 11D). Except for the inhibitors containing statine and glycine (Pepstatin analogs, Chapter 2, 2.1), which do not have any groups occupying the protease subsite S1′, almost all of the described inhibitors include hydrophobic moieties at P1 and P1′. In spite of the fact that the S2 and S2' pockets are hydrophobic, hydrophilic and hydrophobic residues are equally capable of filling these positions. When looking at the various inhibitors, the hydrophobic side chains P2 and P2' were found to be oriented in a variety of different ways, which allowed them to create contacts with a wide variety of distinct groups in the enzyme-binding pocket (Figure 11D). [140]

Roche drug Saquinavir (Ro 31-8959, Saquinavir) was the first FDA approved HIV-1 retroviral protease inhibitor. A crystal structure of Saquinavir with the HIV-1 retroviral protease revealed, as was hypothesized, that this inhibitor binds in an extended conformation and favored R stereochemistry at the carbon containing the hydroxyl group. [141] [142] By replacing the proline in the P1' subsite of oligopeptide inhibitors, (S, S, S)-decahydro-isoquinoline-3-carbonyl (DIQ), the resultant molecule, which was eventually given the name Ro 31-8959, reached to a Ki value of 0.12 at pH 5.5 effective against HIV-1 retroviral protease (Figure 11E). [138] [143]

Because of the successful crystallography-driven development of HIV protease inhibitors, structure-based drug design (SBDD) has become an essential paradigm of drug discovery and development. Along with the development of computer-aided drug design (CADD), SBDD becomes an iterative procedure that progresses through several cycles to arrive at a drug candidate that has been refined and is ready for clinical studies. [144] The identification of ligand binding sites is very essential for SBDD. The initial stage of the process involves cloning, purifying, and determining the structure of the target protein as well as the potential binding sites. Compounds or fragments of chemicals from a database are positioned into binding sites of the structure using computer algorithms. These compounds are given a score and a ranking based on the steric and electrostatic interactions that they have with the target site, and then the biochemical experiments are performed on the compounds that received the highest scores. During the second cycle, the further optimization of the lead compound is conducted according to the structure determination of the target in complex with a potential lead (at least micromolar inhibition in vitro) from the first cycle. After multiple iterations of the drug design process, the improved molecules often exhibit a significant increase in binding affinity and target specificity. [145] (Figure 12)

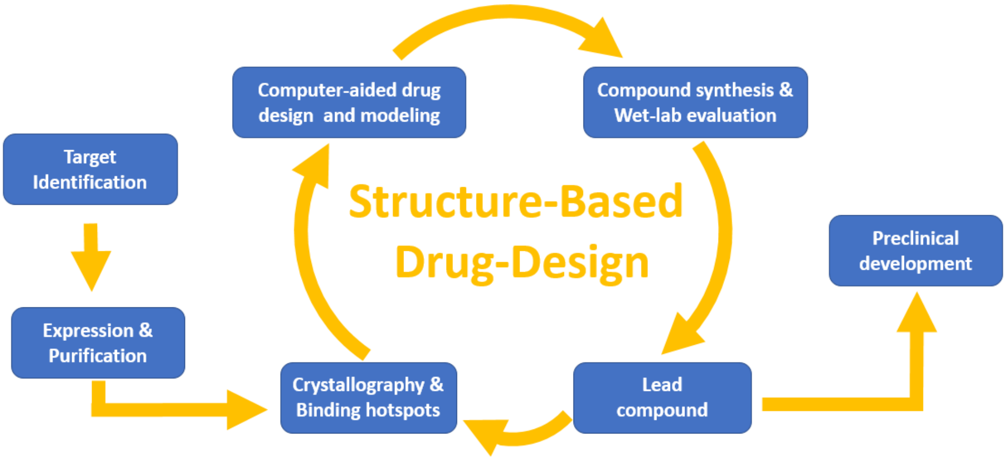

Target Identification
Expression & Purification
Computer-aided drug design and modeling
Compound synthesis & Wet-lab evaluation
Structure-Based Drug-Design
Preclinical development
Crystallography & Binding hotspots
Lead compound

Figure 12. Workflow of structure-based drug design.

Nowadays, various drugs, such as thymidylate synthase inhibitor raltitrexed, [145] diabetic neuropathy drug Epalrestat, [146] and antibiotic norfloxacin, [147] have been developed via the structure-based drug design (SBDD). The knowledge of binding hotspots of drug candidates plays an important role in SBDD and can facilitate drug discovery.

1.8 Mass-spectrometry-based structural biology and structural proteomics

For decades, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy have been used to study protein-ligand and protein-protein complexes. More recently, technological improvements have made electron microscopy (EM) more powerful for obtaining structures at 50-100 ug scale with better resolution. [148] [149] Nevertheless, EM takes snapshots of each protein complex but homogenizes data points in the output. With highly conformational and/or chemical heterogeneity, EM is limited to provide less precise read-out. In addition, in vitro investigations raise the issue of whether they can produce knowledge that is relevant to the structure and function of living organisms. Due to highly conformational and/or chemical heterogeneity, EM is also limited.

In order to investigate the conformational features of proteins on a proteomic scale, a new toolbox consisting of methods based on mass spectrometry (MS) has been developed over the course of the previous decade. [150] These methods include various combinations of protease digestion, chemical modification, protein precipitation, chemical denaturation, and thermal denaturation methodologies with quantitative mass spectrometry-based proteomics read-out. With a whole-cell approach, this toolbox of structural proteomics methods has made it possible to investigate the conformational characteristics of proteins as well as the interactions between proteins and ligands on a scale relevant to proteomics. [151] Crosslinking, photoaffinity labeling, limited proteolysis, hydroxyl-radical footprinting, and hydrogen/deuterium exchange, which comprise the main components of structural proteomics, can be used to characterize protein structures and protein–protein interactions. Indirect information about the inside of the protein and inter-protein interaction areas can be investigate by footprint chemical modification, limited proteolysis, and deuterium exchange. While chemical crosslinking is able to offer knowledge on the residue distances derived from intra- and inter-protein crosslinks, photoaffinity labeling is able to determine the places where ligands bind with proteins. [152]

1.8.1 Hydroxyl radical protein footprinting (HRPF)

HRPF is a technique that makes use of hydroxyl radicals derived from Fenton Chemistry in order to oxidatively modify the side chains of 14 out of the 20 natural amino acids in the protein (Figure 13A). [153] [154] Using high-resolution mass spectrometry coupled to liquid chromatography (LC-MS/MS), these labels are analyzed to determine the interaction sites and areas of conformational change. Because of its small size and strong reactivity, the hydroxyl radical is an effective probe that can identify of solvents exposed surface of biological macromolecules (Figure 13B). Since the early use of HRPF by Tulius and Dombrowski to discover DNA–protein interactions by using Fenton chemistry with gel electrophoresis as the analytical output, [155] the technology has become a powerful solution for investigation of the structure, conformational changes, and binding events of biological macromolecules.

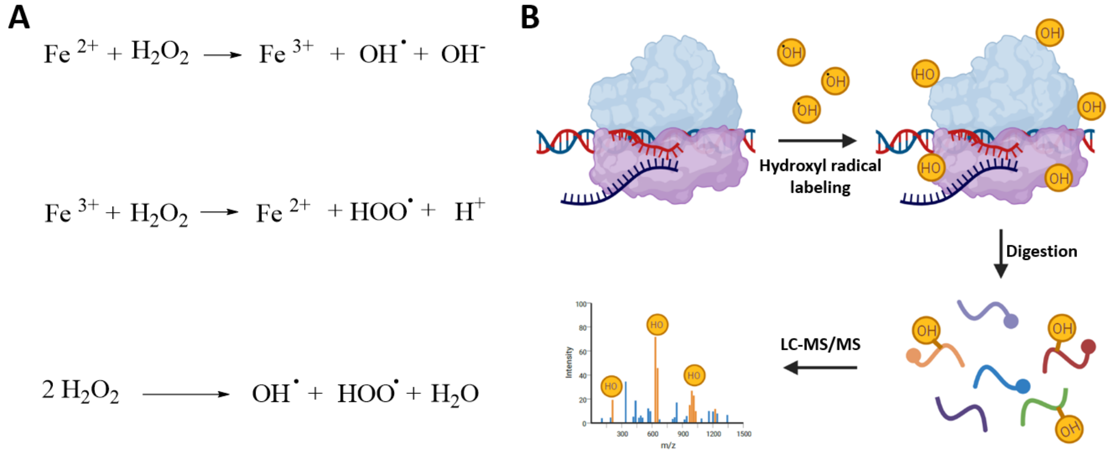

A
$Fe^{2+} + H_2O_2 \longrightarrow Fe^{3+} + OH^{\bullet} + OH^-$
$Fe^{3+} + H_2O_2 \longrightarrow Fe^{2+} + HOO^{\bullet} + H^+$
$2 H_2O_2 \longrightarrow OH^{\bullet} + HOO^{\bullet} + H_2O$
B
OH
HO
Hydroxyl radical labeling
Digestion
LC-MS/MS
Intensity
m/z
100
80
60
40
20
0
300
600
900
1200
1500

Figure 13. **(A)** Fenton Chemistry. **(B)** Hydroxyl radical protein footprinting (HRPF) workflow

1.8.2 Hydrogen-Deuterium Exchange-Mass Spectrometry (HDX-MS)

HDX-MS provides insight into structural variances and changes of proteins and protein complexes by detecting hydrogen-deuterium variations of solvent-accessible amide hydrogens of proteins. [156] To this end, a protein of interest is diluted in a deuterated buffer, and the deuterium from the solvent is allowed to exchange at neutral pH with the backbone amide hydrogens for a predetermined amount of time. The process is stopped when a buffer with a low pH is added, and after that, an acid-stable protease (usually pepsin) is used to digest the protein. Last but not the least, the amount of deuterium that is taken up by each peptide is determined using liquid chromatography mass spectrometry. This enables the detection of parts of the protein sequence that are more or less solvent accessible (Figure 14).

HDX-MS can be utilized to investigate a wide variety of facets related to the structure and dynamics of proteins. [157] Protein conformational states, such as those caused by contact with ligands or other proteins, as well as folding and aggregation, can be compared and investigated in a variety of contexts. In addition to sequence confirmation and the characterization of artifactual and post-translational modifications, the analysis of biopharmaceuticals is a notable application that makes use of individual proteins. Within this context, HDX contributes an additional level of information that is derived from MS. [158] Protein–ligand interactions, including the binding of drugs, and protein–protein interactions, such the formation of antibody–antigen complexes, are also often investigated. [158] [157] As a direct result of this, the methodology is widely recognized and well-established in the pharmaceutical industry.

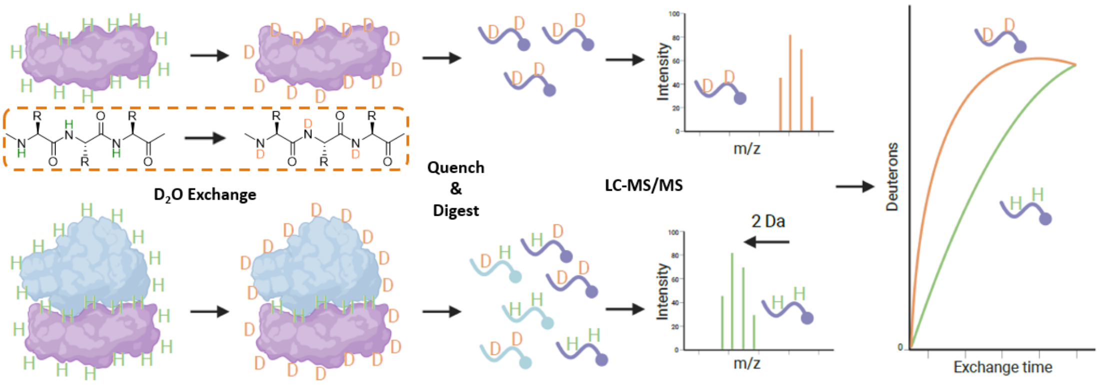


Figure 14. Hydrogen-Deuterium Exchange-Mass Spectrometry (HDX-MS) workflow

1.8.3 Limited proteolysis-based Mass Spectrometry

Drug Affinity Responsive Target Stability (DARTS) was developed to identify protein targets of small molecules based on the premise that drug binding induces conformational changes in proteins, which either increase or decrease the susceptibility of the protein to proteolytic digestion with a nonspecific protease (for example, thermolysin or proteinase K). [159] The utilization of gel-based or LC-MS based proteomics enables the identification of proteins that exhibit distinct cleavage patterns in the presence and absence of ligand (Figure 15).

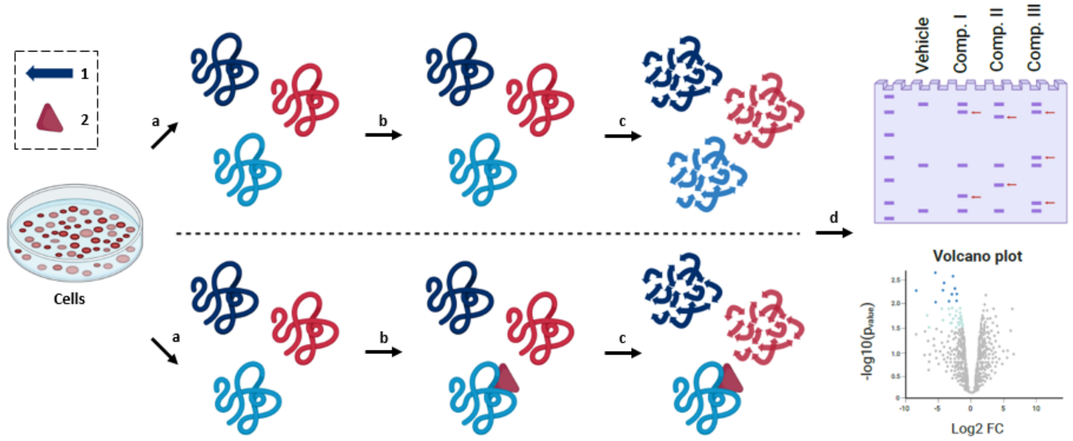


Figure 15. Drug Affinity Responsive Target Stability (DARTS) workflow. **1.** Cleavage site of thermolysin or proteinase K. **2.** Active compound. **a.** Lysis. **b.** Treatment: vehicle or active compound. **c.** Proteolytic digestion. **d.** SDS page read-out: remaining bands can be corresponding targets. Quantitative proteomics read-out: negative fold-changed proteins can be corresponding targets.

Similar as DARTS, Limited proteolysis-coupled mass spectrometry (LiP-MS) was developed to identify changes in protein structure on a proteome-wide scale directly in complex biological systems. After the treatment of interest, proteome extracts are first digested for a short time with a nonspecific protease under native conditions, then further completely digested using the specific protease trypsin under denaturing conditions. After that, structure-dependent proteolytic patterns of the proteome extract are measured directly using a proteomics approach that includes shotgun or targeted MS and label-free quantification (Figure 16). [160] LiP-MS can be used to analyze protein aggregation in real time in biological samples, find therapeutic targets, and identify protein structure states related with diseases. [161] [162] [163] Furthermore, this method may also be used to determine which parts of the protein undergo a structural change or are influenced by a binding event. [161]

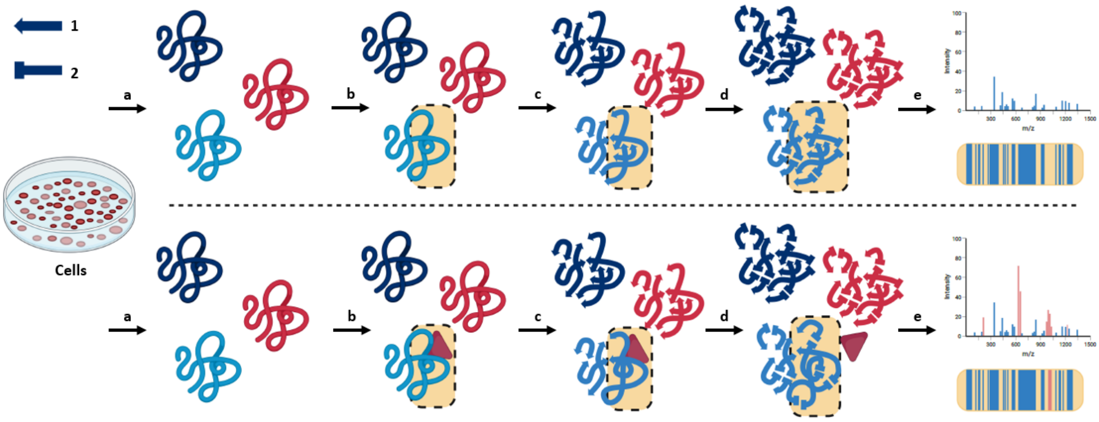


Figure 16. Limited proteolysis-coupled mass spectrometry (LiP-MS) workflow. **1.** Cleavage site of proteinase K. **2.** Cleavage site of trypsin. **3.** Active compound. **a.** Lysis. **b.** Treatment: vehicle or active compound. **c.** Short time proteinase K digestion under native conditions. **d.** Denaturation and trypsin digestion. **e.** LC-MS/MS analysis.

1.8.4 Chemical crosslinking mass spectrometry (XL-MS)

Chemical crosslinking mass spectrometry (XL-MS), provides insight into the three-dimensional structure of proteins and protein complexes by detecting residue pairs that are located in close spatial proximity to one another. [164] [165] In a typical XL-MS approach, the protein assembly of interest is first incubated with an appropriate crosslinking reagent. This results in the formation of covalent bonds between the proximal residues that are targeted by the reactive groups of the crosslinker. A subsequence LC-MS/MS analysis of crosslinking peptides after digestion can give a read-out of protein-protein interaction (PPI) and structural mapping of proteins (Figure 17). The majority of the commercially available crosslinkers consist of two *N*-hydroxysuccinimide- (NHS-) ester functional groups that are joined by a spacer arm. These crosslinkers predominantly react with Lys side-chains, and to a lesser degree, those of Ser, Thr, and Tyr. [166] [167]

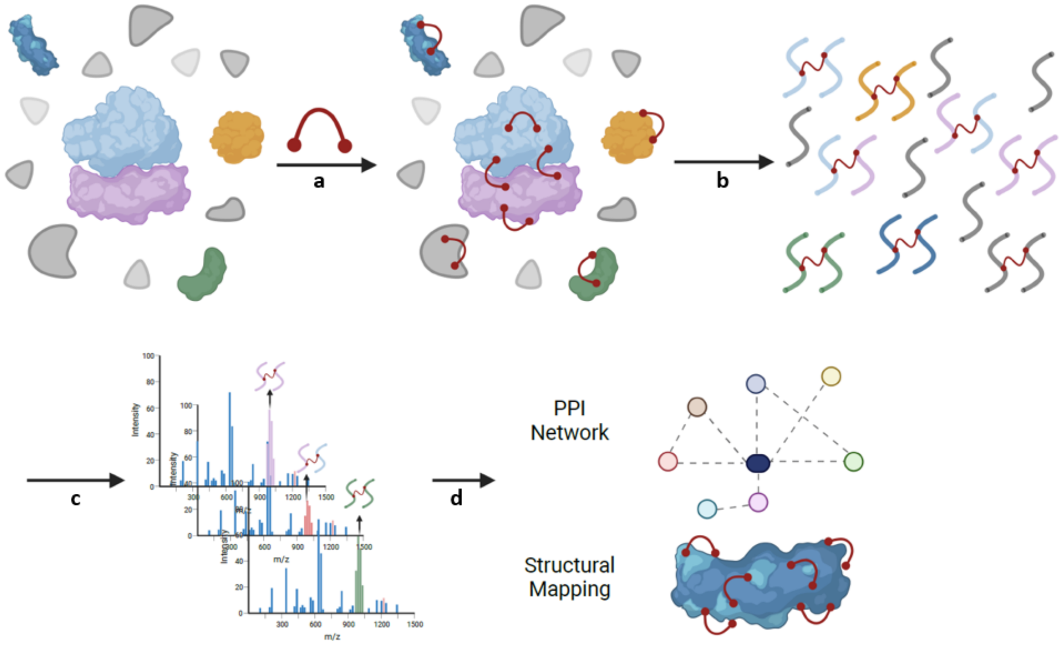
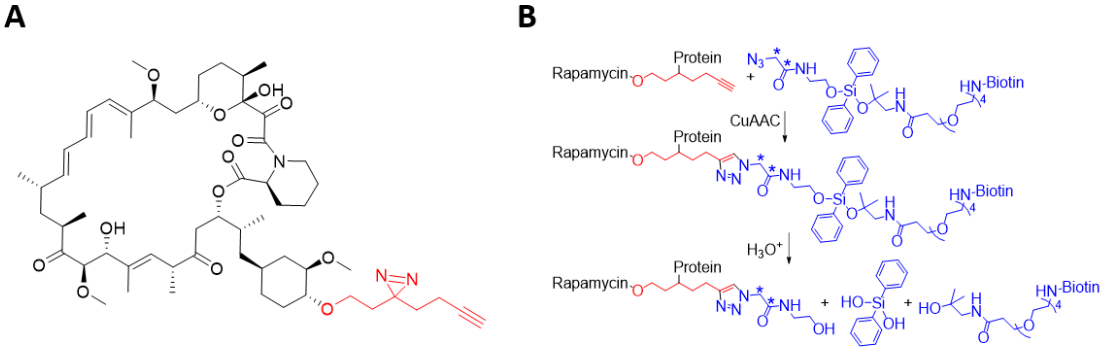


Figure 17. Chemical crosslinking mass spectrometry (XL-MS) workflow. **a.** Crosslinking reagent incubation. **b.** Digestion. **c.** LC-MS/MS analysis. **d.** Data analysis.

1.9 Photoaffinity labeling (PAL) enables precise binding hotspot mapping

Different from indirect mapping approaches like HRPF, HDX and LiP, PAL coupled with LC-MS/MS enables a direct identification of interaction site via installing a photo-crosslinking group on the ligand and labeling the interaction residues covalently. The follow-up LC-MS/MS analysis can give a precise readout of the binding hotspots.

PAL was used by Flaxman and colleagues in order to determine the location of the binding site of the macrocyclic lactone rapamycin (Figure 18). The protein-protein interaction between Tacrolimus (FK506) binding protein 12 (FKBP12) and the FKBP-rapamycin binding (FRB) domain of Mammalian target of rapamycin (mTOR) was previously recognized to be stabilized by rapamycin. By introduce a diazirine handle on C40 of rapamycin, the authors identified alterations to residues 75-110 in FKBP12 and 10-22 in FRB. D79 of FKBP12 and E18 of the FRB domain were the predicted residues of interaction, and their mutation to alanine significantly reduced staining, suggesting that these residues were important sites of interaction of the rapamycin probe. [168]



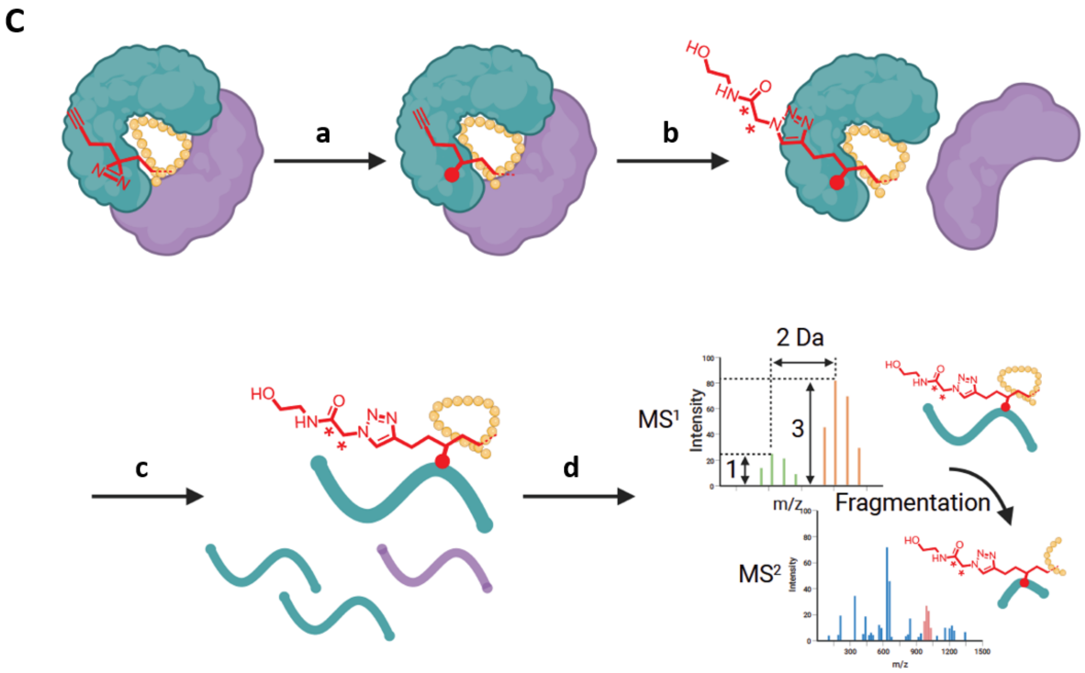


Figure 18. **(A)** Chemical structure of Photo-Rapamycin. **(B)** Click and cleave: isotopic labeling via copper (I) -catalyzed azide alkyne cycloaddition (CuAAC) and cleave biotin under acid condition. *isotope ratio: $^{13}C_2$:$^{12}C_2$ = 3:1. **(C)** Workflow: binding hotspot mapping via PAL. **a.** UV irradiation. **b.** Click and cleave. **c.** Digestion. **d.** LC-MS/MS analysis.

1.9.1 Enrichment of photo-crosslinked peptides

Direct detection of photo-crosslinked peptides by MS is difficult because of the unknown photo-crosslinking efficiency with the target protein, as well as the variable ionization efficiency of the small molecule-conjugated digested peptides. Due to both of these effects, the MS abundance of a photo-crosslinked peptide is lower compared to that of unmodified peptides. Typical shotgun proteomics procedures begin with the collection of a full scan mass spectra (MS^1), then proceed to make an iterative selection of the most abundant species from MS^1 for tandem mass spectrometry sequencing (MS^2) This procedure is known as data-dependent acquisition (DDA). [169] The depth of MS^2 is limited by the instrument scanning speed and the complexity of analytes. Therefore, the assignment of photo-crosslinking sites might be hindered if low-abundance species are not properly chosen for tandem MS sequencing.

To facilitate detection and enrichment of the photo-crosslinked peptides, the small molecule can be functionalized with a reporter group that serves as a handle for further enrichment and purification. Enrichment handles are often connected following chemical conjugation of the photo-crosslinked peptides using biocompatible click chemistry. This prevents that a bulky tag may affect the probe's inherent interactions. [170] As clickable tags, biotin is most often used, as biotinylated proteins can be efficiently enriched by immobilized streptavidin, and in this particular case will allow a better identification of photo-crosslinked peptides.

1.9.2 Application of cleavable linker

The process of mapping particular interaction sites using PAL may look simple at first glance: all that is required is the enrichment and MS/MS analysis of tryptic peptides that have been crosslinked. Nevertheless, only a few selected PAL probes have been shown to be compatible with interaction site identification. [171] These kinds of investigations are often challenging due to the poor efficiency of photo-crosslinking, the unexpected fragmentation of crosslinked peptides, and the increased hydrophobicity of tagged peptides, all of which make enrichment and identification even more difficult.

In order to increase PAL enrichment and to recover the probe modified peptides, cleavable linkers have been used frequently. Bogyo and co-workers were the ones who first developed diazobenzene as a cleavable functional group for MS-based proteome profiling. [172] Sodium dithionite is used to reductively cleave the diazobenzene, allowing the modified peptide or protein to be recovered from bead and removing sodium dithionite salts itself via desalting. The diazobenzene linker was used by Weerapana and co-workers in order to perform site-specific identification of reactive cysteines throughout the whole proteome. [173]

Woo and co-workers developed a technique called small molecule interactome mapping by photo-Affinity Labeling (SIM-PAL) using an isotopically coded biotin picolyl azide incorporating an acid-cleavable linker to help tackle the challenge of assigning binding sites. [174] Following functionalization of a small molecule with diazirine and alkyne, the SIM-PAL process includes cellular treatment, photo-crosslinking, enrichment using a cleavable biotin azide, on-bead digestion, and recovery of photo-crosslinked peptides before analysis by LC-MS/MS. The use of the multifunctional, acid-cleavable, isotopically coded biotin picolyl azide was essential to the success of this approach. This compound makes use of a picolyl group to chelate Cu(I) in close proximity to the reaction site of azide, which speeds up the kinetics of the CuAAC reaction. [175] Moreover, the acid-cleavable diphenyl silane allows for the straightforward recovery of conjugated peptides from beads at acidic conditions that are compatible with further LC-MS analysis. The isotope code, which consists of two carbon atoms implanted with a $^{13}C_2$:$^{12}C_2$ ratio of 3:1, creates a unique pattern in the full-scan MS (MS^1), lending more credence to spectral matches of photo-crosslinked peptides (Figure 19). [174]

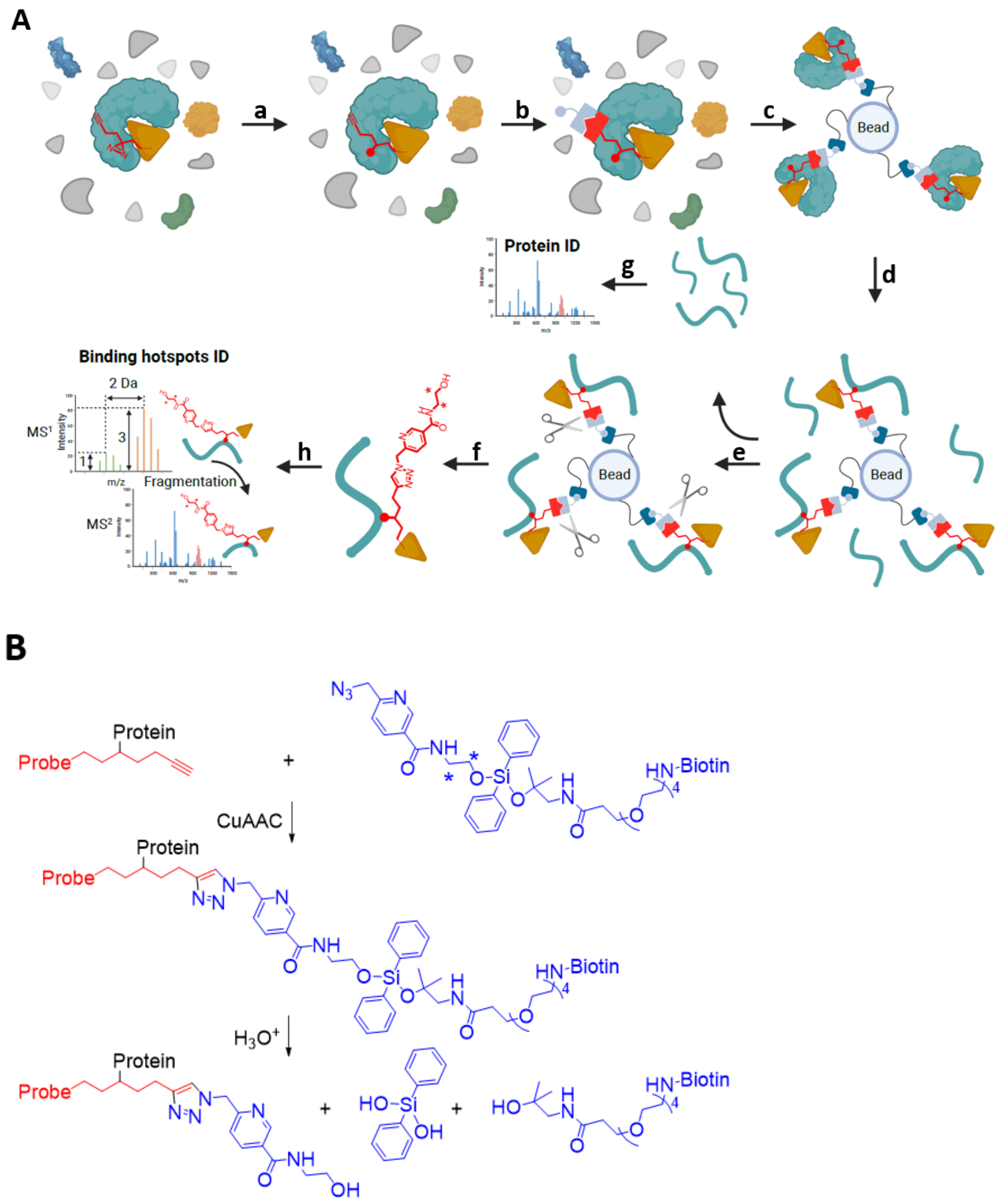


Figure 19. **(A)** Small molecule interactome mapping by Photo-Affinity Labeling (SIM-PAL) workflow. **a.** UV irradiation. **b.** Biotin labeling via CuAAC. **c.** Pull-down using streptavidin beads. **d.** On-bead digestion. **e.** Wash. **f.** Releasement of photo-crosslinked peptides. **g.** Protein identification (ID) via LC-MS/MS analysis. h. Binding hotspot ID via LC-MS/MS analysis. **(B)** Click and cleavage of multifunctional, acid-cleavable, isotopically coded biotin picolyl azide. *isotope ratio: $^{13}C_2$:$^{12}C_2$ = 3:1.

Only very recently, Weiss and co-workers developed ligand-footprinting mass spectrometry (LiF-MS) to reduce the unexpected fragmentation of crosslinked peptides. This was done by introducing an acid cleavable sulfamide linker next to the photo-crosslinking moiety (diazirine) of the PAL probes. [176] After UV irradiation and enrichment, the further acid release can yield a “mini-tag” which is a 72-Da butanol modification on crosslinking amino acid residues. Due to the removing of the majority of the probe, the less complex fragments of the crosslinking peptides allow a better deconvolution of the binding sites. (Figure 20)

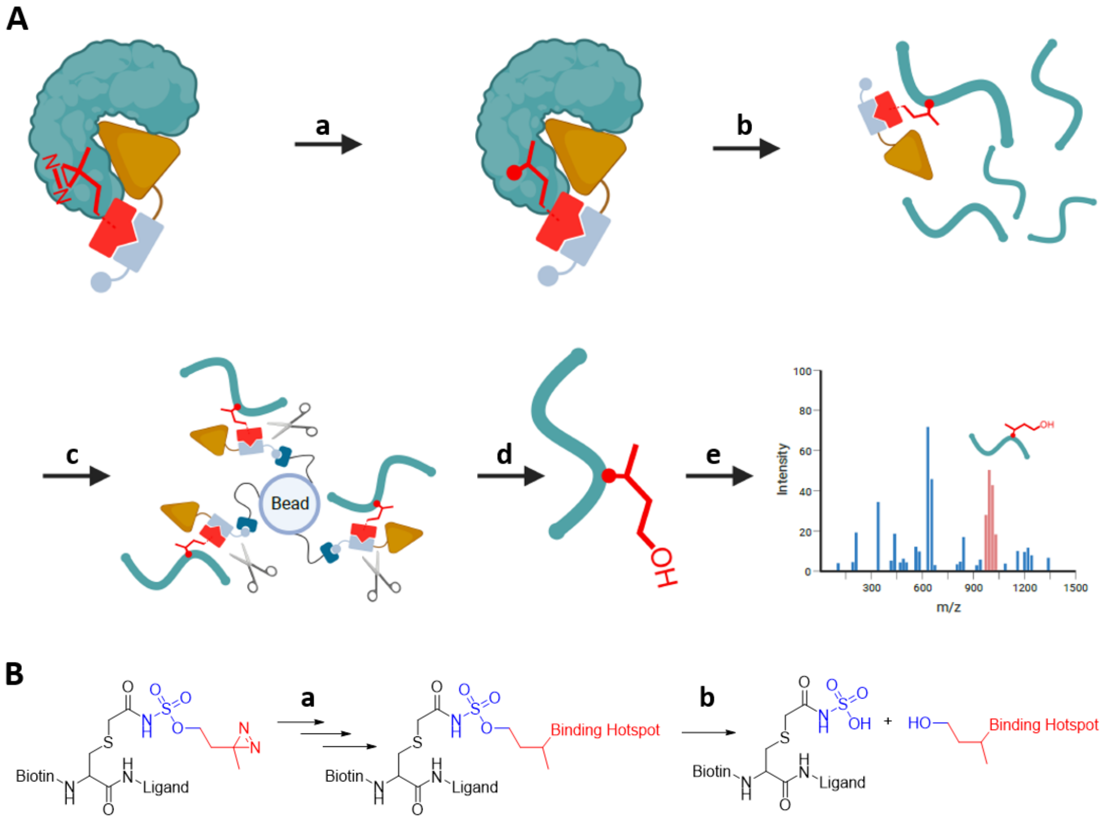


Figure 20. **(A)** Ligand-footprinting mass spectrometry (LiF-MS) workflow. **a.** UV irradiation. **b.** Digestion. **c.** Pull-down using streptavidin bead. **d.** Releasement of photo-crosslinked peptides under acid condition. **e.** Binding hotspot ID via LC-MS/MS analysis. **(B)** Cleavage of acid-cleavable linker. **a.** UV irradiation, Digestion, Pull-down. **b.** Releasement under acid condition (pH 1-2).

2 Research objectives

2.1 Expand the probe set of aspartic proteases

ABPs are naturally restricted to serine, cysteine, and threonine proteases, as other kinds of proteases employ an activated water molecule in their action, which would result in hydrolysis of the warhead. Because of this enzymatic mechanism, aspartic proteases need a different probe design. Currently, there is a lack of chemical tools for aspartic proteases research. Therefore, we would like expand the probe set of aspartic proteases by introducing a photoreactive group to aspartic protease inhibitors, which will result in covalent binding.

2.2 Profile the targets of Pepstatin A

Pepstatin A is a natural inhibitor of aspartic proteases.[177] It has a broad spectrum of inhibition among aspartic proteases (such as pepsin, cathepsins D and E) and has been widely used in life science research. It is therefore ideal as a starting point for the development of general aspartic protease AfBPs. By conducting ABPP with Pepstatin-based probes, we would like to go further to profile the targets of Pepstatin A.

2.3 Enable precise binding hotspots mapping via mass cleavable affinity-based probes

Photo-affinity labeling (PAL) combined with tandem mass spectrometry (MS^n) can reveal noncovalent interactions between small molecule drugs and protein in biological environments. However, the direct detection of the ‘hotspots’ of the photo-crosslinked binding sites by MS is challenging because of the unknown photo-crosslinking efficiency with the target protein, as well as the unexpected fragmentation of small molecule drugs, especially when these are small peptides. Advances in mass spectrometry technology have extended the range of fragmentation methods, including collision-induced dissociation (CID), higher-energy C-trap dissociation (HCD), electron transfer dissociation (ETD), electron-transfer/higher-energy collision dissociation (EThcD), and ultraviolet photodissociation (UVPD), which have substantially facilitated the identification of cross-linked peptides. [178] [179]

By introducing MS-cleavable and photoreactive sulfoxide diazirine (SODA) building blocks to peptide-like probes, we would like to utilize the MS-cleavage of sulfoxide in the MS^2 event to generate a probe-derived reporter ion and a minimal fragment on the modified peptide. Following a subsequent MS^3 fragmentation event and MS data analysis, we would like to finally achieve the unbiased identification of the modification sites of PAL probes and map the binding hotspots of peptide-like bio-active molecules.

3 Materials and methods

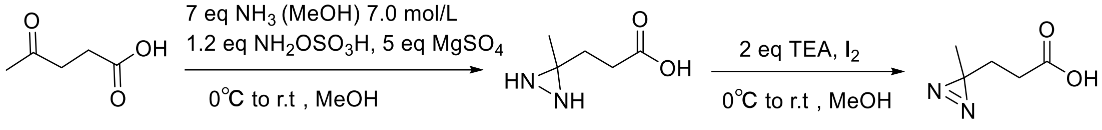
3.1 Pepstatin-based probes for photoaffinity labeling of aspartic proteases

3.1.1 General methods

Unless otherwise noted, all reagents were purchased from commercial suppliers and used without further purification. TLC analysis was performed on pre-coated ALUGRAM SIL G plates (Carl Roth) with detection by a handheld UV lamp (254 nm) and subsequent staining with cerium ammonium molybdate solution followed by heating. Low resolution LC-MS analysis was performed on an Dionex Ultimate 3000 (Thermo Scientific) coupled to MSQ Plus Mass Spectrometer (Thermo Scientific) with a Waters xBridge C18 (2.1 x 150 mm, 5 µm) column with a linear gradient of acetonitrile in water with 0.1% trifluoroacetic acid. High resolution LC-MS analysis was performed on a Dionex Ultimate 3000 (Thermo Scientific) coupled to Velos Pro Mass Spectrometer (Thermo Scientific) with 75µm × 20 mm C18 pre column using 0.1% TFA at a flowrate of 12 µl/min. Following sample separation was accomplished on a reversed phase column (Acclaim C18 PepMap100, 75 µm × 150 mm) at 50 °C using a linear gradient: 3%-58% solvent B (84% ACN with 0.1% FA). Preparative HPLC purification was performed on a using a Thermo Scientific BioBasic-18 C18 column (2 × 15 cm, 5 um). Purifications were performed at room temperature and compounds. were eluted with increasing concentration of acetonitrile (solvent A: 0.1% TFA in water, solvent B: 0.1% TFA in 84% acetonitrile). NMR spectra were recorded on a Bruker UltraShield 600MHz NMR Spectrometer. Silica column chromatography was performed using 230-400 mesh silica (Kieselgel 60).

3.1.2 Synthesis of (3-(3-methyl-3H-diazirin-3-yl)propanoic acid) (Compound 3)

Levulinic acid (3 mmol, Sigma-Aldrich) and $MgSO_4$ (5 equiv, Carl Roth) were cooled in an ice bath with stirring under argon. NH_3 in MeOH (7 N, 7 equiv, Alfa Aesar) was added dropwise. The mixture was slowly warmed up to RT and allowed to stir for 5 h. The mixture was then cooled to -78 °C. A solution of hydroxylamine-O-sulfonic acid (1.2 equiv, Tokyo Chemical Industry) in MeOH (0.5 mL, Carl Roth) was added. The mixture was slowly warmed up to RT and allowed to stir overnight. After centrifugation, the liquid phase was collected and volatiles were evaporated under reduced pressure. The residue was dissolved in 10 mL MeOH and cooled in an ice bath. TEA (2 equiv, Acros Organics) was slowly added. Next, I_2 (Sigma-Aldrich) was slowly added until the solution took the color of I_2 and didn't fade away after 1h. The reaction was quenched by 5% HCl and extracted by ethyl acetate (Carl Roth). The photo-crosslinking building block **3** was purified by silica gel chromatography to give the title compound (150 mg; yield: 30.1%). ^{1}H NMR (600 MHz, CDCl3): δ 2.24 (t, J = 7.7 Hz, 2H), 1.73 (t, J = 7.7 Hz, 2H), 1.05 (s, 3H); ^{13}C NMR (151 MHz, CDCl3): δ 178.56, 29.34, 28.50, 25.05, 19.69.


7 eq NH_3 (MeOH) 7.0 mol/L
1.2 eq NH_2OSO_3H, 5 eq $MgSO_4$
0°C to r.t , MeOH
2 eq TEA, I_2
0°C to r.t , MeOH
O
OH
HN
NH
N
N

3

Scheme 1. Synthesis of compound **3** (3-(3-methyl-3H-diazirin-3-yl)propanoic acid)

3.1.3 Solid Synthesis of probes 4-7

The BAL resin (1 equiv, 450 μmol, 528 mg, Iris Biotech) was added to a solid-phase cartridge in NMP (Biosolve) and gently shaken for 20 minutes. After draining NMP, propargylamine hydrochloride (10 equiv., 4.75 mmol, 436 mg, Sigma-Aldrich) and AcOH (1%, 60 uL, Carl Roth) in 6 ml NMP (0.8 mol/mL) was added to the resin and shaken for 20 minutes at room temperature. Sodium cyanoborohydride ($NaBH_3CN$, 10 equiv, 4.75 mmol, 300 mg) was added to the resin and the reaction was shaken for 15-20 hours at room temperature. The resin was then washed with DMF (3x, Biosolve), DCM (3x, Carl Roth) and MeOH (3x, Carl Roth) and dried. The resin was stored in an argon atmosphere at -20 °C for further use.

N-Fmoc-protected amino acid (3 equiv, Fmoc-(3S,4S)-Sta-OH from Iris Biotech, Fmoc-L-Ala-OH.H_2O from PolyPeptide, Fmoc-L-Val-OH from CREOSALUS), HBTU (3 equiv, CREOSALUS) and DIEA (6 equiv, Carl Roth) were dissolved in DMF (0.2 M final concentration). For elongation of the resin, the solution of the activated amino acid was shaken with the resin at room temperature for 30 min and washed three times with DMF and DCM. The N-terminal Fmoc group was removed by incubating the resin with 20% Piperidine (Biosolve) in DMF (15 min). Next, the resin was washed three times with DMF and DCM, ending the elongation cycle. For each subsequent step of the solid-phase peptide synthesis, the same deprotection and coupling reactions were used. The last N-terminal was blocked by adding carboxylic acid (3 equiv, isovaleric acid from Sigma-Aldrich or homemade 3-(3-methyl-3H-diazirin-3-yl)propanoic acid), HBTU (3 equiv) and DIEA (6 equiv). Coupling reactions were monitored by the Kaiser test for primary amines. The product was cleaved off the resin with a TFA/TIPS/H_2O mixture (v/v/v, 95:2.5:2.5, TFA from Fluorochem, TIPS from Sigma-Aldrich), the liquid was collected and the volatiles removed under reduced pressure. The final product was purified by reversed-phase HPLC. Fractions containing product were pooled and lyophilized. HRMS: **4** $[M+H]^+$ 749.4913 ($C_{37}H_{65}N_8O_8$, theoretical mass: 749.4925, mass difference: -0.0012 Da, -1.6 ppm), **5** $[M+H]^+$ 749.4918 ($C_{37}H_{65}N_8O_8$, theoretical mass: 749.4925, mass difference: -0.0007 Da, -0.9 ppm), **6** $[M+H]^+$ 749.4896 ($C_{37}H_{65}N_8O_8$, theoretical mass: 749.4925, mass difference: -0.0029 Da, -3.8 ppm), **7** $[M+H]^+$ 691.4454 ($C_{34}H_{59}N_8O_7$, theoretical mass: 691.4507, mass difference: -0.0053 Da, -7.6 ppm).

3.1.4 Cell culture

Cells were grown to more than 70% confluency in T175-flask at 37 °C under a humidified 5% CO2 atmosphere. MCF-7, HT29 and HeLa cells were grown in DMEM medium (PAN-Biotech). All media were supplemented with 10% FBS (Sigma-Aldrich) and 100 U/mL penicillin/streptomycin (PAN-Biotech). Medium change was done every two days. When cells were at 80% confluence, they were split by using trypsin-EDTA. For lysis, the cells were washed twice with DPBS (PAN-Biotech), followed by the addition of 1 mL fresh lysate buffer (100 mM sodium acetate buffer, 0.5% NP40, pH 4.5). The cells were harvested by scraping on ice and the mixture was aliquoted into two 1.5 ml Eppendorf tube. Afterwards, 15-20 beads (Diagenode Protein Extraction Bead, diameter < 1 mm) were added and cells were lysed on a Bioruptor (Diagenode) for 10 minutes (30 seconds on, 30 seconds off, 10 cycles) at 4 °C. Solid residues were spun down at 4 °C (5000 rcf, Eppendorf Centrifuge 5424 R). Supernatant (cell lysate) was snap frozen and stored at -80 °C until usage. The protein concentration was determined on lysate dilutions according to the manufacturer's instructions (Pierce BCA protein assay kit, Thermo Scientific).

3.1.5 Gel-based Labeling and competition experiments of Pepsin, Chymosin and cell lysates

Probe concentration titration were performed on 5 pmol of pepsin (Sus scrofa, Sigma-Aldrich: P6887) in a volume of 10 µL of reaction buffer (10 mM HCl, pH 2.0) per condition, 5 pmol of Chymosin (Bos taurus, Sigma-Aldrich: R4877) in a volume of 10 µL of reaction buffer (100 mM sodium acetate buffer, pH 5.6) per condition, 25 ug of cell lysates (MCF-7, HT29 and HeLa) in a volume of 20 µL of reaction buffer (100 mM sodium acetate buffer, pH 4.5) per condition. Different concentrations of Pepstatin A (Cayman Chemical) were applied for probe competition. After 30 min incubation at RT, UV irradiation was performed at RT with a handheld UV lamp (Herolab UV-8 S/L) at 365 nm for 30 min, by placing the samples approximately 2 cm under the lamp. Click reaction was performed using the following conditions: 25 µM of TAMRA-azide (Carl Roth), 200 µM of THPTA (Sigma Aldrich), 4 mM of $CuSO_4$ (Sigma-Aldrich, freshly prepared) and 4 mM of sodium ascorbate (Carl Roth, freshly prepared). Click reaction was incubated for 1 hour at RT, follow by addition of 1/4th volume of 5× Laemmli buffer. Samples were heated at 95 °C for 5 min and resolved by 12% SDS-PAGE. Gels were scanned using a Typhoon Trio+ fluorescent scanner with excitation at 532 nm and an emission filter of 580 nm and stained with Coomassie Brilliant Blue (Carl Roth).

