**Limited Proteolysis-Mass Spectrometry to Identify Metabolite-Protein Interactions**

Aleš Holfeld,1, 4 Jan-Philipp Quast,1, 4 Roland Bruderer,2 Lukas Reiter,2 Natalie de Souza,1, 3 and Paola Picotti1, \*

1Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, Zurich, Switzerland

2Biognosys AG, Schlieren, Switzerland  
3Department of Quantitative Biomedicine, University of Zurich, Zurich, Switzerland

4These authors contributed equally: Aleš Holfeld, Jan-Philipp Quast

\*Correspondence: [picotti@imsb.biol.ethz.ch](mailto:picotti@imsb.biol.ethz.ch)

**Abstract**

Metabolite-protein interactions regulate diverse cellular processes, prompting the development of methods to investigate the metabolite-protein interactome at a global scale. One such method is our previously developed structural proteomics approach, limited proteolysis-mass spectrometry (LiP-MS), which detects proteome-wide metabolite-protein and drug-protein interactions in native bacterial, yeast and mammalian systems, and allows identification of binding sites without chemical modification. Here we describe a detailed experimental and analytical workflow for conducting a LiP-MS experiment to detect small molecule-protein interactions, either in a single-dose (LiP-SMap) or a multiple-dose (LiP-Quant) format. LiP-Quant analysis combines the peptide-level resolution of LiP-MS with a machine-learning based framework to prioritize true protein targets of a small molecule of interest. We provide an updated R script for LiP-Quant analysis via a GitHub repository accessible at <https://github.com/RolandBruderer/MiMB-LiP-Quant>.

**Keywords**: Metabolite-protein interactions, Limited proteolysis, Mass spectrometry, LiP-SMap, LiP-Quant, Proteomics, Structural proteomics, Machine learning

# Introduction

Metabolites play multiple essential roles in living systems. In particular, interactions of metabolites with proteins mediate numerous cellular events, including signaling, enzyme activity, and assembly of protein complexes [1]. However, the complexity and vast number of cellular metabolites, as well as the transient and low-affinity nature of their interactions with proteins, make systematic mapping of metabolite-protein interactions a substantial challenge [2, 3].

Several technological advances have enabled identification of metabolite- and small molecule-binding proteins [4–10]. In our own work, we have adapted our structural proteomics approach termed limited proteolysis-mass spectrometry (LiP-MS) to systematically describe metabolite-protein interactions in multiple systems[9]. In this method (Figure 1), a metabolite is incubated at either a single concentration or multiple concentrations with a native-like cellular lysate and then briefly exposed to a sequence-unspecific protease. Since metabolite binding to a protein can result in altered protease accessibility of the protein, at the binding site itself and/or at other regions that undergo binding-dependent conformational changes, metabolite binding-specific fragments of target proteins are generated upon limited proteolysis. Protein fragments are further digested with trypsin to peptides compatible with quantitative MS analysis, allowing the identification of differentially abundant peptides in metabolite-treated relative to untreated samples.

We have applied this principle to profile metabolite-protein binding in two different ways. Using the LiP-SMap approach, in which a native-like lysate is incubated with a single metabolite concentration, we have systematically mapped 1,678 metabolite-protein interactions and 7,345 putative binding sites in *E. coli* [9]. We recently extended this principle in the LiP-Quant approach [11], in which a native-like cellular lysate is incubated with a concentration range of a small molecule. This enables prioritization of true protein targets based on a composite score (termed the LiP-Quant score) that favors metabolite-dependent LiP peptides where the intensity changes show a sigmoidal dose-response behavior with changing metabolite concentration. The LiP-Quant score also incorporates other machine learning-derived measures, and serves overall to reduce false positives and facilitate identification of true protein targets in complex proteomes. We have shown that LiP-Quant identifies known targets of both specific and promiscuous small molecule binders in complex eukaryotic (i.e., human and yeast) proteomes, distinguishes relative binding affinities of a small molecule to different proteins, including homologous proteins, present in the lysate, and have used the approach to identify a previously unknown target of a fungicide [11]. While the LiP-Quant pipeline more robustly identifies true positive interactions, it also requires more samples and MS runs; for exploratory analyses in particular, the LiP-SMap approach may be sufficient. Users should select which of LiP-SMap or LiP-Quant are more appropriate based on the goal of their experiment and sample/instrument availability. Both LiP-MS approaches provide a list of putative small molecule binding proteins; identified targets must be further validated using orthogonal methods.