3.1.6 Pull down targets for LC-MS/MS Analysis

MCF-7 lysates (250 µg per sample) were incubated in triplicate with DMSO (Biosolve), 2 µM probe **4** or competition cocktail (2 µM probe 4 and 20 µM Pepstatin A) in the dark at room temperature for 30 min. UV irradiation was performed at RT with a handheld UV lamp at 365 nm for 30 min, by placing the samples approximately 2 cm under the lamp. Click reaction was performed using the following conditions: 25 µM of TAMRA-azide-PEG-biotin (Bio Connect), 200 µM of THPTA (Sigma Aldrich), 4 mM of $CuSO_4$ (Sigma-Aldrich, freshly prepared) and 4 mM of sodium ascorbate (Carl Roth, freshly prepared). Click reaction was incubated for 1 hour at RT, follow by addition of 5 volumes of cold ethanol. The proteome of each sample was precipitated at -28 °C for 1 hour. After centrifugation (4 °C, 18k rcf, 30 min, Eppendorf Centrifuge

5424 R), the proteome pellets were resolubilized in 200 µL pull down buffer (50 mM EDTA, 100 mM phosphate buffer, pH 7.4) by sonicating at 4 °C for 15 min. 10 µL of each sample was taken for quality control on SDS-page. 5 µL of streptavidin beads (Streptavidin Sepharose, GE Healthcare) was added into the resolubilized proteome and incubated for 1 hour at RT. After a second incubation (fresh streptavidin beads, 5 µL, 1 h, RT), the supernatant was discarded and the beads were washed 5 times with 200 µL phosphate buffer (100 mM, pH 7.4) and 5 times with 200 µL ammonium bicarbonate (50 mM, pH 7.8). Afterwards, the beads were resuspended in 200 µL ammonium bicarbonate. 10 µL of the resuspended bead solution was taken for quality control on SDS-page. On-bead digestion was performed by adding 200 ng of Trypsin (Mass Spectrometry Grade, Promega: V528A) to the remaining beads. Following a 14 h digestion at 37 °C in a thermomixer (750 rpm, Eppendorf ThermoMixer C), reduction and carbamidomethylation of the supernatant were carried out by using 10 mM dithiothreitol (DTT, VRW) for 30 min at 56 °C followed by application of 30 mM iodoacetamide (IAA, Sigma-Aldrich) for another 30 min at RT. The carbamidomethylated samples were acidified with 10% trifluoroacetic acid (TFA, Biosolve) and desalted by using Poros Oligo R3 reversed-phase material. Bound peptides were washed twice with 0.1% of formic acid (FA) in water and eluted with 100 µL of 60% (v/v) ACN in water with 0.1% FA. After drying under vacuum, peptides were resolubilized in 15 µL of 0.1% FA. 10% of each sample was taken for nano-LC-MS/MS.

3.1.7 LC-MS Analysis

Samples were submitted to an UltiMate 3000 RSLC nano System (Dionex) coupled to an Orbitrap Elite Hybrid Mass Spectrometer (Thermo Fisher Scientific) for label-free analysis. After initial loading, peptides were concentrated on a 75 µm × 2 cm C18 pre column using 0.1% TFA at a flowrate of 20 µL/min. Sample separation was accomplished on a reversed phase column (Acclaim C18 PepMap100, 75 µm 50 cm) at 50 °C using a binary gradient (A: 0.1% FA, B: 84% ACN with 0.1% FA) at a flowrate of 250 nL/min: 3% solvent B for 5 min, a linear increase of solvent B to 38% for 120 min, a linear increase of solvent B to 95% for 3 min followed by washing with 95% solvent B for 3 min and a linear decrease of solvent B to 3% for 1 min. Peptides were ionized by using a nanospray ESI-source. MS survey scans were acquired on the Orbitrap Elite using settings as follows: mass spectrometer was operated in data dependent acquisition mode (DDA) with full MS scans from 300 to 2000 m/z at a resolution of 120,000 (Orbitrap) using the polysiloxane ion at 371.101236 m/z as lock mass. The automatic gain control (AGC) was set to 1E6 and the maximum injection time to 100 milliseconds. The top 15 intense ions above a threshold ion count of 2000 were selected for fragmentation at a normalized collision energy (nCE) of 35% (CID) in each cycle of the acquisition analysis, following each survey scan. The dynamic exclusion time was set to 30 seconds. Fragment ions were acquired in the linear ion trap with an AGC of 5E3 and a maximum injection time of 100 milliseconds.

3.1.8 Database search and Label-Free Data Analysis

Raw data were searched against the Uniprot human database (July 2018, 20,312 target sequences) using the Mascot search algorithms (Version 2.6.1, Matrix Science) and Sequest HT on Proteome Discoverer v2.3. Precursor mass tolerance was limited to 10 ppm and fragment mass tolerance to 0.5 Da. Cleavage specificity was set to fully tryptic, allowing for a maximum of three missed cleavages. Carbamidomethylation of cysteines (+57.0214 Da) was defined as a fixed modification and oxidation of methionine (+15.9949 Da) as a variable modification for all searches. The results were evaluated with Percolator [180] for false discovery rate (FDR) estimation and data were filtered at ≤1% FDR on the PSM and peptide level and filtered ‘master’ proteins at protein level. Unique and razor peptides (except modified peptides) were taken for Label free quantification. Normalization was performed by applying a global rank-invariant set normalization. [181] Next, the mean of the pairwise peptide ratios (obtained from three replicates) were calculated to determine the protein ratio between the probe and dmso samples or the probe and competition samples. *P*-values were determined by a two-sided *t*-test with Benjamini-Hochberg (non-negative) correction. Common contaminating proteins were removed from the hit lists. [182] The significance cut-offs employed were p-value < 0.05 and log2(ratio probe**4**/DMSO or probe**4**/Competition) > 1. LFQ Data was plotted and visualized by R-4.0.3.

The raw data of LC-MS runs are available via ProteomeXchange with identifier PXD037891 (Username: reviewer_pxd037891@ebi.ac.uk; password: tmqTunIl).

3.1.9 Deep learning for cathepsin D substrates prediction

For the deep learning methods, the Bepler & Berger embedding was used. [183] As a training dataset, we utilized a partial dataset based on the demonstration human model from the D-Script website [183] and incorporated human CatD PPIs retrieved from Merops, [184] GPS-Prot, [185] IntAct, [186] iRefWeb[187] and BioGrid[188] databases, as well as from an N-terminomics study. [189] This combined dataset was used as input for the generator. After 30 epochs of training on the embedding, we used the generated model for predictions. Here we applied two methods, D-SCRIPT and Topsy-turvy separately. The training effort was evaluated with a test dataset including 1000 randomly sampled PPIs and was illustrated in a ROC curve (receiver operating characteristic curve) and histograms. (Figure S2)

3.1.10 Western blotting for cathepsin D substrate validation

20 ng of Sequstosome-1 (Human, abbexa: abx069037, His-tagged) and 5 ng of cathepsin D (Human liver, Athens Research & Technology: 16-12-03014) was incubated at 37 °C for 0-12 h. After adding 1/4th volume of 5× Laemmli buffer, samples were heated at 95 °C for 5 min and resolved by 10% SDS-PAGE, transferred to nitrocellulose-membrane (GE Healthcare/Amersham-Biosciences) using a Trans-Blot Turbo Transfer System according to the manufacturer’s manual (Bio-Rad). After blocking with BSA (1 % BSA in TBS buffer), the membrane was

incubated at 4 °C overnight with the primary antibody (1:1000, mouse Penta·His Antibody, QIAGEN, 34660). After washing with TBST buffer, the membrane was incubated for 1 h at room temperature with the secondary antibody (1:40000, Alexa Fluor 680, Goat Anti-Mouse IgG H&L, ab175775) and washed again. Hydrolysis of Sequstosome-1 was visualized on an Amersham Typhoon fluorescent scanner with excitation at 685 nm and an emission filter of IRshort.

3.1.11 In cellulo analysis of the SQSTM1 protein levels

Flp-In T-REx 293 cells were seeded in 6-well plates and cultured in DMEM medium (Sigma, D6429) with 10% FBS. The next day, the cells destined for treatment were washed twice with EBSS (Sigma, 2888) and starved in the same buffer solution for 24 h in the absence or presence of 100 µM pepstatin A (Carl Roth, 2936). The non-treated control cells were kept in DMEM with 10% FBS. After incubation, the cells were collected in RIPA lysis buffer (Sigma, R0278) supplemented with a protease inhibitor mix (Sigma, P2714), and processed for immunoblotting as described previously. [190] Experiments were performed in quadruplicates. Blots were quantified with ImageJ and subsequent analyses were conducted using GraphPad Prism 9 (GraphPad Software, California, USA).

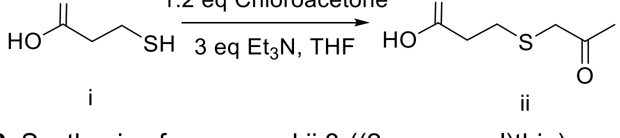
3.2 Mass cleavable affinity-based probes for precise mapping of binding hotspots

3.2.1 General methods

Unless otherwise noted, all reagents were purchased from commercial suppliers and used without further purification. TLC analysis was performed on pre-coated ALUGRAM SIL G plates (Carl Roth) with detection by a handheld UV lamp (254 nm) and subsequent staining with cerium ammonium molybdate solution followed by heating. Low resolution LC-MS analysis was performed on an Dionex Ultimate 3000 (Thermo Scientific) coupled to MSQ Plus Mass Spectrometer (Thermo Scientific) with a Waters xBridge C18 (2.1 x 150 mm, 5 µm) column with a linear gradient of acetonitrile in water with 0.1% trifluoroacetic acid. High resolution LC-MS analysis was performed on a Dionex Ultimate 3000 (Thermo Scientific) coupled to Velos Pro Mass Spectrometer (Thermo Scientific) with 75µm × 20 mm C18 pre column using 0.1% TFA at a flowrate of 12 µl/min. Following sample separation was accomplished on a reversed phase column (Acclaim C18 PepMap100, 75 µm × 150 mm) at 50 °C using a linear gradient: 3%-58% solvent B (84% ACN with 0.1% FA). Preparative HPLC purification was performed on a using a Thermo Scientific BioBasic-18 C18 column (2 × 15 cm, 5 µm). Purifications were performed at room temperature and compounds were eluted with increasing concentration of acetonitrile (solvent A: 0.1% TFA in water, solvent B: 0.1% TFA in 84% acetonitrile). NMR spectra were recorded on a Bruker UltraShield 600MHz NMR Spectrometer. Silica column chromatography was performed using 230-400 mesh silica (Kieselgel 60).

3.2.2 Synthesis of compound ii 3-((2-oxopropyl)thio)propanoic acid

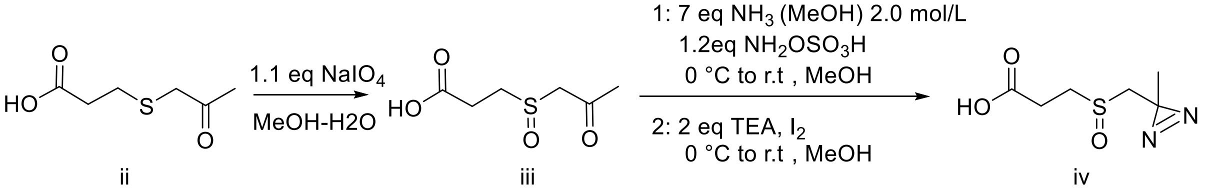
3-mercaptopropanoic acid (6.0 mmol, Sigma-Aldrich) was dissolved in 10 mL THF. TEA (3 equiv, Acros Organics) was added dropwise, while the solution was cooled to 0 °C in an ice bath. After stirring at 0 °C for 10 min, chloroacetone (1.2 equiv, Sigma-Aldrich) was slowly added dropwise. The reaction was quenched by 5% HCl after 4h stirring. Following extraction at pH 2 using ethyl acetate (Carl Roth), compound ii was purified by silica gel chromatography to give the title compound (926.48 mg; yield: 95.2%). ^{1}H NMR (600 MHz, CDCl3): *δ* 3.27 (s, 2H), 2.77 (t, J = 6.9 Hz, 2H), 2.67 (t, J = 7.3 Hz, 2H), 2.31 (s, 3H). ^{13}C NMR (151 MHz, CDCl3): *δ* 203.86, 177.12, 41.76, 33.89, 27.83, 26.55.



Scheme 2. Synthesis of compound ii 3-((2-oxopropyl)thio)propanoic acid

3.2.3 Synthesis of compound iv 3-(((3-methyl-3H-diazirin-3-yl)methyl)sulfinyl)propanoic acid

Compound ii (926.48 mg) was dissolved in methanol-water (methanol: 7.2 mL, water 1.7 mL). An aqueous solution of $NaIO_4$ (1.1 equiv, Sigma-Aldrich, 1.7 mL) was added dropwise, while the solution was cooled at 0 °C in an ice bath. The reaction was quenched by 5% HCl after overnight stirring at room temperature. Following extraction at pH 2 using ethyl acetate (Carl Roth), the compound iii was purified by silica gel chromatography. ESI-MS: $[M+H]^+$ 191.1. Compound iii and $MgSO_4$ (5 equiv, Carl Roth) were cooled in an ice bath with stirring under argon. NH_3 in MeOH (7 N, 7 equiv, Alfa Aesar) was added dropwise. The mixture was slowly warmed up to RT and allowed to stir for 5 h. The mixture was then cooled to -78 °C. A solution of hydroxylamine-O-sulfonic acid (1.2 equiv, Tokyo Chemical Industry) in MeOH (Carl Roth) was added. The mixture was slowly warmed up to RT and allowed to stir overnight. After centrifugation, the liquid phase was collected and volatiles were evaporated under reduced pressure. The residue was dissolved in 10 mL MeOH and cooled in an ice bath. TEA (2 equiv, Acros Organics) was slowly added. Next, I_2 (Sigma-Aldrich) was slowly added until the solution took the color of I_2 and the color didn’t fade away after 1h. The reaction was quenched by 5% HCl and extracted by ethyl acetate (Carl Roth). The organic phase was concentrated under reduced pressure and purified by silica gel chromatography to give the title compound (150 mg; yield: 30.1%). ESI-MS: $[M+H]^+$ 191.1 (theoretical mass: 191.1). ^{1}H NMR (600 MHz, CDCl3): δ 3.27 (s, 2H), 2.77 (t, J = 6.9 Hz, 2H), 2.67 (t, J = 7.3 Hz, 2H), 2.31 (s, 3H). ^{13}C NMR (151 MHz, CDCl3): δ 203.86 (s), 177.12 (s), 41.76 (s), 33.89 (s), 27.83 (s), 26.55 (s).

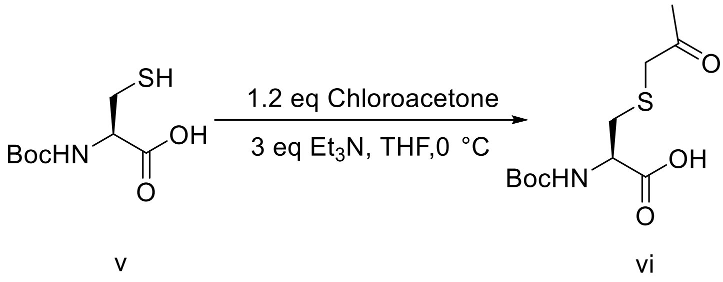


Scheme 3. Synthesis of compound iv 3-(((3-methyl-3H-diazirin-3-yl)methyl)sulfinyl)propanoic acid

3.2.4 Synthesis of compound vi *N*-(tert-butoxycarbonyl)-S-(2-oxopropyl)-L-cysteine

(tert-butoxycarbonyl)-*L*-cysteine (3.0 mmol, Iris Biotech) was dissolved in 10 mL THF. TEA (3 equiv, Acros Organics) was added dropwise, while the solution was cooled to 0 °C in an ice bath. After stirring at 0 °C for 10 min, chloroacetone (1.2 equiv, Sigma-Aldrich) was added dropwise. The reaction was quenched by 5% HCl after 4h stirring. Following extraction at pH 2 using ethyl acetate (Carl Roth), the organic phase was concentrated under reduced pressure and was purified by silica gel chromatography to give the title compound (801.2 mg; yield: 96.3%). ESI-MS: $[M+H]^+$ 278.1 (theoretical mass: 278.1). ^{1}H NMR (500 MHz, CDCl3) δ 5.52 (d, J = 7.5 Hz, 1H), 4.52 (m, 1H), 3.38 (s, 2H), 3.03 (dd, J = 14.0, 5.0 Hz, 1H), 2.95 (dd, J =

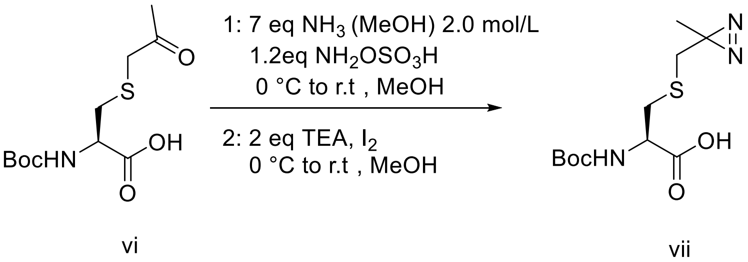
14.0, 5.9 Hz, 1H), 2.29 (s, 3H), 1.46 (s, 9H). ^{13}C NMR (126 MHz, CDCl3) δ 204.22, 174.40, 155.58, 80.71, 53.08, 42.57, 34.26, 28.30, 20.76.



Scheme 4. Synthesis of compound vi *N*-(tert-butoxycarbonyl)-S-(2-oxopropyl)-L-cysteine

3.2.5 Synthesis of compound vii *N*-(tert-butoxycarbonyl)-S-((3-methyl-3H-diazirin-3-yl)methyl)-L-cysteine

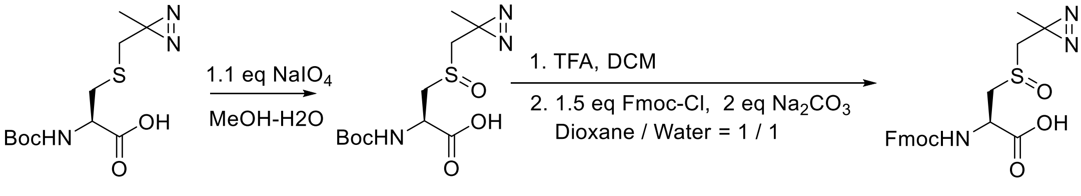
Compound vi (801.2 mg) and $MgSO_4$ (5 equiv, Carl Roth) were cooled in an ice bath with stirring under argon. NH_3 in MeOH (7 N, 7 equiv, Alfa Aesar) was added dropwise. The mixture was slowly warmed up to RT and allowed to stir for 5 h. The mixture was then cooled to -78 °C. A solution of hydroxylamine-O-sulfonic acid (1.2 equiv, Tokyo Chemical Industry) in MeOH (Carl Roth) was added. The mixture was slowly warmed up to RT and allowed to stir overnight. After centrifugation, the liquid phase was collected and volatiles were evaporated under reduced pressure. The residue was dissolved in 10 mL MeOH and cooled in an ice bath. TEA (2 equiv, Acros Organics) was slowly added. Next, I_2 (Sigma-Aldrich) was slowly added in until the solution took the color of I_2 and didn't fade away after 1h. The reaction was quenched by 5% HCl, and extracted by ethyl acetate (Carl Roth). The organic phase was concentrated under reduced pressure and purified by silica gel chromatography to give the compound vii (560.3 mg; yield: 25.9 %). ESI-MS: $[M-H]^-$ 288.0 (theoretical mass: 288.1). ^{1}H NMR (400 MHz, $CDCl_3$) *δ* 9.75 (s, 1H), 5.40 (d, J = 7.6 Hz, 1H), 4.51 (bs, 1H), 3.09 (dd, J = 13.8, 3.8 Hz, 1H), 2.96 (dd, J = 12.9, 4.2 Hz, 1H), 2.48 (s, 3H), 1.41 (s, 9H), 1.10 (s, 3H). ^{13}C NMR (101 MHz, $CDCl_3$) *δ* 174.48, 155.38, 80.54, 53.04, 38.01, 34.13, 28.24, 25.36, 18.86.



Scheme 5. Synthesis of compound vii *N*-(tert-butoxycarbonyl)-S-((3-methyl-3H-diazirin-3-yl)methyl)-L-cysteine

3.2.6 Synthesis of compound ix (((9H-fluoren-9-yl)methoxy)carbonyl)(((3-methyl-3H-diazirin-3-yl)methyl)sulfinyl)-D-alanine

Compound vii was dissolved in methanol-water (methanol: 7.2 mL, water 1.7 mL). $NaIO_4$ aqueous solution (1.1 equiv, Acros Organics, 1.7 mL) was added dropwise, while the solution was cooled to 0 °C in an ice bath. The reaction was quenched by 5% HCl after overnight stirring at room temperature. Following extraction at pH 2 using ethyl acetate (Carl Roth), the organic phase was concentrated under reduced pressure and purified by silica gel chromatography (455.9 mg, yield: 77.1 %). ESI-MS: $[M+H]^+$ 306.1. Compound viii (455.9 mg) was dissolved in 2 mL dichloromethane. 0.3 mL TFA (Fluorochem) was added with stirring under argon. After 0.5 h, the residues were dissolved in 5 mL dioxane (Carl Roth) / water (1 / 1). Na_2CO_3 (5 equiv, Carl Roth) was added at 0 °C. Fluorenylmethyloxycarbonyl chloride (1.5 equiv, Carbolution Chemicals) was added and the mixture was slowly warmed up to RT and allowed to stir overnight. The reaction was quenched by 5% $KHSO_4$. Following extraction at pH 2 using ethyl acetate (Carl Roth), the organic phase was concentrated under reduced pressure and purified by silica gel chromatography to yield the compound ix (499.7 mg, yield: 78.3 %). ESI-MS: $[M+H]^+$ 428.1 (theoretical mass: 428.1).



vii viii ix

Scheme 6. Synthesis of compound ix (((9H-fluoren-9-yl)methoxy)carbonyl)(((3-methyl-3H-diazirin-3-yl)methyl)sulfinyl)-D-alanine

3.2.7 Solid Synthesis of probes 10-13

BAL resin (1 equiv, 450 µmol, 528 mg, Iris Biotech) was added to a solid-phase cartridge in NMP (Biosolve) and gently shaken for 20 minutes. After draining NMP, propargylamine hydrochloride (10 equiv., 4.75 mmol, 436 mg, Sigma-Aldrich) and AcOH (60 µL, Carl Roth) in 6 mL NMP (0.8 mol/mL) were added to the resin and shaken for 20 minutes at room temperature. Sodium cyanoborohydride ($NaBH_3CN$, 10 equiv, 4.75 mmol, 300 mg) was added to the resin and the reaction was shaken for 15-20 hours at room temperature. The resin was then washed with DMF (3x, Biosolve), DCM (3x, Carl Roth) and MeOH (3x, Carl Roth) and dried. The resin was stored in an argon atmosphere at -20 °C for further use.

N-Fmoc-protected amino acid (3 equiv, Fmoc-(3S,4S)-Sta-OH from Iris Biotech, Fmoc-L-Ala-OH.H_2O from PolyPeptide, Fmoc-L-Val-OH from CREOSALUS, homemade compound ix (((9H-fluoren-9-yl)methoxy)carbonyl)(((3-methyl-3H-diazirin-3-yl)methyl)sulfinyl)-*D*-alanine), HBTU (3 equiv, CREOSALUS) and DIEA (6 equiv, Carl Roth) were dissolved in DMF (0.2 M final concentration). For elongation of the resin, the solution of the activated amino acid was shaken with the resin at room temperature for 30 min and washed three times with DMF and DCM. The N-terminal Fmoc group was removed by incubating the resin with 20% Piperidine (Biosolve) in DMF (15 min). Next, the resin was washed three times with DMF and DCM, ending the elongation cycle. For each subsequent step of the solid-phase peptide synthesis, the same deprotection and coupling reactions were used. The N-terminus was blocked by adding a carboxylic acid (3 equiv, isovaleric acid from Sigma-Aldrich or homemade compound iv 3-(((3-methyl-3H-diazirin-3-yl)methyl)sulfinyl)propanoic acid), HBTU (3 equiv) and DIEA (6 equiv). Coupling reactions were monitored by the Kaiser test for primary amines. The product was cleaved from the resin with a TFA/TIPS/H_2O mixture (v/v/v, 95:2.5:2.5, TFA from Fluorochem, TIPS from Sigma-Aldrich), the liquid was collected and the volatiles removed under reduced pressure. The final product was purified by reversed-phase HPLC. Fractions containing product were pooled and lyophilized. ESI-MS: **10** $[M+H]^+$ 811.5 (theoretical mass: 811.5), **11** $[M+H]^+$ 811.3 (theoretical mass: 811.5), **12** $[M+H]^+$ 811.2 (theoretical mass: 811.5), **13** $[M+H]^+$ 753.4 (theoretical mass: 753.4).

3.2.8 Gel-based Labeling and competition experiments of Chymosin

Probe concentration titration was performed on 5 pmol of Chymosin (Bos taurus, Sigma-Aldrich: R4877) in a volume of 10 µL of reaction buffer (100 mM sodium acetate buffer, pH 5.6) per condition. Different concentrations of Pepstatin A (Cayman Chemical) were applied for probe competition. After 30 min incubation at RT, UV irradiation was performed at RT with a handheld UV lamp (Herolab UV-8 S/L) at 365 nm for 30 min, by placing the samples approximately 2 cm under the lamp. Click reaction was performed using the following conditions: 25 µM of TAMRA-azide (Carl Roth), 200 µM of THPTA (Sigma Aldrich), 4 mM of $CuSO_4$ (Sigma-Aldrich, freshly prepared) and 4 mM of sodium ascorbate (Carl Roth, freshly prepared). Click reaction was incubated for 1 hour at RT, follow by addition of 1/4th volume of 5× Laemmli buffer. Samples were heated at 95 °C for 5 min and resolved by 12% SDS-PAGE. Gels were scanned using a Typhoon Trio+ fluorescent scanner with excitation at 532 nm and an emission filter of 580 nm and stained with Coomassie Brilliant Blue (Carl Roth).

3.2.9 Sample preparation of probe irradiation for LC-MS/MS Analysis

Probe **11** (2.25 µL, 1 mM) was diluted in 147.75 µL reaction buffer (100 mM sodium acetate buffer, pH 5.6). The probe solution (15 µM, 150 µL) was irradiated at RT with a handheld UV lamp at 365 nm for 30 min, by placing the samples approximately 2 cm under the lamp. The irradiated samples were acidified with 10% trifluoroacetic acid (TFA, Biosolve) and desalted by using Poros Oligo R3 reversed-phase material. Bound irradiation products were washed twice with 0.1% of formic acid (FA) in water and eluted with 100 µL of 95% (v/v) ACN in water with 0.1% FA. After drying under vacuum, products were resolubilized in 450 µL of 0.1% FA. 15 µL of each sample was taken for nano-LC-MS/MS.

3.2.10 Probe irradiation product analysis using LC-MS/MS

Samples were submitted to an UltiMate 3000 RSLC nano System (Dionex) coupled to an Orbitrap Velos Pro Mass Spectrometer (Thermo Fisher Scientific). After initial loading, peptides were concentrated on a 75 µm × 2 cm C18 pre column using 0.1% TFA at a flowrate of 20 µL/min. Sample separation was accomplished on a reversed phase column (Acclaim C18 PepMap100, 75 µm 50 cm) at 50 °C using a binary gradient (A: 0.1% FA, B: 84% ACN with 0.1% FA) at a flowrate of 250 nL/min: 3% solvent B for 5 min, a linear increase of solvent B to 95% for 35 min followed by washing with 95% solvent B for 5 min and a linear decrease of solvent B to 3% for 1 min. Irradiation products were ionized by using a nanospray ESI-source. MS survey scans were acquired on the Orbitrap Velos Pro using settings as follows: mass spectrometer was operated in data dependent acquisition mode (DDA) with full MS scans from 300 to 2000 m/z at a resolution of 60,000 at 400 m/z (Orbitrap) using the polysiloxane ion at 371.101236 m/z as lock mass. The automatic gain control (AGC) was set to 1E6 and the maximum injection time to 500 milliseconds. The top 5 intense ions above a threshold ion count of 500 were selected for fragmentation at a normalized collision energy (nCE) of 35% (CID) in each cycle of the acquisition analysis, following each survey scan. The dynamic exclusion time was set to 10 seconds. Fragment ions were acquired in the linear ion trap with an AGC of 1E4 and a maximum injection time of 10 milliseconds.

3.2.11 Probe collision energy investigation using LC-MS/MS

Products of probe irradiation (in triplicate) were submitted to an UltiMate 3000 RSLC nano System (Dionex) coupled to an Orbitrap Velos Pro Mass Spectrometer (Thermo Fisher Scientific). After initial loading, peptides were concentrated on a 75 µm × 2 cm C18 pre column using 0.1% TFA at a flowrate of 20 µL/min. Sample separation was accomplished on a reversed phase column (Acclaim C18 PepMap100, 75 µm 50 cm) at 50 °C using a binary gradient (A: 0.1% FA, B: 84% ACN with 0.1% FA) at a flowrate of 250 nL/min: 3% solvent B for 5 min, a linear increase of solvent B to 95% for 35 min followed by washing with 95% solvent B for 5 min and a linear decrease of solvent B to 3% for 1 min. Irradiation products were ionized by using a nanospray ESI-source. MS survey scans were acquired on the Orbitrap Velos Pro using settings as follows: mass spectrometer was operated in scheduled Parallel reaction monitoring (PRM) with full MS scans from 300 to 2000 m/z at a resolution of 30,000 at 400 m/z (Orbitrap) using the polysiloxane ion at 371.101236 m/z as lock mass. The automatic gain control (AGC) was set to 1E6 and the maximum injection time to 500 milliseconds. The transitions were selected in 0.8 m/z mass isolation window for fragmentation at different normalized collision energy (nCE) of 19%, 23%, 25%, 27% 29% and 31% (CID). Fragment ions were acquired in the orbitrap at a resolution of 30,000 at 400 m/z. The automatic gain control (AGC) was set to 1E6 and the maximum injection time to 500 milliseconds. The eliminated product ($C_{38}H_{66}N_6O_9S$, HRMS: $[M+H]^+$ 783.4677, theoretical mass: 783.4690, mass difference: -0.0013 Da, -1.7 ppm), hydrolyzed product ($C_{38}H_{68}N_6O_{10}S$, HRMS: $[M+H]^+$

801.4779, theoretical mass: 801.4795, mass difference: -0.0016 Da, -2.0 ppm) and acetylated product ($C_{40}H_{70}N_6O_{11}S$, HRMS: $[M+H]^+$ 843.4889, theoretical mass: 843.4901, mass difference: -0.0012 Da, -1.4 ppm) with same sulfoxide cleavage products (theoretical mass: $C_{35}H_{58}N_6O_7S$ $[M+H]^+$ 675.4445, $C_{35}H_{60}N_6O_8S$ $[M+H]^+$ 693.4554) were quantified by Skyline 21.2.

3.2.12 Sample preparation of chymosin for binding hotspots mapping

Chymosin (50 µg, 10 µM) was incubated in reaction buffer (100 mM sodium acetate buffer, pH 5.6) with 15 µM probe in the dark at room temperature for 30 min. UV irradiation was performed at RT with a handheld UV lamp at 365 nm for 30 min, by placing the samples approximately 2 cm under the lamp. After the denaturation at 95 °C for 5 min, reduction and carbamidomethylation of the photo-crosslinked chymosin were carried out by using 10 mM dithiothreitol (DTT, VWR) for 30 min at 56 °C followed by application of 30 mM iodoacetamide (IAA, Sigma-Aldrich) for another 30 min at RT. The photo-crosslinked chymosin was precipitated in ethanol at -28 °C for 1 hour. After centrifugation (4 °C, 18k rcf, 30 min, Eppendorf Centrifuge 5424 R), the protein pellets were resolubilized in 100 µL digestion buffer (ammonium bicarbonate, 50 mM, pH 7.8) by sonicating at RT for 5 min. In solution digestion was performed at 37 °C in a thermomixer (750 rpm, Eppendorf ThermoMixer C) by adding 250 ng of Glu-C (Mass Spectrometry Grade, Roche). Following a 8 h Glu-C digestion, a secondary digestion was performed under the same conditions by adding 250 ng of Trypsin (Mass Spectrometry Grade, Promega). After 8 h of secondary digestion the digested sample was acidified with 10% trifluoroacetic acid (TFA, Biosolve) and desalted by using Poros Oligo R3 reversed-phase material. Bound peptides were washed twice with 0.1% of formic acid (FA) in water and eluted with 100 µL of 95% (v/v) ACN in water with 0.1% FA. After drying under vacuum, peptides were resolubilized in 100 µL of 0.1% FA for nano-LC-MS^n.

3.2.13 Identification of photo-crosslinked peptides using data dependent acquisition square (DDA^2)

Digested samples (1% of total volume) were submitted to an UltiMate 3000 RSLC nano System (Dionex) coupled to an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher Scientific. After initial loading, peptides were concentrated on a 75 µm × 2 cm C18 pre column using 0.1% TFA at a flowrate of 20 µL/min. Sample separation was accomplished on a reversed phase column (Acclaim C18 PepMap100, 75 µm 50 cm) at 60 °C using a binary gradient (A: 0.1% FA, B: 84% ACN with 0.1% FA) at a flowrate of 250 nL/min: 3% solvent B for 5 min, a linear increase of solvent B to 42% for 95 min followed by washing with 95% solvent B for 3 min and a linear decrease of solvent B to 3% for 1 min. Peptides were ionized by using a nanospray ESI-source. MS survey scans were acquired on the Orbitrap Eclipse Tribrid using settings as follows: mass spectrometer was operated in data dependent acquisition mode (DDA) with full MS scans from 350 to 2000 m/z at a resolution of 120,000 at 200 m/z (Orbitrap) using the polysiloxane ion at 445.12002 m/z as lock mass. The automatic gain control (AGC) was set to 4E5, the maximum injection time to 150 milliseconds and the RF amplitude was set to 35%. The top 15 intense ions (charge state >= 2) were selected within 2 m/z mass isolation window for fragmentation at a normalized collision energy (nCE) of 21, 23, 25, 27% (CID) in each cycle of the acquisition analysis, following each survey scan. The dynamic exclusion time was set to 30 seconds. Fragment ions were acquired in the Orbitrap at a resolution of 30,000 at 200 m/z with an AGC of 5E4 and a maximum injection time of 150 milliseconds. A secondary DDA was triggered by the detection of one of the probe fragments (C35H58N6O7S, $[M+H]^+$ 675.4445, C35H60N6O8S $[M+H]^+$ 693.4554) within top 15 intense ions in each cycle of the MS^2 scan at the mass tolerance of 20 ppm. The top 10 intense MS^2 fragment ions were selected within 2 m/z mass isolation window for secondary fragmentation at a normalized collision energy (nCE) of 35% (HCD) in each cycle of the acquisition analysis, following each survey scan. Fragment ions were acquired in the linear ion trap with an AGC of 1E4 and a maximum injection time of 100 milliseconds.

3.2.14 Optimization of key parameters of parallel reaction monitoring acquisition square (PRM^2)

2.5 % of digested samples were submitted to an UltiMate 3000 RSLC nano System (Dionex) coupled to an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher Scientific. After initial loading, peptides were concentrated on a 75 µm × 2 cm C18 pre column using 0.1% TFA at a flowrate of 20 µL/min. Sample separation was accomplished on a reversed phase column (Acclaim C18 PepMap100, 75 µm 50 cm) at 60 °C using a binary gradient (A: 0.1% FA, B: 84% ACN with 0.1% FA) at a flowrate of 250 nL/min: 3% solvent B for 5 min, a linear increase of solvent B to 42% for 95 min followed by washing with 95% solvent B for 3 min and a linear decrease of solvent B to 3% for 1 min. Peptides were ionized by using a nanospray ESI-source. MS survey scans were acquired on the Orbitrap Eclipse Tribrid using settings as follows: mass spectrometer was operated in scheduled parallel reaction monitoring (PRM) with direct MS^2 scans. The transitions (Table 1) were selected in 2 m/z mass isolation window for fragmentation at different normalized collision energy (nCE) of 19%, 21%, 23%, 25%, 27% (CID). Fragment ions were acquired in the orbitrap at a resolution of 30,000 at 200 m/z with an AGC of 5E4 and a maximum injection time of 150 milliseconds. A secondary scheduled parallel reaction monitoring PRM was triggered by the detection of one of the probe fragments ($C_{35}H_{58}N_6O_7S$ $[M+H]^+$ 675.4445, $C_{35}H_{60}N_6O_8S$ $[M+H]^+$ 693.4554) within top 10 intense ions in each cycle of the MS^2 scan at the mass tolerance of 20 ppm. The targeted MS^2 fragment ions (Table 2) were selected within 2 m/z mass isolation window for secondary fragmentation at a normalized collision energy (nCE) of 35% (HCD) in each cycle of the acquisition analysis, following each survey scan. Fragment ions were acquired in the linear ion trap with an AGC of 1E4, 2.5E4, a maximum injection time of 100, 250 milliseconds and 1 or 5 times of microscans in each cycle of the MS^3 acquisition.

Table 1, MS^2 transitions list of probe **11** label sample.

Compound	m/z	z	t start (min)	t stop (min)
VASVPLTNYLDSQYFGK[+782.461199]	895.4800	3	66.45	72.45
VASVPLTNYLDSQYFGK[+782.461199]	895.4800	3	65.63	70.63
VASVPLTNYLDSQYFGK[+782.461199]	895.4800	3	57.39	62.39
VASVPLTNYLDSQYFGK[+782.461199]	895.4800	3	56.78	61.78
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1020.477	4	47.54	52.54
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1020.477	4	48.87	53.87
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1020.477	4	47.54	52.54
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1366.301	3	49.34	55.34

Table 2, MS^3 transitions list of probe **11** label sample.

Compound	m/z	t start (min)	t stop (min)
VASVPLTNYLDSQYFGK[+782.461199]	996.4925	66.45	72.45
VASVPLTNYLDSQYFGK[+782.461199]	996.4925	65.63	70.63
VASVPLTNYLDSQYFGK[+782.461199]	996.4925	57.39	62.39
VASVPLTNYLDSQYFGK[+782.461199]	996.4925	56.78	61.78
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1129.485	47.54	52.54
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1129.485	48.87	53.87
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1129.485	47.54	52.54
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1702.724	49.34	55.34
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1701.721	49.34	55.34
VASVPLTNYLDSQYFGK[+782.461199]	987.4872	66.45	72.45
VASVPLTNYLDSQYFGK[+782.461199]	987.4872	65.63	70.63
VASVPLTNYLDSQYFGK[+782.461199]	987.4872	57.39	62.39
VASVPLTNYLDSQYFGK[+782.461199]	987.4872	56.78	61.78
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1123.481	47.54	52.54
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1123.481	48.87	53.87
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1123.481	47.54	52.54
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1693.719	49.34	55.34
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1692.715	49.34	55.34
VASVPLTNYLDSQYFGK[+782.461199]	951.4855	66.45	72.45
VASVPLTNYLDSQYFGK[+782.461199]	951.4855	65.63	70.63
VASVPLTNYLDSQYFGK[+782.461199]	951.4855	57.39	62.39
VASVPLTNYLDSQYFGK[+782.461199]	951.4855	56.78	61.78
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1099.48	47.54	52.54
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENH	1099.48	48.87	53.87

SQK[+782.461199]			
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1099.48	47.54	52.54
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1657.717	49.34	55.34
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1656.714	49.34	55.34
VASVPLTNYLDSQYFGK[+782.461199]	942.4802	66.45	72.45
VASVPLTNYLDSQYFGK[+782.461199]	942.4802	65.63	70.63
VASVPLTNYLDSQYFGK[+782.461199]	942.4802	57.39	62.39
VASVPLTNYLDSQYFGK[+782.461199]	942.4802	56.78	61.78
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1093.476	47.54	52.54
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1093.476	48.87	53.87
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1093.476	47.54	52.54
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1648.712	49.34	55.34
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1647.709	49.34	55.34

3.2.15 Mapping of binding hotspots of different probes

Chymosin (50 µg, 10 µM) was incubated in reaction buffer (100 mM sodium acetate buffer, pH 5.6) with 15 µM of probe **10**, **11** and **12** in the dark at room temperature for 30 min. Following UV irradiation and sample preparation, 1 % of triplicate samples were submitted to LC-MS for DDA^2 analysis (25% CID collision energy at MS^2). After photo-crosslinked peptides were identified by DDA^2 analysis, 2.5 % of triplicate samples were submitted to PRM^2 analysis (25% CID collision energy at MS^2, ion trap microscan: 5, ion trap AGC target: 250%, ion trap maximum injection time: 250 ms).

3.2.16 Data analysis of Probe collision energy investigation

MS^1 and MS^2 data were filtered according to the transition list (Table 3 and Table 4) with orbitrap resolution (30,000 at 400 m/z) in Skyline 21.2. The peak area of precursors and fragment ion was integrated and visualized by R-4.0.3.

Table 3, Transitions list of probe **11** after UV irradiation in acetate buffer.

Name	Precursor (P)	P (m/z)	P (z)	Fragment (F)	F (m/z)	F (z)
Reporter	Compound I	801.4790	1	C35H58N6O7	675.4440	1
Reporter - H_2O	Compound I	801.4790	1	C35H60N6O8	693.4545	1
Reporter	Compound II	783.4685	1	C35H58N6O7	675.4440	1
Reporter - H_2O	Compound II	783.4685	1	C35H60N6O8	693.4545	1
Reporter	Compound III	843.4896	1	C35H58N6O7	675.4440	1
Reporter - H_2O	Compound III	843.4896	1	C35H60N6O8	693.4545	1
Compound I	Compound I	801.4790	1	C38H68N6O10S	801.4790	1
Compound II	Compound II	783.4685	1	C38H66N6O9S	783.4685	1
Compound III	Compound III	843.4896	1	C40H70N6O11S	843.4896	1

Table 4, Transitions list of probe **11** label sample.

Name	Precursor (P)	P (m/z)	P (z)	Fragment (F)	F (m/z)	F (z)
R	VAS***FGK	895.4800	3	C35H58N6O7	675.4440	1
R - H2O	VAS***FGK	895.4800	3	C35H60N6O8	693.4545	1
P-R(- H2O)	VAS***FGK	895.4800	3	P-R- H2O	987.4857	2
R	MYP***SQK	1020.4772	4	R	675.4440	1
R - H2O	MYP***SQK	1020.4772	4	R - H2O	693.4545	1
P-R(- H2O)	MYP***SQK	1020.4772	4	P-R	1129.4848	3

P: precursor, probe-modified peptide VAS***FGK, probe-modified peptide MYP***SQK; R - H_2O: Reporter - H_2O; P - R (- H_2O): Precursor - Reporter - H_2O (VAS***FGK), Precursor - Reporter (MYP***SQK).

3.2.17 Database search via Proteome Discoverer

MS^3 data were extracted from the raw data. MS^2 precursors were selected and searched against the Uniprot bovine database (March 2018, 37,512 target sequences) using the Sequest HT on Proteome Discoverer v2.3. Precursor mass tolerance was limited to 20 ppm and fragment mass tolerance to 0.5 Da. Cleavage specificity was set to fully Glu C/ trypsin (Cleave at the C-terminal of Lys, Arg and Glu), allowing for a maximum of three missed cleavages. Carbamidomethylation of cysteines (+57.0214 Da), oxidation of methionine (+15.9949 Da) and cleaved probe modification (+90.014 Da and +72.003 Da (neutral loss of H_2O)) of 20 natural amino acids were defined as a variable modification for all searches. The results were evaluated with Target Decoy PSM validator for false discovery rate (FDR) estimation and data were filtered at ≤1% FDR on the PSM and peptide level and filtered ‘master’ proteins at protein level. Different isoforms of modified peptides were sorted and visualized by R-4.0.3.

3.2.18 Transitions selection of PRM^2

Chymosin was in-silico digested with three missed cleavages and intact probe modification (782.461199 for probe **10**, **11**, **12**) was set at the N-terminus of each in-silico peptides in Skyline 21.2. MS^1 was filtered according to the precursors list from the database search (Table 5) with orbitrap resolution (120,000 at 200 m/z). The precursor transitions at different retention time were selected for MS^2 experiment after (isotope dot product) idotp (>0.85) filtration (Table 1). Following the m/z calculation of probe fragmentation, the MS^2 transitions were defined for the secondary PRM (MS^3) experiment. (Table 2)

Table 5, Precursors list of probe **10-12** label sample.

Compound	m/z	z
VASVPLTNYLDSQYFGK[+782.461199]	895.4800	3
MYPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1020.477	4
M[+15.994915]YPLTPSAYTSQDQGFC[+57.021464]TSGFQSENHSQK[+782.461199]	1366.301	3

4 Results

4.1 Pepstatin-based probes for photoaffinity labeling of aspartic proteases

4.1.1 Probe design and Synthesis

Pepstatin is a natural product obtained from Actinomycetes [191] that inhibits aspartic proteases in a broad spectrum and reversible manner. [192] Three amino acids, two statine residues, and an N-terminal isovaleroyl cap make up its structure (Figure 21A). Pepstatin's inhibitory effect is based on imitating a substrate's tetrahedral intermediate, which is created when a water molecule attacks the scissile bond (Figure 21B). According to published crystal structures (Figure 21C), the binding mechanism of pepstatin A is similar for different aspartic proteases. Importantly, adequate space seems to be available for a diazirine moiety as a 'minimal photocrosslinker' inside the pepstatin structure for different residues. As a result, we decided to replace different amino acids in the pepstatin structure, as well as the C-terminal statine residue, with a 'photoleucine' building block (**2**, Figure 21D), and the N-terminal acyl group with a reported diazirine building block **3**. The use of an alkyne as a C-terminal tag enables biorthogonal click chemistry to detect the photo-crosslinked probe-protease complexes.