In comparison to other proteomic approaches, the LiP-MS based methods combine the ability to identify metabolite-protein interactions in complex lysates, to identify metabolite binding sites without chemical modification, and to determine the affinity of a metabolite for its target protein(s). Nevertheless, identifying metabolite interactions with proteins of low abundance has remained challenging, especially in complex proteomes, such as human. Fractionation techniques can be employed to increase the coverage of a given proteome, including of membrane proteins [11, 12].

In this chapter, we describe a step-by-step workflow to identify metabolite-protein interactions in a complex cellular environment based on the LiP-MS principle. We discuss all steps in the LiP-MS pipeline including experimental design, analytical procedures and data analysis for single- and multiple-dosage LiP experiments. We focus in particular on the LiP-Quant approach since it more robustly identifies true positive interactions, and we provide a revised R script for these multiple dose LiP-Quant assays.

# Materials

Prepare all solutions using ultrapure water (18 MΩ-cm at 25 °C). Prepare and store all reagents at room temperature (unless stated otherwise).

## Cell lysis under native conditions

1. HEPES BioPerformance, certified 99.5% (Sigma-Aldrich, cat. no. H4034).
2. Potassium chloride (Merck, cat. no. 104.936.1000).
3. Magnesium chloride hexahydrate, puriss. p.a. (Fluka, cat. no. 63072).
4. LiP buffer: 100 mM HEPES pH 7.4, 150 mM KCl, 1 mM MgCl2. The buffer should be freshly prepared. The pH can be adjusted with 1 M KOH or 37% HCl.
5. Acid-washed glass beads 425-600 μm (Sigma Aldrich, cat. no. G8772).
6. FastPrep-24™ 5G bead beating grinder (MP Biomedicals).
7. Kimble® pellet pestle cordless motor (Sigma Aldrich, cat. no. Z359971).
8. Kimble® pellet pestle CTFE/stainless steel (Sigma Aldrich, cat. no. Z359963).
9. Zeba™ Spin Desalting Columns, 7K MWCO, 0.5 mL (Sigma Aldrich).
10. Pierce BCA Protein Assay Kit (Thermo Scientific).

## Preparation of stock solutions of metabolites

1. 100 mM HEPES, pH 7.4 (Sigma Aldrich, cat. no. H3375).

## Limited proteolysis under native conditions

1. Thin-walled PCR tubes (Thermo Scientific, cat. no. AB-1182).
2. Proteinase K (PK) from *Tritirachium album* (Sigma Aldrich, cat. no. P2308). Prepare a stock by dissolving the lyophilized enzyme in ultrapure water to reach a final concentration of 1.0 µg/µL. Freeze in liquid nitrogen and store at -20 ºC prior to further use (see ***Note 1***).
3. 10 % (wt/vol) sodium deoxycholate (DOC) (Sigma Aldrich, cat. no. D6750). The solution can be stored at room temperature (20-24 °C) for at least 1 month.
4. Multichannel pipette (Eppendorf, cat. No. 3125000010 and 3125000052).
5. Biometra TRIO thermal cycler (Analytik Jena GmbH).

## Sample preparation prior to MS analysis

1. Reduction buffer: 300 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl). (ThermoFisher Scientific, cat. no. 20490). Dissolve TCEP-HCl in 1 M HEPES pH 7.4 (see ***Note 2***).
2. Alkylation buffer: 1 M iodoacetamide (IAA) (Sigma Aldrich, cat. no. I1149).
3. 100 mM ammonium bicarbonate (Fluka Analytical, cat. no. 40867).
4. 0.5 µg/µL sequencing-grade porcine trypsin (Promega, cat. no. V5113).
5. Endoproteinase Lys-C from *Lysobacter enzymogenes* (FUJIFILM Wako Pure Chemical Corporation, cat. no. 129-02541). Dissolve the lyophilized enzyme in ultrapure water to reach a final concentration of 1.0 µg/µL. Freeze in liquid nitrogen and store at -20 ºC prior to further use.
6. 50 % (vol/vol) formic acid (Fluka Analytica, cat. no. 94318).
7. 96 well filter plate, 0.2 μm PVDF filter (Corning, cat. no. CLS3508).
8. 96-well MACROSpin Plate (The Nest Group).
9. SpeedVac (Eppendorf, cat. no. 5305000304).
10. 0.1 % (vol/vol) formic acid.
11. Sonicator (Sono Swiss, SW 12 H).
12. iRT Kit (Biognosys, Switzerland).

## LC-MS/MS data acquisition

1. Buffer A: HPLC-grade water with 0.1 % (vol/vol) formic acid.
2. Buffer B: 95 % (vol/vol) HPLC-grade acetonitrile in water with 0.1 % (vol/vol) formic acid.
3. A high-resolution mass spectrometer (e.g., Thermo Scientific Orbitrap Eclipse) equipped with a nano-electrospray ion source coupled to a liquid chromatography system (e.g., Thermo Scientific EASY-nLC 1200) operating in the nanoliter/min flow rate range.
4. A 40 cm x 0.75 mm i.d. chromatographic analytical column (New Objective, PF360-75-10-N-5) for nano-LC separation, packed with 1.9 μm Reprosil-Pur Basic C18 beads (Dr. Maisch).