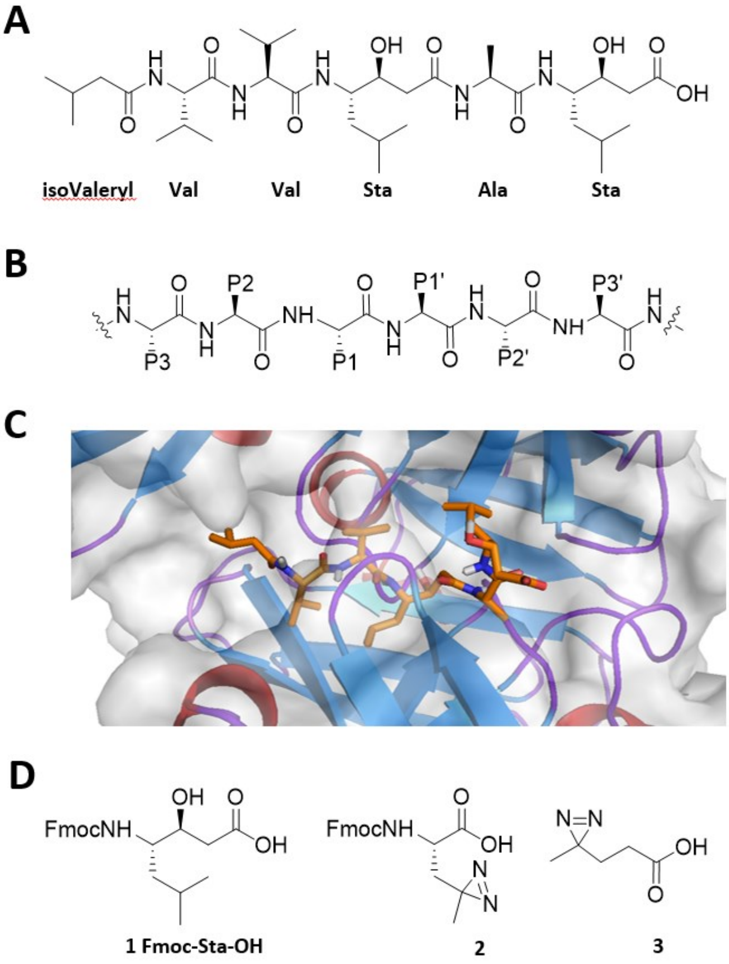
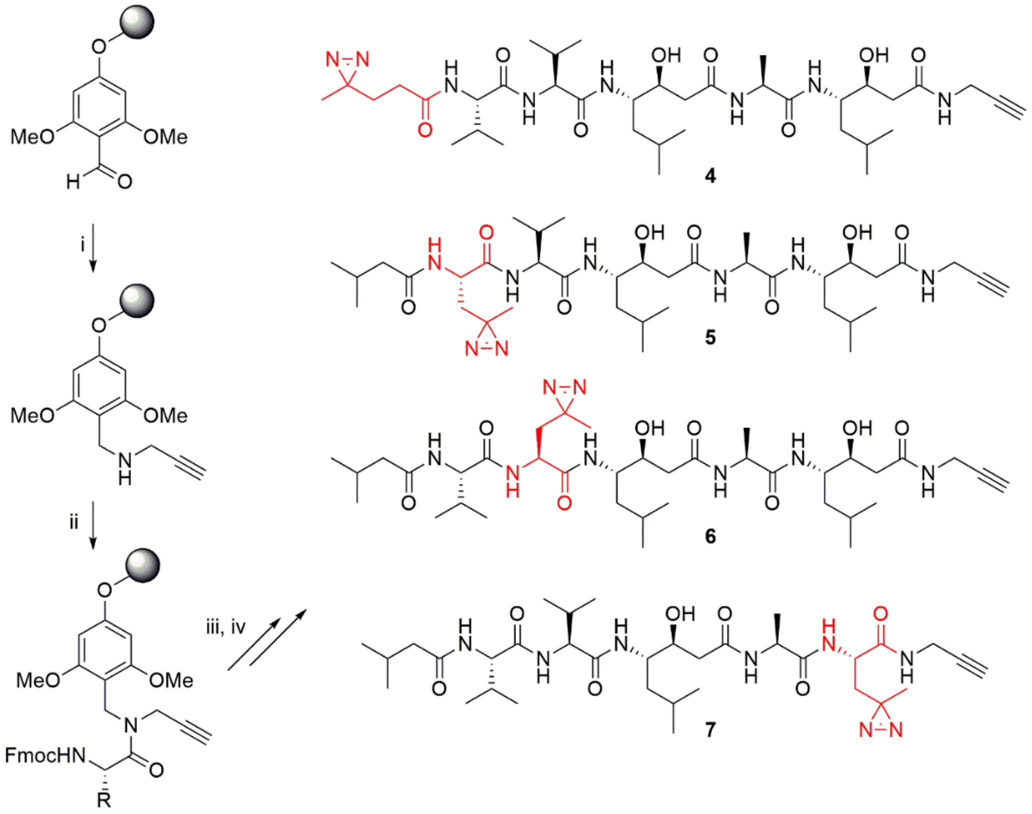


Figure 21. The general aspartic protease inhibitor Pepstatin. (A) Chemical structure of pepstatin. **(B)** Chemical structure of a protease substrate. Amino acid residues at the N-terminal side are named

P1, P2 etc., whereas the residues at the C-terminal side are denoted with an apostrophe (P1' etc). **(C)** Crystal structure of pepstatin in the active site of Cathepsin D (PDB code: 1LYB). [193] Protein depicted in cartoon mode with sheets in blue, helices in red and random coils in purple. The protein surface is indicated in transparent grey. Pepstatin is depicted in stick model (orange). Picture rendered with PyMol. [6] **(D)** Chemical structures of Fmoc-protected statine (**1**), Fmoc-protected photoleucine (**2**) and diazirine building block **3**

While Fmoc-statine (**1**) and Fmoc-photoleucine (**2**) are commercially accessible, diazirine building block **3** was synthesized in a two-step procedure from levulinic acid by following previously reported method. [194] Conveniently, the use of a backbone amide linker (BAL) resin enabled the elongation of the C-terminally-tagged pepstatin probes on a solid support.[195] To summarize, reductive amination was employed to attach a propargylamine to the BAL resin. Followed by Fmoc-based solid phase peptide elongation and cleavage via trifluoroacetic acid, four pepstatin probes with a diazirine in different locations were obtained (**4-7**) (Scheme 7).



Scheme 7. Solid phase synthesis of pepstatin probes **4-7**. Photoreactive residue is indicated in red. *Reagents and conditions*: (i) propargylamine hydrochloride (10 eq.), $NaBH_3CN$ (10 eq.), AcOH (1%) in NMP. (ii) Fmoc-Sta-OH or Fmoc-photoLeu-OH (3 eq.), HBTU (3 eq), DIEA (6 eq), DMF. (iii) Elongation of the peptide chain by repeated Fmoc-deprotection and coupling of building blocks: a. 20% piperidine in DMF; b. Fmoc-Val-OH, Fmoc-Ala-OH, Fmoc-Sta-OH, Fmoc-photoLeu-OH or building block **3** (3 eq.), HBTU (3 eq), DIEA (6 eq), DMF. (iv) TFA/TIPS/H_2O (95/2.5/2.5).

4.1.2 Labeling of pure porcine pepsin and bovine chymosin

We performed labeling experiments on pure porcine pepsin and bovine chymosin as two model aspartic proteases by using the four different probes we had at hand. Titrations of the probes were carried out to identify the best concentration.

To achieve this, 50 pmol of pure enzyme was incubated with varying concentrations of probes **4-7** under 365 nm irradiation (Figure 22A), followed by the introduction of a TAMRA-fluorophore via Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC). Pepsin is labeled by all of the probes (Figure 22A). While probe **5** is producing the most intense labeling on pepsin, probes **5** and **6** are giving the strongest signal on chymosin. As expected, the position of the photo-crosslinking group (diazirine) has an impact on the labeling capacity. Nevertheless, we also observed a target depended labeling intensity, which indicated some delicate difference between porcine pepsin and bovine chymosin (Figure 22A). UV radiation was required for labeling (Figure S1), demonstrating that a covalent modification is created through photoaffinity labeling. We were able to detect as little as 10–50 ng of pepsin and 50–100 ng of chymosin after titrating down the enzyme concentration (Figure 22B).

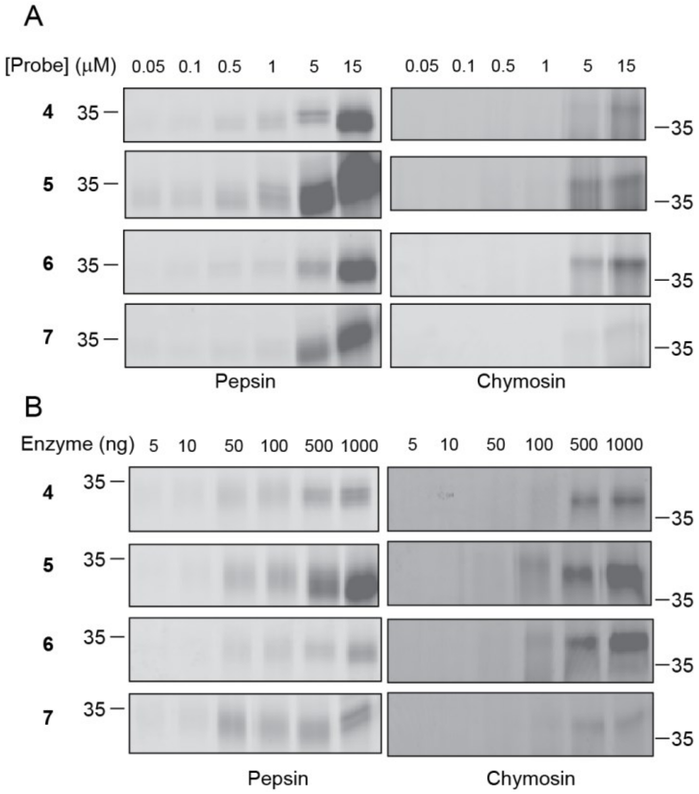


Figure 22. Labeling of purified proteases by pepstatin-based AfBPs **4-7**. All PAL reactions were performed for 30 min with a handheld UV lamp. **(A)** Photoaffinity labeling of 50 pmol of porcine pepsin (left panels) or bovine chymosin (right panels) by increasing concentrations of probe (0.05, 0.1, 0.5, 1, 5, 15 μM) under 365 nm UV irradiation with visualization by CuAAC with a TAMRA-azide tag. **(B)** Photoaffinity labeling of increasing amounts of porcine pepsin (left panels) or bovine trypsin (right panels; 5, 10, 50, 100, 500, 1000 ng) with 5 μM of the indicated probe under 365 nm UV irradiation with visualization by CuAAC with a TAMRA-azide tag.

4.1.3 Labeling of aspartic protease cathepsin D in human cancer cell lines

Following the demonstration of effective labeling on two model aspartic proteases, we applied the Pepstatin A-derived AfBPs **4-7** to human cancer cell lines. Lysates from MCF-7 (human breast cancer), HT-29 (human colorectal adenocarcinoma), and HeLa (human cervical cancer) were chosen since they have different expression amounts of the aspartic protease cathepsin D (Figure 23). [196] [197] These PAL experiments resulted in the selective labeling of bands at 45 (faint) and 30 kDa (Figure 24A), which correspond to the proprotein and heavy chain of cathepsin D, respectively, as confirmed by western blot (Figure S2). MCF-7 had the strongest labeled bands, HT-29 had the weakest, while none could be observed with HeLa lysates (Figure 24A). Titration of the probe indicated that the cathepsin D target was saturated at probe concentrations as low as 1 μM. (Figure 24B, Figure S3). This is in contrast to pure pepsin and chymosin data, and might be due to self-digestion of these latter two proteases at lower probe concentrations. Additionally, the decreasing of cathepsin D labeling with increasing amounts of the parent pepstatin demonstrates that the parent compound's binding site is identical by the pepstatin-based probes (Figure 24B, Figure S4).

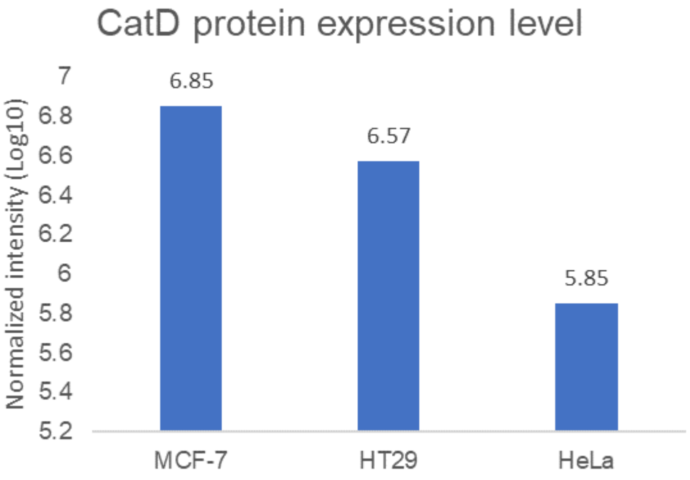


Figure 23. Protein expression level of CatD in various cell lines with normalized intensity-based absolute quantification. Data taken from proteomicsdb.org.

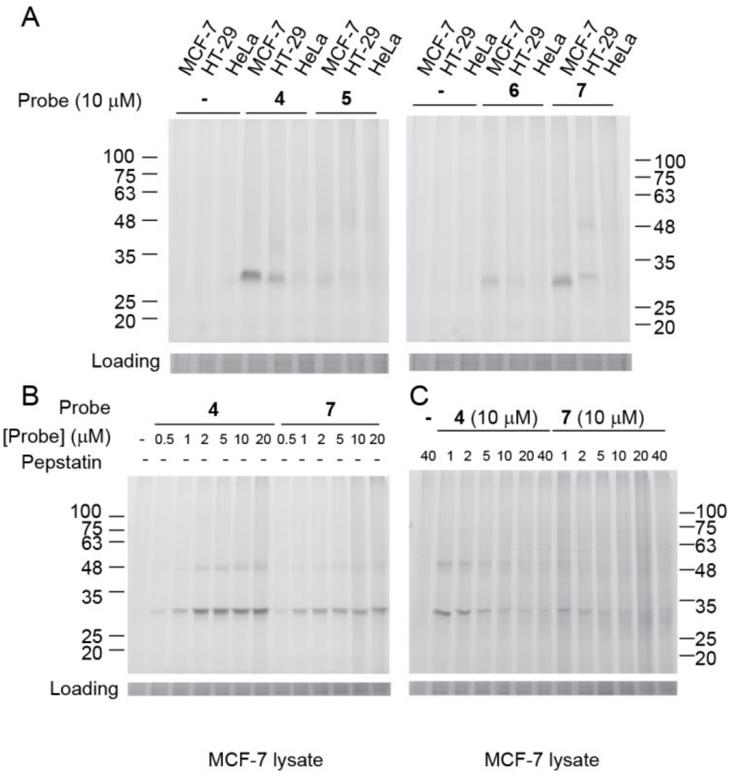


Figure 24. Labeling of cell lysates by pepstatin-based AfBPs. **(A)** Labeling of lysates of three different cancer cell lines reveals highest band intensity in MCF-7 with probes **4** and **7**. **(B)** Labeling of targets with increasing concentration of probe (stepwise: 0.5, 1, 2, 5, 10, 20 μM) reveals saturation of labeling at 1-2 μM probe concentration). **(C)** Addition of increasing concentrations of Pepstatin A as competitor for labeling (stepwise: 1, 2, 5, 10, 20, 40 μM competitor with constant probe concentration) shows that labeling of bands in MCF-7 lysates is outcompeted, illustrating the specificity of the binding event. Coomassie stains of full gels of Figure 14A-4C are provided in the supporting information.

4.1.4 Proteomics analysis of AfBP 4 affinity-enrichment

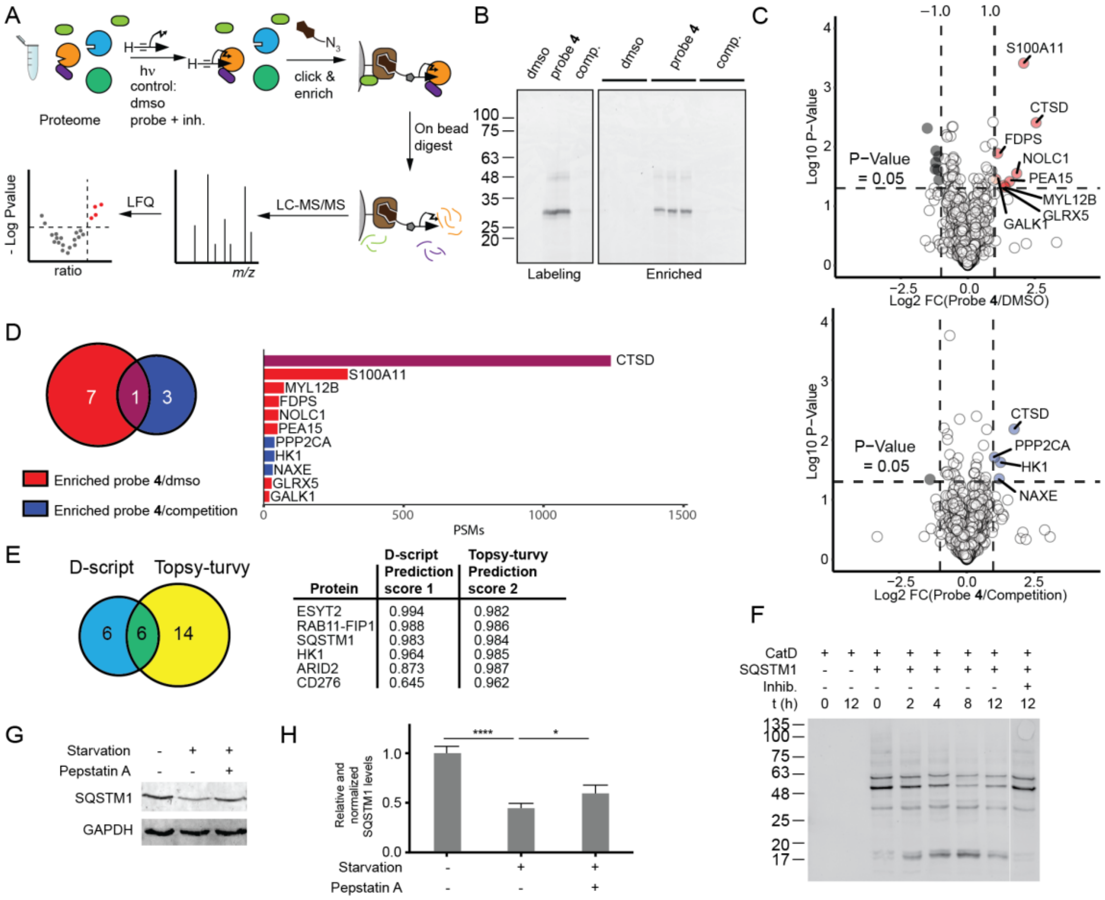
A proteomics affinity-enrichment procedure was used, as shown in Figure 15A, to confirm the cathepsin D identity of the labeled target band in MCF-7 lysates and to find any more potential targets below the gel-based detection method. AfBP **4** was used for the analysis, with DMSO (blank) and AfBP plus parent Pepstatin A (competition) serving as the controls. After doing a quality check on SDS-PAGE (Figure 15B), it was observed that samples were efficiently labeled in triplicate, whereas labeling was absent in controls. Captured proteins were digested on bead. Quality control of label-free quantification (LFQ) revealed a good correlation between runs (Figure S5). In total, 2,067 proteins were quantified. We plotted the fold enrichment over control vs the p-value in two volcano plots (Figure 25C), indicating 8 enriched proteins versus DMSO control (Table 6) and 4 enriched versus the competition experiment (Table 7). Only the protein cathepsin D (CTSD) was present in both (Figure 25D). In addition, among all enriched proteins, cathepsin D had by far the most peptide spectrum matches (PSMs), indicating that it was the primary probe target (Figure 25D).

Table 6 Enriched proteins versus DMSO control

Genes	PSMs	Unique Peptides	Probe/DMSO. log2.FC	Probe/DMSO. p.Value
GALK1	20	2	1.037539876	0.03497864
GLRX5	28	2	1.349367661	0.044724149
PEA15	49	3	1.560388046	0.038465313
FDPS	54	3	1.112136611	0.012915689
MYL12B	72	4	1.395639973	0.047847038
Q14978	52	6	1.827611513	0.028068668
P31949	300	8	2.086565736	0.000389236
CTSD	1240	21	2.550749388	0.003921729

Table 7 Enriched proteins versus the competition experiment

Genes	PSMs	Unique Peptides	Probe/Competition log2.FC	Probe/Competition p.Value
Q8NCW5	32	5	1.204162	0.04479
P19367	38	6	1.25369	0.023839
P67775	38	6	1.028489	0.019647
CTSD	1240	21	1.75658	0.006636



Protein	D-script Prediction score 1	Topsy-turvy Prediction score 2
ESYT2	0.994	0.982
RAB11-FIP1	0.988	0.986
SQSTM1	0.983	0.984
HK1	0.964	0.985
ARID2	0.873	0.987
CD276	0.645	0.962

Figure 25. Chemical proteomics identification of pepstatin-based probe targets. **(A)** Schematic representation of the followed chemical proteomics: UV irradiation of (365 nm, 0.5 h) of the proteome (MCF-7 lysate) in presence of probe (2 µM) probe 4 and 20 µM Pepstatin A, DMSO (control 1, triplicates) or probe (2 µM) + pepstatin parent inhibitor (control 2, 20 µM Pepstatin A, triplicates) is followed by click chemistry mediated introduction of a TAMRA-biotin tag, ethanol precipitation of proteins and target enrichment after protein resolubilization. Tryptic peptides generated by on-bead digest are analyzed by LC-MS/MS and analyzed by label-free quantification. Data was processed by Proteome Discoverer 2.3. See materials and methods for details on LC-MS/MS methods and data analysis parameters (charpter 2). **(B)** Fluorescent gel analysis as quality control of enrichment. Samples were unlabeled (DMSO, triplicates), labeled with probe **4** (triplicates) or labeled with probe **4** in competition with pepstatin A (comp, triplicates). Input samples are shown in the left panel. The right panel displays labeled proteins after enrichment. **(C)** Volcano plots of probe **4**/DMSO and probe **4**/competition. A p-value of 0.05 and a Log2 fold change>1 (2-fold enrichment) were taken as cut-off values. The plot indicates that 8 proteins were enriched versus DMSO and 4 proteins versus competition with the parent compound. **(D)** Left: Venn diagram of the enriched proteins compared with the DMSO control (in red) and compared with the competition control (in blue). Right: Bargraph of PSMs for each of the 11 identified hits. CTSD = cathepsin D; FDPS = farnesyl pyrophosphate synthase, GALK1 = galactokinase, GLRX5 = glutaredoxin-related protein 5, HK1 = hexokinase-1, NAXE = NAD(P)H-hydrate epimerase, NOLC1 = nucleolar and coiled-body phosphoprotein 1, MYL12B = myosin regulatory light chain 12B, PEA15 = astrocytic phosphoprotein PEA15, PPP2CA = serine/threonine-protein phosphatase 2a catalytic subunit alpha, S100A11 = protein S100-A11. **(E)** Left: Venn diagram of proteins above the cutoff value from D-SCRIPT (in blue) and Topsy-Turvy (in yellow). Right: Prediction scores resulted from the deep learning algorithms. **(F)** Incubation of His$_6$-tagged SQSTM-1 with cathepsin D for increasing amounts of time reveals degradation and appearance of degradation products, which is inhibited by pepstatin A (detected by anti-His$_6$ western blot). **(G)** Pepstatin A partially counteracts the starvation-induced decrease of SQSTM1 in T-REx 293 cells. A quantification of 4 biological replicates is provided in panel H. **(H)** Quantification of the

SQSTM1 levels in T-REx 293 cells upon starvation in the absence or presence of pepstatin A. The results are normalized to GAPDH, and presented as the average (± the standard deviation) of 4 biological replicates. ****, $p < 0.0001$; *, $p < 0.05$. Significance was tested by One-way ANOVA with Tukey post hoc test.

4.1.5 Deep learning prediction of cathepsin D substrates

In addition to cathepsin D, 10 other proteins were identified as enriched when compared to controls. We hypothesized that cathepsin D interaction partners and substrates were co-enriched via protein-protein interactions (PPIs) near the active site or at a possible exosite. We used qualitative data analysis (Table S1) as a less strict filtering criterion to determine whether our dataset contained unknown interacting proteins as potential substrates since PPIs can be weak and may not be detected in all replicates. Two recently developed sequence-based deep learning algorithms (D-SCRIPT [183] and Topsy-Turvy[198]), were then used to this expanded list (Table S1) to determine if these proteins might interact with cathepsin D. Following model training, quality control of the generated models revealed that both were of sufficient quality for prediction (Figure S6). We identified 6 predicted PPIs using the models that earned scores meeting both D-SCRIPT and Topsy-Turvy cut-offs (Figure 15E; see also Table S2).

4.1.6 In-vitro validation suggest SQSTM1 as a substrate of cathepsin D

Sequestosome-1 (SQSTM1, also known as p62), which is predicted to interact with cathepsin D, drew our attention because both have a function in autophagy. SQSTM1 is a multidomain protein that functions as both a signaling hub for mTORC1, which is a regulator of autophagy among other cellular processes, and an adapter for ubiquitinylated proteins intended for autophagosome degradation. [199] SQSTM1 is degraded during autophagy, hence its accumulation has been utilized as a marker of autophagy disorders. We first checked whether SQSTM1 is a PAL target, either directly as a target of the probe or indirectly through cathepsin D interaction, but no labeling of SQSTM1 was observed in presence or absence of cathepsin D (Figure S7). We then incubated purified recombinant, His-tagged SQSTM1 with or without purified cathepsin D to see if SQSTM1 would be a substrate of cathepsin D. In the presence of cathepsin D, we observed that SQSTM1 was gradually digested, resulting in lower-running digestion products (Figure 25F, Figure S8). Pepstatin A inhibited this digestion, indicating that SQSTM1 is a cathepsin D substrate. Then, we conducted tests on living cells. After being starved for 24 hours, T-REx 293 cells demonstrated decreased SQSTM1 levels. Notably, the cathepsin D inhibitor pepstatin A partially but significantly reduced this decrease (Figure 25G-25H). Our findings suggest that SQSTM1 can act as a substrate of cathepsin D. This is consistent with a previous study that found increased amounts of SQSTM1 in several cell lines after pepstatin A incubation and knock-down or knock-out of cathepsin D. [200] Overall, our results add to the evidence that SQSTM1 is a cathepsin D substrate.

4.2 Mass cleavable affinity-based probes for precise mapping of binding hotspots

4.2.1 Probe design and Synthesis

The carbonyl sulfoxide-type linker is a typical CID-cleavable linker which undergoes a McLafferty-type rearrangement at a lower fragmentation energy than the peptide backbone (Chapter 1.6, Figure 10A). The unexpected fragmentation of crosslinked peptides can be potentially reduced by installing a CID-cleavable linker next to the photo-crosslinking moiety of a PAL probe. In addition, the unique cleavage of carbonyl sulfoxide during MS^2 allows a supplementary validation of photo-crosslinked peptides and enables a better crosslinking site assignment in the following MS^3 event.

Pepstatin is a natural substance obtained from Actinomycetes [191] that inhibits aspartic proteases in a broad spectrum and reversible manner [192]. Three amino acids, two statine residues, and an N-terminal isovaleroyl cap make up its structure (Figure 26A). Pepstatin's inhibitory effect is based on imitating a substrate's tetrahedral intermediate, which is created when a water molecule attacks the scissile link (Figure 26B). According to published crystal structures (Figure 26C), the binding mechanism of pepstatin A is similar for different aspartic proteases. Importantly, adequate space seems to be available for a carbonyl sulfoxide diazirine moiety as a 'minimal MS-cleavable photocrosslinker' inside the pepstatin structure for different residues. As a result, we decided to replace different amino acids in the pepstatin structure, as well as the C-terminal statine residue, with a homemade building block (**4**, Figure 26D), and the N-terminal acyl group with a homemade building block **7**. The use of an alkyne as a C-terminal tag enables biorthogonal click chemistry to detect the photo-crosslinked probe-protease complexes.

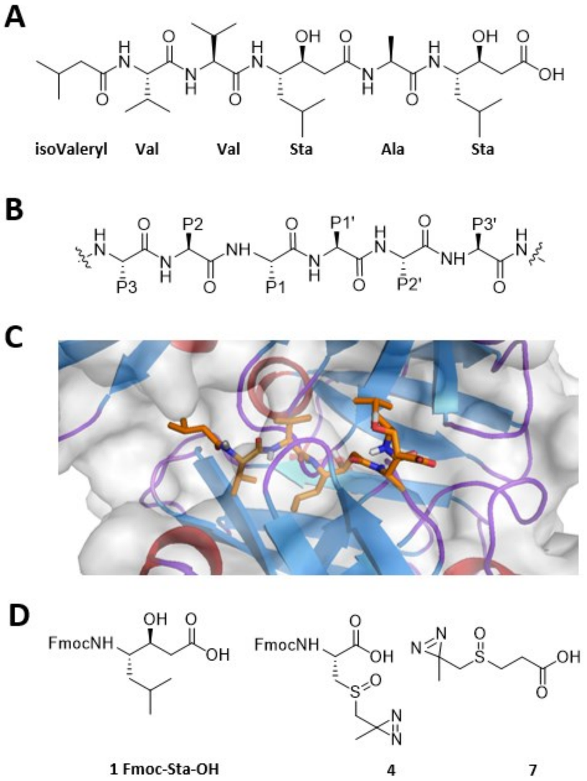
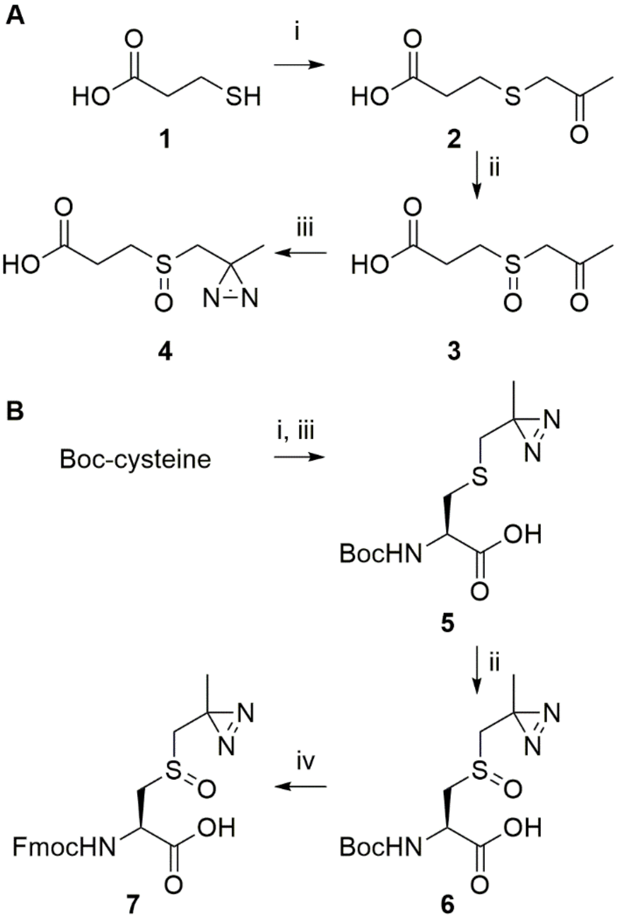


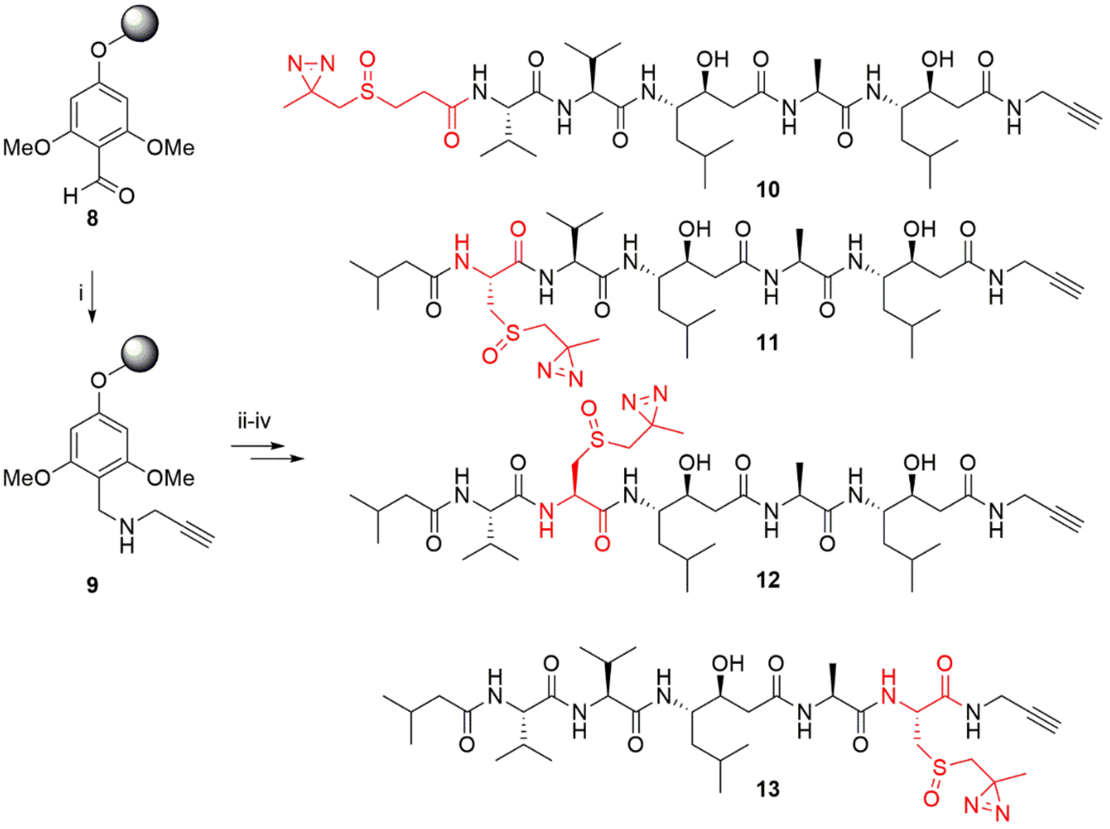
Figure 26. The general aspartic protease inhibitor Pepstatin. (A) Chemical structure of pepstatin. **(B)** Chemical structure of a protease substrate. Amino acid residues at the N-terminal side are named P1, P2 etc., whereas the residues at the C-terminal side are denoted with an apostrophe (P1' etc). **(C)** Crystal structure of pepstatin in the active site of Cathepsin D (PDB code: 1LYB).17 Protein depicted in cartoon mode with sheets in blue, helices in red and random coil in purple. Protein surface is indicated in transparent grey. Pepstatin is depicted in stick model (orange). Picture rendered with PyMol. [6] **(D)** Chemical structures of Fmoc-protected statine **1,** Fmoc-protected diazirine building block **4** and diazirine building block **7**.

The two building blocks **4** and **7** were easily accessed by a straightforward synthesis from mercaptoproprionic acid or Boc-cysteine (Scheme 8). In the first step, the thiol groups are alkylated with chloroacetone. Next, the sulfur is oxidized to the sulfoxide and the ketone moiety is converted into a diazirine. For the cysteine-based building block, the Boc protective group is then replaced by the Fmoc group to enable Fmoc-based solid phase peptide synthesis (SPPS). The protecting group manipulation was necessary, as the Fmoc is incompatible with diazirine formation, which happens under strong basic conditions.



Scheme 8. Reagents and conditions: (i) chloroacetone, Et_3N, THF, 0 °C. (ii) $NaIO_4$, MeOH/H_2O = 7:3 . (iii) 1. NH_3 - MeOH, then NH_2OSO_3H, **solvent**; 2. I_2, Et_3N, MeOH. (iv) 1. TFA, DCM; 2. Fmoc-Cl, Na_2CO_3, Dioxane/H2O = 1:1.

Conveniently, the use of a backbone amide linker (BAL) resin [195] enabled the elongation of the C-terminally-tagged pepstatin probes on a solid support. To summarize, a reductive amination was employed to attach a propargylamine to the BAL resin. Followed by the Fmoc-based solid phase peptide elongation and cleavage via trifluoroacetic acid, four pepstatin probes with a different diazirine location were obtained (**10-13**) (Scheme 9).



Scheme 9. Solid phase synthesis of pepstatin probes **10-13**. Photoreactive residue indicated in red. *Reagents and conditions*: (i) propargylamine hydrochloride (10 eq.), $NaBH_3CN$ (10 eq.), AcOH (1%) in NMP. (ii) Fmoc-Sta-OH or building block **7** (3 eq.), HBTU (3 eq), DIEA (6 eq), DMF. (iii) Elongation of the peptide chain by repeated Fmoc-deprotection and coupling of building blocks: a. 20% piperidine in DMF; b. Fmoc-Val-OH, Fmoc-Ala-OH, Fmoc-Sta-OH, building block **7** or building block **4** (3 eq.), HBTU (3 eq), DIEA (6 eq), DMF. (iv) TFA/TIPS/H_2O (95/2.5/2.5).

4.2.2 Labeling of pure porcine pepsin and bovine chymosin

We performed labeling experiments on pure porcine pepsin and bovine chymosin as two model aspartic proteases by using the four different probes we had at hand. Titrations of the probes were carried out to identify the best concentration for use. To achieve this, 50 pmol of pure enzyme was incubated with varying concentrations of probes **10-13** under 365 nm irradiation, followed by the introduction of a TAMRA-fluorophore via Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) (Figure 27). The labeling of pepsin is observed in all probes, and the most intense labeling is given by probe **11** and **13**. For chymosin, probe **13** gave a sensitive labeling at 1 μM while probes **12** and **13** provided the most intense around 10 μM. As anticipated, the position of the photoreactive group significantly affects the labeling capacity, which is also influenced by the target protein as illustrated in Figure 27.

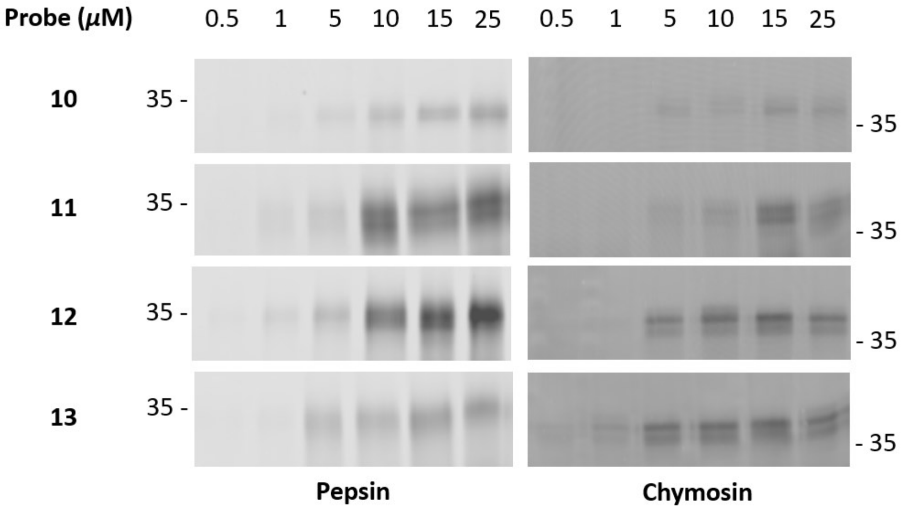


Figure 27. Labeling of purified proteases by pepstatin-based AfBPs **4-7**. Photoaffinity labeling of 50 pmol of porcine pepsin (left panels) or bovine chymosin (right panels) by increasing concentrations of probe (0.5, 1, 5, 10, 15, 25 μM) under 365 nm UV irradiation with visualization by CuAAC with a TAMRA-azide tag.

4.2.3 Mass spectrometry (MS) method development Fragmentation energy investigation at probe level

Fragmentation energy is one of the key parameters for the MS method. Having shown that the pepstatin A based probes lead to PAL of their aspartic protease targets, we next investigated the cleavage of the sulfoxide moiety within these probes in order to optimize MS conditions. Irradiation of probe **11** in buffer led to three different species (Figure 28A, (I)-(III)). Stepwise increase of the collision energy from 19% to 31% gave rise to sulfoxide linker cleavage with the concomitant appearance of two different reporter ions, one of which included a neutral loss of H_2O (Figure 28A). As expected, with increasing collision energy, the precursor ion became less abundant and the two reporter ions became more intense (Figure 28B, Figure S9).

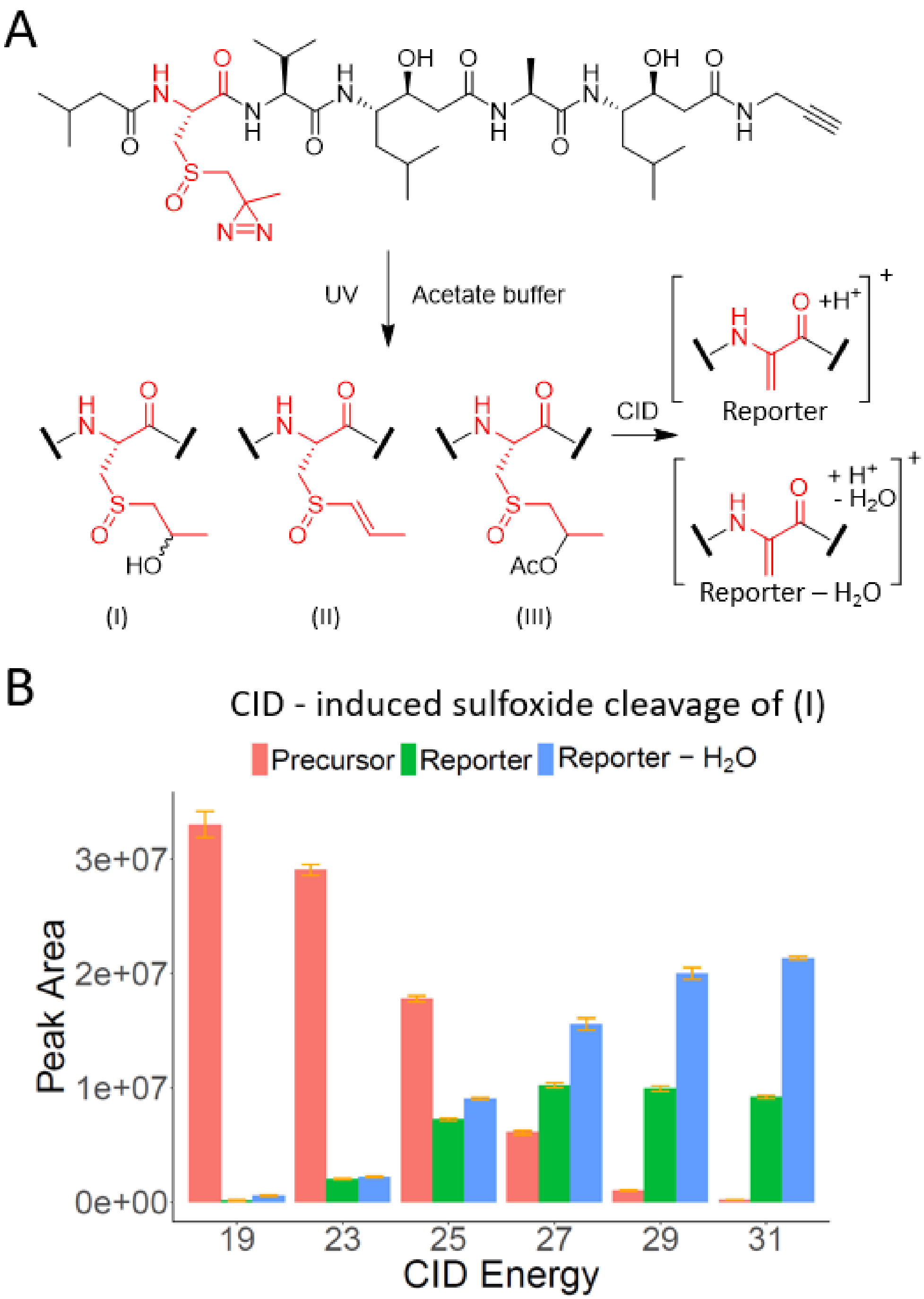


Figure 28. **(A)** Irradiation of probe 11 in acetate buffer gave (I), (II) and (II) as major product. Collision-induced dissociation (CID) in mass spectrometry gave [Reporter] and [Reporter – H_2O] as major fragment. **(B)** Peak area (Error bar: mean of 3 replicates, standard deviation of peak area of (I) and its fragments ([Reporter] ($C_{35}H_{61}N_6O_8$, theoretical m/z: 693.4550, z: 1, mass tolerance: +/- 5ppm), [Reporter – H_2O] ($C_{35}H_{59}N_6O_7$, theoretical m/z: 675.4445, z: 1, mass tolerance: +/- 5ppm)) at different normalized CID energy.

Identification of crosslinking peptides

Reporter ions were detected in a good intensity when 25 % - 31% of CID energy was applied. To identify crosslinking peptides, we then created a squared data-dependent acquisition (DDA^2) method using reporter ions as mass trigger of a second data-dependent acquisition on the MS^2 fragments. (Figure 29) To test this, chymosin was incubated with probe **11** and crosslinked under UV irradiation, followed by a Glu C / tryptic digestion. Ammonium bicarbonate buffer (50 mM, pH 7.8) was used as digestion buffer to promote preferential hydrolases of GluC C-terminally of glutamic acid. [201] We then analyzed digested peptides with the DDA^2 method at different collision energies.

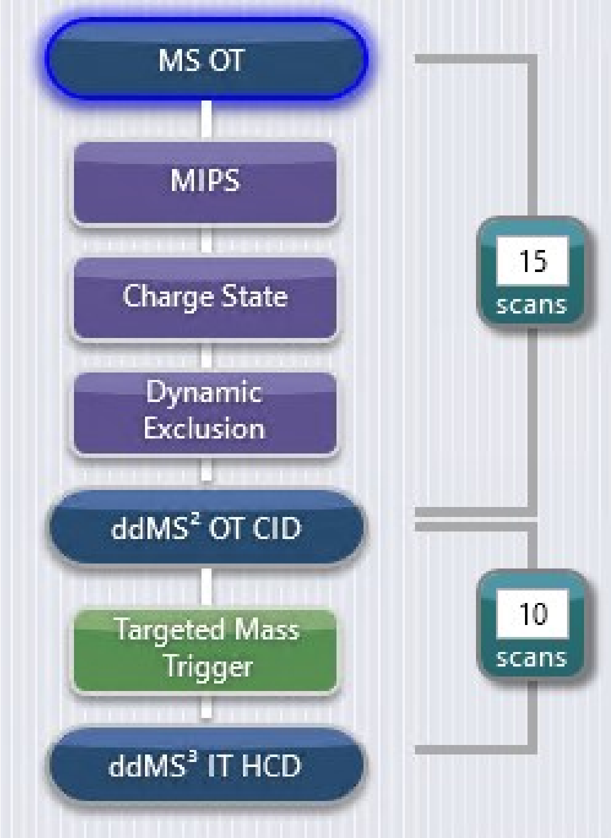


Figure 29. Method workflow of squared data-dependent acquisition (DDA^2). **MS OT**: Orbitrap Resolution: 120K, Scan Range (m/z): 350-2000, RF Lens (%): 35, Maximum Injection Time (ms): 150. **MIPS** (Monoisotopic Peak Determination): Peptide. **Charge state**: 2-7. **Dynamic Exclusion** (of MS^2 event): Exclude after 1 Time, Exclusion Duration (S): 30, Mass Tolerance: +/- 10 ppm, exclude Isotopes. **ddMS^2 OT CID** (Data-Dependent MS^2 scan): Isolation Window (m/z): 2, Activation Type: CID, Collision Energy Mode: Fixed, CID Collision Energy (%): 19/21/23/25/27, Detector Type Orbitrap, Orbitrap Resolution: 30K, Maximum Injection Time (ms): 150, Microscans: 1. **15 Scans** (MS^2): Scan top 15 most abundant ions in the MS^1 spectrum. **Targeted Mass Trigger**: Mass List Type m/z, Mass List ($C_{35}H_{58}N_6O_7S$ [M+H]+ 675.4445, $C_{35}H_{60}N_6O_8S$ [M+H]+ 693.4554), Mass Tolerance: +/- 10 ppm, Trigger Only with Detection of at Least 1 Ion from the list, Trigger Only Ion(s) within Top 15 Most Intense. **ddMS^3 IT HCD** (Data-Dependent MS^3 scan): MS^n Level 3, MS Isolation Window (m/z): 2.5, MS^2 Isolation Window (m/z): 2, Activation Type: HCD, HCD Collision Energy (%): 35, Detector Type: Ion Trap, Ion Trap Scan Rate: Normal, Maximum Injection Time (ms): 100, Microscans: 1. **10 Scans** (MS^3): Scan top 10 most abundant ions in the MS^2 spectrum.

MS^3 data were extracted and initially searched against the bovine proteome. Two peptides modified with a ‘mini-tag’ (C_3H_6OS or C_3H_6OS-H_2O) were identified as photo-crosslinked peptides from bovine chymosin (CYM). (Tabel 8, Entry 1-2) Some peptides from other proteins were given by the search engine with modifications by multiple tags (Tabel 8, Entry 3-6). However, their corresponding precursor m/z at

MS^1 didn't match to search results. And these identifications turned out to be false positive.

Tabel 8, Peptides identification after database search against bovine proteome.

Entry	Gene	Accessions	Annotated Sequence	Modifications	PSMs*
1	**CYM**	**P00794**	**[K].MYPLTPSAYTSQDQGFCTSGFQSENHSQK.[W]**	**1xCarbamidomethyl [C17]; 1xOxidation [M1]; 1xC3H6OS [E/F/G/T]**	**3**
2	**CYM**	**P00794**	**[E].VASVPLTNYLDSQYFGK.[I]**	**1xC3H6OS-H2O [D/L/Q/S/T/Y]**	**3**
3	SSUH2	F1N504	[E].ALLSFVNSKCCYGSAAASDLVILELKQQNLCR.[Y]	1xCarbamidomethyl [C]; 2xC3H6OS [L25; L30]; 1xC3H6OS-H2O [Q27]	2
4	SELENOT	A6QP01	[K].LESGHLPSMQQLVQILDNEMKLNVHMDSIPHHR.[S]	1xOxidation [M26]; 1xC3H6OS [L22]; 2xC3H6OS-H2O [H25; H31]	1
5	SLC39A4	A0A3Q1NDU1	[K].TGLATSLAVFCHEVPHELGEPCGVPAGRR.[R]	1xCarbamidomethyl [C11]; 1xC3H6OS [A8]; 2xC3H6OS-H2O [P15; L18]	1
6	BPIFB4	A0A3Q1M7V6	[E].VMVSQPNDVETTICLIDVVSGGGR.[S]	1xC3H6OS [C14]; 3xC3H6OS-H2O [I13; G22; G23]	1

* Total PSMs from triple measurements of 5 CID conditions.

After UV irradiation, diazirine can couple with any amino acid residue that is in spatial proximity. Therefore, we took every amino acid residue into consideration when database searches were performed. Note that this creates a huge search space for the search engine and this might be the reason of the above described false positive identifications. MS^3 data were extracted and searched against the single sequence of CYM. The same modified peptides with 'mini-tag' (C_3H_6OS or C_3H_6OS-H_2O) were identified as photo-crosslinked peptides (Tabel 9, Entry 1-3), while peptides modified by multiple tags (Tabel 9, Entry 4-5), as well as non-modified peptides (Tabel 9, Entry 1-2) were found to be false positive identifications. Benefiting from the single sequence searching, tag-modified peptides had more peptide-spectrum matches (PSMs) and one more additional tag modified peptide was identified (Tabel 9, Entry 1). Despite this, some Glu C cleavage events C-terminally to aspartic acid were also found when doing specificity validation of Glu C cleavage. However, the identified peptide (6 amino acid with probe modification) was too short for further investigation. (Table S3, Entry 1 and 2)

Tabel 9, Peptides identification after database search against bovine chymosin (CYM).

Entry	Gene	Accessions	Annotated Sequence	Modifications	#PSMs*
1	**CYM**	**P00794**	**[K].MYPLTPSAYTSQDQGFCTSGFQSENHSQK.[W]**	**1xCarbamidomethyl [C17]; 1xOxidation [M1]; 1xC3H6OS-H2O [A/D/G/H/P/Q/S/T]**	4
2	**CYM**	**P00794**	**[K].MYPLTPSAYTSQDQGFCTSGFQSENHSQK.[W]**	**1xCarbamidomethyl [C17]; 1xOxidation [M1]; 1xC3H6OS [A/D/E/F/G/K/L/N/P/Q/S/T/Y]**	13
3	**CYM**	**P00794**	**[E].VASVPLTNYLDSQYFGK.[I]**	**1xC3H6OS-H2O [A/D/F/G/K/L/N/P/Q/S/T/V/Y]**	86
4	CYM	P00794	[E].ITRIPLYK.[G]	2xC3H6OS [R3; P5]; 1xC3H6OS-H2O [I1]	1
5	CYM	P00794	[E].VASVPLTNYLDSQYFGK.[I]	3xC3H6OS-H2O [D11; F15; G16]	1
6	CYM	P00794	[K].WILGDVFIR.[E]	None	8
7	CYM	P00794	[K].WILGDVFIRE.[Y]	None	44

*Total PSMs from triple measurements of 5 CID conditions.

As expected, with increasing collision energy, the MS^3 event was triggered in the DDA^2 by the increasing abundance of reporter ions and resulted in more PSMs of tag-modified peptide after database search. Taking peptide VAS***FGK as an example, a substantial increase of PSMs was observed when CID energy was increased from 19%-25% (Figure 30). Nevertheless, the amount of PSMs was slightly decreased when CID energy increased to 27%. We observed similar behavior with MYP***SQK peptide. Due to the lower abundance of its precursor, however, much less PSMs were identified (Figure S10).

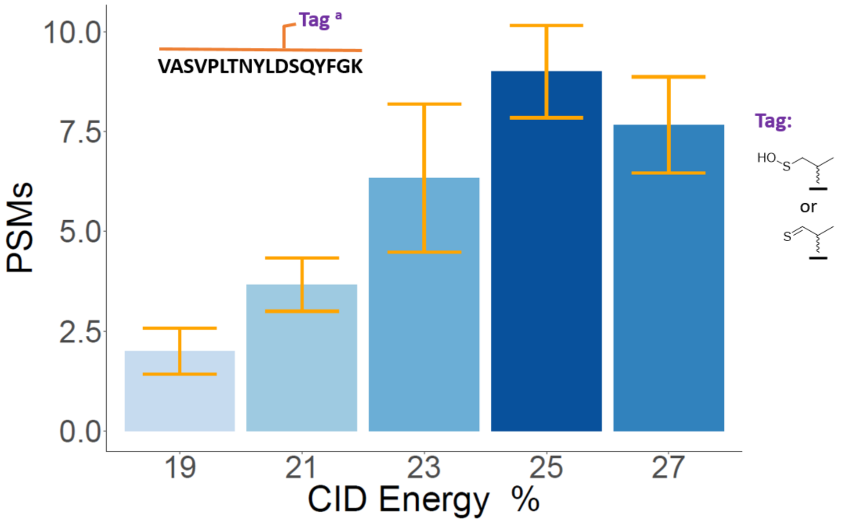


Figure 30. PSMs of tag-modified peptide VAS***FGK from data-dependent acquisition square (DDA^2) analysis. Error bar: mean of 3 replicates, standard deviation of PSMs.

In addition, different isomers of crosslinked peptides (probe-modified VAS***FGK) were observed at different retention time in MS^1 (Figure 31A). Due to the structure of sulfoxides and the mechanism of diazirine photo-crosslinking, we suspect that these arise because of the chiral center at the sulfoxide as well as the chiral center at the photo-crosslinking site (see Figure 31F) or different crosslinking species on different amino residues. Although lower abundant photo-crosslinked isomers (Peaks at 61.0 min, 61.5 min and 70.3 min) were observed in the MS^1 scan, only a few MS^3 spectra of the highly abundant peak were acquired in the DDA^2 measurements. To solve this problem, we chose for a targeted MS method for increasing the number of PSMs of modified peptides.

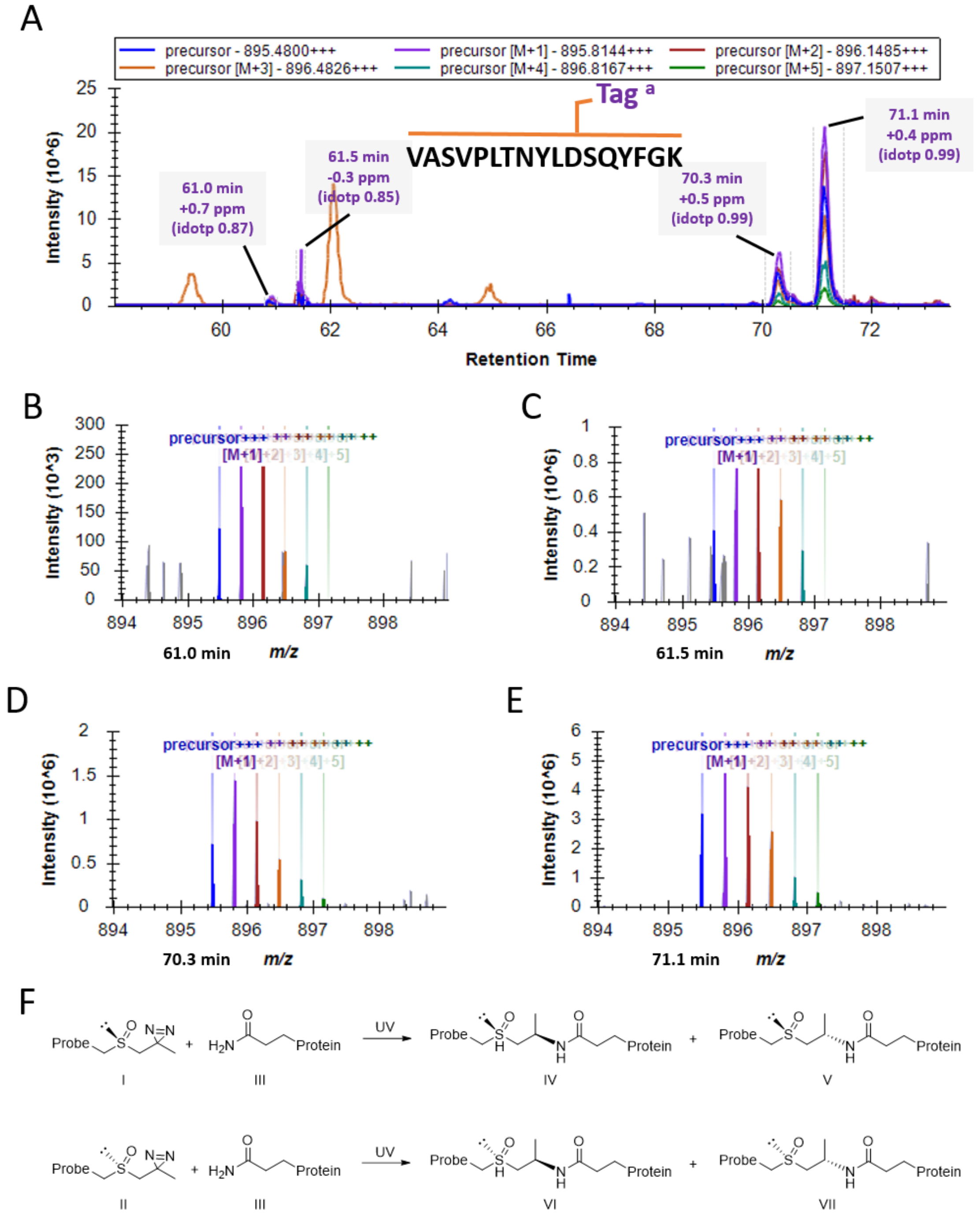


Figure 31. **(A)** Extract ion chromatogram of probe-modified VAS***FGK (theoretical m/z: 895.4800, z: 3, mass tolerance: +/- 5ppm). **(B)** Mass spectrum of peak at 61.0 min. **(C)** Mass spectrum of peak at 61.5 min. **(D)** Mass spectrum of peak at 70.3 min. **(E)** Mass spectrum of peak at 71.1 min. **(F)** Asymmetric chemistry of SODA probe. **I:** S-sulfoxide configuration of SODA; **II:** R sulfoxide configuration of SODA; **III:** Glutamine residue of protein; **IV:** S-sulfoxide, R configuration of photo-crosslinking product; **V:** S-sulfoxide, S configuration of photo-crosslinking product; **VI:** R-sulfoxide, R configuration of photo-crosslinking product; **VII:** R-sulfoxide, S configuration of photo-crosslinking product.

Optimization of CID energy

The results from DDA^2 analysis indicated that the PSMs of photo-crosslinking can be increased by raising the CID energy, however, a decrease of PSMs can also be observed when the CID energy was increased above 25% (Figure 30). Fragmentation of peptide bonds is one of the main reasons. It would be desired that the CID-induced sulfoxide cleavage at MS^2 level would take place selectively over peptide bond fragmentation. To find an optimal balance between these two processes (cleavage of sulfoxide and fragmentation of peptide bonds in the MS^2 event), we then analyzed digested peptides with a scheduled parallel reaction monitoring square (PRM^2) method using reporter ions as mass trigger of a secondary parallel reaction monitoring and subjected crosslinked peptides to different collision energies (Figure 32).

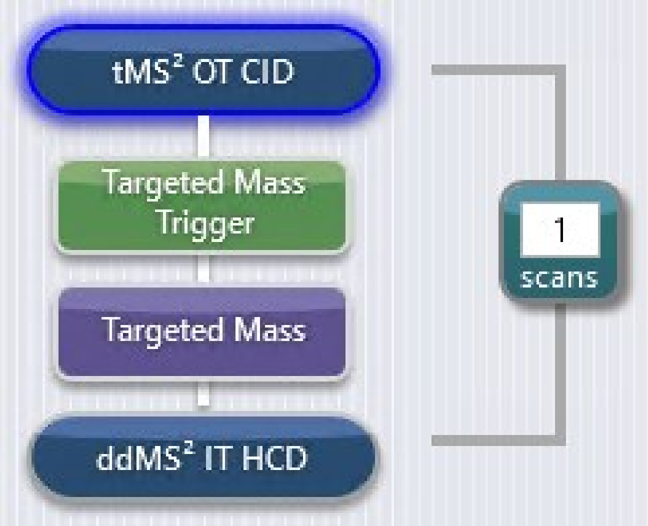
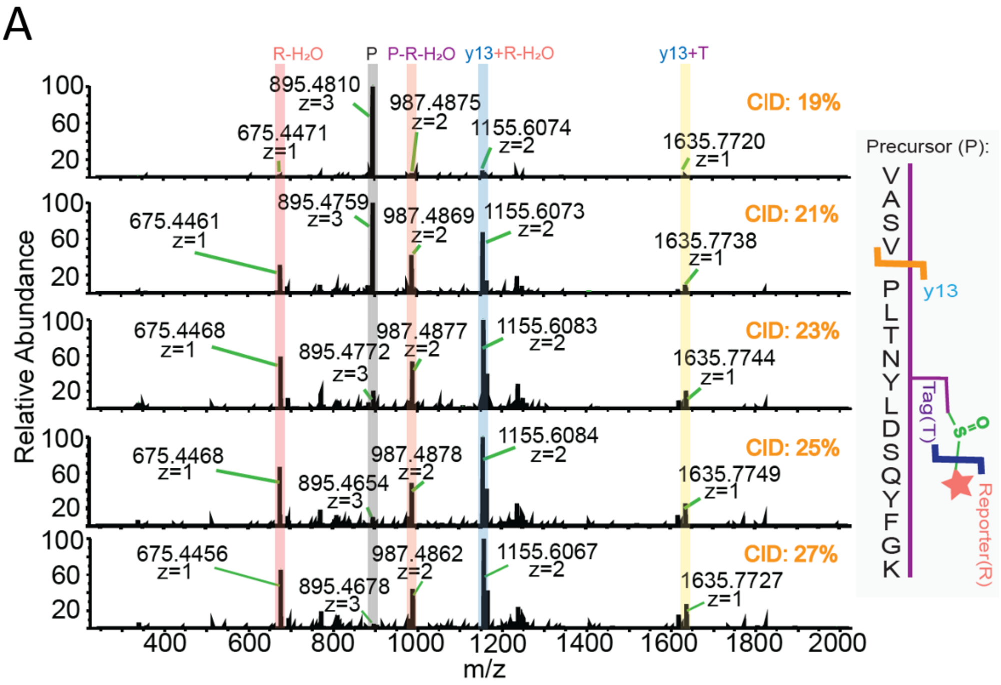


Figure 32. Method workflow of parallel reaction monitoring square (PRM^2). **tMS^2 OT CID** (Targeted MS^2 scan): Isolation Window (m/z): 2, Activation Type: CID, CID Collision Energy (%): 19/21/23/25/27, Detector Type: Orbitrap, Orbitrap Resolution: 30K, RF lens (%): 35, Maximum Injection Time (ms): 150, Mass List (Tabel 1). **Targeted Mass Trigger**: Mass List Type: m/z, Mass List ($C_{35}H_{58}N_6O_7S$ [M+H]+ 675.4445, $C_{35}H_{60}N_6O_8S$ [M+H]+ 693.4554), Mass Tolerance: +/- 10 ppm, Trigger Only with Detection of at Least 1 Ion from the list, Trigger Only Ion(s) within Top 10 Most Intense, Ignore Charge State Requirement for Unassigned Ions. **Targeted Mass**: Mass List Type: m/z & z, Time mode: Start/End Time, Mass list (Table 2), Mass Tolerance: +/- 10 ppm, Ignore Charge State Requirement for Unassigned Ions. **ddMS^2 IT HCD** (Data-Dependent MS^2 scan): MS^n Level: 2, MS Isolation Window (m/z): 2, MS2 Isolation Window (m/z): 2, Activation Type: HCD, Collision Energy Mode: Fixed, HCD Collision Energy (%): 35, Detector Type: Ion Trap, Ion Trap Scan Rate: Normal, Maximum Injection

Time (ms): 250, Normalized AGC Target (%): 100 or 250, Microscans: 1 or 5. **1 Scans** (MS^3): Scan top 1 most abundant ions in the MS^2 spectrum.

Taking probe-modified peptide VAS***FGK as an example, the precursor ion (P; Figure 33A) stayed mostly intact under 19% CID collision energy. We found that around 99% of modified precursors were cleaved under 25% CID collision energy and resulted in formation of reporter ions (figure 33A, R-H_2O) as well as peptide fragment ions without the probe (Figure 33A, P-R-H_2O), but with the 'mini-tag' modification (C_3H_6OS or C_3H_6OS-H_2O) (Figure 33B). We also noticed that some premature peptide bond fragmentation occurred, but mostly N-terminal to proline due to the "Proline Effect". [202] [203] (Figure 33A, y13+R-H_2O, y13+T) Moreover, we found the most of PSMs of peptide VAS***FGK under 23% CID collision energy after database searching (Figure 33C).



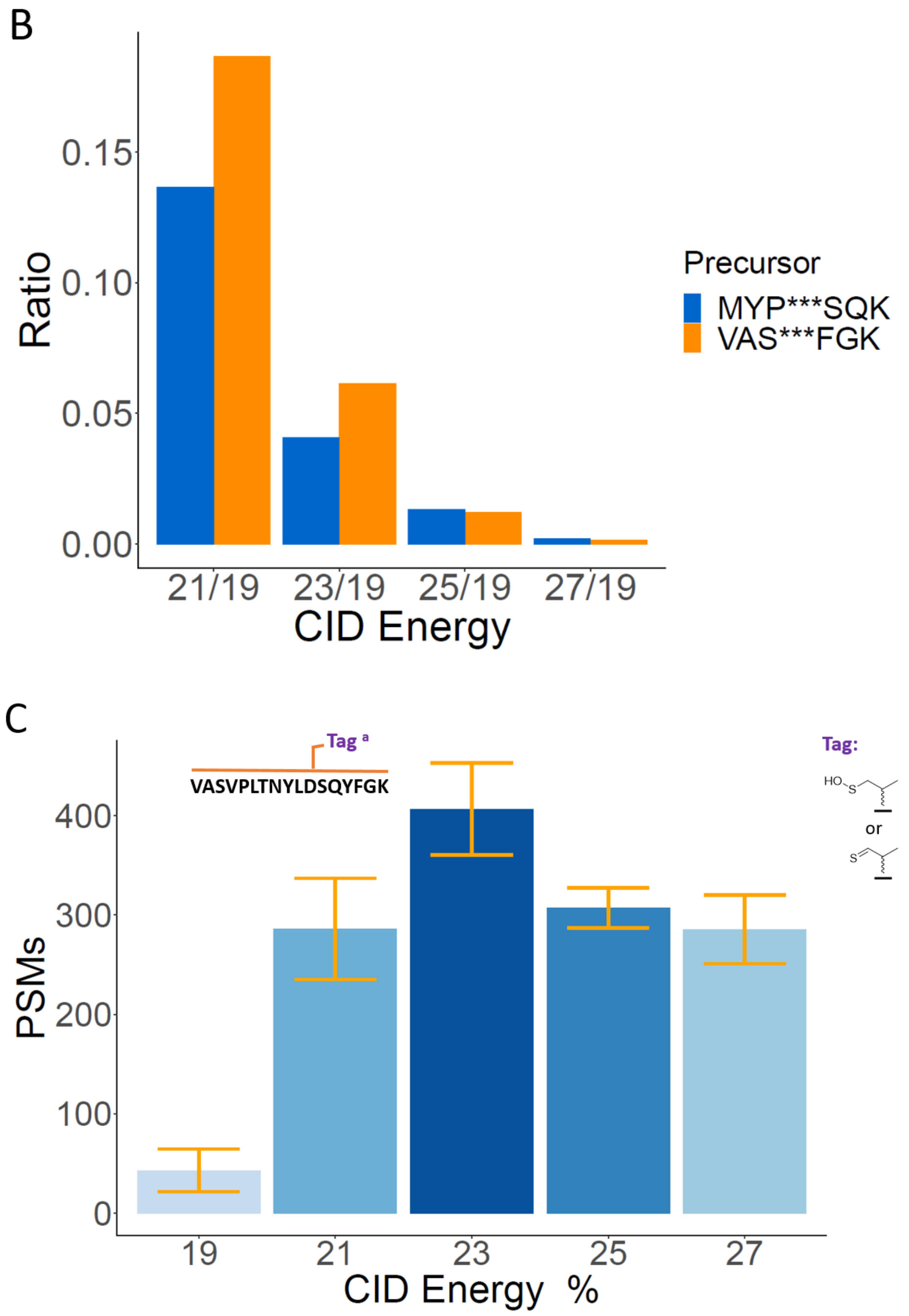


Figure 33. **(A)** MS^2 spectrum of probe-modified peptide VAS***FGK (m/z: 895.4800, z: 3) at 19, 21, 23, 25, 27 % of CID activation energy; P: precursor, probe-modified peptide VAS***FGK (theoretical m/z: 895.4800, z: 3); R - H_2O: Reporter - H_2O ($C_{35}H_{58}N_6O_7S$, theoretical m/z: 675.4445, z: 1); P - R - H_2O: Precursor - Reporter - H_2O (theoretical m/z: 987.4872, z: 2); y13+R-H_2O: PLT*** FGK + Reporter - H_2O (theoretical m/z: 1155.6081, z: 2); y13+T: PLT*** FGK + Tag (theoretical m/z: 1635.7723, z: 1). **(B)** Peak area ratio of precursors. Ratio = mean(peak area of triplicates) $_i$ / mean(peak area of triplicates) $_j$; i: CID 21, 23, 25 or 27 %; j: CID 19 %. **(C)** PSMs of tag-modified peptide VAS***FGK from parallel reaction monitoring square (PRM^2) analysis. Error bar: mean of 3 replicates, standard deviation of PSMs.

The degree of fragmentation is substantially influenced by the activation energy, as well as the molecular size. [204] A large molecule fragments much slower than a small molecule. If the activation energy of the fragmentation reaction is very low, there will be a large amount of fragmentation even at modest collision energies; if the activation energy of the fragmentation reaction is high, then higher collision energy is needed to observe fragments. Fragmentation processes of peptides are likely characterized by similar activation energy. In such a case, peptides of similar size will require similar amount of internal energy to observe analogous MS/MS spectra. However, with a labile functional group or increasing charge states of the precursor, the necessary collision energy will be lower or much lower. [205]

Although the charge dependence is taken into account by the "normalized collision energy" (nCE) setting of Thermo Fisher Orbitrap mass spectrometers, [206] [207] we observed different behavior on different size and different charge states of probe modified peptides. Desired fragment ions of probe-modified peptide MYP***SQK (theoretical m/z = 1020.4772, z = 4), reporter ion (R-H_2O) and 'mini-tag' modified peptide ion (P-R-H_2O), reached to the maximum intensity under 21% CID collision energy, while fragment ions of probe-modified peptide VAS***FGK (theoretical m/z = 895.4800, z = 3) reached to the maximum intensity under 23% CID collision energy (Figure 34, Figure S11). To get good fragmentation of sulfoxide from the probe-modified peptide, it was decided to utilize 23% collision energy at the MS^2 fragmentation as a general parameter.

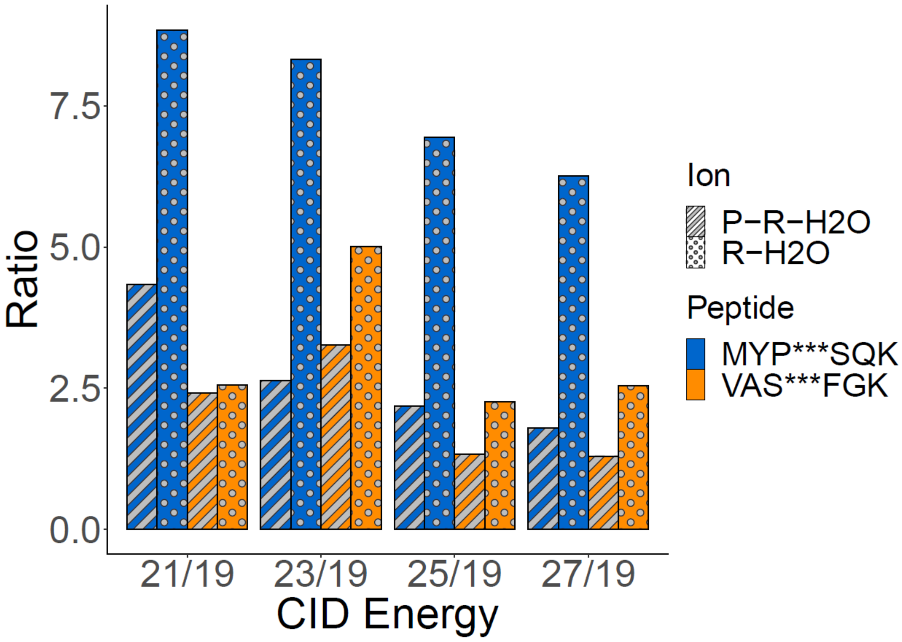


Figure 34. Peak area ratio of product ions. Ratio = mean(peak area of triplicates)$_i$ / mean(peak area of triplicates)$_j$; i: CID 21, 23, 25 or 27 %; j: CID 19 %. P: precursor, probe-modified peptide VAS***FGK (theoretical m/z: 895.4800, z: 3), probe-modified peptide MYP***SQK (theoretical m/z: 1020.4772, z: 4); R - H_2O: Reporter - H_2O ($C_{35}H_{58}N_6O_7S$, theoretical m/z: 675.4445, z: 1); P - R (- H_2O): Precursor - Reporter - H_2O (VAS***FGK , theoretical m/z: 987.4872, z: 2), Precursor - Reporter (MYP***SQK, theoretical m/z: 1129.4848, z: 3).

Optimization of other key parameters

Better MS^3 spectra quality can lead to a better identification of photo-crosslinked peptides. The desired 'mini-tag' modified peptides were obtained by two times of m/z isolation and one time CID fragmentation (m/z isolation - CID fragmentation - m/z isolation) in the tribrid mass spectrometer. Therefore, the MS^3 scan is usually dealing with low abundant ions that are obtained after the isolation and fragmentation. The linear ion trap of tribrid mass spectrometer allows several 'microscans' to average several spectra together and improve signal-to-noise ratio.[208] When the 'microscans' parameter was increased from 1 to 5 in the PRM^2 measurements, the spectral quality of low abundant 'mini-tag' modified peptides was substantially increased. The percentage of high quality PSMs (sum of triplicates, Xcorr >= 2) of 'mini-tag' modified peptides VAS***FGK_1 (rt, 59.5-62.0 min) was increased from 15.87% to 82.32% and the percentage of high quality PSMs of 'mini-tag' modified peptides MYP***SQK was increased from 6.48% to 43.33%. For the highly abundant modified peptide VAS***FGK_2 (rt, 70.0 - 72.0 min), high quality PSMs could reach to 60.47% with the default setting of the linear ion trap, but were increased to 92.93% when 'microscans' was increased to 5 (Figure 35).

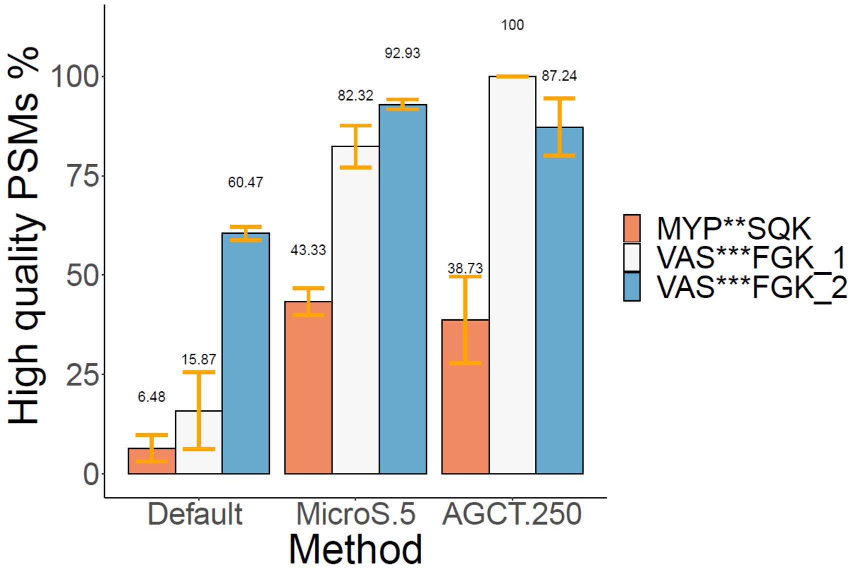


Figure 35. High quality PSMs of PRM^2 analysis. High quality PSMs % = PSMs (Xcorr >= 2) / PSMs × 100%. Error bar: mean of 3 replicates, standard deviation of High quality PSMs %. Default **(**in $ddMS^2$ IT HCD setting): Normalized AGC Target (%) 100, Microscans 1. MicroS.5 **(**in $ddMS^2$ IT HCD setting): Normalized AGC Target (%) 100, Microscans 5. AGCT.250 **(**in $ddMS^2$ IT HCD setting): Normalized AGC Target (%) 250, Microscans 5.

The identification of the crosslinking site of low abundant peptide is challenging. As shown in figure 36, only 3 PSMs (Xcorr >= 2) were identified for the peptide VAS***FGK_1 (rt, 59.5-62.0 min) with 'mini-tag' modification on the serine (S12), and none of them were unambiguous PSMs (Figure 36 A, B). A substantial improvement of PSMs quality can be achieved when the 'microscans' parameter was increased to 5, which enables the spectral counting comparison between potential photo-crosslinking sites. (Figure 36 C, D) Because most of the unambiguous PSMs were identified as VAS***FGK_1 with 'mini-tag' modification on the Glutamine 'Q13' modification, we think that the most efficient photo-crosslinking potentially takes place at 'Q13' (Figure 36 D).

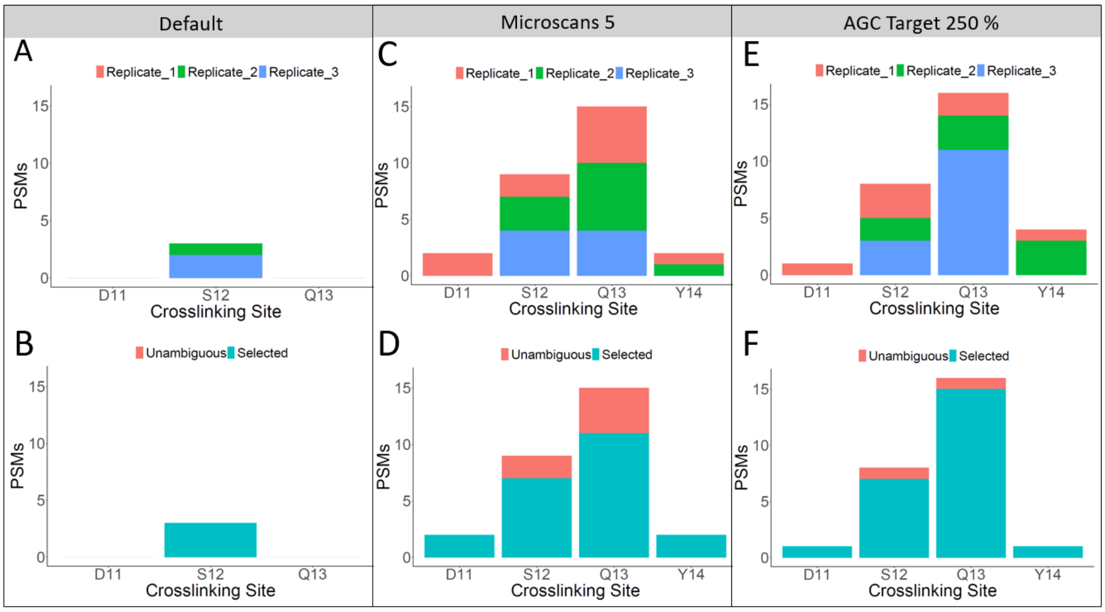


Figure 36. Crosslinking site of PSMs ('mini-tag' modified VAS***FGK_1). **Default (in ddMS^2 IT HCD setting):** Normalized AGC Target (%) 100, Microscans 1. Microscans 5 **(**in ddMS^2 IT HCD setting): Normalized AGC Target (%) 100, Microscans 5. **AGC Targent 250 (in ddMS^2 IT HCD setting):** Normalized AGC Target (%) 250, Microscans 5. **Unambiguous:** Indicates that this PSM is the only match that the application considered for this spectrum; there is no ambiguity that it needs to resolve. **Selected:** Indicates that the application selected this PSM from a set of two or more matches that it considered for the protein group inference process.

The identification of crosslinking site of low abundant peptide MYP***SQK is also benefiting from the increase of 'microscans' (figure 37 G,H,I,J). However, the number of unambiguous PSMs of the modification at A8, T18 and E24 were the same (figure 37 J).

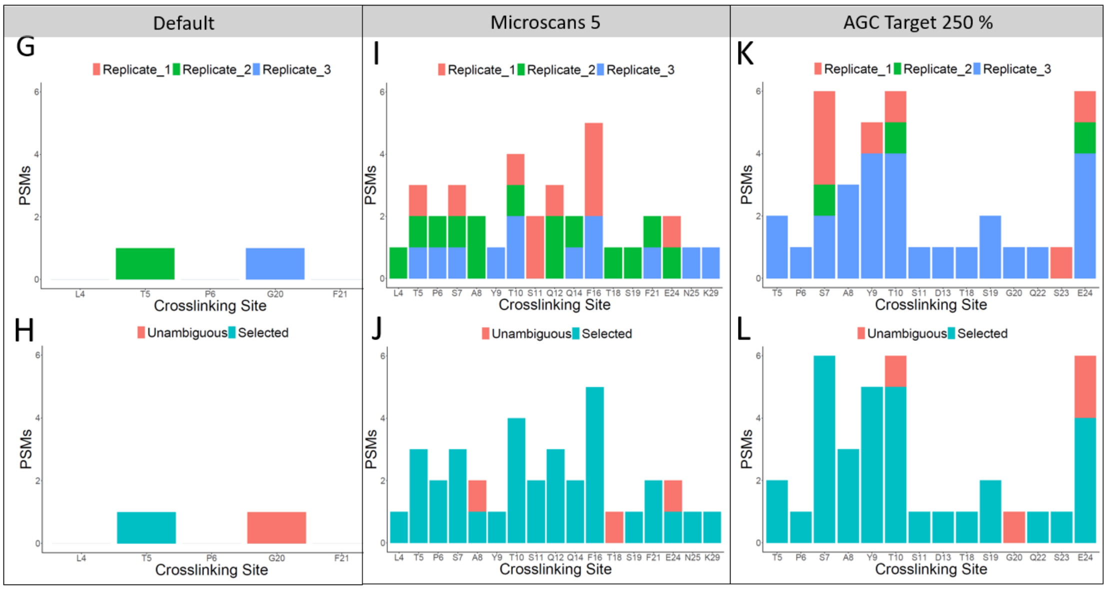


Figure 37. Crosslinking site of PSMs (‘mini-tag’ modified MYP***SQK). **Default (in ddMS^2 IT HCD setting):** Normalized AGC Target (%) 100, Microscans 1. **Microscans 5 (in ddMS^2 IT HCD setting):** Normalized AGC Target (%) 100, Microscans 5. **AGC Targent 250 (in ddMS^2 IT HCD setting):** Normalized AGC Target (%) 250, Microscans 5. **Unambiguous:** Indicates that this PSM is the only match that the application considered for this spectrum; there is no ambiguity that it needs to resolve. **Selected:** Indicates that the application selected this PSM from a set of two or more matches that it considered for the protein group inference process.

Injecting more fragment ions to the linear ion trap by adjusting automatic gain control target (AGC target) can also be helpful. When the AGC target was subsequently increased from 100% to 250% in the MS^3 acquisition of PRM^2 measurements, an increasing number of PSMs of ‘mini-tag’ modified MYP***SQK could be observed (Figure 37 K,I). Given that most of the unambiguous PSMs were identified as MYP***SQK with ‘mini-tag’ modification on the glutamic acid E24, we think that the most efficient photo-crosslinking potentially takes place at E24. Nevertheless, we observed a very slight increase in PSMs of the low abundant probe-modified peptide VAS***FGK_1, when the ‘AGC target’ was increased to 250%. The number of unambiguous PSMs was even decreasing (Figure 37 E, F). In addition, the percentage of high quality PSMs of modified peptide VAS***FGK_1 was increased to 100%, while modified peptide MYP***SQK was slightly decreased to 39%, and modified peptide VAS***FGK_2 was decreased to 87% (Figure 36, AGCT.250).

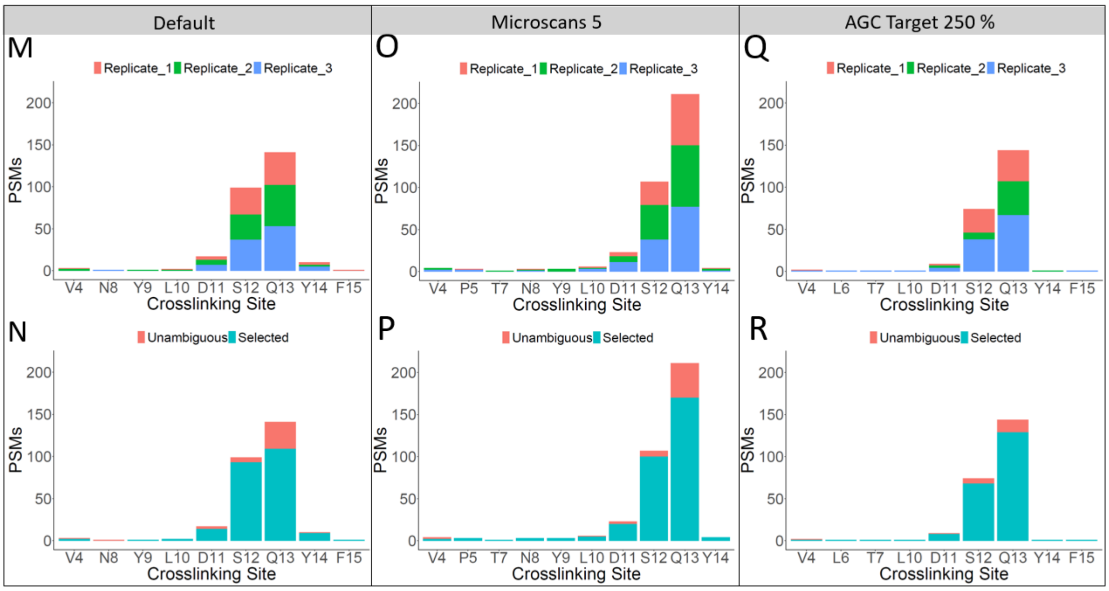


Figure 38. Crosslinking site of PSMs (‘mini-tag’ modified VAS***FGK_2). **Default (in ddMS^2 IT HCD setting):** Normalized AGC Target (%) 100, Microscans 1. **Microscans 5 (in ddMS^2 IT HCD setting):** Normalized AGC Target (%) 100, Microscans 5. **AGC Targent 250 (in ddMS^2 IT HCD setting):** Normalized AGC Target (%) 250, Microscans 5. **Unambiguous:** Indicates that this PSM is the only match that the application considered for this spectrum; there is no ambiguity that it needs to resolve. **Selected:** Indicates that the application selected this PSM from a set of two or more matches that it considered for the protein group inference process.

In the case of the highly abundant modified peptide VAS***FGK_2 (rt, 70.0 - 72.0 min), 'microscans' and 'AGC Target' have no impact on the photo-crosslinking site assignment. The highest number of unambiguous PSMs suggested that the photo-crosslinking happened at Glutamine Q13. The increase in PSMs was also observed on VAS***FGK_2 when the 'microscans' was increased to 5. However, a significant decrease of PSMs took place when 'AGC Target' was increased to 250% (Figure 38).

4.2.4 Summarization of MS methods

In short, data-dependent acquisition square (DDA^2) and parallel reaction monitoring square (PRM^2) were developed to identify the sulfoxide diazirine (SODA) crosslinking peptides and map the binding hotspots of the bio-active peptides. Overall, DDA^2 is the discovery method for the initial identification of the crosslinked peptides. After the full scan in the orbitrap (MS^1 scan) of the DDA^2, the sulfoxide is cleaved at MS^2 event using 25% of CID fragmentation energy. Once the reporter ions are detected in the orbitrap, a secondary DDA on the MS^2 fragments will be triggered and MS^3 fragments will be analysis in the ion trap. In addition, the database search against the MS^3 data can provide the ID of the photo-crosslinked peptides (Figure 39A).

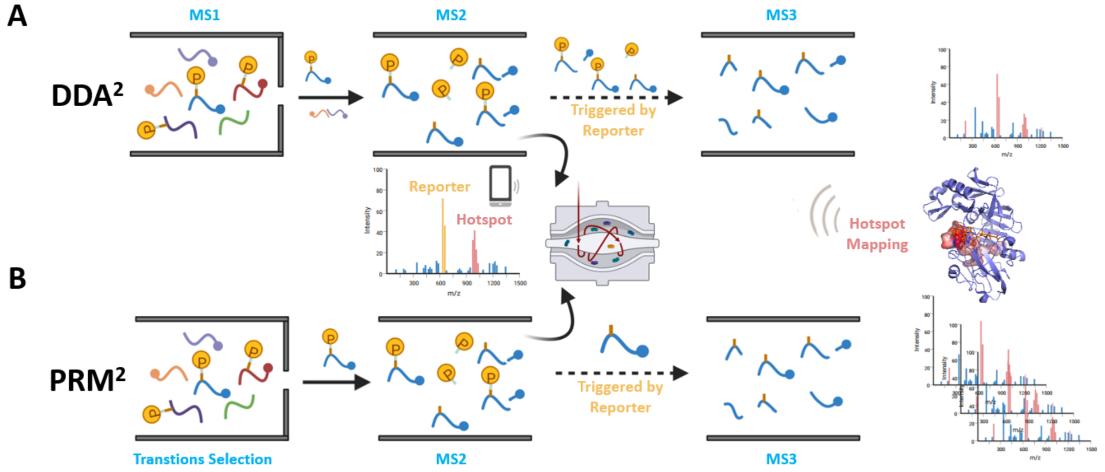


Figure 39. **(A)** Data-dependent acquisition square (DDA^2) workflow. **(B)** Parallel reaction monitoring square (PRM^2) workflow.

In summary, parallel reaction monitoring square (PRM^2) is a validation method for mapping the photo-crosslinking site after the DDA^2 discovery phase. In the PRM^{2,} photo-crosslinking transitions are scheduled according to their retention time, isolated based on their m/z and directly fragmented using 23% of CID fragmentation energy. Once the reporter ions are detected in the orbitrap, a secondary PRM on the desired MS^2 fragments will be triggered and MS^3 fragments will be analyzed in the ion trap with the optimum parameters (microscans:5, Maximum Injection Time: 250 ms, Normalized AGC Target: 250%) (Figure 39B).

Note that we observed different behavior of sulfoxide fragmentation and unequal performance of PSM quality due to the difference of abundance, m/z and charge state of the precursors. The optimal parameters that we applied is a balanced solution to obtain the best results for most peptides.

4.2.5 Mapping of the binding hotspots of different probes

Chymosin was incubated in reaction buffer with 15 μM of probe **10**, **11** or **12**. Following UV irradiation and sample preparation, photo-crosslinked peptides were identified by the DDA^2 analysis and binding hotspots were validated and mapped by PRM^2 analysis. Photo-crosslinked peptide VAS***FGK was identified by probe **10**, and photo-crosslinked peptides VAS***FGK and MYP***SQK were both captured by probe **11** and **12**. VAS***FGK and MYP***SQK are both located the binding pocket of chymosin, which was disclosed by the crystal structure of bovine chymosin in complex with pepstatin A (PDB ID: 4AUC). The number of PSMs suggested that probe **11** may photo-crosslinked on the Gln13 of peptide VAS***FGK and Glu24 of peptide MYP***SQK (Figure 40). Probe **12** may photo-crosslinked on the Gln13 of peptide VAS***FGK as well (Figure 41). Nevertheless, insufficient of PSMs may lead to an uncertain read-out of photo-crosslinking sites, such as Lys29 of peptide MYP***SQK (Figure 27) and Val4 on peptide VAS***FGK (Figure 42) which were a bit far away from the original binding interface of pepstatin A.

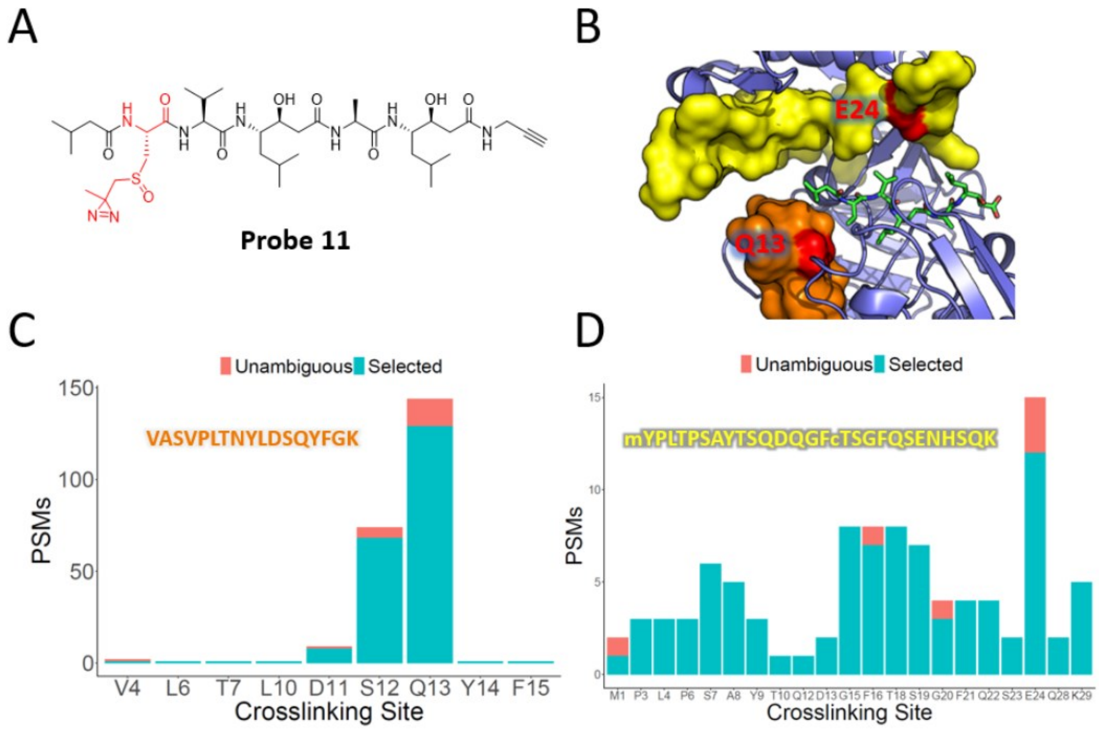


Figure 40. Binding hotspots of probes **11**. **(A)** Chemical structure of probe **11**. **(B)** Crystal structure of bovine chymosin in complex with pepstatin A (PDB ID: 4AUC). **(C)** PSMs of photo-crosslinked peptide VAS***FGK. **(D)** PSMs of photo-crosslinked peptide MYP***SQK. **Unambiguous:** Indicates that this PSM is the only match that the application considered for this spectrum; there is no ambiguity that it needs to resolve. **Selected:** Indicates that the application selected this PSM from a set of two or more matches that it considered for the protein group inference process.

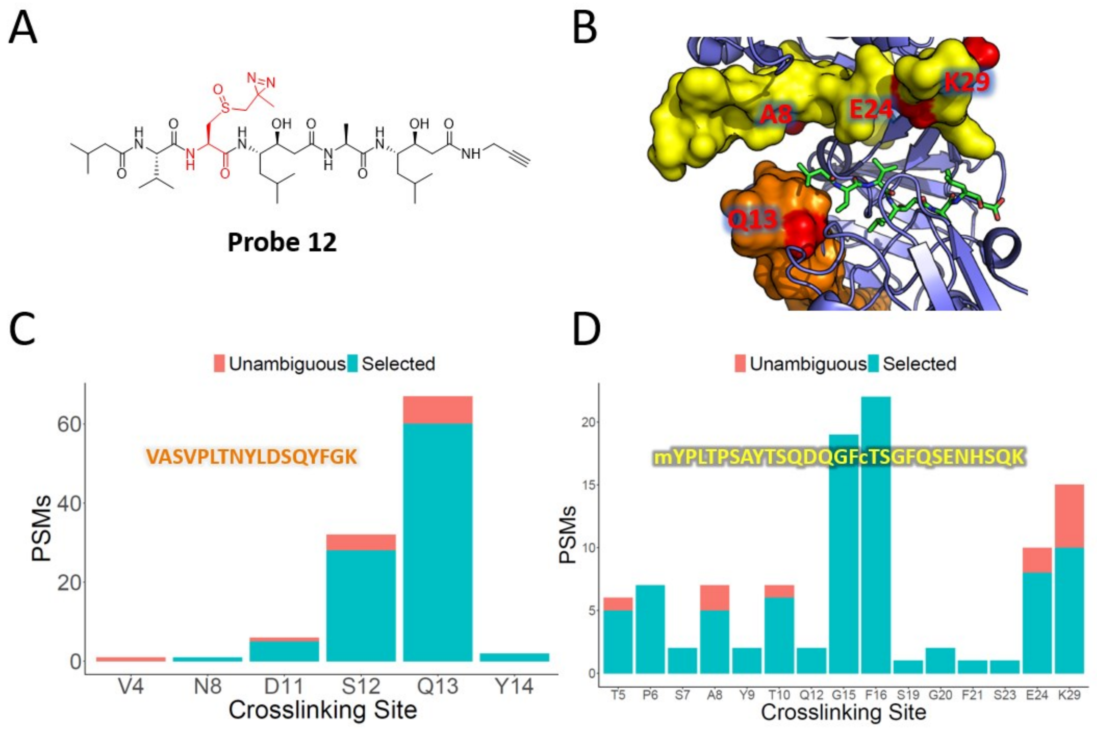


Figure 41. Binding hotspots of probes **12**. **(A)** Chemical structure of probe **12**. **(B)** Crystal structure of bovine chymosin in complex with pepstatin A (PDB ID: 4AUC). **(C)** PSMs of photo-crosslinked peptide VAS***FGK. **(D)** PSMs of photo-crosslinked peptide MYP***SQK. **Unambiguous:** Indicates that this PSM is the only match that the application considered for this spectrum; there is no ambiguity that it needs to resolve. **Selected:** Indicates that the application selected this PSM from a set of two or more matches that it considered for the protein group inference process.