## Software

To perform the label-free DDA- and DIA-MS analysis presented here, download and install a licensed copy of the Spectronaut software (Biognosys, Switzerland) [13]. A user manual can be found on <http://files.biognosys.ch/058_Spectronaut/ReleaseMaterial/00_Manual/Spectronaut15_UserManual.pdf>. For the LiP-Quant analysis, download and install R (version 4.1.1) and the RStudio software (version 1.4.1717). Obtain a copy of the LiP-Quant script and files from the GitHub repository at <https://github.com/RolandBruderer/MiMB-LiP-Quant>.

# Methods

Here we describe both experimental procedures and data analysis for identifying metabolite-protein interactions with LiP-MS. Wear gloves at all times during sample handling to avoid sample contamination.

## Cell lysis under native conditions

Depending on the sample type under investigation (e.g., bacteria, yeast, mammalian cells, human tissue, or plasma) different extraction procedures are used [14]. Always use fresh extracts and, if not otherwise indicated, keep them on ice throughout sample preparation.

1. Resuspend cell pellets in approximately 4 pellet volumes of ice-cold LiP buffer (see ***Note 3***).
2. Add the same volume of 425-600 μm acid-washed beads.
3. For *E. coli* and *S. cerevisiae*, lyse cells at 4 ºC using a FastPrep-24™ 5G bead-beating grinder or similar instrument in 8 cycles, each consisting of 30 seconds of bead-beating and 200 seconds pause at a speed of 5.5.

For mammalian cells, lyse cells at 4 ºC using a FastPrep-24™ 5G bead-beating grinder or similar instrument by 40 seconds of bead-beating at a speed of 5.5. Alternatively, a pellet pestle can be used to disrupt mammalian cells in LiP buffer in 10 cycles of 10 seconds homogenization and 1-minute pause at 4 ºC. In this case, steps 2 and 4 are skipped.

1. Transfer the sample into a fresh 1.5 ml Eppendorf tube as follows: pierce the bottom of the sample tube with a hot needle, stack the sample tube onto the fresh tube, centrifuge at 1,000 g for 1 minute at 4 ºC.
2. Remove cell debris by centrifugation at 1,000 g for 5 minutes at 4 ºC and transfer the supernatant to a fresh 1.5 mL Eppendorf tube.
3. Remove endogenous metabolites and nucleic acids from the lysate by size-exclusion chromatography using Zeba Spin desalting columns according to the manufacturer's instructions (see ***Note 4***). Always use fresh LiP-buffer for buffer exchange.
4. Determine protein concentration of the lysate with the bicinchoninic acid (BCA) assay following the manufacturer's instructions. Adjust the protein concentration of the cleared cell lysate to 2 µg/µL using LiP buffer.

## Preparation of stock solutions of metabolites

1. Solubilize ultra-pure metabolite powder in 100 mM HEPES, pH 7.4.
2. Measure pH with a micro-pH electrode and double-check with pH strips. If necessary, pH can be adjusted to 7.4 with 1 M KOH or 37% (vol/vol) HCl.
3. Store stock solutions at -20 ºC prior to further use.

## Limited proteolysis

This step is applied to the cleared native cell lysates. We recommend using 100 µg of total protein per sample, and four replicates per condition. The metabolite of interest can be added at a single concentration or at multiple concentrations. Exact timing of the following steps is crucial in order to obtain reproducible results.

1. Pre-heat/cool the two blocks of the thermocycler to 25 °C and 99 °C. The thermocycler block set to 25 °C should have the lid open. The thermocycler block set to 99 °C should have the lid pre-heated to the same temperature and be firmly closed after adding the samples to avoid evaporation.
2. Transfer 50 μL of cleared lysate (2 µg/µL) into 0.2 mL thin-walled tubes and keep at 4 °C until used.
3. Prepare a 0.2 µg/µL PK solution by diluting the frozen PK stock (1 µg/µL) in ultrapure water (see ***Note 5***). Transfer into 0.2 mL thin-walled tubes and keep at 4 °C until used.
4. Incubate the cleared lysate with a metabolite of interest at a single concentration or at different concentrations (see ***Note 6***) for 5 minutes at 25 ºC. Make sure to add the metabolite simultaneously into each tube using a multichannel pipet.
5. Add 5 µL PK solution (0.2 µg/µL) to the samples to reach an enzyme-to-protein ratio of 1:100 (wt/wt). Make sure to add the PK solution simultaneously into each tube using a multichannel pipet. Mix well by pipetting up and down, vortex and quickly spin down to collect the sample content at the bottom of the tube. Incubate at 25 °C for 5 minutes. Exact timing of this step is crucial (see ***Note 7***).
6. Transfer the sample tubes rapidly to the heat-block set to 99 °C and incubate for 5 minutes to inactivate PK. This step is critical (see ***Note 8***).
7. Cool down the samples by setting the temperature of the block to 4 °C and incubate for 5 minutes (see ***Note 9***).
8. Transfer the samples to an equal volume of 10% (wt/vol) DOC to reach a final concentration of 5% DOC. This facilitates protein denaturation. At this point, samples can be stored at -20 °C.