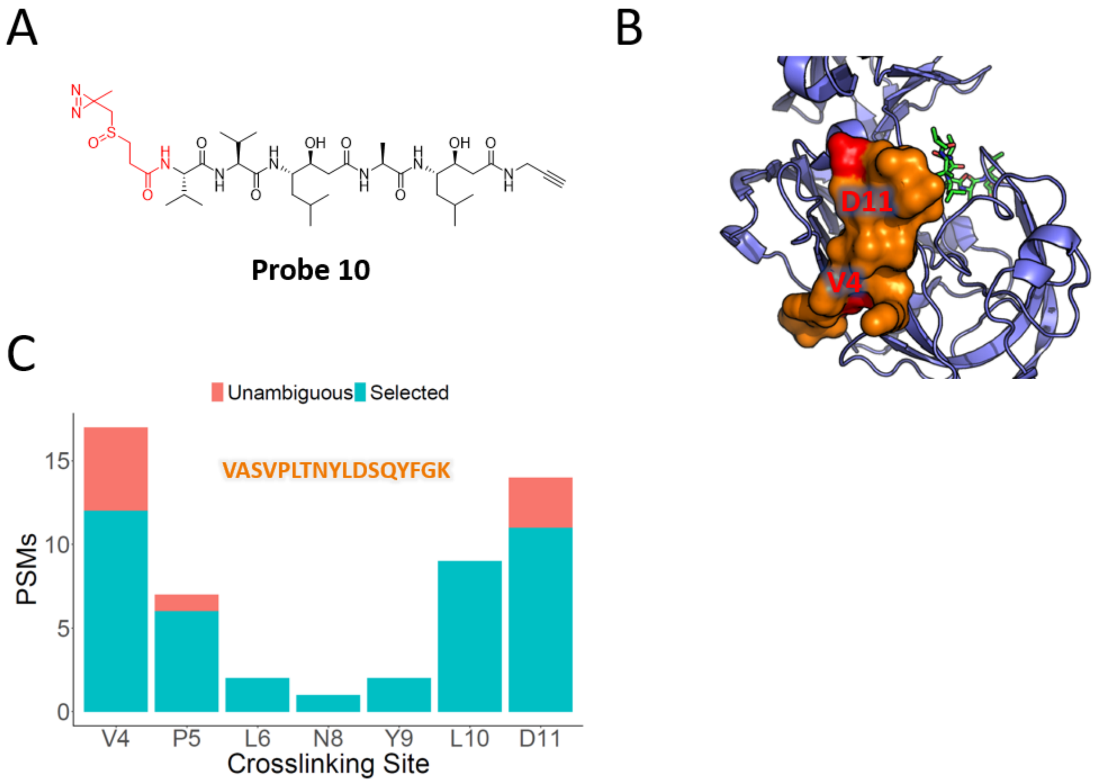


Figure 42. Binding hotspots of probes **10**. **(A)** Chemical structure of probe **10**. **(B)** Crystal structure of bovine chymosin in complex with pepstatin A (PDB ID: 4AUC). **(C)** PSMs of photo-crosslinked peptide VAS***FGK. **Unambiguous:** Indicates that this PSM is the only match that the application considered for this spectrum; there is no ambiguity that it needs to resolve. **Selected:** Indicates that the application selected this PSM from a set of two or more matches that it considered for the protein group inference process.

5 Conclusion

5.1 Pepstatin-based probes for photoaffinity labeling of aspartic proteases

There are relatively few reported covalent chemical aspartic protease probes. In order to profile aspartic proteases by PAL, we developed clickable probes based on pepstatin in this study. Even though a few publications have described the solid phase production of aspartic protease probes, they either used specially synthesized building blocks [209] or positioned heavy crosslinkers far from the inhibitory scaffold that binds the active site cleft. [210] Here, we revealed that a minimal diazirine photoreactive group can be incorporated into the universal aspartic protease inhibitor pepstatin. Conveniently, the probes can be fully synthesized on solid support utilizing building blocks that are readily accessible. In the future, it may allow the incorporation of multiple other natural or non-natural amino acids into a diazirine photoreactive peptide library formation. This approach may facilitate the rapid optimization of selective probes for a desired aspartic protease.

The efficacy of PAL depended on where the diazirine photoreactive group was located, and probes with the diazirine at the N- or C-terminal side of the pepstatin scaffold often demonstrated more efficient PAL. *Gertsik* and co-workers have utilized benzophenone as photo-reactive groups at different positions within a γ-secretase inhibitor to investigate probes for intramembrane aspartic proteases such as gamma-secretase and signal peptide peptidase (SPP).[211] In addition to discovering varying degrees of efficiency in PAL with different positions of benzophenone groups, they also observed distinct effects of allosteric modulators on probe labeling of gamma-secretase and SPP. This observation suggests that these modulators have diverse impacts on the subsite pockets surrounding the active site, where the probes bind. We anticipate that the pepstatin-based probes described here may be useful in identifying allosteric modulators on soluble aspartic proteases as well as intramembrane aspartic proteases.

Cathepsin D has been suggested as a histopathological biomarker for disease progression since it is overexpressed in several malignancies, particularly breast cancer. [212] The probes developed here enable covalently label and detect cathepsin D in a breast cancer cell lysate, as detected on gel and with MS-based proteomics following target enrichment. Our findings further point out the importance of in-depth data analysis in chemical proteomics. A deep learning algorithm suggested various proteins as cathepsin D interaction partners and possible substrates from a list of co-enriched proteins. SQSTM1 was found to be degraded by cathepsin D in a biochemical experiment. As a result, we propose that cathepsin D is responsible for SQSTM1 degradation in autophagosomes during autophagy, and the techniques described here could aid in the future elucidation of this degradation mechanism.

5.2 Mass cleavable affinity-based probes for precise mapping of binding hotspots

We report on sulfoxide diazirine (SODA) building blocks that can be easily incorporated into peptide-like probes for photo-affinity labeling. The cleavable photo-reactive groups allow for a MS^2 cleavage event, generating a probe-derived reporter ion and a minimal fragment on the modified peptide, which is the same for any probe (Figure 1). With tailored methods DDA^2 and PRM^2, we show that this strategy can be utilized to identify the modification sites of PAL probes. We believe that the building blocks represent valuable reagents for MS analysis of photoaffinity probes.

6 Future prospects

6.1 Pepstatin-based probes enable future study of biology and drug discovery

We revealed that the universal aspartic protease inhibitor pepstatin can be properly incorporated with a minimal diazirine photoreactive group and conveniently synthesized on solid support using homemade and commercially available building blocks. In the future, the incorporation of multiple non-natural amino acids and diazirine building blocks (such as C-terminal diazirines, N-terminal diazirines, and diazirine amino acids) may allow a varied chemical space of the photo-reactive peptide library (Figure 43A). [213] [214] This approach may facilitate the rapid optimization of selective probes for desired proteases such as γ-secretase probes, [215] [216] [217] sunflower trypsin inhibitor probes [218] (Figure 43B-43C). These may eventually be used in parallel when different tags are attached (Figure 44A).

Pepstatin-based probes are able to detect as little as 10–50 ng of pepsin and 50–100 ng of chymosin, as visualized in gel-based titration experiments, as well as endogenously expressed cathepsin D in cell lysates. Combined with an enzyme-linked immunosorbent assay (ELISA), pepstatin-based probes also have potential to achieve a highly sensitive detection for diagnoses application and a high-throughput screening assay for drug discovery (Figure 44B). [219] [220]

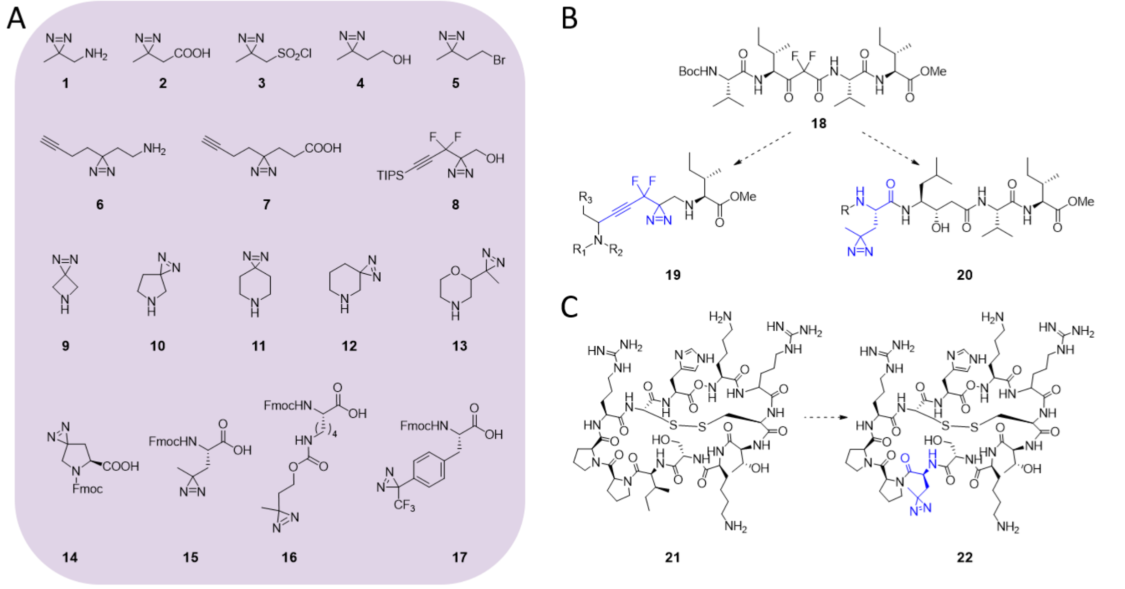


Figure 43. **(A)** Examples of commercialized diazirine building blocks. **1-7, 9-13:** from Enamine; **8:** from Sigma-Aldrich; **14-17:** from Iris Biotech. **(B)** Design of γ-secretase probes. **18:** transition-state analog inhibitor (TSA) of γ-secretase. **19-20:** proposed γ-secretase probes. **(C)** Design of sunflower trypsin inhibitor probes. **21:** sunflower trypsin inhibitor. **22:** proposed sunflower trypsin inhibitor probe.

Overexpressed pro-cathepsin-D can be secreted from the cell and subsequently endocytosed by both cancer cells and fibroblasts through mannose 6-phosphate (M6P) receptors and additionally by yet undiscovered receptors. [221] [222] Furthermore, similar to pepsinogen, pro-cathepsin-D can self-activate at low pH in vitro, and produce a catalytically active pseudo-cathepsin-D, in which 18 residues

(27–44) of the pro-segment remain. [223] Pepstatin-based probes have the potential to facilitate breast cancer research by distinguishing and profiling the active cathepsin D (Figure 44C). Our chemical proteomics experiments and deep learning prediction suggest that sequestosome-1 (SQSTM1), an important player in autophagy, is a direct interaction partner and substrate of cathepsin D. Pepstatin-based probes enable the tracking of cathepsin D activity using gel-based experiments as well as co-localization imaging experiment and offer an access to investigate the function of cathepsin D in autophagy (Figure 44C).

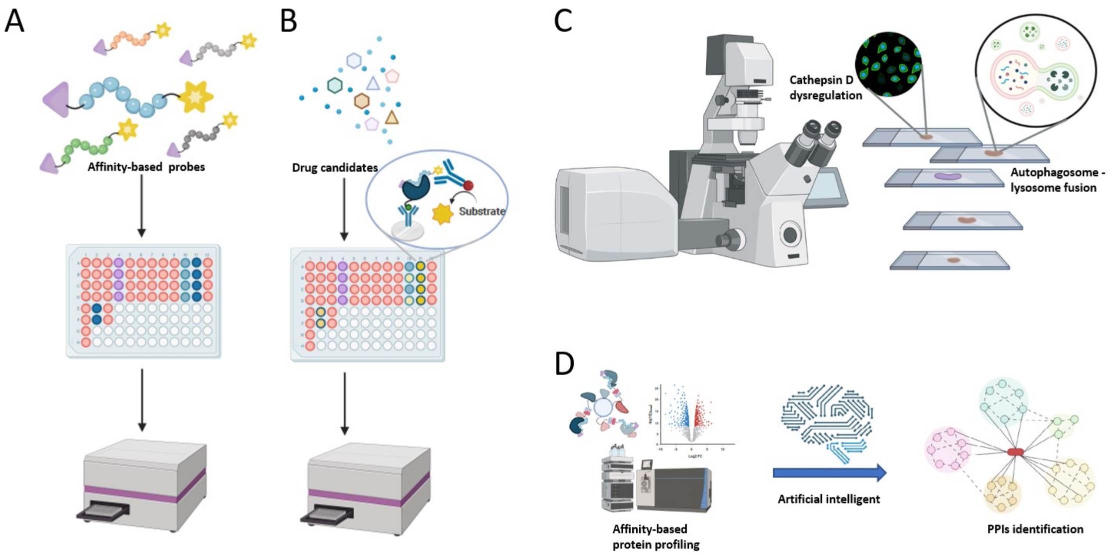


Figure 44. **(A)** Optimization workflow of selective probes for desired proteases. **(B)** Affinity based enzyme-linked immunosorbent assay (AfB-ELISA) for drug discovery. **(C)** Bio-image study of cathepsin D dysregulation and its role in autophagy. The SQSTM1 will be recombined with mCherry fluorescent protein and cathespsin D will be label by pepstatin-based probes (or anti-cathepsin D fluorescent antibody). **(D)** Deeper data analyses using artificial intelligent for PPIs identification.

Our work also highlights the importance of deeper data analysis in chemical proteomics. From a list of co-enriched proteins, a deep learning algorithm predicted various proteins as cathepsin D interaction partners and potential substrates. SQSTM1 was a successful example that we found as direct interaction partner and substrate of cathepsin D. Other co-enriched proteins also have the potential to be the interaction partner or substrate of cathepsin D, and further experiments along the same lines as describe in Figure 25 may be performed for these targets as well.

Affinity-based protein profiling (AfBPP) is a powerful approach to study the targets and off-targets of bio-active small molecules. However, it is challenging to assign a small molecule directly to the proteins that are enriched in the quantitative proteomics experiment. Incorporating artificial intelligent (AI), deeper data analysis of AfBPP experiments may help to distinguish the protein candidates that may engage in protein-protein interaction and potentially facilitate the identification of targets, interactors of these targets, and off-targets of bio-active small molecules (Figure 44D).

6.2 Sulfoxide diazirine (SODA) allows the mapping of binding hotspots

Sulfoxide diazirine (SODA) building blocks can be easily incorporated into bio-active small molecules, especially bio-active peptides and short linear motifs (SLiMs). [176] They can be used for photoaffinity labeing and identification of binding partners as for the diazirine-based molecules from Project 1. Importantly, with tailored data dependent acquisition square (DDA^2) and parallel reaction monitoring square (PRM^2), this strategy can be utilized to identify the modification sites of photo-affinity labeling (PAL) probes. We have established a data-base search workflow on Proteome Discoverer, which can extract the corresponding MS^3 and search against the 'mini-tag' of SODA probes. However, our current workflow still encounters the challenge of precisely assigning the photo-crosslinking to its corresponding modification site. The non-site specific manner of photo-crosslinking plus the exclusion of MS^1 information enormously increased the search space of the database search and results in false positive identifications (Table 8), in which the m/z of the corresponding precursor at MS^1 didn't match to the search results. More detailed MS^1 and MS^2 information can only be observed manually at the current stage. Our high-quality MS^n dataset sets may form a starting point for developing customized bioinformatics tools that can automate the data analysis and facilitate a broad implementation of the SODA chemical tools in standard chemical proteomics target identification workflows (Figure 45A).

The crosslinking efficiency of diazirine is generally around 15-20%, which is challenging the sample preparation and LC-MS^3 detection. [224] Therefore, a further enrichment and better separation of PAL labeled proteins or peptides are highly demanded. In the future, the incorporation of solid phase-enhanced sample-preparation (SP3) and high-field asymmetric waveform ion mobility spectrometry (FAIMS) may allow a better sample quality after enrichment and an additional on-line ion separation using the ion mobility device in front of the mass spectrometry, which can promote a better detection of PAL modified peptides (Figure 45B). [225] [226]

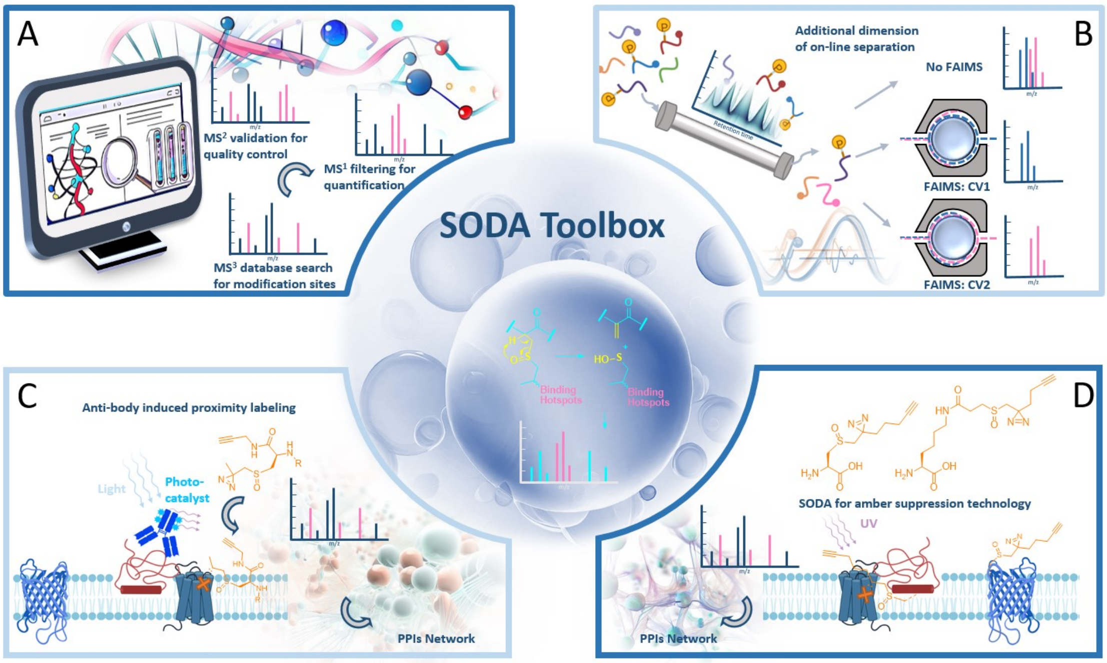

A
MS² validation for quality control
MS¹ filtering for quantification
MS³ database search for modification sites
m/z
SODA Toolbox
Binding Hotspots
Binding Hotspots
B
Additional dimension of on-line separation
Retention time
No FAIMS
FAIMS: CV1
FAIMS: CV2
C
Anti-body induced proximity labeling
Light
Photo-catalyst
PPIs Network
D
SODA for amber suppression technology
UV
PPIs Network

Figure 45. **(A)** Proposed customized bioinformatics tools, in which the MS^3 database search will be orthogonally validated by the reporter ion from MS^2 spectra. The filtering of MS^1 spectra will facilitate quantification application. **(B)** Proposed high-field asymmetric waveform ion mobility spectrometry (FAIMS) workflow for a better detection of PAL modified peptides. Compensation voltage (CV) is a voltage applied to the FAIMS device that allows selective transmission of ions based on their mobility. **(C)** SODA proximity labeling using antibody induced photo-catalytic approach. **(D)** SODA proximity labeling via introducing of SODA into protein of interest.

The identification of SODA modified peptides not only enables the unbiased profiling of targets and off-targets of bio-active small molecules, but also allows the mapping of binding hotspots and potentially facilitate the drug development. Furthermore, SODA building blocks also have potential to be incorporated into the platform of protein-protein interactions (PPIs) identification, in which the photo-reactive group can be activated by a photo-catalyst conjugated antibody and label the interacting partners. [227] [228] Nevertheless, the application of the SODA building block is limited to small-to-medium sized peptides as it can only be introduced by solid-phase peptide synthesis (SPPS). [229] In the future, amber suppression technology may allow introducing a photo-affinity labeling residue onto the protein of interest and facilitate the research of protein-protein interactions as well (Figure 45C-45D). [230]

7 References

1. Rawlings, N. D.; Alan, J.; Thomas, P. D.; Huang, X. D.; Bateman, A.; Finn, R. D., The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research* **2018,** *46* (D1), D624-D632.

2. Eder, J.; Hommel, U.; Cumin, F.; Martoglio, B.; Gerhartz, B., Aspartic proteases in drug discovery. *Curr Pharm Design* **2007,** *13* (3), 271-285.

3. Ostermann, N.; Gerhartz, B.; Worpenberg, S.; Trappe, J.; Eder, J., Crystal structure of an activation intermediate of cathepsin E. *J Mol Biol* **2004,** *342* (3), 889-899.

4. Andreeva, N. S.; Rumsh, L. D., Analysis of crystal structures of aspartic proteinases: On the role of amino acid residues adjacent to the catalytic site of pepsin-like enzymes. *Protein Sci* **2001,** *10* (12), 2439-2450.

5. Ostermann, N.; Eder, J.; Eidhoff, U.; Zink, F.; Hassiepen, U.; Worpenberg, S.; Maibaum, J.; Simic, O.; Hommel, U.; Gerhartz, B., Crystal structure of human BACE2 in complex with a hydroxyethylamine transition-state inhibitor. *J Mol Biol* **2006,** *355* (2), 249-261.

6. DeLano, W. L. The Pymol Molecular Graphics Systems. Http://Www.Pymol.Org.

7. Northrop, D. B., Follow the protons: A low-barrier hydrogen bond unifies the mechanisms of the aspartic proteases. *Accounts Chem Res* **2001,** *34* (10), 790-797.

8. Veerapandian, B.; Cooper, J. B.; Sali, A.; Blundell, T. L.; Rosati, R. L.; Dominy, B. W.; Damon, D. B.; Hoover, D. J., Direct Observation by X-Ray-Analysis of the Tetrahedral Intermediate of Aspartic Proteinases. *Protein Sci* **1992,** *1* (3), 322-328.

9. Masson, O.; Bach, A. S.; Derocq, D.; Prebois, C.; Laurent-Matha, V.; Pattingre, S.; Liaudet-Coopman, E., Pathophysiological functions of cathepsin D: Targeting its catalytic activity versus its protein binding activity? *Biochimie* **2010,** *92* (11), 1635-1643.

10. Wolfe, M. S., Dysfunctional -Secretase in Familial Alzheimer's Disease. *Neurochem Res* **2019,** *44* (1), 5-11.

11. Ranganathan, P.; Weaver, K. L.; Capobianco, A. J., Notch signalling in solid tumours: a little bit of everything but not all the time. *Nat Rev Cancer* **2011,** *11* (5), 338-351.

12. Brik, A.; Wong, C. H., HIV-1 protease: mechanism and drug discovery. *Org Biomol Chem* **2003,** *1* (1), 5-14.

13. Cheuka, P. M.; Dziwornu, G.; Okombo, J.; Chibale, K., Plasmepsin Inhibitors in Antimalarial Drug Discovery: Medicinal Chemistry and Target Validation (2000 to Present). *J Med Chem* **2020,** *63* (9), 4445-4467.

14. Barrett, A. J., Cathepsin-D - Purification of Isoenzymes from Human and Chicken Liver. *Biochem J* **1970,** *117* (3), 601-&.

15. Diment, S.; Martin, K. J.; Stahl, P. D., Cleavage of Parathyroid-Hormone in Macrophage Endosomes Illustrates a Novel Pathway for Intracellular Processing of Proteins. *J Biol Chem* **1989,** *264* (23), 13403-13406.

16. Saftig, P.; Hetman, M.; Schmahl, W.; Weber, K.; Heine, L.; Mossmann, H.; Koster, A.; Hess, B.; Evers, M.; Vonfigura, K.; Peters, C., Mice Deficient for the Lysosomal Proteinase Cathepsin-D Exhibit Progressive Atrophy of the Intestinal-Mucosa and Profound Destruction of Lymphoid-Cells. *Embo J* **1995,** *14* (15), 3599-3608.

17. Richo, G.; Conner, G. E., Proteolytic Activation of Human Procathepsin-D. *Adv Exp Med Biol* **1991,** *306*, 289-296.

18. Hasilik, A.; Neufeld, E. F., Biosynthesis of Lysosomal-Enzymes in Fibroblasts -

Synthesis as Precursors of Higher Molecular-Weight. *J Biol Chem* **1980,** *255* (10), 4937-4945.

19. Richo, G. R.; Conner, G. E., Structural Requirements of Procathepsin-D Activation and Maturation. *J Biol Chem* **1994,** *269* (20), 14806-14812.

20. Vonfigura, K.; Hasilik, A., Lysosomal-Enzymes and Their Receptors. *Annual Review of Biochemistry* **1986,** *55*, 167-193.

21. Kornfeld, S., Lysosomal-Enzyme Targeting. *Biochem Soc T* **1990,** *18* (3), 367-374.

22. Tang, J.; Wong, R. N. S., Evolution in the Structure and Function of Aspartic Proteases. *J Cell Biochem* **1987,** *33* (1), 53-63.

23. Conner, G. E.; Richo, G., Isolation and Characterization of a Stable Activation Intermediate of the Lysosomal Aspartyl Protease Cathepsin-D. *Biochemistry-Us* **1992,** *31* (4), 1142-1147.

24. Larsen, L. B.; Boisen, A.; Petersen, T. E., Procathepsin-D Cannot Autoactivate to Cathepsin-D at Acid Ph. *Febs Lett* **1993,** *319* (1-2), 54-58.

25. Felbor, U.; Kessler, B.; Mothes, W.; Goebel, H. H.; Ploegh, H. L.; Bronson, R. T.; Olsen, B. R., Neuronal loss and brain atrophy in mice lacking cathepsins B and L. *P Natl Acad Sci USA* **2002,** *99* (12), 7883-7888.

26. Duffy, M. J., Proteases as prognostic markers in cancer. *Clin Cancer Res* **1996,** *2* (4), 613-618.

27. Vignon, F.; Capony, F.; Chambon, M.; Freiss, G.; Garcia, M.; Rochefort, H., Autocrine Growth-Stimulation of the Mcf-7 Breast-Cancer Cells by the Estrogen-Regulated 52-K Protein. *Endocrinology* **1986,** *118* (4), 1537-1545.

28. Fusek, M.; Vetvicka, V., Mitogenic Function of Human Procathepsin-D - the Role of the Propeptide. *Biochem J* **1994,** *303*, 775-780.

29. Vetvicka, V.; Vektvickova, J.; Fusek, M., Effect of Human Procathepsin-D on Proliferation of Human Cell-Lines. *Cancer Lett* **1994,** *79* (2), 131-135.

30. Vetvicka, V.; Vetvickova, J.; Fusek, M., Effect of procathepsin D and its activation peptide on prostate cancer cells. *Cancer Lett* **1998,** *129* (1), 55-59.

31. Garcia, M.; Derocq, D.; Pujol, P.; Rochefort, H., Overexpression of Transfected Cathepsin-D in Transformed-Cells Increases Their Malignant Phenotype and Metastatic Potency. *Oncogene* **1990,** *5* (12), 1809-1814.

32. Liaudet, E.; Garcia, M.; Rochefort, H., Cathepsin-D Maturation and Its Stimulatory Effect on Metastasis Are Prevented by Addition of Kdel Retention Signal. *Oncogene* **1994,** *9* (4), 1145-1154.

33. Liaudet, E.; Derocq, D.; Rochefort, H.; Garcia, M., Transfected Cathepsin-D Stimulates High-Density Cancer Cell-Growth by Inactivating Secreted Growth-Inhibitors. *Cell Growth Differ* **1995,** *6* (9), 1045-1052.

34. Berchem, G.; Glondu, M.; Gleizes, M.; Brouillet, J. P.; Vignon, F.; Garcia, M.; Liaudet-Coopman, E., Cathepsin-D affects multiple tumor progression steps in vivo: proliferation, angiogenesis and apoptosis. *Oncogene* **2002,** *21* (38), 5951-5955.

35. Gonzalez-Vela, M. C.; Garijo, M. F.; Fernandez, F.; Buelta, L.; Val-Bernal, J. F., Cathepsin D in host stromal cells is associated with more highly vascular and aggressive invasive breast carcinoma. *Histopathology* **1999,** *34* (1), 35-42.

36. Briozzo, P.; Badet, J.; Capony, F.; Pieri, I.; Montcourrier, P.; Barritault, D.; Rochefort, H., Mcf7 Mammary-Cancer Cells Respond to Bfgf and Internalize It Following Its Release from Extracellular-Matrix - a Permissive Role of Cathepsin-D. *Exp Cell Res* **1991,** *194* (2), 252-259.

37. Morikawa, W.; Yamamoto, K.; Ishikawa, S.; Takemoto, S.; Ono, M.; Fukushi, J.; Naito, S.; Nozaki, C.; Iwanaga, S.; Kuwano, M., Angiostatin generation by cathepsin D

secreted by human prostate carcinoma cells. *J Biol Chem* **2000,** *275* (49), 38912-38920.

38. Piwnica, D.; Touraine, P.; Struman, I.; Tabruyn, S.; Bolbach, G.; Clapp, C.; Martial, J. A.; Kelly, P. A.; Goffin, V., Cathepsin D processes human prolactin into multiple 16K-like N-terminal fragments: Study of their antiangiogenic properties and physiological relevance. *Mol Endocrinol* **2004,** *18* (10), 2522-2542.

39. Glondu, M.; Coopman, P.; Laurent-Matha, V.; Garcia, M.; Rochefort, H.; Liaudet-Coopman, E., A mutated cathepsin-D devoid of its catalytic activity stimulates the growth of cancer cells. *Oncogene* **2001,** *20* (47), 6920-6929.

40. Glondu, M.; Liaudet-Coopman, E.; Derocq, D.; Platet, N.; Rochefort, H.; Garcia, M., Down-regulation of cathepsin-D expression by antisense gene transfer inhibits tumor growth and experimental lung metastasis of human breast cancer cells. *Oncogene* **2002,** *21* (33), 5127-5134.

41. Westley, B.; Rochefort, H., A Secreted Glycoprotein Induced by Estrogen in Human-Breast Cancer Cell-Lines. *Cell* **1980,** *20* (2), 353-362.

42. Capony, F.; Morisset, M.; Barrett, A. J.; Capony, J. P.; Broquet, P.; Vignon, F.; Chambon, M.; Louisot, P.; Rochefort, H., Phosphorylation, Glycosylation, and Proteolytic Activity of the 52-Kd Estrogen-Induced Protein Secreted by Mcf7 Cells. *J Cell Biol* **1987,** *104* (2), 253-262.

43. Capony, F.; Rougeot, C.; Montcourrier, P.; Cavailles, V.; Salazar, G.; Rochefort, H., Increased Secretion, Altered Processing, and Glycosylation of Pro-Cathepsin D in Human Mammary-Cancer Cells. *Cancer Res* **1989,** *49* (14), 3904-3909.

44. Augereau, P.; Garcia, M.; Mattei, M. G.; Cavailles, V.; Depadova, F.; Derocq, D.; Capony, F.; Ferrara, P.; Rochefort, H., Cloning and Sequencing of the 52k Cathepsin-D Complementary Deoxyribonucleic-Acid of Mcf7 Breast-Cancer Cells and Mapping on Chromosome-11. *Mol Endocrinol* **1988,** *2* (2), 186-192.

45. Rochefort, H.; Cavailles, V.; Augereau, P.; Capony, F.; Maudelonde, T.; Touitou, I.; Garcia, M., Overexpression and Hormonal-Regulation of Pro-Cathepsin-D in Mammary and Endometrial Cancer. *J Steroid Biochem* **1989,** *34* (1-6), 177-182.

46. Cavailles, V.; Augereau, P.; Rochefort, H., Cathepsin-D Gene Is Controlled by a Mixed Promoter, and Estrogens Stimulate Only Tata-Dependent Transcription in Breast-Cancer Cells. *P Natl Acad Sci USA* **1993,** *90* (1), 203-207.

47. Westley, B. R.; May, F. E. B., Estrogen Regulates Cathepsin-D Messenger-Rna Levels in Estrogen Responsive Human-Breast Cancer-Cells. *Nucleic Acids Research* **1987,** *15* (9), 3773-3786.

48. Cavailles, V.; Garcia, M.; Rochefort, H., Regulation of Cathepsin-D and Ps2 Gene-Expression by Growth-Factors in Mcf7 Human-Breast Cancer-Cells. *Mol Endocrinol* **1989,** *3* (3), 552-558.

49. Garcia, M.; Salazarretana, G.; Richer, G.; Domergue, J.; Capony, F.; Pujol, H.; Laffargue, F.; Pau, B.; Rochefort, H., Immunohistochemical Detection of the Estrogen-Regulated 52,000 Mol Wt Protein in Primary Breast Cancers but Not in Normal Breast and Uterus. *J Clin Endocr Metab* **1984,** *59* (3), 564-566.

50. Rochefort, H., Cathepsin D in Breast-Cancer - a Tissue Marker Associated with Metastasis. *Eur J Cancer* **1992,** *28A* (11), 1780-1783.

51. Westley, B. R.; May, F. E. B., Prognostic value of cathepsin D in breast cancer. *Brit J Cancer* **1999,** *79* (2), 189-190.

52. Ferrandina, G.; Scambia, G.; Bardelli, F.; Panici, P. B.; Mancuso, S.; Messori, A., Relationship between cathepsin-D content and disease free survival in node-negative breast cancer patients: A meta-analysis. *Brit J Cancer* **1997,** *76* (5), 661-666.

53. Foekens, J. A.; Look, M. P.; Bolt-de Vries, J.; Meijer-van Gelder, M. E.; van Putten, W. L. J.; Klijn, J. G. M., Cathepsin-D in primary breast cancer: prognostic evaluation involving 2810 patients. *Brit J Cancer* **1999,** *79* (2), 300-307.

54. Abbott, D. E.; Margaryan, N. V.; Jeruss, J. S.; Khan, S.; Kaklamani, V.; Winchester, D. J.; Hansen, N.; Rademaker, A.; Khalkhali-Ellis, Z.; Hendrix, M. J. C., Reevaluating cathepsin D as a biomarker for breast cancer Serum activity levels versus histopathology. *Cancer Biol Ther* **2010,** *9* (1), 23-30.

55. Liu, Y. S.; Patricelli, M. P.; Cravatt, B. F., Activity-based protein profiling: The serine hydrolases. *P Natl Acad Sci USA* **1999,** *96* (26), 14694-14699.

56. Greenbaum, D.; Medzihradszky, K. F.; Burlingame, A.; Bogyo, M., Epoxide electrophiles as activity-dependent cysteine protease profiling and discovery tools. *Chem Biol* **2000,** *7* (8), 569-581.

57. Sanman, L. E.; Bogyo, M., Activity-Based Profiling of Proteases. *Annu Rev Biochem* **2014,** *83*, 249-273.

58. Niphakis, M. J.; Cravatt, B. F., Enzyme Inhibitor Discovery by Activity-Based Protein Profiling. *Annu Rev Biochem* **2014,** *83*, 341-377.

59. Chakrabarty, S.; Kahler, J. P.; van de Plassche, M. A. T.; Vanhoutte, R.; Verhelst, S. H. L., Recent Advances in Activity-Based Protein Profiling of Proteases. *Curr Top Microbiol* **2019,** *420*, 253-281.

60. Saghatelian, A.; Jessani, N.; Joseph, A.; Humphrey, M.; Cravatt, B. F., Activity-based probes for the proteomic profiling of metalloproteases. *P Natl Acad Sci USA* **2004,** *101* (27), 10000-10005.

61. Sieber, S. A.; Niessen, S.; Hoover, H. S.; Cravatt, B. F., Proteomic profiling of metalloprotease activities with cocktails of active-site probes. *Nat Chem Biol* **2006,** *2* (5), 274-81.

62. Li, Y. M.; Xu, M.; Lai, M. T.; Huang, Q.; Castro, J. L.; DiMuzio-Mower, J.; Harrison, T.; Lellis, C.; Nadin, J. L.; Neduvelil, J. G.; Register, R. B.; Sardana, M. K.; Shearman, M. S.; Smith, A. L.; Shi, X. P.; Yin, K. C.; Shafer, J. A.; Gardell, S. J., Photoactivated gamma-secretase inhibitors directed to the active site covalently label presenilin 1. *Nature* **2000,** *405* (6787), 689-694.

63. Shi, H. B.; Zhang, C. J.; Chen, G. Y. J.; Yao, S. Q., Cell-Based Proteome Profiling of Potential Dasatinib Targets by Use of Affinity-Based Probes. *J Am Chem Soc* **2012,** *134* (6), 3001-3014.

64. Singh, A.; Westheimer, F. H.; Thornton, E. R., Photolysis of Diazoacetylchymotrypsin. *J Biol Chem* **1962,** *237* (9), 3006-&.

65. Fleet, G. W. J.; Porter, R. R.; Knowles, J. R., Affinity Labelling of Antibodies with Aryl Nitrene as Reactive Group. *Nature* **1969,** *224* (5218), 511-&.

66. Preston, G. W.; Wilson, A. J., Photo-induced covalent cross-linking for the analysis of biomolecular interactions. *Chem Soc Rev* **2013,** *42* (8), 3289-3301.

67. Fleming, S. A., Chemical Reagents in Photoaffinity-Labeling. *Tetrahedron* **1995,** *51* (46), 12479-12520.

68. Geurink, P. P.; Prely, L. M.; van der Marel, G. A.; Bischoff, R.; Overkleeft, H. S., Photoaffinity Labeling in Activity-Based Protein Profiling. *Curr Top Microbiol* **2012,** *324*, 85-113.

69. Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G., Addition of p-azido-L-phenylaianine to the genetic code of Escherichia coli. *J Am Chem Soc* **2002,** *124* (31), 9026-9027.

70. Tam, E. K. W.; Li, Z. Q.; Goh, Y. L.; Cheng, X. M.; Wong, S. Y.;

Santhanakrishnan, S.; Chai, C. L. L.; Yao, S. Q., Cell-Based Proteome Profiling Using an Affinity-Based Probe (AfBP) Derived from 3-Deazaneplanocin A (DzNep). *Chem-Asian J* **2013,** *8* (8), 1818-1828.

71. Galardy, R. E.; Craig, L. C.; Printz, M. P., Benzophenone Triplet - New Photochemical Probe of Biological Ligand-Receptor Interactions. *Nature-New Biol* **1973,** *242* (117), 127-128.

72. Wagner, P. J.; Truman, R. J.; Scaiano, J. C., Substituent Effects on Hydrogen Abstraction by Phenyl Ketone Triplets. *J Am Chem Soc* **1985,** *107* (24), 7093-7097.

73. Riga, E. K.; Saar, J. S.; Erath, R.; Hechenbichler, M.; Lienkamp, K., On the Limits of Benzophenone as Cross-Linker for Surface-Attached Polymer Hydrogels. *Polymers-Basel* **2017,** *9* (12).

74. Abendroth, H.; Henrich, G., Über ein isomeres Acetonhydrazon. *Angewandte Chemie* **1959,** *71* (8), 283-283.

75. Church, R. F.; Weiss, M. J., Diazirines. II. Synthesis and properties of small functionalized diazirine molecules. Observations on the reaction of a diaziridine with the iodine-iodide ion system. *The Journal of Organic Chemistry* **1970,** *35* (8), 2465-2471.

76. Schmitz, E.; Ohme, R., Neue Diaziridin‐ Synthese. *Angewandte Chemie* **1961,** *73* (6), 220-221.

77. Smith, R. A.; Knowles, J. R., Aryldiazirines. Potential reagents for photolabeling of biological receptor sites. *J Am Chem Soc* **1973,** *95* (15), 5072-5073.

78. Ge, S. S.; Chen, B. A.; Wu, Y. Y.; Long, Q. S.; Zhao, Y. L.; Wang, P. Y.; Yang, S., Current advances of carbene-mediated photoaffinity labeling in medicinal chemistry. *Rsc Adv* **2018,** *8* (51), 29428-29454.

79. Das, J., Aliphatic Diazirines as Photoaffinity Probes for Proteins: Recent Developments. *Chem Rev* **2011,** *111* (8), 4405-4417.

80. Blencowe, A.; Hayes, W., Development and application of diazirines in biological and synthetic macromolecular systems. *Soft Matter* **2005,** *1* (3), 178-205.

81. Pan, S. J.; Zhang, H. L.; Wang, C. Y.; Yao, S. C. L.; Yao, S. Q., Target identification of natural products and bioactive compounds using affinity-based probes. *Nat Prod Rep* **2016,** *33* (5), 612-620.

82. Platz, M.; Admasu, A. S.; Kwiatkowski, S.; Crocker, P. J.; Imai, N.; Watt, D. S., Photolysis of 3-Aryl-3-(Trifluoromethyl)Diazirines - a Caveat Regarding Their Use in Photoaffinity Probes. *Bioconjugate Chem* **1991,** *2* (5), 337-341.

83. Li, Z. Q.; Hao, P. L.; Li, L.; Tan, C. Y. J.; Cheng, X. M.; Chen, G. Y. J.; Sze, S. K.; Shen, H. M.; Yao, S. Q., Design and Synthesis of Minimalist Terminal Alkyne-Containing Diazirine Photo-Crosslinkers and Their Incorporation into Kinase Inhibitors for Cell- and Tissue-Based Proteome Profiling. *Angew Chem Int Edit* **2013,** *52* (33), 8551-8556.

84. Chou, C. J.; Uprety, R.; Davis, L.; Chin, J. W.; Deiters, A., Genetically encoding an aliphatic diazirine for protein photocrosslinking. *Chem Sci* **2011,** *2* (3), 480-483.

85. Stoll, D.; Templin, M. F.; Schrenk, M.; Traub, P. C.; Vohringer, C. F.; Joos, T. O., Protein microarray technology. *Front Biosci-Landmrk* **2002,** *7,* C13-C32.

86. Wilson, D. S.; Nock, S., Recent developments in protein microarray technology. *Angew Chem Int Edit* **2003,** *42* (5), 494-500.

87. Gharahdaghi, F.; Weinberg, C. R.; Meagher, D. A.; Imai, B. S.; Mische, S. M., Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: A method for the removal of silver ions to enhance sensitivity. *Electrophoresis* **1999,** *20* (3), 601-605.

88. Raikos, V.; Hansen, R.; Campbell, L.; Euston, S. R., Separation and identification of hen egg protein isoforms using SDS-PAGE and 2D gel electrophoresis with MALDI-TOF mass

spectrometry. *Food Chem* **2006,** *99* (4), 702-710.
89. Ong, S. E.; Foster, L. J.; Mann, M., Mass spectrometric-based approaches in quantitative proteomics. *Methods* **2003,** *29* (2), 124-130.
90. Mann, M., Functional and quantitative proteomics using SILAC. *Nat Rev Mol Cell Bio* **2006,** *7* (12), 952-958.
91. Zhu, W. H.; Smith, J. W.; Huang, C. M., Mass Spectrometry-Based Label-Free Quantitative Proteomics. *J Biomed Biotechnol* **2010**.
92. van Rooden, E. J.; Florea, B. I.; Deng, H.; Baggelaar, M. P.; van Esbroeck, A. C. M.; Zhou, J.; Overkleeft, H. S.; van der Stelt, M., Mapping in vivo target interaction profiles of covalent inhibitors using chemical proteomics with label-free quantification. *Nat Protoc* **2018,** *13* (4), 752-767.
93. Kleiner, R. E.; Hang, L. E.; Molloy, K. R.; Chait, B. T.; Kapoor, T. M., A Chemical Proteomics Approach to Reveal Direct Protein-Protein Interactions in Living Cells. *Cell Chem Biol* **2018,** *25* (1), 110-+.
94. Yang, F.; Gao, J. J.; Che, J. T.; Jia, G. G.; Wang, C., A Dimethyl-Labeling-Based Strategy for Site-Specifically Quantitative Chemical Proteomics. *Anal Chem* **2018,** *90* (15), 9576-9582.
95. Chen, Y.; Liu, Y.; Hou, X. M.; Ye, Z.; Wang, C., Quantitative and Site-Specific Chemoproteomic Profiling of Targets of Acrolein. *Chem Res Toxicol* **2019,** *32* (3), 467-473.
96. Ross, P. L.; Huang, Y. L. N.; Marchese, J. N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; Purkayastha, S.; Juhasz, P.; Martin, S.; Bartlet-Jones, M.; He, F.; Jacobson, A.; Pappin, D. J., Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol Cell Proteomics* **2004,** *3* (12), 1154-1169.
97. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C., Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal Chem* **2003,** *75* (8), 1895-1904.
98. Bantscheff, M.; Lemeer, S.; Savitski, M. M.; Kuster, B., Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal Bioanal Chem* **2012,** *404* (4), 939-965.
99. Neilson, K. A.; Ali, N. A.; Muralidharan, S.; Mirzaei, M.; Mariani, M.; Assadourian, G.; Lee, A.; van Sluyter, S. C.; Haynes, P. A., Less label, more free: Approaches in label-free quantitative mass spectrometry. *Proteomics* **2011,** *11* (4), 535-553.
100. Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M., Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol Cell Proteomics* **2002,** *1* (5), 376-386.
101. Galmozzi, A.; Dominguez, E.; Cravatt, B. F.; Saez, E., Application of Activity-Based Protein Profiling to Study Enzyme Function in Adipocytes. *Method Enzymol* **2014,** *538*, 151-169.
102. Van Hoof, D.; Pinkse, M. W.; Oostwaard, D. W.; Mummery, C. L.; Heck, A. J.; Krijgsveld, J., An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat Methods* **2007,** *4* (9), 677-8.
103. Lau, H. T.; Suh, H. W.; Golkowski, M.; Ong, S. E., Comparing SILAC- and Stable Isotope Dimethyl-Labeling Approaches for Quantitative Proteomics. *J Proteome Res* **2014,** *13* (9), 4164-4174.
104. Pappireddi, N.; Martin, L.; Wuhr, M., A Review on Quantitative Multiplexed Proteomics. *Chembiochem* **2019,** *20* (10), 1210-1224.
105. Jones, H. B. L.; Heilig, R.; Fischer, R.; Kessler, B. M.; Pinto-Fernandez, A., ABPP-

HT-High-Throughput Activity-Based Profiling of Deubiquitylating Enzyme Inhibitors in a Cellular Context. *Front Chem* **2021,** *9*.

106. Xu, Q. C.; Lam, K. S., Protein and chemical microarrays - Powerful tools for proteomics. *J Biomed Biotechnol* **2003,** (5), 257-266.

107. Aguilar-Mahecha, A.; Hassan, S.; Ferrario, C.; Basik, M., Microarrays as validation strategies in clinical samples: tissue and protein microarrays. *OMICS* **2006,** *10* (3), 311-26.

108. Hett, E. C.; Kyne, R. E., Jr.; Gopalsamy, A.; Tones, M. A.; Xu, H.; Thio, G. L.; Nolan, E.; Jones, L. H., Selectivity Determination of a Small Molecule Chemical Probe Using Protein Microarray and Affinity Capture Techniques. *Acs Comb Sci* **2016,** *18* (10), 611-615.

109. Green, N. M., Avidin and Streptavidin. *Methods in Enzymology* **1990,** *184*, 51-67.

110. Wright, M. H.; Sieber, S. A., Chemical proteomics approaches for identifying the cellular targets of natural products (vol 33, pg 681, 2016). *Nat Prod Rep* **2016,** *33* (5), 731-733.

111. Yang, Y. L.; Fonovic, M.; Verhelst, S. H. L., Cleavable Linkers in Chemical Proteomics Applications. *Methods Mol Biol* **2017,** *1491*, 185-203.

112. Korovesis, D.; Beard, H.; Chen, S.; Verhelst, S. H., Cleavable linkers and their application in MS-based target identification. *Molecular Omics* **2021**.

113. Steigenberger, B.; Albanese, P.; Heck, A.; Scheltema, R., To cleave or not to cleave in XL-MS? *Journal of the American Society for Mass Spectrometry* **2019,** *31* (2), 196-206.

114. Iacobucci, C.; Goze, M.; Sinz, A., Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Curr Opin Biotech* **2020,** *63*, 48-53.

115. Yu, C.; Huang, L., Cross-Linking Mass Spectrometry: An Emerging Technology for Interactomics and Structural Biology. *Anal Chem* **2018,** *90* (1), 144-165.

116. Steen, H.; Mann, M., The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Bio* **2004,** *5* (9), 699-711.

117. Froelich, J. M.; Reid, G. E., Mechanisms for the proton mobility-dependent gas-phase fragmentation reactions of S-alkyl cysteine sulfoxide-containing peptide ions. *Journal of the American Society for Mass Spectrometry* **2007,** *18* (9), 1690-1705.

118. Reid, G. E.; Roberts, K. D.; Kapp, E. A.; Simpson, R. J., Statistical and mechanistic approaches to understanding the gas-phase fragmentation behavior of methionine sulfoxide containing peptides. *J Proteome Res* **2004,** *3* (4), 751-759.

119. Muller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schafer, M.; Sinz, A., Cleavable Cross-Linker for Protein Structure Analysis: Reliable Identification of Cross-Linking Products by Tandem MS. *Anal Chem* **2010,** *82* (16), 6958-6968.

120. Clifford-Nunn, B.; Showalter, H. D. H.; Andrews, P. C., Quaternary Diamines as Mass Spectrometry Cleavable Crosslinkers for Protein Interactions. *Journal of the American Society for Mass Spectrometry* **2012,** *23* (2), 201-212.

121. Gardner, M. W.; Brodbelt, J. S., Preferential Cleavage of N-N Hydrazone Bonds for Sequencing Bis-arylhydrazone Conjugated Peptides by Electron Transfer Dissociation. *Anal Chem* **2010,** *82* (13), 5751-5759.

122. Chakrabarty, J. K.; Bugarin, A.; Chowdhury, S. M., Evaluating the performance of an ETD-cleavable cross-linking strategy for elucidating protein structures. *J Proteomics* **2020,** *225*.

123. Wu, S. L.; Jiang, H. T.; Lu, Q. Z.; Dai, S. J.; Hancock, W. S.; Karger, B. L., Mass Spectrometric Determination of Disulfide Linkages in Recombinant Therapeutic Proteins Using Online LC-MS with Electron-Transfer Dissociation. *Anal Chem* **2009,** *81* (1), 112-122.

124. Rombouts, I.; Lagrain, B.; Scherf, K. A.; Lambrecht, M. A.; Koehler, P.; Delcour, J. A., Formation and reshuffling of disulfide bonds in bovine serum albumin demonstrated

using tandem mass spectrometry with collision-induced and electron-transfer dissociation (vol 5, 12210, 2015). *Sci Rep-Uk* **2015,** *5*.

125. Kendrew, J. C.; Perutz, M. F., A comparative X-ray study of foetal and adult sheep haemoglobins. *Proc R Soc Lond A Math Phys Sci* **1948,** *194* (1038), 375-98.

126. Perutz, M. F.; Rossmann, M. G.; Cullis, A. F.; Muirhead, H.; Will, G.; North, A. C., Structure of haemoglobin: a three-dimensional Fourier synthesis at 5.5-A. resolution, obtained by X-ray analysis. *Nature* **1960,** *185* (4711), 416-22.

127. Dodson, E.; Harding, M. M.; Hodgkin, D. C.; Rossmann, M. G., Crystal Structure of Insulin .3. Evidence for a 2-Fold Axis in Rhombohedral Zinc Insulin. *J Mol Biol* **1966,** *16* (1), 227-&.

128. Kemp, T. J.; Alcock, N. W., 100 years of X-ray crystallography. *Sci Prog* **2017,** *100* (1), 25-44.

129. Batool, M.; Ahmad, B.; Choi, S., A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci* **2019,** *20* (11).

130. Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Merson, J. R.; Whittle, P. J.; Danley, D. E.; Geoghegan, K. F.; et al., X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes. *Nature* **1989,** *342* (6247), 299-302.

131. Navia, M. A.; Fitzgerald, P. M.; McKeever, B. M.; Leu, C. T.; Heimbach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P., Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* **1989,** *337* (6208), 615-20.

132. Spinelli, S.; Liu, Q. Z.; Alzari, P. M.; Hirel, P. H.; Poljak, R. J., The 3-Dimensional Structure of the Aspartyl Protease from the Hiv-1 Isolate Bru. *Biochimie* **1991,** *73* (11), 1391-1396.

133. Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B., Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. *Science* **1989,** *245* (4918), 616-21.

134. Ratner, L.; Haseltine, W.; Patarca, R.; Livak, K. J.; Starcich, B.; Josephs, S. F.; Doran, E. R.; Rafalski, J. A.; Whitehorn, E. A.; Baumeister, K.; et al., Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* **1985,** *313* (6000), 277-84.

135. Mellor, J.; Fulton, S. M.; Dobson, M. J.; Wilson, W.; Kingsman, S. M.; Kingsman, A. J., Retroviral Protease-Like Sequence in the Yeast Transposon Ty1 - Reply. *Nature* **1985,** *315* (6021), 691-692.

136. Rao, J. K.; Erickson, J. W.; Wlodawer, A., Structural and evolutionary relationships between retroviral and eucaryotic aspartic proteinases. *Biochemistry-Us* **1991,** *30* (19), 4663-71.

137. Davies, D. R., The structure and function of the aspartic proteinases. *Annu Rev Biophys Biophys Chem* **1990,** *19*, 189-215.

138. Roberts, N. A.; Redshaw, S., Discovery and Development of the HIV Proteinase Inhibitor Ro 31-8959. In *The Search for Antiviral Drugs: Case Histories from Concept to Clinic*, Adams, J.; Merluzzi, V. J., Eds. Birkhäuser Boston: Boston, MA, 1993; pp 129-151.

139. Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Selk, L.; Kent, S. B.; Wlodawer, A., Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. *Science* **1989,** *246* (4934), 1149-52.

140. Wlodawer, A.; Vondrasek, J., Inhibitors of HIV-1 protease: a major success of

structure-assisted drug design. *Annu Rev Biophys Biomol Struct* **1998,** *27*, 249-84.

141. Graves, B. J.; Hatada, M. H.; Miller, J. K.; Graves, M. C.; Roy, S.; Cook, C. M.; Krohn, A.; Martin, J. A.; Roberts, N. A., The three-dimensional x-ray crystal structure of HIV-1 protease complexed with a hydroxyethylene inhibitor. *Adv Exp Med Biol* **1991,** *306*, 455-60.

142. Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H., Novel binding mode of highly potent HIV-proteinase inhibitors incorporating the (R)-hydroxyethylamine isostere. *J Med Chem* **1991,** *34* (11), 3340-2.

143. Roberts, N. A.; Martin, J. A.; Kinchington, D.; Broadhurst, A. V.; Craig, J. C.; Duncan, I. B.; Galpin, S. A.; Handa, B. K.; Kay, J.; Krohn, A.; et al., Rational design of peptide-based HIV proteinase inhibitors. *Science* **1990,** *248* (4953), 358-61.

144. Reich, S. H.; Webber, S. E., Structure-based drug design (SBDD): Every structure tells a story. *Perspectives in Drug Discovery and Design* **1993,** *1* (2), 371-390.

145. Anderson, A. C., The process of structure-based drug design. *Chem Biol* **2003,** *10* (9), 787-97.

146. Wang, L.; Gu, Q.; Zheng, X.; Ye, J.; Liu, Z.; Li, J.; Hu, X.; Hagler, A.; Xu, J., Discovery of new selective human aldose reductase inhibitors through virtual screening multiple binding pocket conformations. *J Chem Inf Model* **2013,** *53* (9), 2409-22.

147. Rutenber, E. E.; Stroud, R. M., Binding of the anticancer drug ZD1694 to E. coli thymidylate synthase: assessing specificity and affinity. *Structure* **1996,** *4* (11), 1317-24.

148. Bai, X. C.; McMullan, G.; Scheres, S. H., How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci* **2015,** *40* (1), 49-57.

149. Bhella, D., Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility. *Biophys Rev* **2019,** *11* (4), 515-519.

150. Leitner, A., Cross-linking and other structural proteomics techniques: how chemistry is enabling mass spectrometry applications in structural biology. *Chem Sci* **2016,** *7* (8), 4792-4803.

151. Kaur, U.; Meng, H.; Lui, F.; Ma, R.; Ogburn, R. N.; Johnson, J. H. R.; Fitzgerald, M. C.; Jones, L. M., Proteome-Wide Structural Biology: An Emerging Field for the Structural Analysis of Proteins on the Proteomic Scale. *J Proteome Res* **2018,** *17* (11), 3614-3627.

152. Serpa, J. J.; Parker, C. E.; Petrotchenko, E. V.; Han, J.; Pan, J.; Borchers, C. H., Mass spectrometry-based structural proteomics. *Eur J Mass Spectrom (Chichester)* **2012,** *18* (2), 251-67.

153. McKenzie-Coe, A.; Montes, N. S.; Jones, L. M., Hydroxyl Radical Protein Footprinting: A Mass Spectrometry-Based Structural Method for Studying the Higher Order Structure of Proteins. *Chem Rev* **2022,** *122* (8), 7532-7561.

154. Takamoto, K.; Chance, M. R., Radiolytic protein footprinting with mass spectrometry to probe the structure of macromolecular complexes. *Annu Rev Biophys Biomol Struct* **2006,** *35*, 251-76.

155. Ermacora, M. R.; Delfino, J. M.; Cuenoud, B.; Schepartz, A.; Fox, R. O., Conformation-dependent cleavage of staphylococcal nuclease with a disulfide-linked iron chelate. *Proc Natl Acad Sci U S A* **1992,** *89* (14), 6383-7.

156. Trabjerg, E.; Nazari, Z. E.; Rand, K. D., Conformational analysis of complex protein states by hydrogen/deuterium exchange mass spectrometry (HDX-MS): Challenges and emerging solutions. *Trac-Trend Anal Chem* **2018,** *106*, 125-138.

157. Pirrone, G. F.; Iacob, R. E.; Engen, J. R., Applications of hydrogen/deuterium exchange MS from 2012 to 2014. *Anal Chem* **2015,** *87* (1), 99-118.

158. Percy, A. J.; Rey, M.; Burns, K. M.; Schriemer, D. C., Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry-a review. *Anal Chim Acta* **2012,** *721*, 7-21.

159. Lomenick, B.; Hao, R.; Jonai, N.; Chin, R. M.; Aghajan, M.; Warburton, S.; Wang, J.; Wu, R. P.; Gomez, F.; Loo, J. A.; Wohlschlegel, J. A.; Vondriska, T. M.; Pelletier, J.; Herschman, H. R.; Clardy, J.; Clarke, C. F.; Huang, J., Target identification using drug affinity responsive target stability (DARTS). *Proc Natl Acad Sci U S A* **2009,** *106* (51), 21984-9.

160. Schopper, S.; Kahraman, A.; Leuenberger, P.; Feng, Y.; Piazza, I.; Muller, O.; Boersema, P. J.; Picotti, P., Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nat Protoc* **2017,** *12* (11), 2391-2410.

161. Feng, Y.; De Franceschi, G.; Kahraman, A.; Soste, M.; Melnik, A.; Boersema, P. J.; de Laureto, P. P.; Nikolaev, Y.; Oliveira, A. P.; Picotti, P., Global analysis of protein structural changes in complex proteomes. *Nat Biotechnol* **2014,** *32* (10), 1036-44.

162. Stauffer, S.; Feng, Y.; Nebioglu, F.; Heilig, R.; Picotti, P.; Helenius, A., Stepwise priming by acidic pH and a high K+ concentration is required for efficient uncoating of influenza A virus cores after penetration. *J Virol* **2014,** *88* (22), 13029-46.

163. Geiger, R.; Rieckmann, J. C.; Wolf, T.; Basso, C.; Feng, Y.; Fuhrer, T.; Kogadeeva, M.; Picotti, P.; Meissner, F.; Mann, M.; Zamboni, N.; Sallusto, F.; Lanzavecchia, A., L-Arginine Modulates T Cell Metabolism and Enhances Survival and Anti-tumor Activity. *Cell* **2016,** *167* (3), 829-842 e13.

164. Calabrese, A. N.; Radford, S. E., Mass spectrometry-enabled structural biology of membrane proteins. *Methods* **2018,** *147*, 187-205.

165. Iacobucci, C.; Gotze, M.; Sinz, A., Cross-linking/mass spectrometry to get a closer view on protein interaction networks. *Curr Opin Biotechnol* **2020,** *63*, 48-53.

166. Kalkhof, S.; Sinz, A., Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. *Anal Bioanal Chem* **2008,** *392* (1-2), 305-12.

167. Madler, S.; Bich, C.; Touboul, D.; Zenobi, R., Chemical cross-linking with NHS esters: a systematic study on amino acid reactivities. *J Mass Spectrom* **2009,** *44* (5), 694-706.

168. Flaxman, H. A.; Chang, C. F.; Wu, H. Y.; Nakamoto, C. H.; Woo, C. M., A Binding Site Hotspot Map of the FKBP12-Rapamycin-FRB Ternary Complex by Photoaffinity Labeling and Mass Spectrometry-Based Proteomics. *J Am Chem Soc* **2019,** *141* (30), 11759-11764.

169. Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd, Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013,** *113* (4), 2343-94.

170. Li, Z.; Hao, P.; Li, L.; Tan, C. Y.; Cheng, X.; Chen, G. Y.; Sze, S. K.; Shen, H. M.; Yao, S. Q., Design and synthesis of minimalist terminal alkyne-containing diazirine photo-crosslinkers and their incorporation into kinase inhibitors for cell- and tissue-based proteome profiling. *Angew Chem Int Ed Engl* **2013,** *52* (33), 8551-6.

171. Burton, N. R.; Kim, P.; Backus, K. M., Photoaffinity labelling strategies for mapping the small molecule-protein interactome. *Org Biomol Chem* **2021,** *19* (36), 7792-7809.

172. Verhelst, S. H.; Fonovic, M.; Bogyo, M., A mild chemically cleavable linker system for functional proteomic applications. *Angew Chem Int Ed Engl* **2007,** *46* (8), 1284-6.

173. Qian, Y.; Martell, J.; Pace, N. J.; Ballard, T. E.; Johnson, D. S.; Weerapana, E., An isotopically tagged azobenzene-based cleavable linker for quantitative proteomics. *Chembiochem* **2013,** *14* (12), 1410-4.

174. Flaxman, H. A.; Miyamoto, D. K.; Woo, C. M., Small Molecule Interactome Mapping by Photo-Affinity Labeling (SIM-PAL) to Identify Binding Sites of Small Molecules on a Proteome-Wide Scale. *Curr Protoc Chem Biol* **2019,** *11* (4), e75.

175. Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y., Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. *Angew Chem Int Ed Engl* **2012,** *51* (24), 5852-6.

176. Parker, B. W.; Goncz, E. J.; Krist, D. T.; Statsyuk, A. V.; Nesvizhskii, A. I.; Weiss, E. L., Mapping low-affinity/high-specificity peptide-protein interactions using ligand-footprinting mass spectrometry. *Proc Natl Acad Sci U S A* **2019,** *116* (42), 21001-21011.

177. Fujinaga, M.; Chernaia, M. M.; Tarasova, N. I.; Mosimann, S. C.; James, M. N., Crystal structure of human pepsin and its complex with pepstatin. *Protein Sci* **1995,** *4* (5), 960-72.

178. Kolbowski, L.; Mendes, M. L.; Rappsilber, J., Optimizing the Parameters Governing the Fragmentation of Cross-Linked Peptides in a Tribrid Mass Spectrometer. *Anal Chem* **2017,** *89* (10), 5311-5318.

179. Mukherjee, S.; Fang, M.; Kok, W. M.; Kapp, E. A.; Thombare, V. J.; Huguet, R.; Hutton, C. A.; Reid, G. E.; Roberts, B. R., Establishing Signature Fragments for Identification and Sequencing of Dityrosine Cross-Linked Peptides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal Chem* **2019,** *91* (19), 12129-12133.

180. Kall, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J., Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* **2007,** *4* (11), 923-5.

181. Pelz, C. R.; Kulesz-Martin, M.; Bagby, G.; Sears, R. C., Global rank-invariant set normalization (GRSN) to reduce systematic distortions in microarray data. *BMC Bioinformatics* **2008,** *9*, 520.

182. Frankenfield, A. M.; Ni, J.; Ahmed, M.; Hao, L., Protein Contaminants Matter: Building Universal Protein Contaminant Libraries for DDA and DIA Proteomics. *J Proteome Res* **2022,** *21* (9), 2104-2113.

183. Sledzieski, S.; Singh, R.; Cowen, L.; Berger, B., D-SCRIPT translates genome to phenome with sequence-based, structure-aware, genome-scale predictions of protein-protein interactions. *Cell Syst* **2021,** *12* (10), 969-982 e6.

184. Rawlings, N. D.; Waller, M.; Barrett, A. J.; Bateman, A., MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **2014,** *42* (Database issue), D503-9.

185. Fahey, M. E.; Bennett, M. J.; Mahon, C.; Jager, S.; Pache, L.; Kumar, D.; Shapiro, A.; Rao, K.; Chanda, S. K.; Craik, C. S.; Frankel, A. D.; Krogan, N. J., GPS-Prot: a web-based visualization platform for integrating host-pathogen interaction data. *BMC Bioinformatics* **2011,** *12*, 298.

186. Orchard, S.; Ammari, M.; Aranda, B.; Breuza, L.; Briganti, L.; Broackes-Carter, F.; Campbell, N. H.; Chavali, G.; Chen, C.; del-Toro, N.; Duesbury, M.; Dumousseau, M.; Galeota, E.; Hinz, U.; Iannuccelli, M.; Jagannathan, S.; Jimenez, R.; Khadake, J.; Lagreid, A.; Licata, L.; Lovering, R. C.; Meldal, B.; Melidoni, A. N.; Milagros, M.; Peluso, D.; Perfetto, L.; Porras, P.; Raghunath, A.; Ricard-Blum, S.; Roechert, B.; Stutz, A.; Tognolli, M.; van Roey, K.; Cesareni, G.; Hermjakob, H., The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. *Nucleic Acids Res* **2014,** *42* (Database issue), D358-63.

187. Turner, B.; Razick, S.; Turinsky, A. L.; Vlasblom, J.; Crowdy, E. K.; Cho, E.; Morrison, K.; Donaldson, I. M.; Wodak, S. J., iRefWeb: interactive analysis of consolidated protein interaction data and their supporting evidence. *Database (Oxford)* **2010,** *2010*, baq023.

188. Stark, C.; Breitkreutz, B. J.; Reguly, T.; Boucher, L.; Breitkreutz, A.; Tyers, M.,

BioGRID: a general repository for interaction datasets. *Nucleic Acids Res* **2006,** *34* (Database issue), D535-9.

189. Alcaraz, L. B.; Mallavialle, A.; David, T.; Derocq, D.; Delolme, F.; Dieryckx, C.; Mollevi, C.; Boissiere-Michot, F.; Simony-Lafontaine, J.; Du Manoir, S.; Huesgen, P. F.; Overall, C. M.; Tartare-Deckert, S.; Jacot, W.; Chardes, T.; Guiu, S.; Roger, P.; Reinheckel, T.; Moali, C.; Liaudet-Coopman, E., A 9-kDa matricellular SPARC fragment released by cathepsin D exhibits pro-tumor activity in the triple-negative breast cancer microenvironment. *Theranostics* **2021,** *11* (13), 6173-6192.

190. Lismont, C.; Nordgren, M.; Brees, C.; Knoops, B.; Van Veldhoven, P. P.; Fransen, M., Peroxisomes as Modulators of Cellular Protein Thiol Oxidation: A New Model System. *Antioxid Redox Signal* **2019,** *30* (1), 22-39.

191. Umezawa, H.; Aoyagi, T.; Morishima, H.; Matsuzaki, M.; Hamada, M., Pepstatin, a new pepsin inhibitor produced by Actinomycetes. *J Antibiot (Tokyo)* **1970,** *23* (5), 259-62.

192. Marciniszyn, J., Jr.; Hartsuck, J. A.; Tang, J., Mode of inhibition of acid proteases by pepstatin. *J Biol Chem* **1976,** *251* (22), 7088-94.

193. Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Hosur, M. V.; Sowder, R. C., 2nd; Cachau, R. E.; Collins, J.; Silva, A. M.; Erickson, J. W., Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. *Proc Natl Acad Sci U S A* **1993,** *90* (14), 6796-800.

194. Vervacke, J. S.; Funk, A. L.; Wang, Y. C.; Strom, M.; Hrycyna, C. A.; Distefano, M. D., Diazirine-containing photoactivatable isoprenoid: synthesis and application in studies with isoprenylcysteine carboxyl methyltransferase. *J Org Chem* **2014,** *79* (5), 1971-8.

195. Jensen, K. J.; Alsina, J.; Songster, M. F.; Vagner, J.; Albericio, F.; Barany, G., Backbone Amide Linker (BAL) strategy for solid-phase synthesis of C-terminal-modified and cyclic peptides. *J Am Chem Soc* **1998,** *120* (22), 5441-5452.

196. Schmidt, T.; Samaras, P.; Frejno, M.; Gessulat, S.; Barnert, M.; Kienegger, H.; Krcmar, H.; Schlegl, J.; Ehrlich, H. C.; Aiche, S.; Kuster, B.; Wilhelm, M., ProteomicsDB. *Nucleic Acids Res* **2018,** *46* (D1), D1271-D1281.

197. Samaras, P.; Schmidt, T.; Frejno, M.; Gessulat, S.; Reinecke, M.; Jarzab, A.; Zecha, J.; Mergner, J.; Giansanti, P.; Ehrlich, H. C.; Aiche, S.; Rank, J.; Kienegger, H.; Krcmar, H.; Kuster, B.; Wilhelm, M., ProteomicsDB: a multi-omics and multi-organism resource for life science research. *Nucleic Acids Res* **2020,** *48* (D1), D1153-D1163.

198. Singh, R.; Devkota, K.; Sledzieski, S.; Berger, B.; Cowen, L., Topsy-Turvy: integrating a global view into sequence-based PPI prediction. *Bioinformatics* **2022,** *38* (Suppl 1), i264-i272.

199. Katsuragi, Y.; Ichimura, Y.; Komatsu, M., p62/SQSTM1 functions as a signaling hub and an autophagy adaptor. *FEBS J* **2015,** *282* (24), 4672-8.

200. Seo, S. U.; Woo, S. M.; Im, S. S.; Jang, Y.; Han, E.; Kim, S. H.; Lee, H.; Lee, H. S.; Nam, J. O.; Gabrielson, E.; Min, K. J.; Kwon, T. K., Cathepsin D as a potential therapeutic target to enhance anticancer drug-induced apoptosis via RNF183-mediated destabilization of Bcl-xL in cancer cells. *Cell Death Dis* **2022,** *13* (2), 115.

201. Vidmar, R.; Vizovisek, M.; Turk, D.; Turk, B.; Fonovic, M., Protease cleavage site fingerprinting by label-free in-gel degradomics reveals pH-dependent specificity switch of legumain. *Embo J* **2017,** *36* (16), 2455-2465.

202. Bleiholder, C.; Suhai, S.; Harrison, A. G.; Paizs, B., Towards understanding the tandem mass spectra of protonated oligopeptides. 2: The proline effect in collision-induced dissociation of protonated Ala-Ala-Xxx-Pro-Ala (Xxx = Ala, Ser, Leu, Val, Phe, and Trp). *J Am Soc Mass Spectrom* **2011,** *22* (6), 1032-9.

203. Breci, L. A.; Tabb, D. L.; Yates, J. R., 3rd; Wysocki, V. H., Cleavage N-terminal to proline: analysis of a database of peptide tandem mass spectra. *Anal Chem* **2003,** *75* (9), 1963-71.

204. Memboeuf, A.; Nasioudis, A.; Indelicato, S.; Pollreisz, F.; Kuki, A.; Keki, S.; van den Brink, O. F.; Vekey, K.; Drahos, L., Size effect on fragmentation in tandem mass spectrometry. *Anal Chem* **2010,** *82* (6), 2294-302.

205. Revesz, A.; Hever, H.; Steckel, A.; Schlosser, G.; Szabo, D.; Vekey, K.; Drahos, L., Collision energies: Optimization strategies for bottom-up proteomics. *Mass Spectrom Rev* **2021**, e21763.

206. Scientific, T. F., Thermo Fisher Scientific Product Support Bulletin 104. In *PRODUCT SUPPORT BULLETIN*.

207. Scientific, T. F., Thermo Fisher Scientific Product Support Bulletin 121. In *PRODUCT SUPPORT BULLETIN*.

208. Scientific, T. F., Glossary. In *Orbitrap Tribrid Series Getting Started Guide*, 2019; p 88.

209. Shi, H.; Uttamchandani, M.; Yao, S. Q., Applying small molecule microarrays and resulting affinity probe cocktails for proteome profiling of mammalian cell lysates. *Chem Asian J* **2011,** *6* (10), 2803-15.

210. Nussbaumerova, M.; Srp, J.; Masa, M.; Hradilek, M.; Sanda, M.; Reinis, M.; Horn, M.; Mares, M., Single- and double-headed chemical probes for detection of active cathepsin D in a cancer cell proteome. *Chembiochem* **2010,** *11* (11), 1538-41.

211. Gertsik, N.; Chau, D. M.; Li, Y. M., gamma-Secretase Inhibitors and Modulators Induce Distinct Conformational Changes in the Active Sites of gamma-Secretase and Signal Peptide Peptidase. *ACS Chem Biol* **2015,** *10* (8), 1925-31.

212. Abbott, D. E.; Margaryan, N. V.; Jeruss, J. S.; Khan, S.; Kaklamani, V.; Winchester, D. J.; Hansen, N.; Rademaker, A.; Khalkhali-Ellis, Z.; Hendrix, M. J., Reevaluating cathepsin D as a biomarker for breast cancer: serum activity levels versus histopathology. *Cancer Biol Ther* **2010,** *9* (1), 23-30.

213. Ge, S. S.; Chen, B.; Wu, Y. Y.; Long, Q. S.; Zhao, Y. L.; Wang, P. Y.; Yang, S., Current advances of carbene-mediated photoaffinity labeling in medicinal chemistry. *Rsc Adv* **2018,** *8* (51), 29428-29454.

214. Halloran, M. W.; Lumb, J. P., Recent Applications of Diazirines in Chemical Proteomics. *Chemistry* **2019,** *25* (19), 4885-4898.

215. Wolfe, M. S.; Citron, M.; Diehl, T. S.; Xia, W.; Donkor, I. O.; Selkoe, D. J., A substrate-based difluoro ketone selectively inhibits Alzheimer's gamma-secretase activity. *J Med Chem* **1998,** *41* (1), 6-9.

216. Philip, A. T.; Devkota, S.; Malvankar, S.; Bhattarai, S.; Meneely, K. M.; Williams, T. D.; Wolfe, M. S., Designed Helical Peptides as Functional Probes for gamma-Secretase. *Biochemistry-Us* **2019,** *58* (44), 4398-4407.

217. Wolfe, M. S., Probing Mechanisms and Therapeutic Potential of gamma-Secretase in Alzheimer's Disease. *Molecules* **2021,** *26* (2).

218. Fittler, H.; Avrutina, O.; Empting, M.; Kolmar, H., Potent inhibitors of human matriptase-1 based on the scaffold of sunflower trypsin inhibitor. *J Pept Sci* **2014,** *20* (6), 415-20.

219. Tam, J.; Henault, M.; Li, L.; Wang, Z.; Partridge, A. W.; Melnyk, R. A., An activity-based probe for high-throughput measurements of triacylglycerol lipases. *Anal Biochem* **2011,** *414* (2), 254-60.

220. Crowley, V. M.; Thielert, M.; Cravatt, B. F., Functionalized Scout Fragments for Site-Specific Covalent Ligand Discovery and Optimization. *ACS Cent Sci* **2021,** *7* (4), 613-623.

221. Laurent-Matha, V.; Maruani-Herrmann, S.; Prebois, C.; Beaujouin, M.; Glondu, M.; Noel, A.; Alvarez-Gonzalez, M. L.; Blacher, S.; Coopman, P.; Baghdiguian, S.; Gilles, C.; Loncarek, J.; Freiss, G.; Vignon, F.; Liaudet-Coopman, E., Catalytically inactive human cathepsin D triggers fibroblast invasive growth. *J Cell Biol* **2005,** *168* (3), 489-99.

222. Laurent-Matha, V.; Farnoud, M. R.; Lucas, A.; Rougeot, C.; Garcia, M.; Rochefort, H., Endocytosis of pro-cathepsin D into breast cancer cells is mostly independent of mannose-6-phosphate receptors. *J Cell Sci* **1998,** *111 (Pt 17)*, 2539-49.

223. Hasilik, A.; von Figura, K.; Conzelmann, E.; Nehrkorn, H.; Sandhoff, K., Lysosomal enzyme precursors in human fibroblasts. Activation of cathepsin D precursor in vitro and activity of beta-hexosaminidase A precursor towards ganglioside GM2. *Eur J Biochem* **1982,** *125* (2), 317-21.

224. Sakurai, K.; Ozawa, S.; Yamada, R.; Yasui, T.; Mizuno, S., Comparison of the reactivity of carbohydrate photoaffinity probes with different photoreactive groups. *Chembiochem* **2014,** *15* (10), 1399-403.

225. Yan, T.; Desai, H. S.; Boatner, L. M.; Yen, S. L.; Cao, J.; Palafox, M. F.; Jami-Alahmadi, Y.; Backus, K. M., SP3-FAIMS Chemoproteomics for High-Coverage Profiling of the Human Cysteinome*. *Chembiochem* **2021,** *22* (10), 1841-1851.

226. Desai, H. S.; Yan, T.; Backus, K. M., SP3-FAIMS-Enabled High-Throughput Quantitative Profiling of the Cysteinome. *Curr Protoc* **2022,** *2* (7), e492.

227. Trowbridge, A. D.; Seath, C. P.; Rodriguez-Rivera, F. P.; Li, B. X.; Dul, B. E.; Schwaid, A. G.; Buksh, B. F.; Geri, J. B.; Oakley, J. V.; Fadeyi, O. O.; Oslund, R. C.; Ryu, K. A.; White, C.; Reyes-Robles, T.; Tawa, P.; Parker, D. L., Jr.; MacMillan, D. W. C., Small molecule photocatalysis enables drug target identification via energy transfer. *Proc Natl Acad Sci U S A* **2022,** *119* (34), e2208077119.

228. Oakley, J. V.; Buksh, B. F.; Fernandez, D. F.; Oblinsky, D. G.; Seath, C. P.; Geri, J. B.; Scholes, G. D.; MacMillan, D. W. C., Radius measurement via super-resolution microscopy enables the development of a variable radii proximity labeling platform. *Proc Natl Acad Sci U S A* **2022,** *119* (32), e2203027119.

229. Buksh, B. F.; Knutson, S. D.; Oakley, J. V.; Bissonnette, N. B.; Oblinsky, D. G.; Schwoerer, M. P.; Seath, C. P.; Geri, J. B.; Rodriguez-Rivera, F. P.; Parker, D. L.; Scholes, G. D.; Ploss, A.; MacMillan, D. W. C., muMap-Red: Proximity Labeling by Red Light Photocatalysis. *J Am Chem Soc* **2022,** *144* (14), 6154-6162.

230. Li, X. M.; Huang, S.; Li, X. D., Photo-ANA enables profiling of host-bacteria protein interactions during infection. *Nat Chem Biol* **2023**.

8 Appendix: Supplementary information

8.1 Pepstatin-based probes for photoaffinity labeling of aspartic proteases

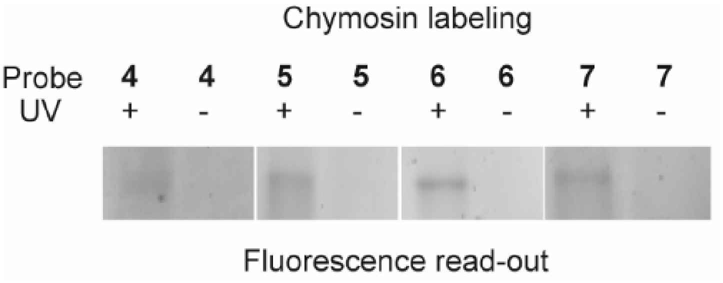


Figure S1. Labeling of purified chymosin by probes **4-7** (2 μM) with or without 365 nm UV irradiation (30 min) followed by click chemistry with an TAMRA-azide dye and in-gel scanning. Note that the absence of bands without UV confirms photoaffinity labeling and specificity of the click chemistry reaction.

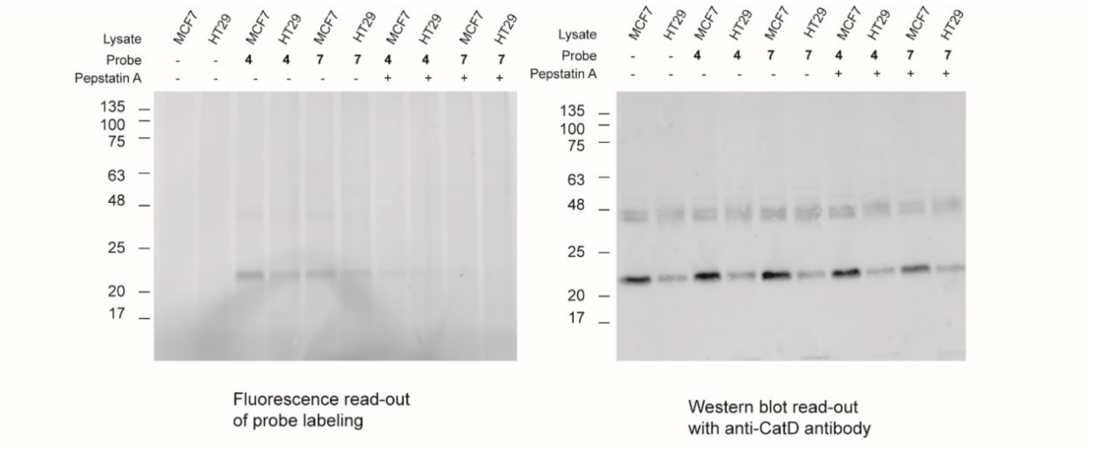

Lysate MCF7 HT29 MCF7 HT29 MCF7 HT29 MCF7 HT29 MCF7 HT29
Probe - - 4 4 7 7 4 4 7 7
Pepstatin A - - - - - - + + + +
135
100
75
63
48
25
20
17
Fluorescence read-out of probe labeling
Western blot read-out with anti-CatD antibody

Figure S2. Probe labeling (left panel) in MCF-7 and HT29 lysates and Western blot (right panel) with an anti-cathepsin D antibody. Note that the band just below 48 kDa is the pro-form of cathepsin D, which is barely labeled by the probes.

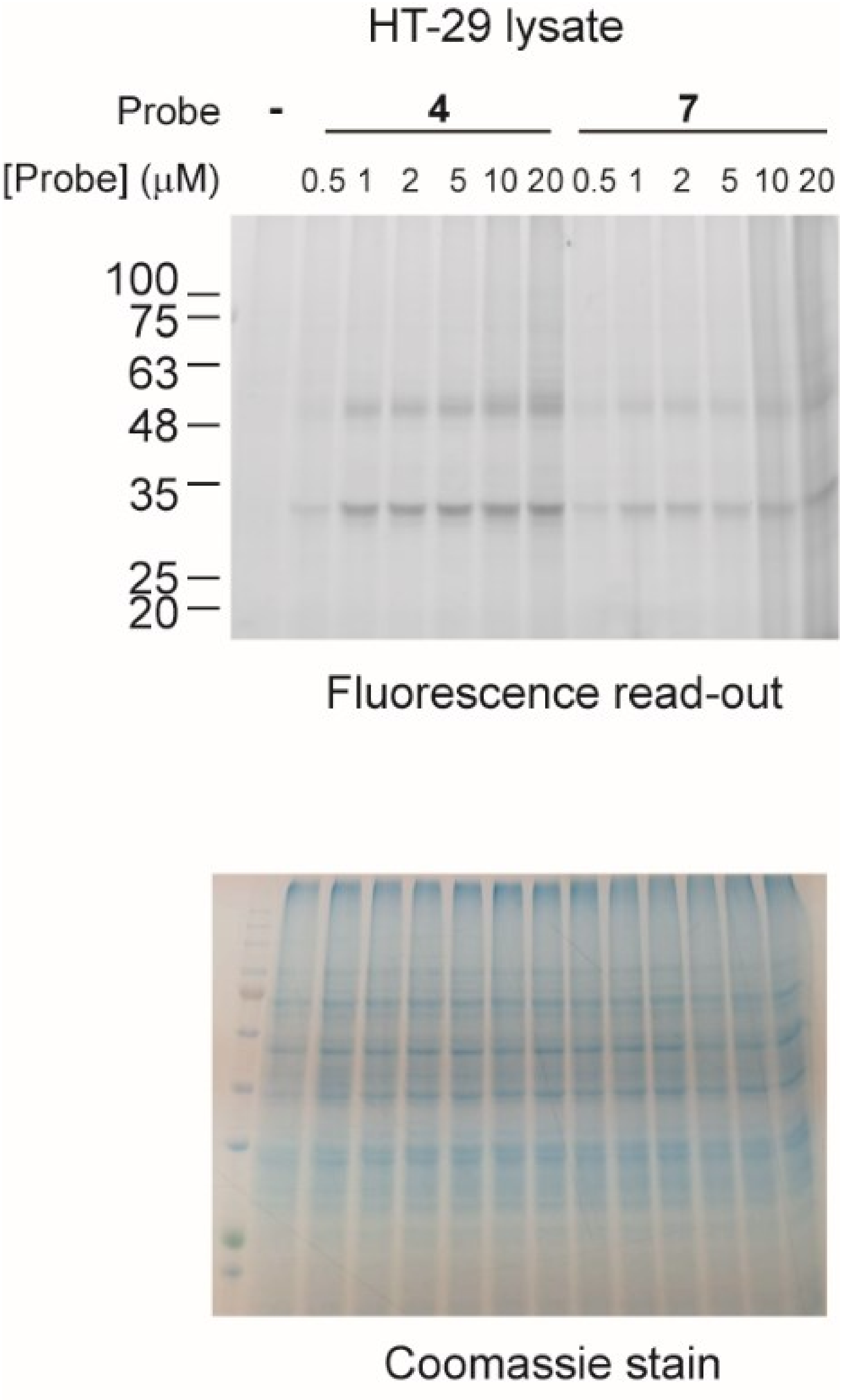


Figure S3. Labeling of HT-29 cell lysates by pepstatin-based AfBPs **4** and **7**. Labeling of targets with increasing concentration of probe reveals saturation of labeling at 1-2 μM probe concentration.

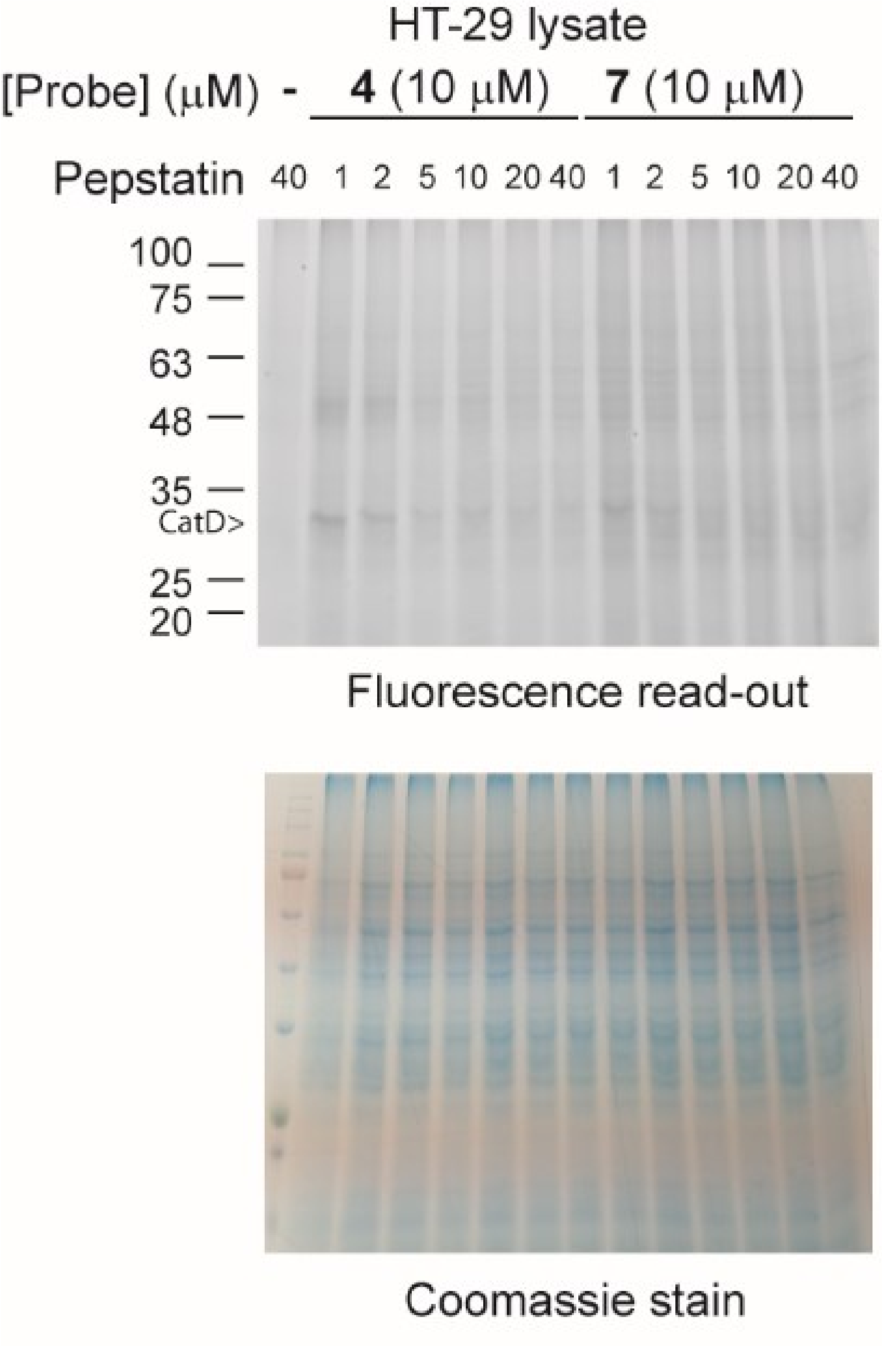

HT-29 lysate
[Probe] (μM) - 4 (10 μM) 7 (10 μM)
Pepstatin 40 1 2 5 10 20 40 1 2 5 10 20 40
100
75
63
48
35
CatD>
25
20
Fluorescence read-out
Coomassie stain

Figure S4. Addition of increasing concentrations of Pepstatin A as competitor for labeling with constant concentration of probe **4** or probe **7** in HT-29 lysates shows that labeling the band at approximately 30 kDa is outcompeted, illustrating the specificity of the binding event and suggesting the same binding pocket for the probes as the parent Pepstatin A.

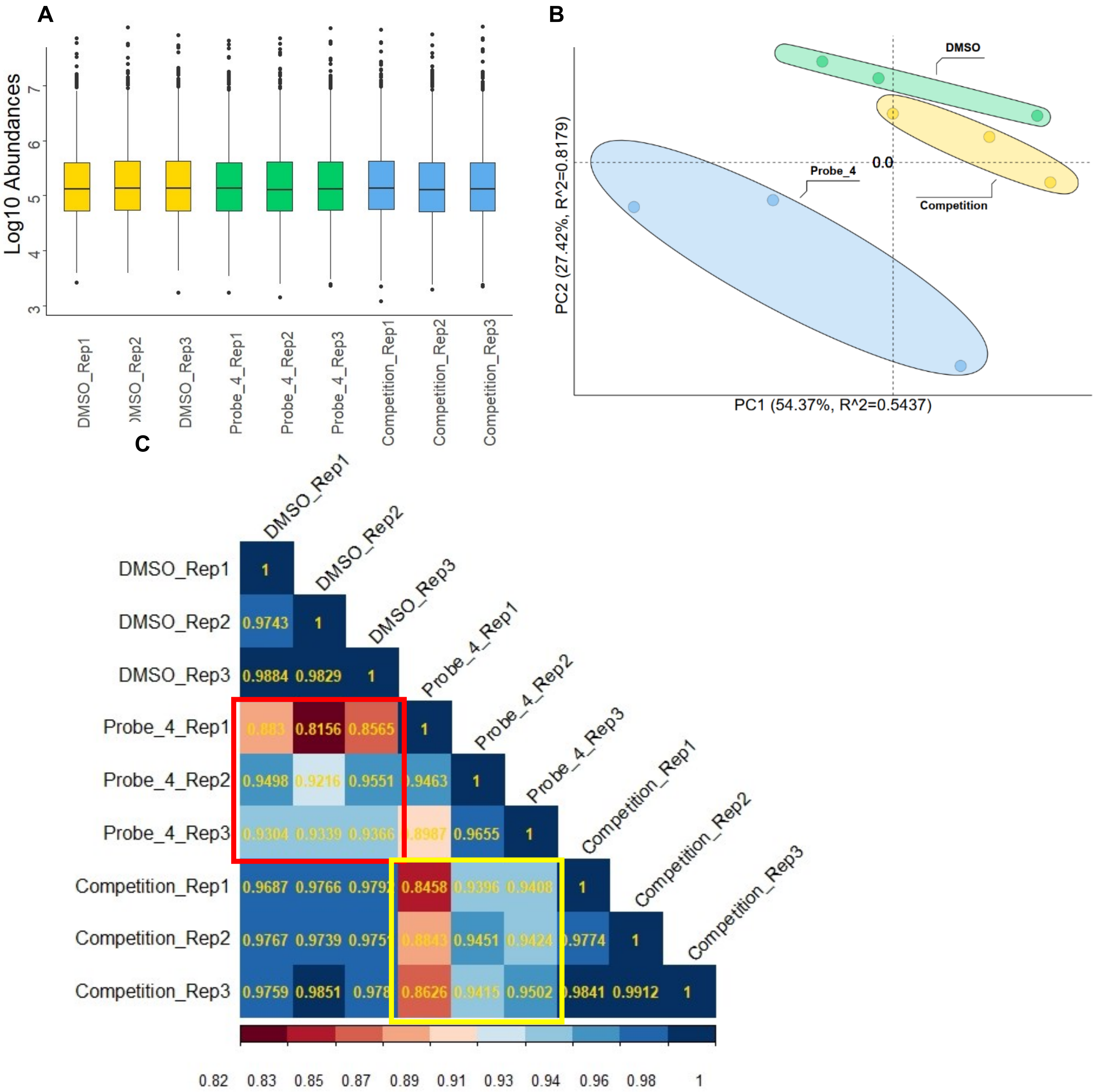
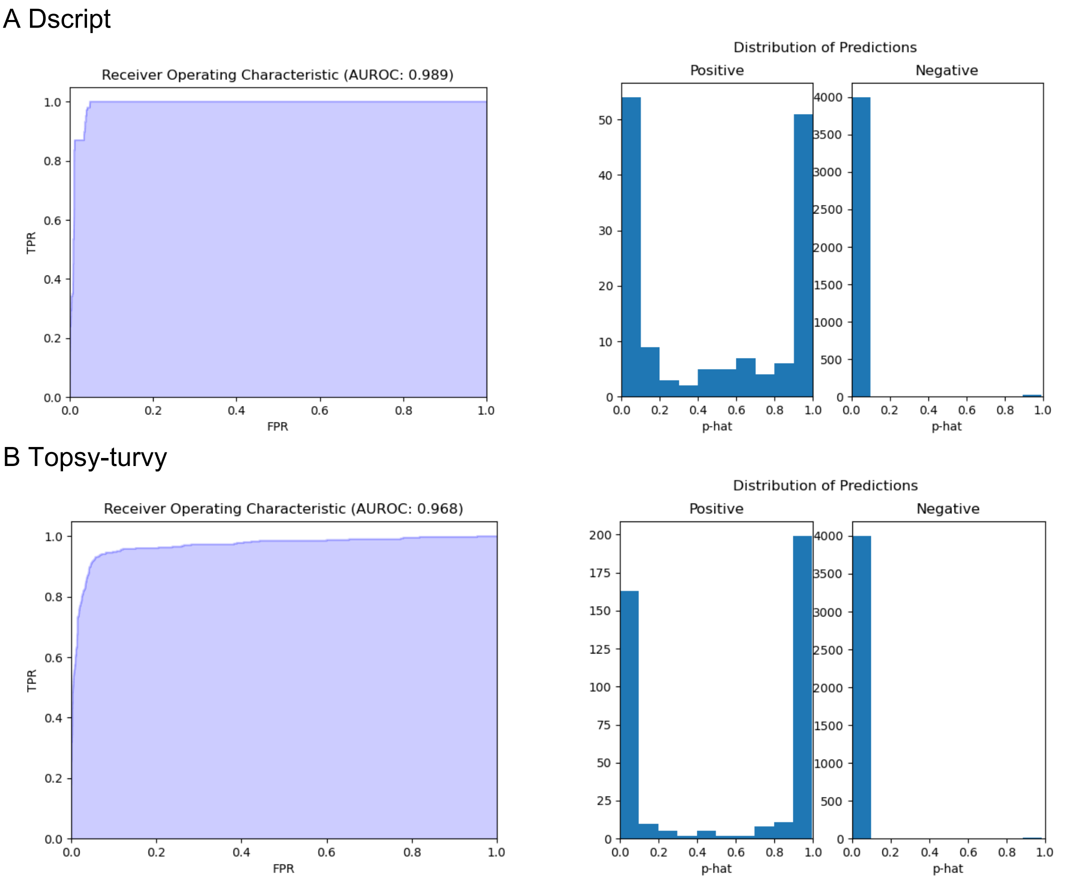


Figure S5. Quality control of LFQ of the different runs. **(A)** Boxplot showing similar protein abundances for the different replicates. **(B)** Principal component analysis. Graphical representation of the first two principal components for the three different treatments show distinct populations with those of the DMSO and competition control being closest to each other. **(C)** Pearson correlation plot generally shows good correlations between replicates and lower correlation between probe and dmso (boxed in red), as well as probe and competition (boxed in yellow).



A Dscript

B Topsy-turvy

Figure S6. Evaluation of models trained by two different methods. Left: Receiver operating characteristic for the models generated by Dscript **(A)** and Topsy-turvy **(B)** respectively. Right: histogram panels illustrate the distribution of prediction scores for the assumed positive and negative datasets.

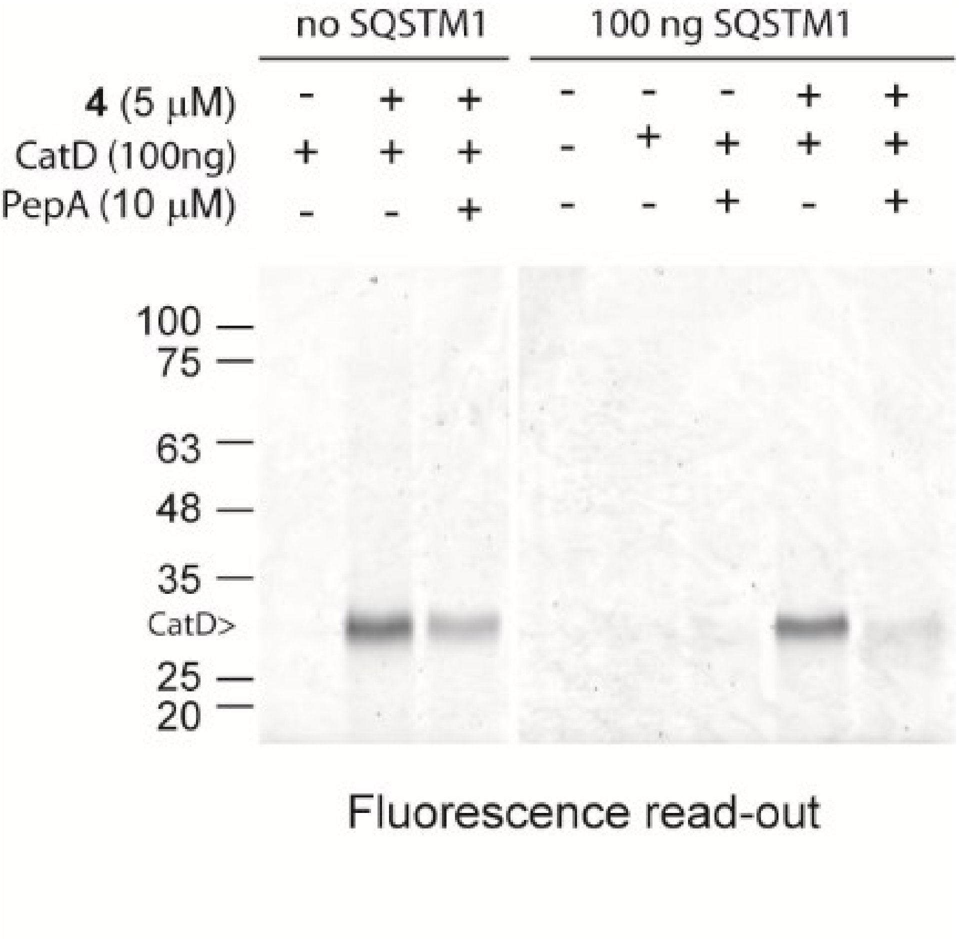


Figure S7. Lack of photoaffinity labeling of SQSTM1 in the absence and presence of cathepsin D reveals that SQSTM1 is not a direct target of pepstatin A probe, and is also not labeled by possible photoaffinity labeling-by-proxy through interaction with cathepsin D.

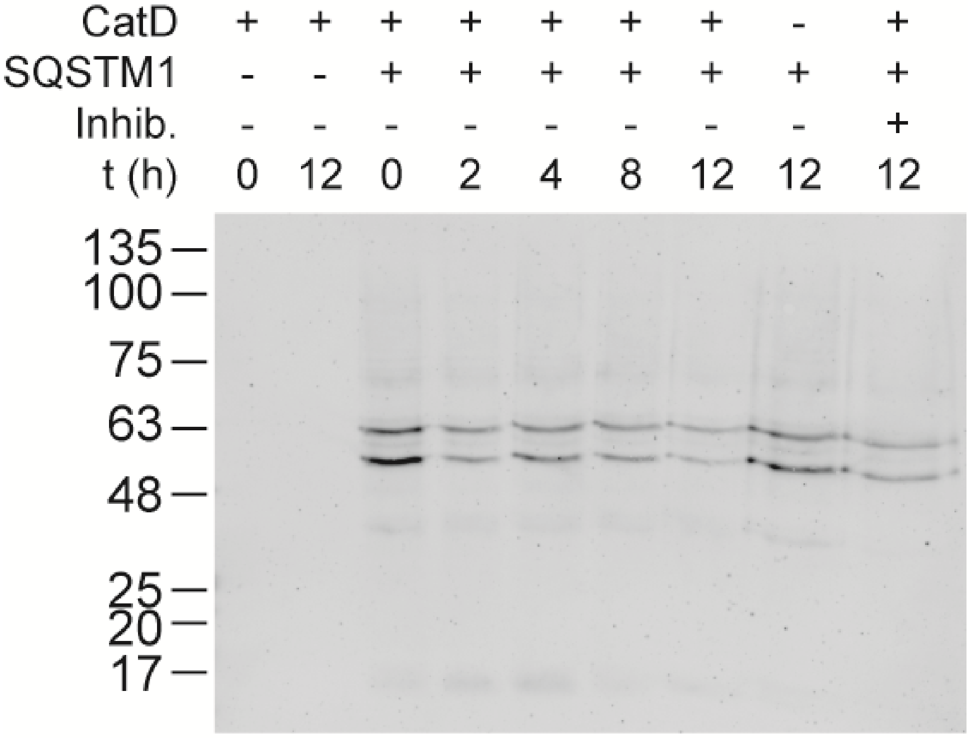


Figure S8. Incubation of SQSTM1 with cathepsin D for an increasing amount of time reveals degradation as detected by Western Blot.

Table S1, Candidates for deep learning prediction

Genes	PSMs	Unique Peptides	Filter criteria
CD276	17	2	c
COPB1	8	2	c
FLOT1	9	2	c
GALK1	20	2	a
GEMIN5	8	2	c
GLRX5	28	2	a
TIMM10B	6	2	c
HMGN2	14	2	c
PEA15	49	3	a
ESYT2	11	3	c
FDPS	54	3	a
GNB1	46	1	c
KRT86	25	2	c
PTPN23	10	3	c
MYL12B	72	4	a
PPP1R7	8	4	c
Q8NCW5	32	5	b
P19367	38	6	b
Q14978	52	6	a
P67775	38	6	b
P31949	300	8	a
CTSD	1240	21	a & b
ARID2	8	1	c
IGLC2	8	1	c
SLC25A11	8	2	c
MYADM	4	1	c
NLN	8	1	c
PSIP1	7	2	c
CTSH	8	1	c
RAB11FIP1	4	1	c
ATP2A3	20	1	c
SQSTM1	4	1	c
YKT6	6	2	c
WDR1	4	1	c

a: Probe **4**/DMSO, p-value ≤ 0.05 and a Log2 fold change ≥ 1 (2-fold enrichment) were taken as cut-off values. b: Probe **4**/competition, p-value ≤ 0.05 and a Log2 fold change ≥ 1 (2-fold enrichment) were taken as cut-off values. c: PSMs ≥ 3, protein confidence is „high", IDs of Probe 4 >0, IDs of DMSO ≤ 1 and IDs of competition ≤ 1; Note: IDs means count number of 'TRUE' identification in triplicates (0-3).

Table S2A: Proteins with Dscript prediction score*

Protein name	Accession	Dscript prediction score
NAXE	Q8NCW5	0.964911401
HK1	P19367	0.96404171
CTSD	P07339	0.827750921
PEA15	Q15121	0.995822906
FDPS	P14324	0.990472853
KRT86	O43790	0.746884525
ARID2	Q68CP9	0.87320292
CD276	Q5ZPR3	0.644954979
ESYT2	A0FGR8	0.994292974
TIMM10B	Q9Y5J6	0.852067649
MYADM	Q96S97	0.97159189
RAB11FIP1	Q6WKZ4	0.987747788
SQSTM1	Q13501	0.982824445

* cut-off score of 0.6 was utilized

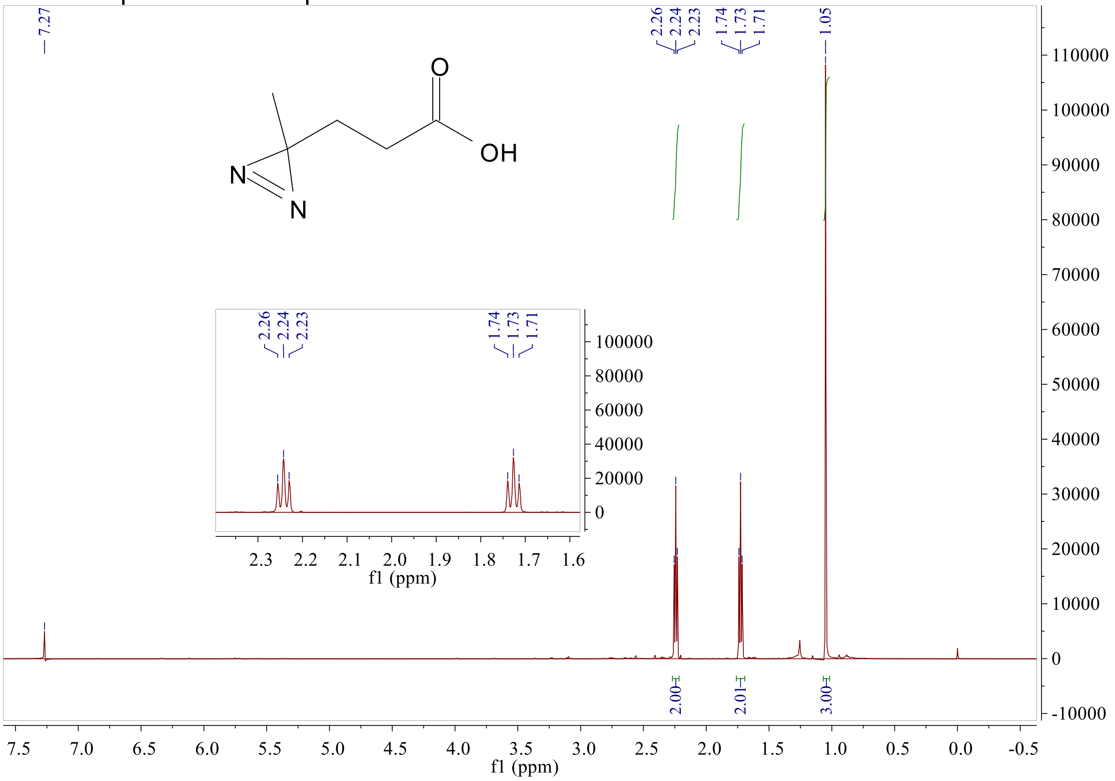
Table S2B: Proteins with Topsy-Turvy prediction score*

protein name	accession	Topsy-turvy prediction score
HK1	P19367	0.985040426
PPP2CA	P67775	0.978475809
GALK1	P51570	0.986124694
GLRX5	Q86SX6	0.957203269
NOLC1	Q14978	0.988221347
ARID2	Q68CP9	0.986543953
CD276	Q5ZPR3	0.962188542
COPB1	P53618	0.987775803
ESYT2	A0FGR8	0.981552303
FLOT1	O75955	0.99014014
GEMIN5	Q8TEQ6	0.974216819
IGLC2	P0DOY2	0.963448584
NLN	Q9BYT8	0.98279804
PSIP1	O75475	0.974285185
CTSH	P09668	0.978896677
RAB11FIP1	Q6WKZ4	0.986103952
ATP2A3	Q93084	0.975800574
SQSTM1	Q13501	0.983634472
YKT6	O15498	0.969990492
PTPN23	Q9H3S7	0.98051399

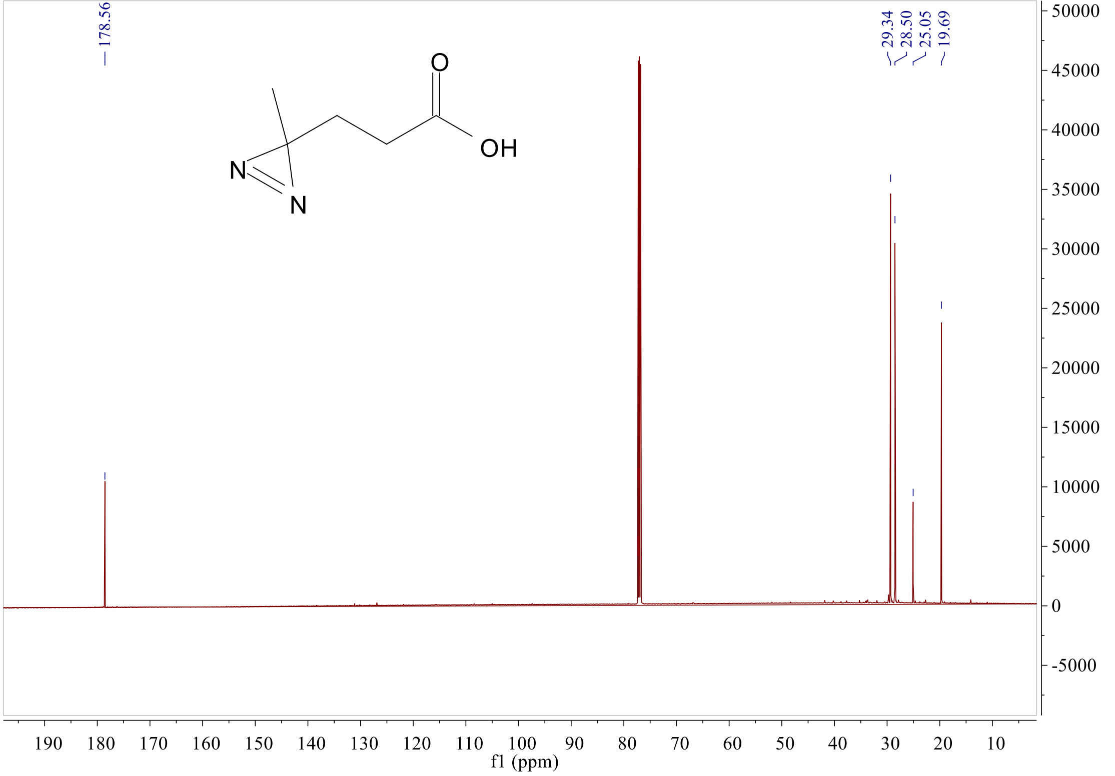
* cut-off score of 0.95 was utilized

Copies of NMR spectra

^{1}H NMR spectrum of compound **3**

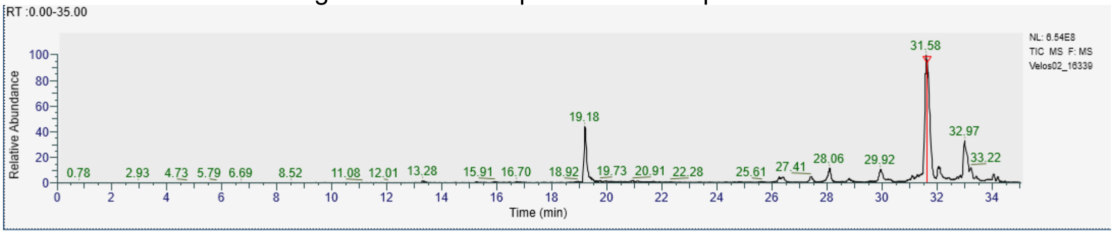


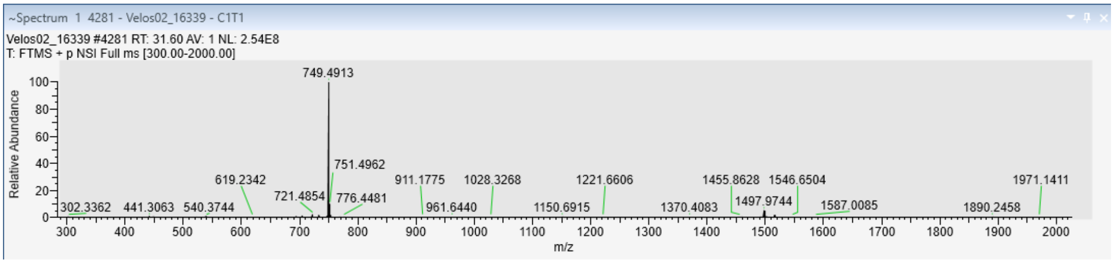
^{13}C NMR spectrum of compound **3**



LC-MS spectra of final probes

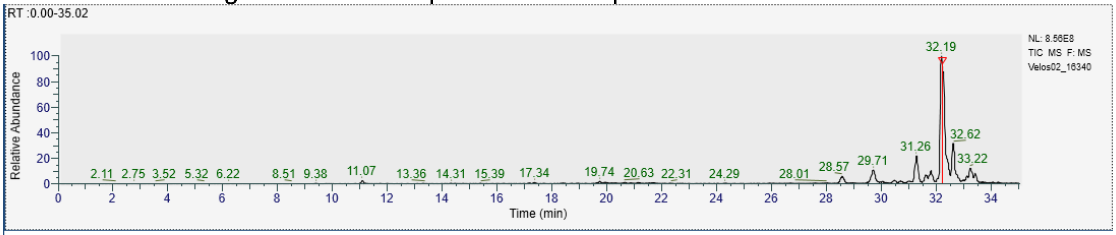
Total ion current chromatogram and mass spectrum of compound **4**

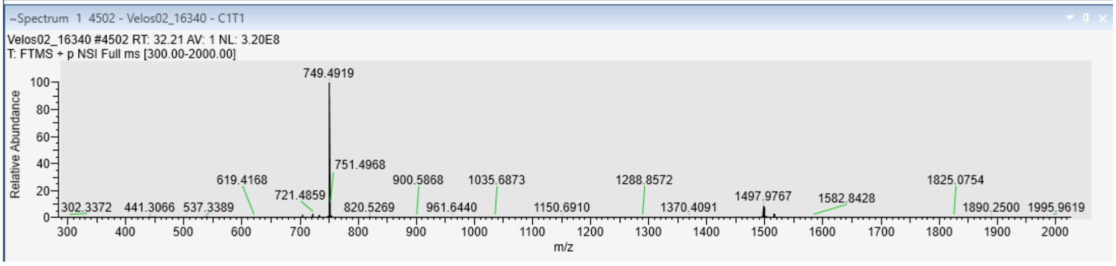



~Spectrum 1 4281 - Velos02_16339 - C1T1

Velos02_16339 #4281 RT: 31.60 AV: 1 NL: 2.54E8
T: FTMS + p NSI Full ms [300.00-2000.00]

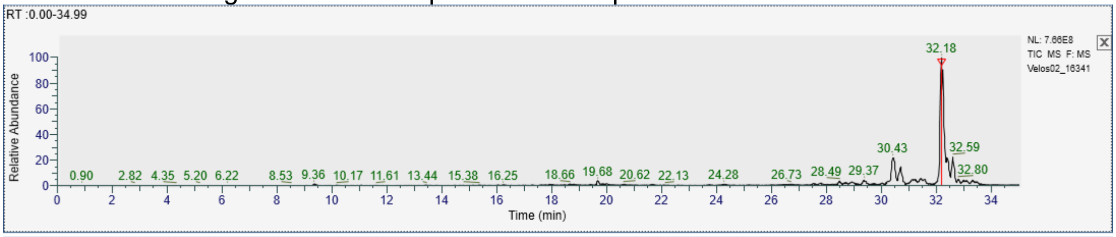
Total ion chromatogram and mass spectrum of compound **5**

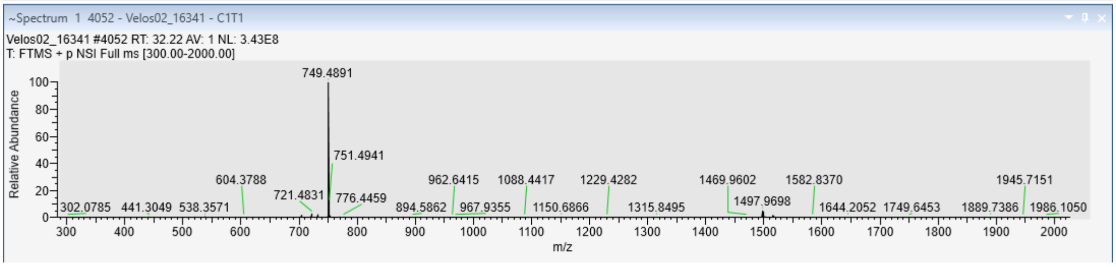



~Spectrum 1 4502 - Velos02_16340 - C1T1

Velos02_16340 #4502 RT: 32.21 AV: 1 NL: 3.20E8
T: FTMS + p NSI Full ms [300.00-2000.00]

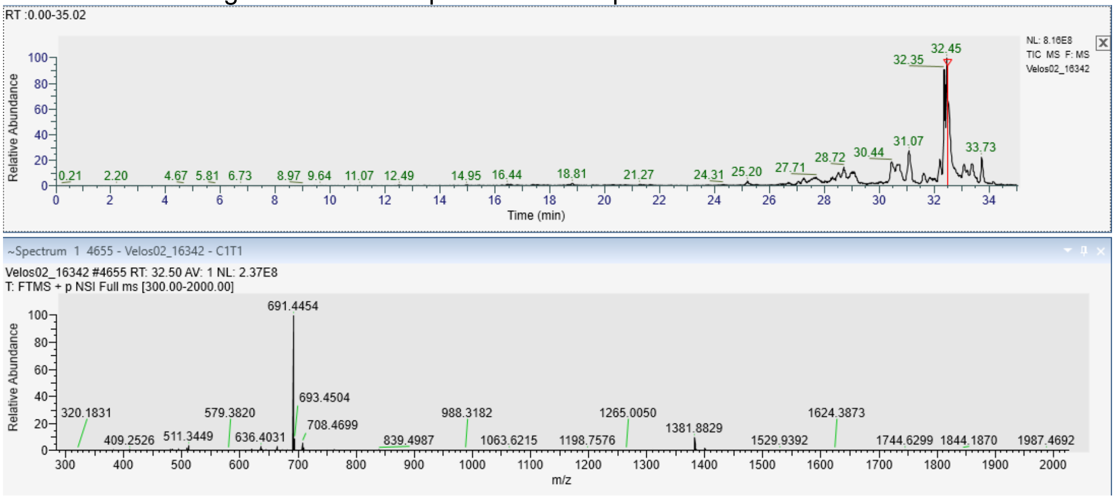
Total ion chromatogram and mass spectrum of compound **6**




~Spectrum 1 4052 - Velos02_16341 - C1T1

Velos02_16341 #4052 RT: 32.22 AV: 1 NL: 3.43E8
T: FTMS + p NSI Full ms [300.00-2000.00]

Total ion chromatogram and mass spectrum of compound **7**



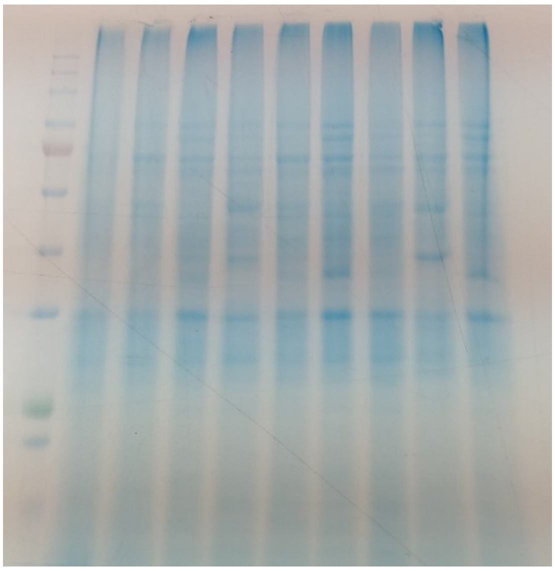
RT :0.00-35.02

NL: 8.16E8
TIC MS F: MS
Velos02_16342

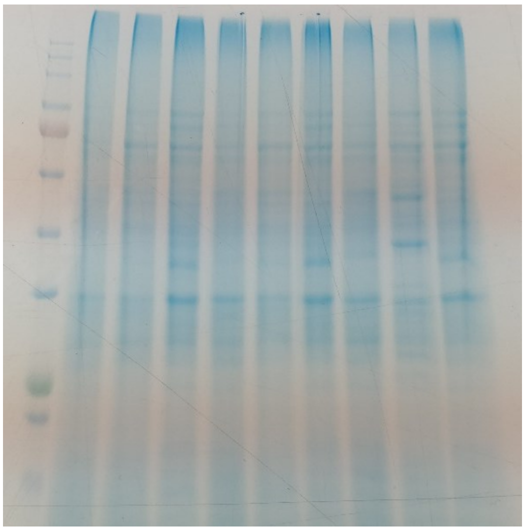
~Spectrum 1 4655 - Velos02_16342 - C1T1

Velos02_16342 #4655 RT: 32.50 AV: 1 NL: 2.37E8
T: FTMS + p NSI Full ms [300.00-2000.00]

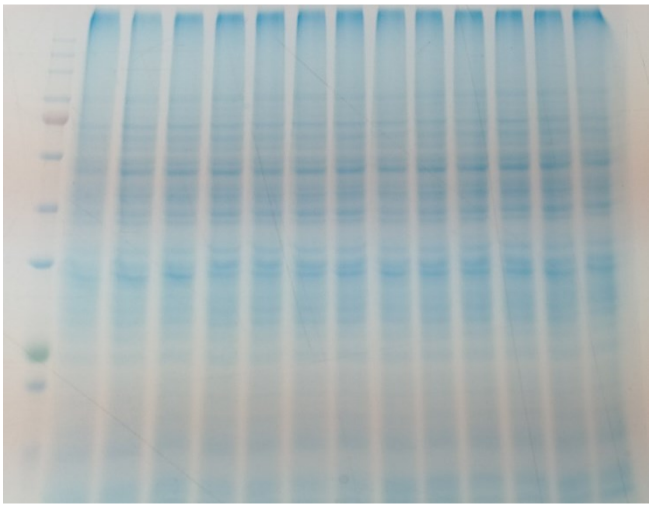
Coomassie stains of gels



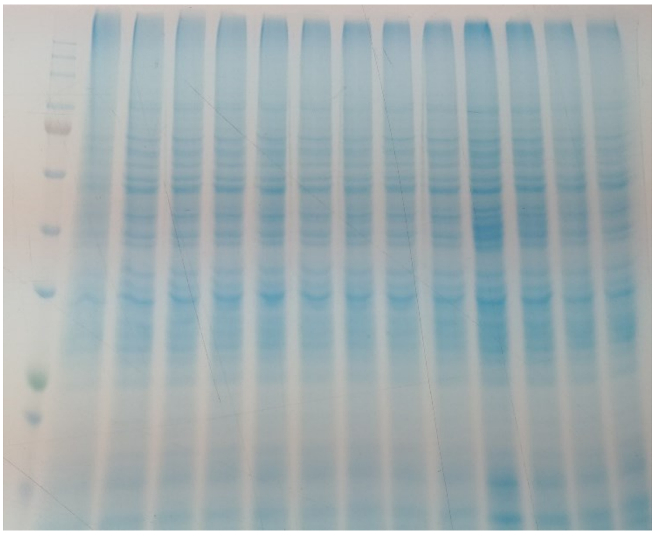
Coomassie stain of gel in Figure 14a, left panel



Coomassie stain of gel in Figure 14a, right panel



Coomassie stain of gel in Figure 14b.



Coomassie stain of gel in Figure 14c.

8.2 Mass cleavable affinity-based probes for precise mapping of binding hotspots

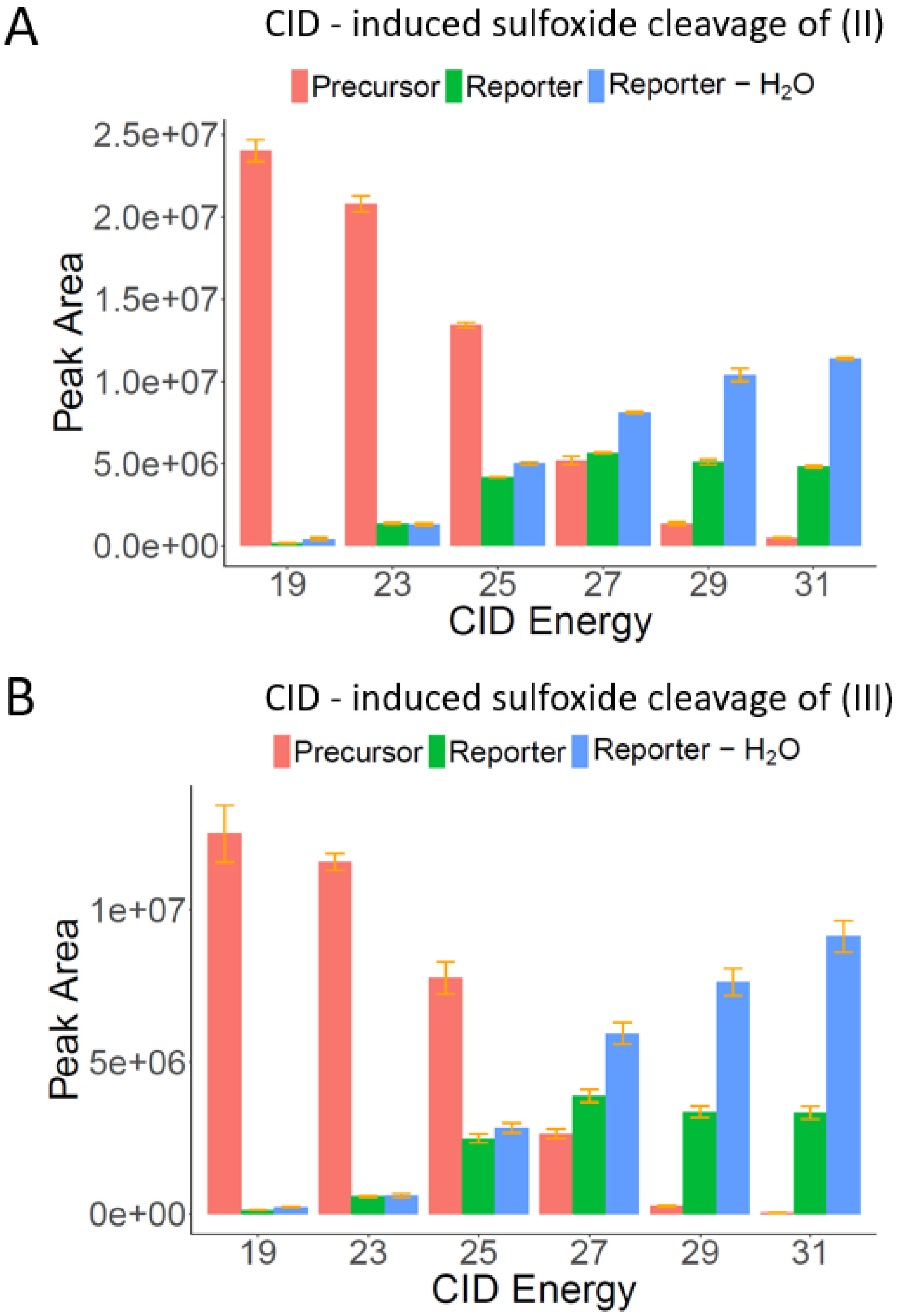


Figure S9. **(A)** Cleavage of (II) at CID energy 19%-31%. **(B)** Cleavage of (III) at CID energy 19%-31%. **Error bar:** mean of of peak area (3 replicates), standard deviation. Fragments: [Reporter] ($C_{35}H_{61}N_6O_8$, theoretical m/z: 693.4550, z: 1), [Reporter – H_2O] ($C_{35}H_{59}N_6O_7$, theoretical m/z: 675.4445, z: 1).

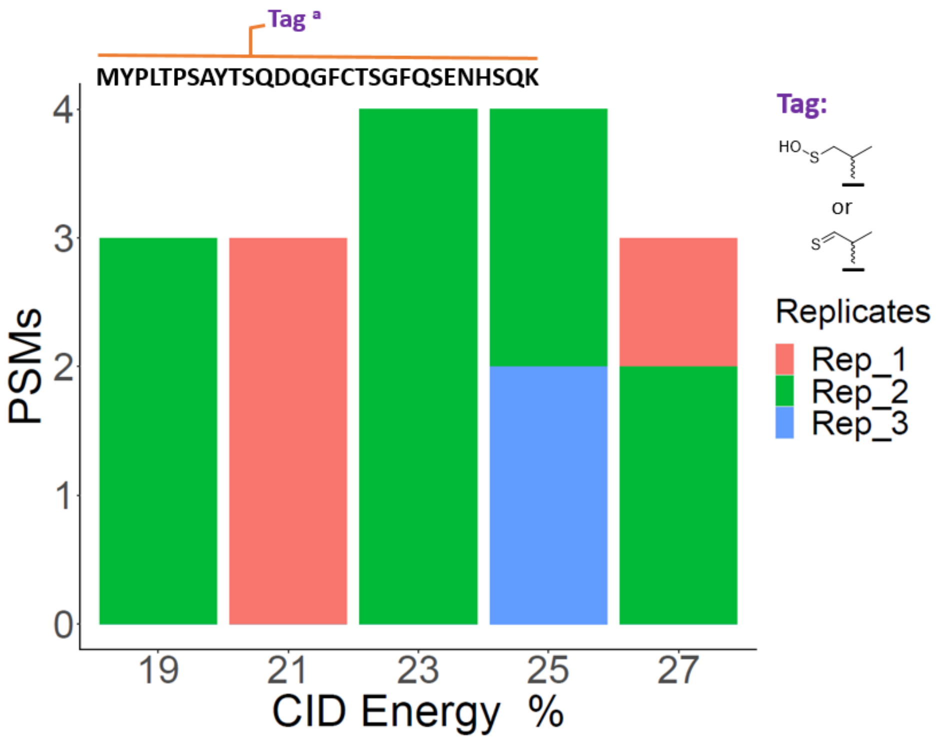


Figure S10. PSMs of tag-modified peptide MYP***SQK from data-dependent acquisition (DDA^2).

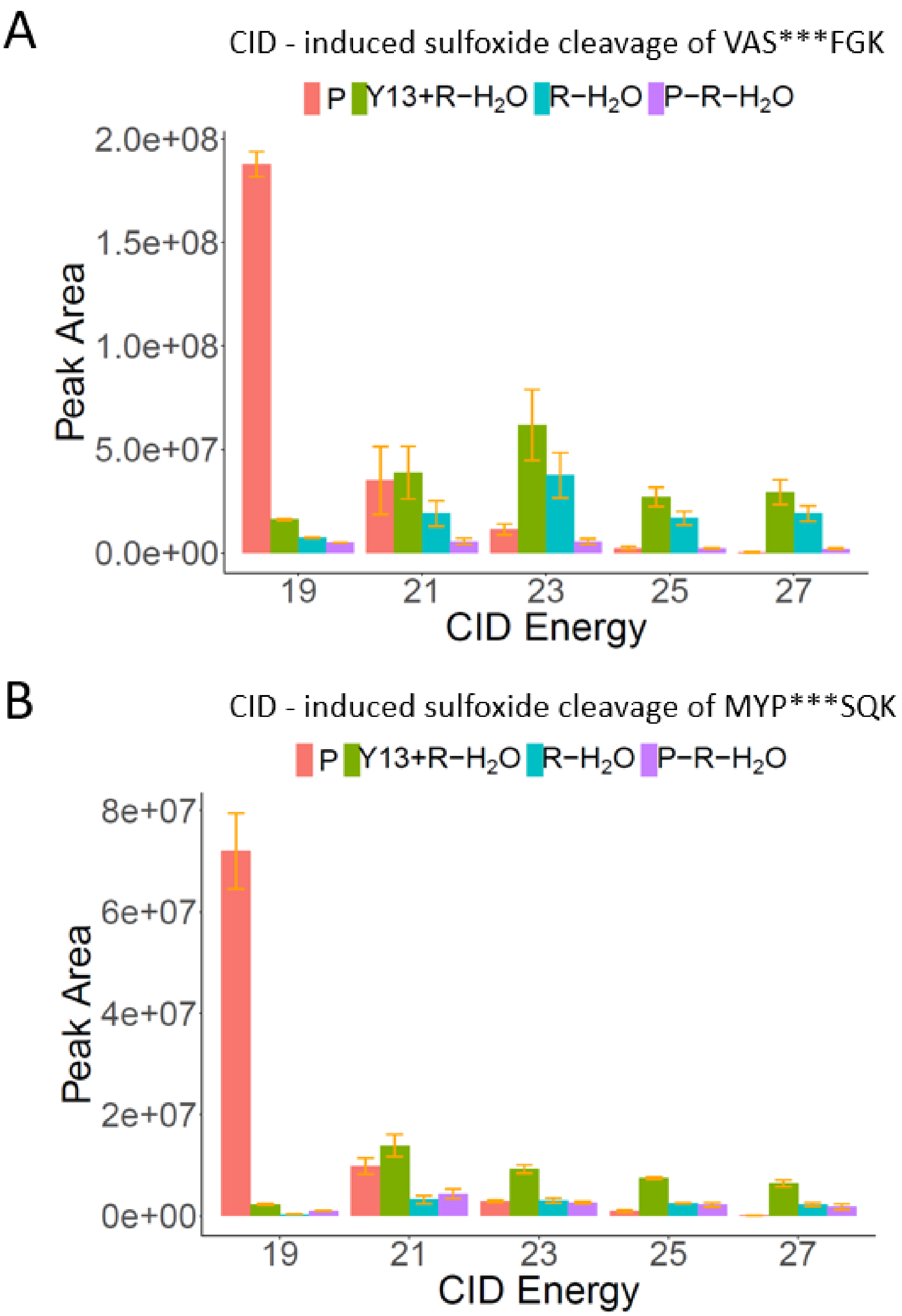


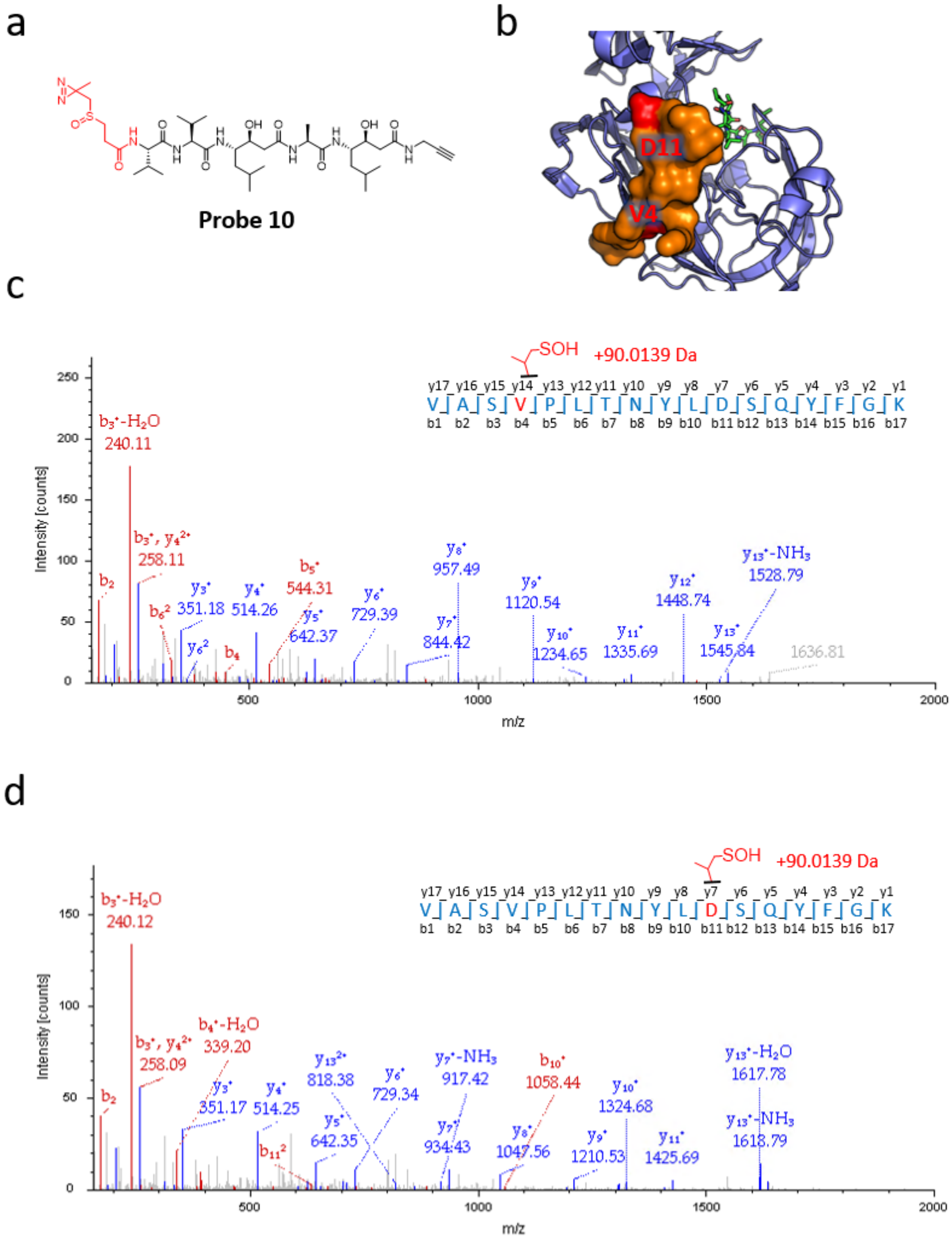
Figure S11. Peak area (mean of 3 replicates), Error bar: standard deviation of peak area of probe-modified peptide VAS***FGK in **Figure S11A** and probe-modified peptide MYP***SQK in **Figure S11B** and their fragments (y13+R-H_2O: PLT*** FGK + Reporter - H_2O (theoretical m/z: 1155.6081, z: 2); R - H_2O: Reporter - H_2O ($C_{35}H_{59}N_6O_7$, theoretical m/z: 675.4445, z: 1); P - R - H_2O: Precursor - Reporter - H_2O (theoretical m/z: 987.4872, z: 2)).

Table S3, Peptides identification after database search against Chymosin*.

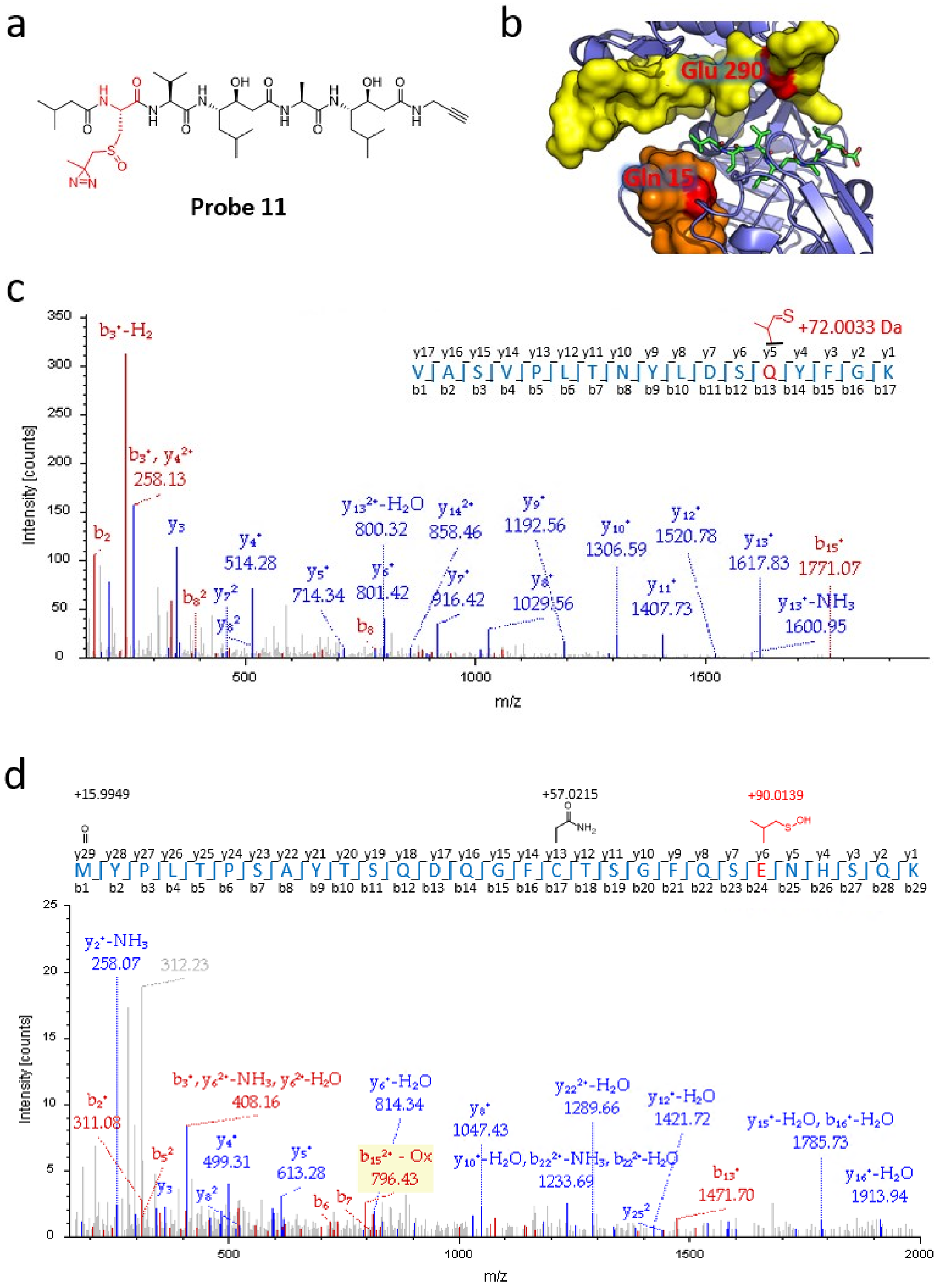
Entry	Gene	Accessions	Annotated Sequence	Modifications	# PSMs**
1	**CYM**	**P00797**	**[D].SQYFGK.[I]**	**1xSY_Sulfox_Clea-H2O [Q/S]**	**89**
2	**CYM**	**P00799**	**[D].SQYFGK.[I]**	**1xSY_Sulfox_Clea [S1]**	**1**
3	**CYM**	**P00798**	**[E].VASVPLTNYLDSQYFGK.[I]**	**1xSY_Sulfox_Clea-H2O [A/D/F/G/K/L/N/P/Q/S/T/V/Y]**	**88**
4	**CYM**	**P00802**	**[E].VASVPLTNYLDSQYFGK.[I]**	**1xSY_Sulfox_Clea [A/D/G/K/L/N/P/Q/S/T/V/Y]**	**20**
5	**CYM**	**P00803**	**[K].MYPLTPSAYTSQDQGFCTSGFQSENHSQK.[W]**	**1xCarbamidomethyl [C17]; 1xOxidation [M1]; 1xSY_Sulfox_Clea-H2O [A/D/G/H/P/Q/S/T]**	**4**
6	**CYM**	**P00804**	**[K].MYPLTPSAYTSQDQGFCTSGFQSENHSQK.[W]**	**1xCarbamidomethyl [C17]; 1xOxidation [M1]; 1xSY_Sulfox_Clea [A/D/E/F/G/K/L/N/P/Q/S/T/Y]**	**12**
7	CYM	P00794	[R].CLVVLLAVFALSQGAE.[I]	3xSY_Sulfox_Clea [L2; F9; G14]; 1xSY_Sulfox_Clea-H2O [C1]	1
8	CYM	P00795	[D].TGSSDFWVPSIYCKSNACK.[N]	3xSY_Sulfox_Clea [K14; N16; K19]	1
9	CYM	P00796	[E].ITRIPLYK.[G]	2xSY_Sulfox_Clea [R3; P5]; 1xSY_Sulfox_Clea-H2O [I1]	1
10	CYM	P00801	[D].RANNLVGLAKAI.[-]	1xSY_Sulfox_Clea [A/G/K/L/N/R/V]; 3xSY_Sulfox_Clea-H2O [A/G/I/K/L/N/R/V]	5
11	CYM	P00800	[R].ANNLVGLAK.[A]	1xSY_Sulfox_Clea [G6]; 1xSY_Sulfox_Clea-H2O [K9]	1
12	CYM	P00805	[K].LVGPSSD.[I]		4
13	CYM	P00806	[D].SQYFGK.[I]		2
14	CYM	P00807	[K].WILGDVFIR.[E]		8
15	CYM	P00808	[K].WILGDVFIRE.[Y]		41

* Cleavage specificity in Proteome Discoverer search was set to fully Glu C/ trypsin (Cleave at the C-terminal of Lys, Arg, Gul and Asp). ** Total PSMs from triple measurements of 5 CID conditions.

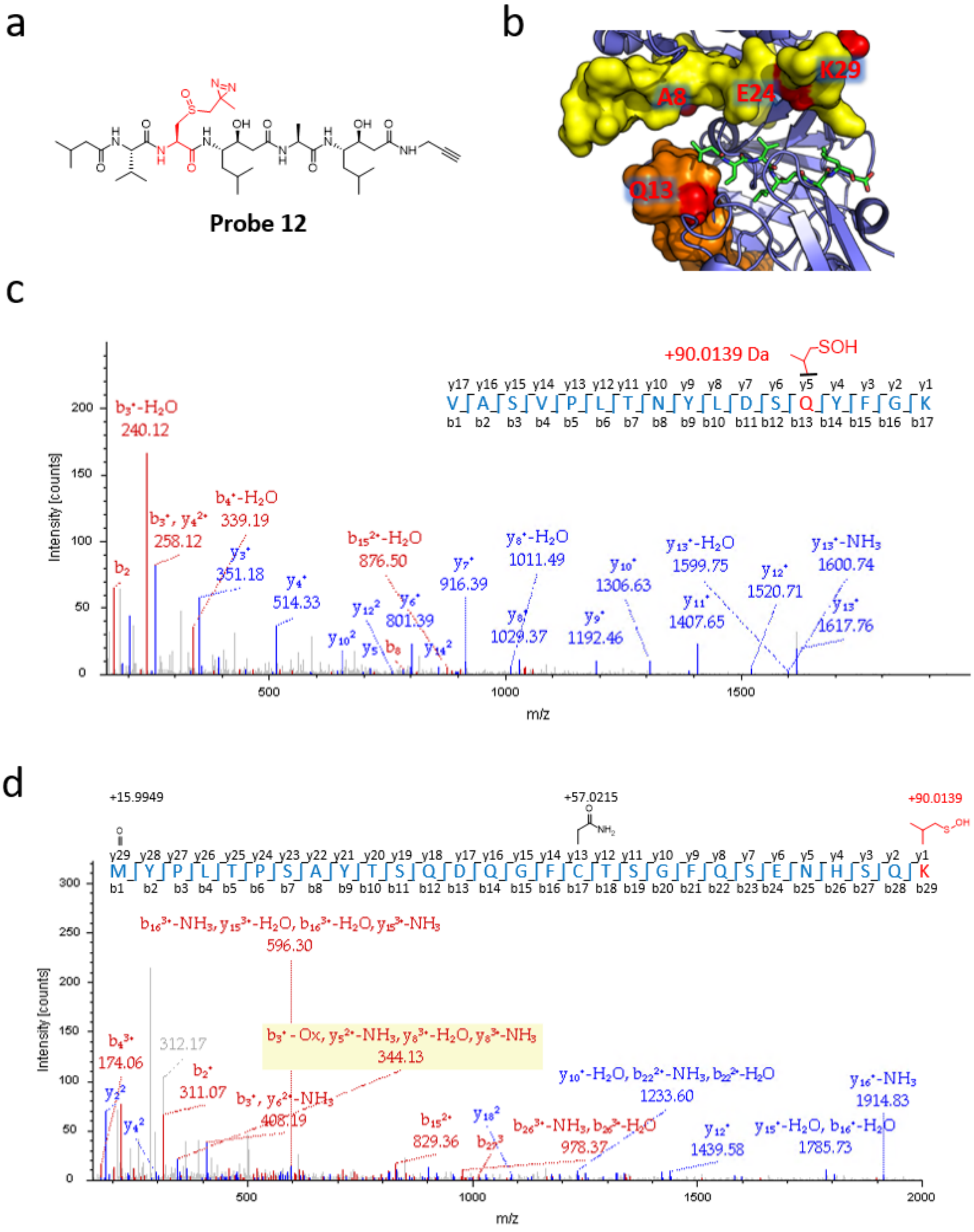
Spectra of a peptide modified by probe 10



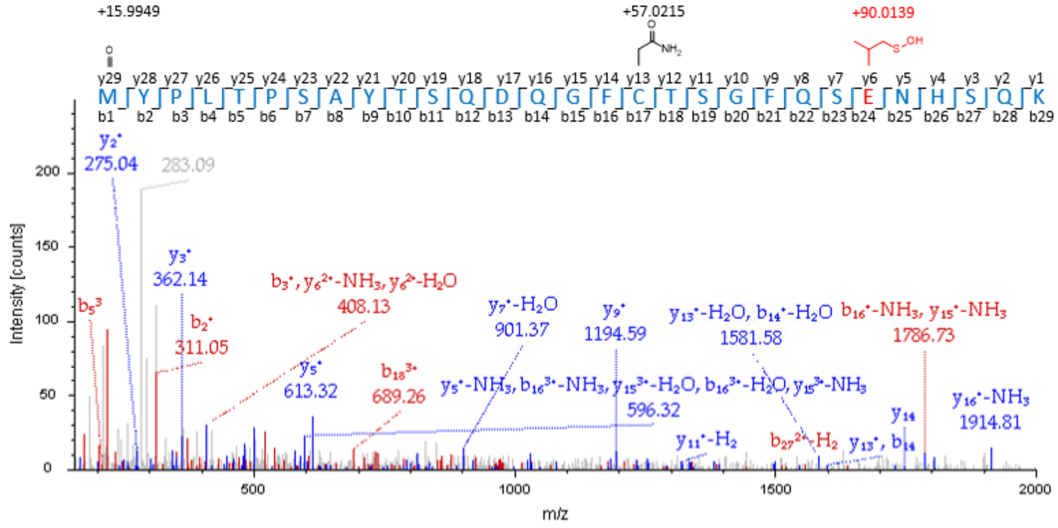
Spectra of a peptide modified by probe 11



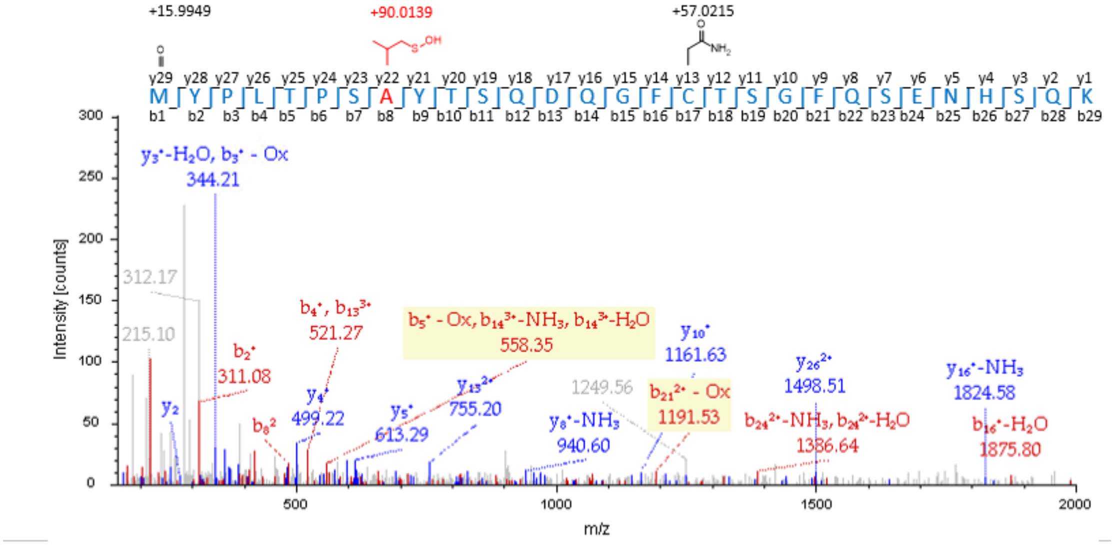
Spectra of a peptide modified by probe 12



e

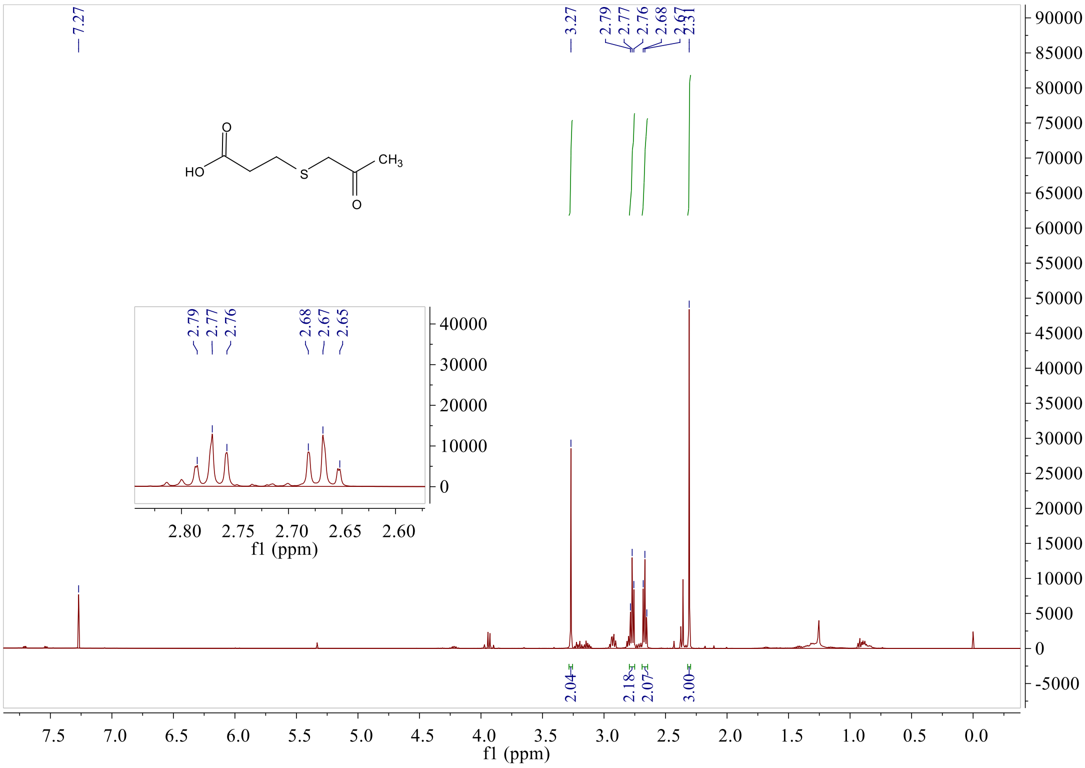

+15.9949
+57.0215
+90.0139
M Y P L T P S A Y T S Q D Q G F C T S G F Q S E N H S Q K
y2+ 275.04
283.09
y3+ 362.14
b5 3
b2+ 311.05
b3+, y6 2+-NH3, y6 2+-H2O 408.13
y5+ 613.32
b18 3+ 689.26
y7+-H2O 901.37
y9+ 1194.59
y13+-H2O, b14+-H2O 1581.58
b16+-NH3, y15+-NH3 1786.73
y5+-NH3, b16 3+-NH3, y15 3+-H2O, b16 3+-H2O, y15 3+-NH3 596.32
y11+-H2
b27 2+-H2
y13+, b14
y14
y16+-NH3 1914.81
Intensity [counts]
m/z

f

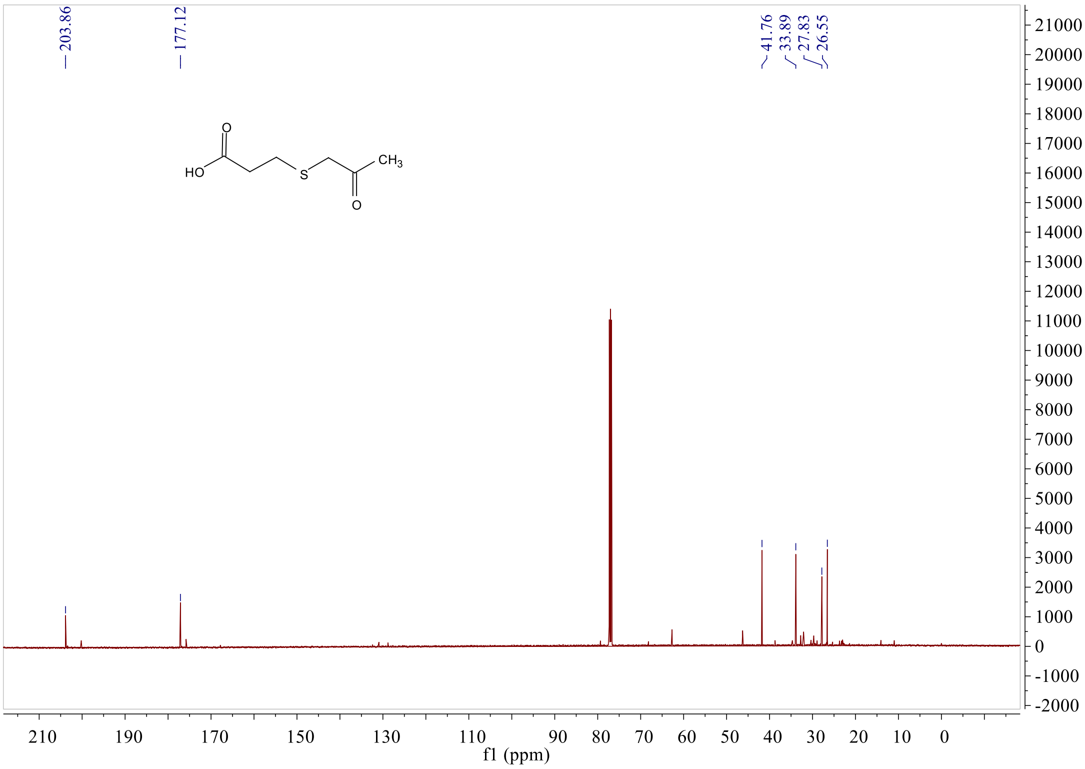

+15.9949
+90.0139
+57.0215
M Y P L T P S A Y T S Q D Q G F C T S G F Q S E N H S Q K
y3+-H2O, b3+ - Ox 344.21
312.17
215.10
b2+ 311.08
y2
b4+, b13 3+ 521.27
y4+ 499.22
b6 2
b5+ - Ox, b14 3+-NH3, b14 3+-H2O 558.35
y5+ 613.29
y13 2+ 755.20
y8+-NH3 940.60
1249.56
y10+ 1161.63
b21 2+ - Ox 1191.53
y26 2+ 1498.51
b24 2+-NH3, b24 2+-H2O 1386.64
y16+-NH3 1824.58
b16+-H2O 1875.80
Intensity [counts]
m/z

Copies of NMR spectra

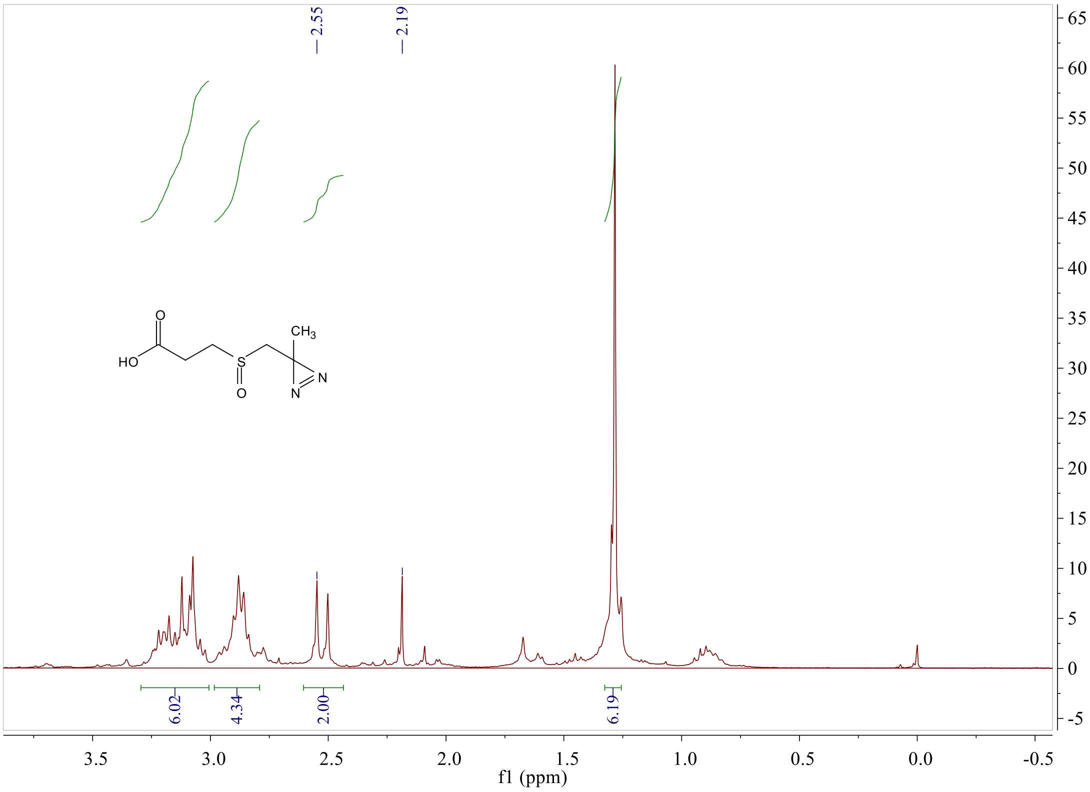
^{1}H NMR spectrum of compound **ii**



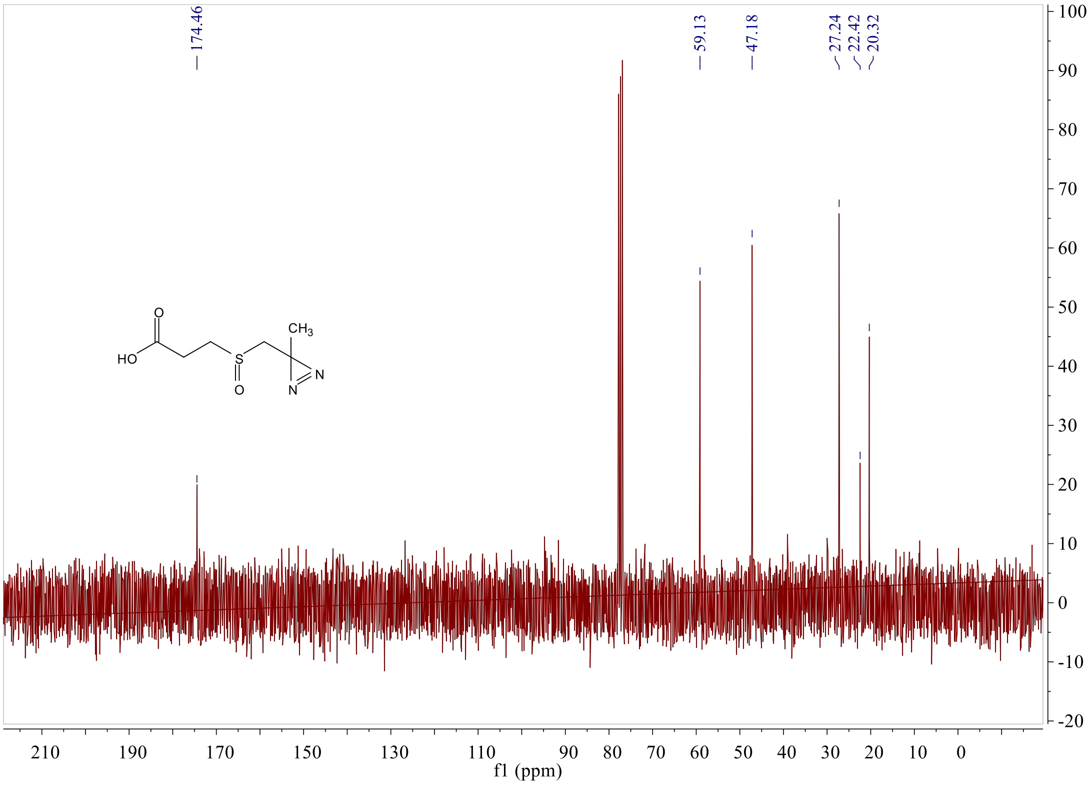
^{13}C NMR spectrum of compound **ii**



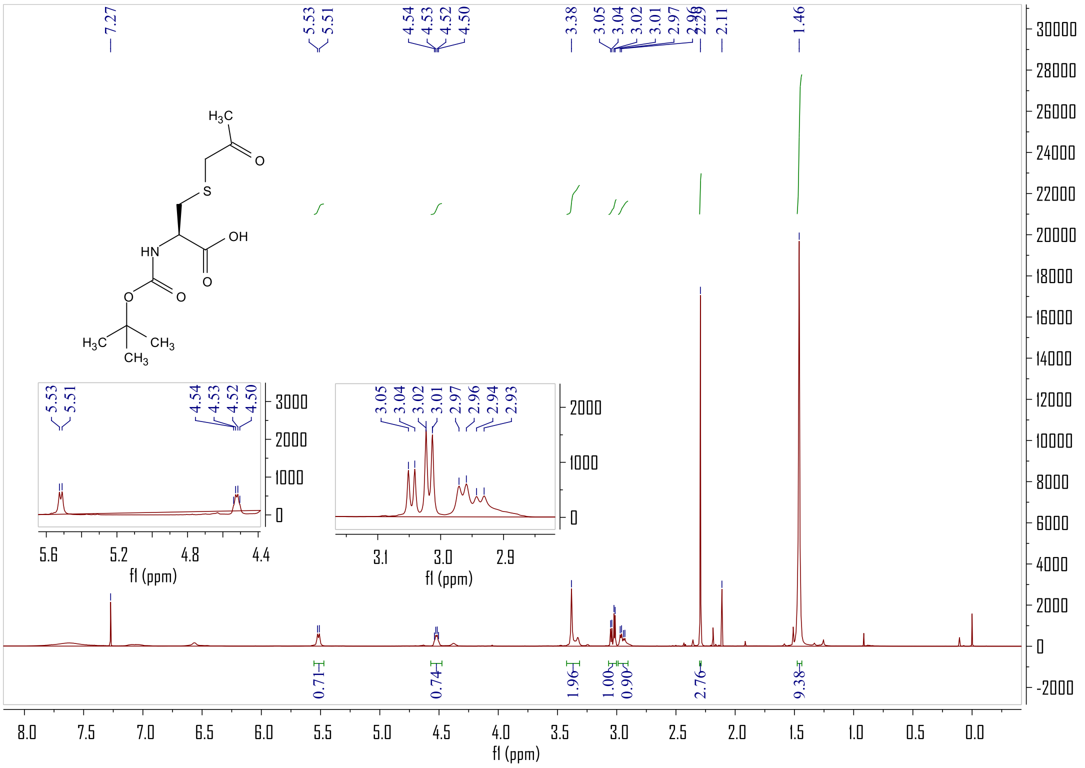
^{1}H NMR spectrum of compound **iv**



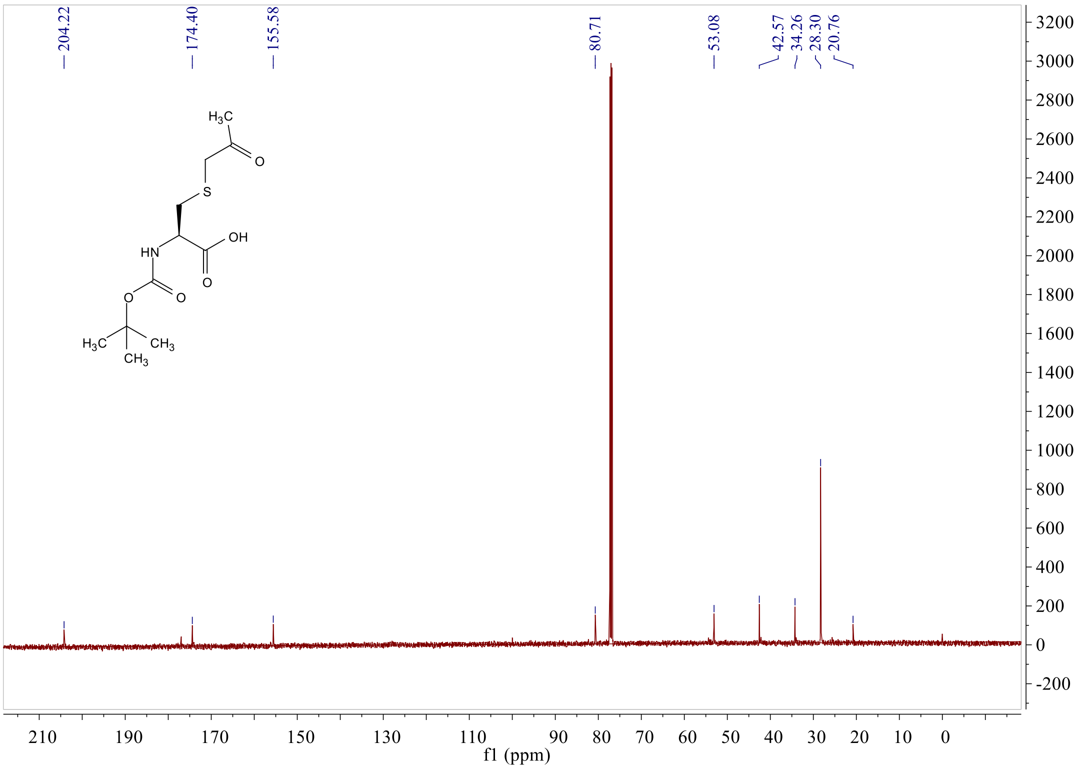
^{13}C NMR spectrum of compound **iv**



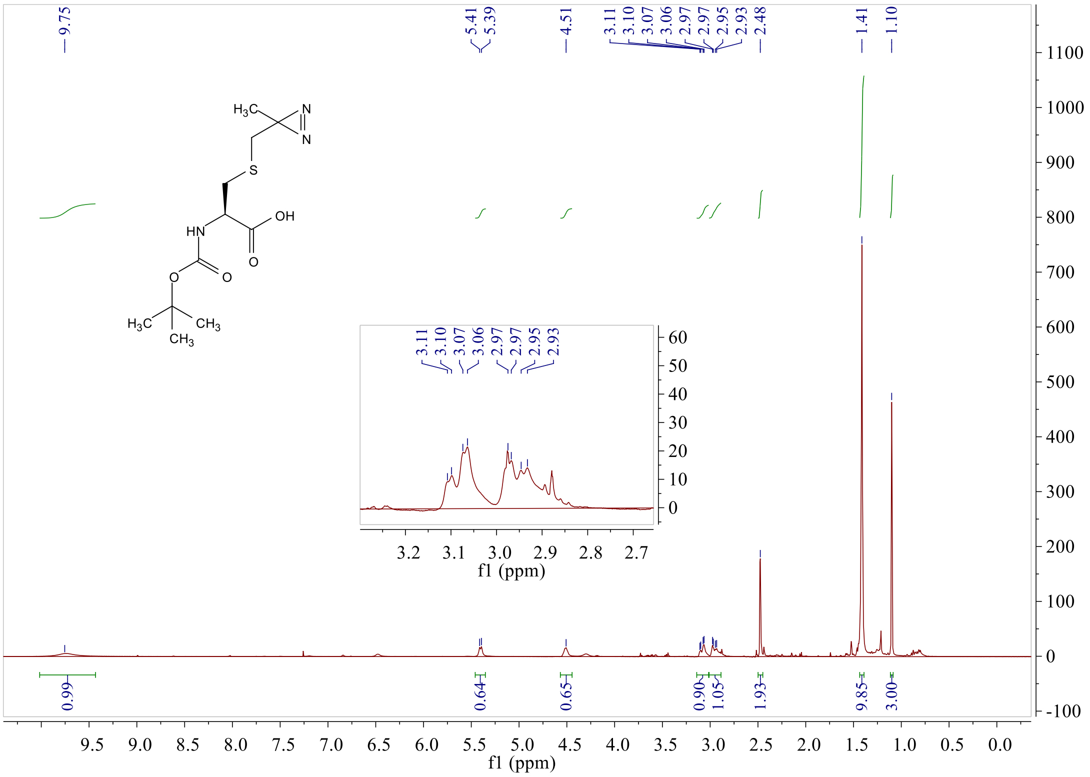
^{1}H NMR spectrum of compound **vi**



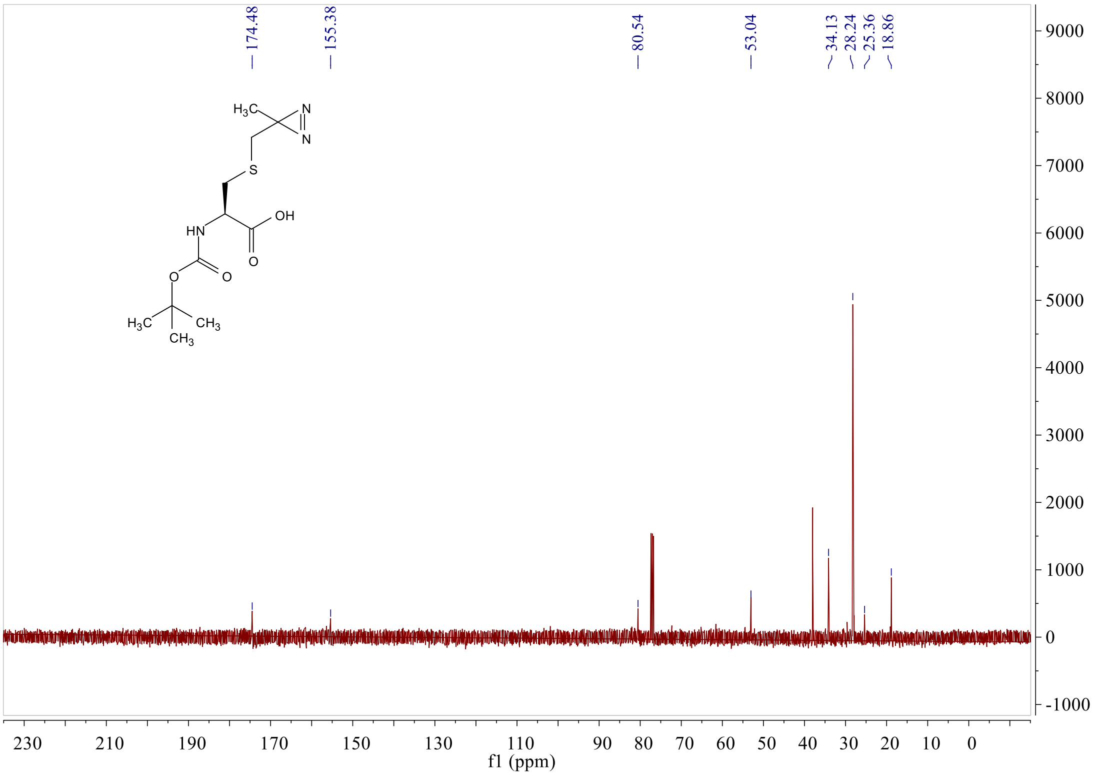
^{13}C NMR spectrum of compound **vi**



^{1}H NMR spectrum of compound **vii**

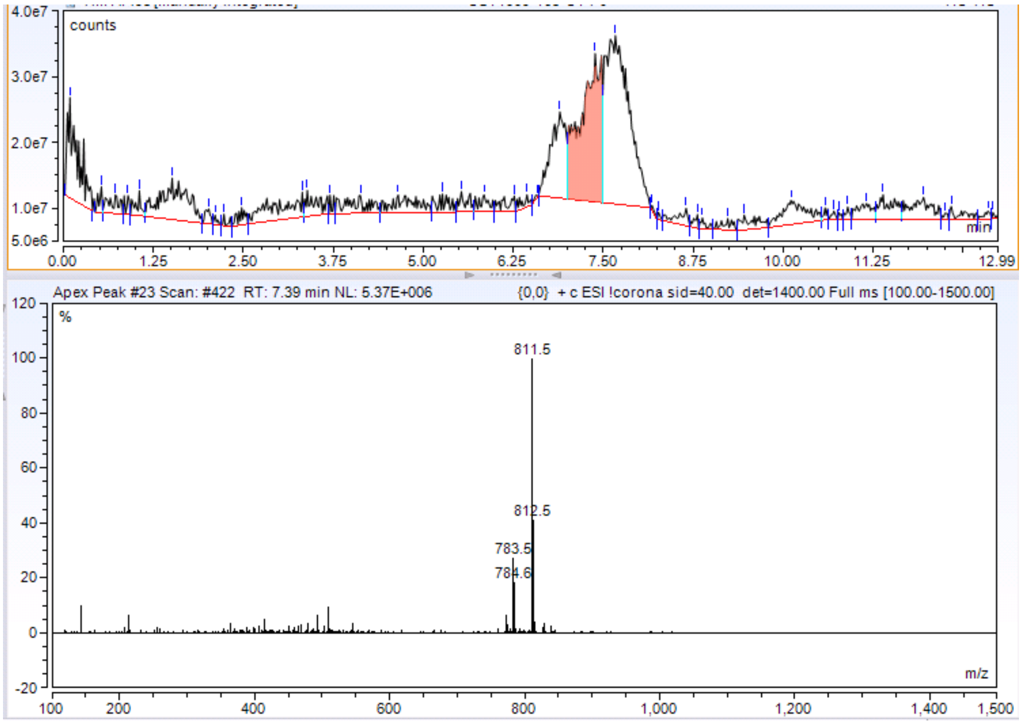


^{13}C NMR spectrum of compound **vii**

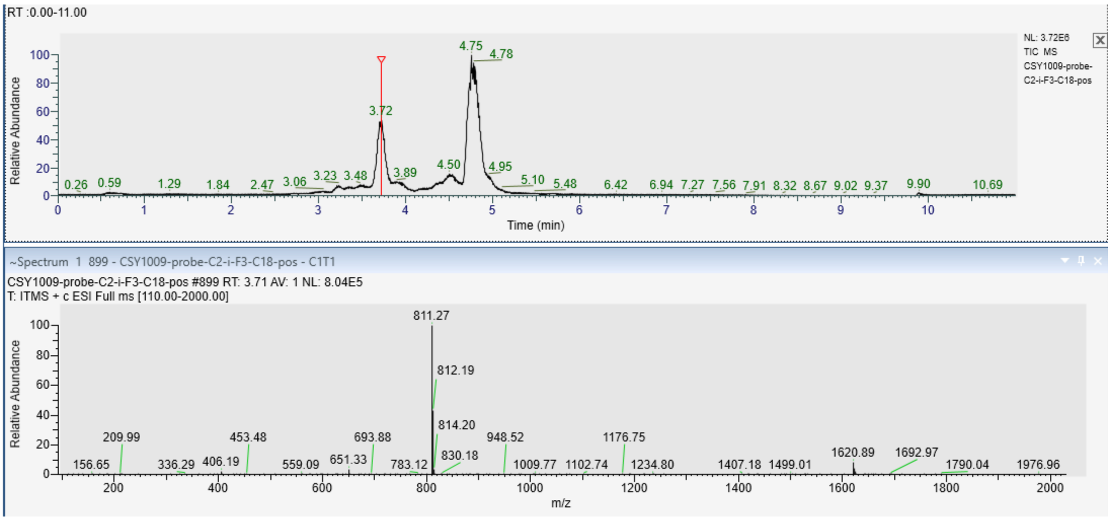


LC-MS spectra of probe 10-13

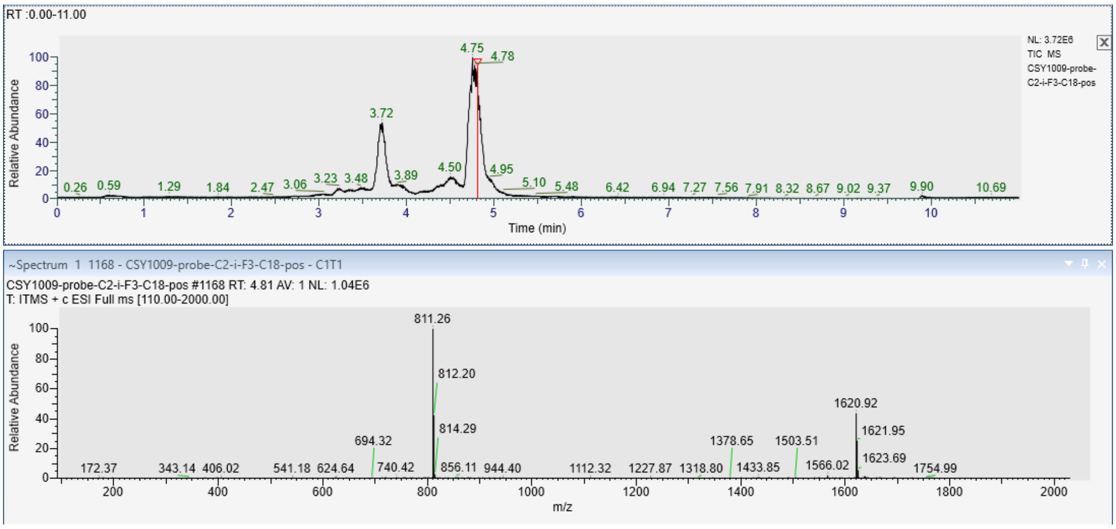
Total ion current chromatogram and mass spectrum of compound **10**



Total ion current chromatogram and mass spectrum of compound **11**

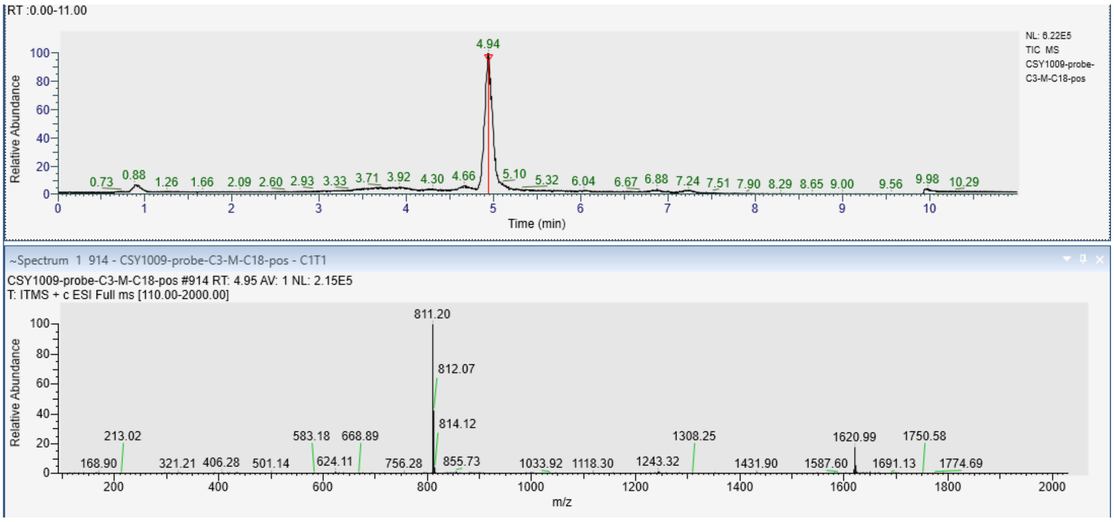


CSY1009-probe-C2-i-F3-C18-pos #899 RT: 3.71 AV: 1 NL: 8.04E5
T: ITMS + c ESI Full ms [110.00-2000.00]

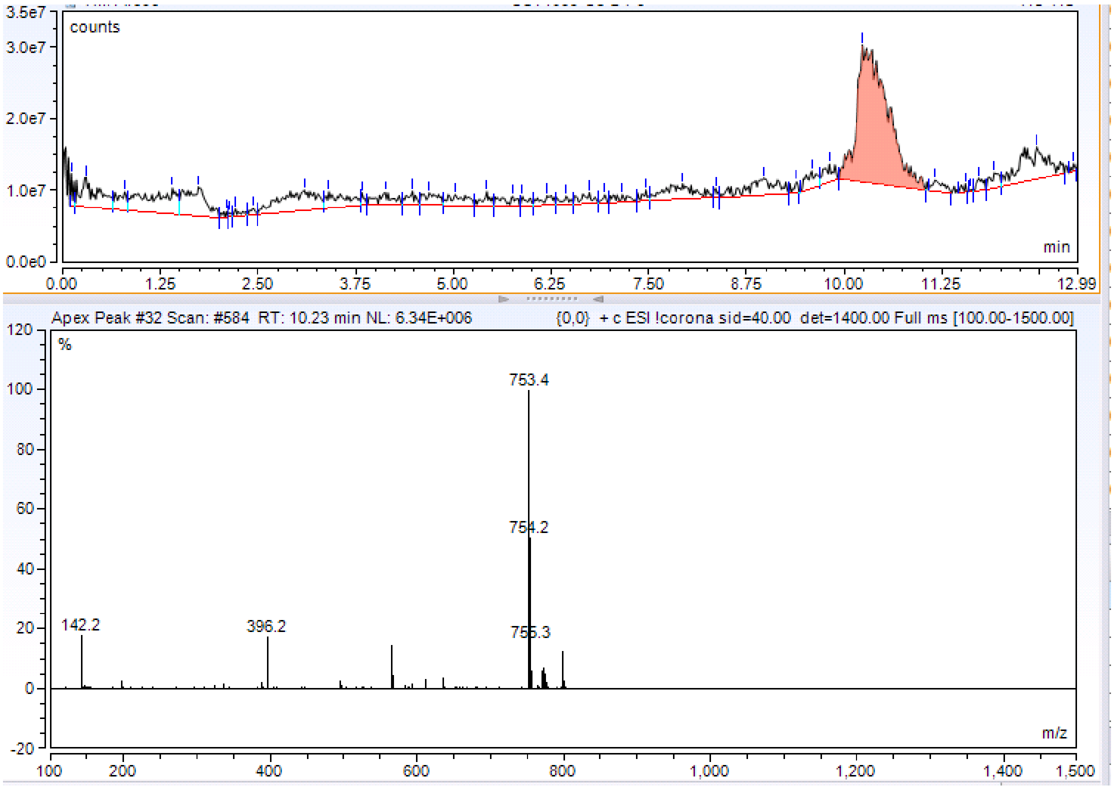


CSY1009-probe-C2-i-F3-C18-pos #1168 RT: 4.81 AV: 1 NL: 1.04E6
T: ITMS + c ESI Full ms [110.00-2000.00]

Total ion current chromatogram and mass spectrum of compound **12**



Total ion current chromatogram and mass spectrum of compound **13**



Apex Peak #32 Scan: #584 RT: 10.23 min NL: 6.34E+006

{0,0} + c ESI !corona sid=40.00 det=1400.00 Full ms [100.00-1500.00]

9 Acknowledgements

I would like to acknowledge funding by the Chinese Scholarship Council (No. 201704910881), the Ministerium für Kultur und Wissenschaft des Landes Nordrhein-Westfalen, the Regierende Bürgermeister von Berlin–inkl. Wissenschaft und Forschung, and the Bundesministerium für Bildung und Forschung. Thank Leibniz-Institut für Analytische Wissenschaften – ISAS – e.V. (ISAS) for usage of their modern facility including state of art mass spectrometry and great interdisciplinary working atmosphere, Max-Planck-Institut für molekulare Physiologie (MPI Dortmund) for usage of their Typhoon scanner and Preparative HPLC, Fakultät für Chemie und Chemische Biologie.an der Technische Universität Dortmund for usage of their NMR.

I would like to extend my appreciation to the individuals who provided invaluable assistance and support, without whom this dissertation would not have been possible:

Sincerely gratitude to my mentor, advisor and supervisor Prof. Dr. Steven H. L. Verhelst, who spares no effort in developing chemical tools to answer biological questions with his wise insight and ability to carry out critical investigations. Thanks for his motivating, inspiring ideas, critical discussions and helpful guidance.

I would like to thank my Doktorvater and thesis supervisor Prof. Dr. Albert Sickmann, who gave me scientific access to bioanalytics field. I will always remember his inspiring ideas and critical discussions on my PhD projects and thesis.

Special thanks to former ISAS colleague Dr. Stefan Loroch, Ms. Luzia Seifert and Dr. Chi Nguyen who were acting as a door opener of the mass spectrometry (MS) world. I'm very appreciate for their unconditional support Thanks Dr. Loroch for his wise insight into mass spectrometry method development and critical discussions on the data analysis. Thank Ms. Seifert for her professional guidance on the hardware maintenance of nano-LC and MS. Thanks Dr. Nguyen for her inspiring discussions on targeted proteomics. Also thanks Ms. Svenja Idel for her professional practical guidance on cell culture. Thank Dr. Laxmikanth Kollipara, Msc. Tingting Li and Dr. Jingnan Huang for their professional practical guidance on proteomics sample preparation. Thank Dr. Bernhard Blank-Landeshammer and Dr. Roman Sakson for their professional practical guidance on data analysis with Skyline.

I would like to thank Dr. Chunguang Liang and Ms. Weimeng Yu from Universität Würzburg for their efforts on the deep learning prediction. Thank MSc Hongli Li, Prof. Dr. Marc Fransen and Dr. Michaela Prothiwa from KU Leuven for their efforts on cellular experiment involving SQSTM1 degradation. Also, thank Dr. Dominik Kopczynski and Dr.

Stefan Loroch from ISAS for their efforts on the statistic in proteomics data analysis. Thank ISAS colleague Dr. Olga Shevchuk for critical discussions on the statistic in proteomics data analysis.

Thank Thibaut Vignane, Dunja Petrovic and Dr. Milos Filipovic from the Sulfaging group at ISAS for the anti-mouse-688 antibody and instrument support involving western blot, Dr. Hefei Wang (MPI Dortmund) for the anti-his antibody.

Many thanks to Dipl.-Ing. Ingo Feildmann, Ms. Luzia Seifert, Ms. Susanne Krois, Ms. Rita Fobbe, Dipl.-Ing. Rolf Bandur, Ms. Cornelia Schumbrutzki and Dipl. Chem. Stefanie Wortelkamp from ISAS for the large amount of work involving instrument maintenance and taking care of the safety in the chemical and instrumental laboratory. And Dr. Raffael Kettler and Ms. Svenja Idel are acknowledged for their work on taking care of the safety and maintaining of the cell culture room at ISAS-Campus. Many thanks to Dr. Ewelina Dutkiewicz, Dr. Fiorella.Solari and MSc Susmita Ghosh from ISAS for instrumental support on timsTOF Pro. Also, thank MSc Jacqueline Katzer from ISAS for instrumental support on Bravo Liquid Handling System. Thank Dr. Jianing Xie from MPI Dortmund for the instrumental support on Preparative HPLC.

Also, I would like to thank my lab-mates and good friends Dr. Daniel Krahn, Dr. Tim Van Kersavond, MSc. Raphael Konopatzki and Dr. Minh Nguyen at ISAS, as well as Dr. Dimitris Korovesis, Dr. Hester Beard, Dr. Jian Yang, Dr. Roeland.vanhoutte, Dr. Jan Pascal Kahler, Dr. Snezhana Akpunarlieva, Dr. Merel van de Plassche, Dr. Suravi Chakrabarty, Dr. Marta Barniol-xicota, MSc. Shanping Ji, Msc. Jonathan Coene, Dr. Michaela Prothiwa and Dr. Joachim Demaerel from KU Leuven for fruitful discussions. Many thanks to colleagues, inparticular from the proteomics group and Technical Service Bioanalytics group for their support and for creating a nice working environment.

Thanks for the proofreading of this thesis from Prof. Dr. Steven H. L. Verhelst, Prof. Dr. Albert Sickmann, Dr. Hester Beard, Dr. Daniel Krahn, Dr. Tobias Grabe and Dr. Stefan Loroch.

Finally, I would like to thank my family for their spiritual support. Thank my parents and my father’s younger sister for their endorsement of oversea education so that I can pursue further achievement in such interdisciplinary project.