## Sample preparation prior to MS analysis

1. Reduce the cysteine residues by adding TCEP-HCl to a final concentration of 5 mM using a stock solution of 300 mM TCEP-HCl. Incubate for 45 minutes at 37 °C with agitation at 800 rpm. The liquid solution might turn into a gel (see ***Note 2***).
2. Alkylate the reduced cysteines residues by adding IAA to 40 mM. Vortex and spin down to collect the sample content at the bottom of the tube. Incubate for 30 minutes at room temperature in darkness (see ***Note 10***).
3. Dilute the sample to 1% (wt/vol) DOC by adding 4 reaction volumes of 100 mM Ambic. Ensure that the pH is ~8 using a pH indicator paper. If the pH is too low, adjust with 100 mM Ambic.
4. Add Lys-C to a 1:100 enzyme/substrate ratio (wt/wt). Incubate at 37 °C for 2 hours with agitation at 800 rpm.
5. Add trypsin to a 1:100 enzyme/substrate ratio (wt/wt). Incubate at 37 °C overnight under agitation at 800 rpm.
6. Stop the digestion and precipitate DOC by adding 50% (vol/vol) FA to a final concentration of 2 % (vol/vol). Verify that the final pH of the sample is lower than 3.
7. DOC will form a white precipitate upon acidification. To remove the precipitate, transfer 200 μL samples onto a filtration plate and centrifuge at 800 g for 1 minute at room temperature. Repeat until all the sample is filtered. Samples can now be stored at -20 °C prior further use.
8. Desalt the samples using a 96-well plate packed with C18 resin, following the manufacturer's instructions.
9. Evaporate the elution buffer from the eluted peptides using a SpeedVac at 45 °C. When dry, peptides can be stored at -20 °C until further use. If desired, peptides can be further fractionated for the generation of a spectral library (see ***Note 11)***.
10. Resuspend dried peptides in 25 μL 0.1% (vol/vol) formic acid and sonicate samples for 10 minutes to dissolve the peptides. Add peptides from the iRT kit according to the manufacturer's instructions.

## LC-MS/MS analysis

A mass spectrometer with high-resolving power, equipped with a nano-electrospray ion source, and interfaced with a nano-LC system is required for data-dependent (DDA) and data-independent acquisition (DIA) mass spectrometric (MS) analysis, according to the setup described in the Materials section. We recommend using both DDA-MS and DIA-MS measurements for spectral library generation. For each sample, DDA-MS measurements can be made once per condition, on a single pool of all replicates.

1. Separate peptide mixtures using a gradient with an increasing concentration of buffer B. We recommend using a linear gradient ranging from 3 to 30 % buffer B. The flow rate is set to 300 nL/min throughout the gradient.
2. For DDA-MS measurements, operate the mass spectrometer in positive ion mode with an electrospray voltage of 2500 V. Acquire the survey MS1 scans over a mass range of 350-1400 *m/z* with an Orbitrap resolution of 120,000 using a normalized automatic gain control (AGC) target of 200 % (maximum injection time: 100 ms). Use a data-dependent mode with a cycle time of 3 seconds to generate MS2 spectra using a 30% HCD collision energy at an Orbitrap resolution of 30,000. Accumulate all multiply charged ions (charge states 2-7) with a normalized AGC target of 200 % for maximally 54 ms and with a dynamic exclusion for 60 s.
3. For DIA-MS measurements, set 41 variable-width DIA isolation windows with a 1 *m/z* overlap between windows (see ***Note 12***). DIA-MS2 spectra are acquired at an Orbitrap resolution of 30,000 over a scan range of 150-2000 *m/z* and a normalized AGC target of 400 % for each window. The maximum injection time is set to 54 ms. Acquire MS1 scans using the same settings as described in step 2.

## Library generation in Spectronaut

Here, we describe how to generate a spectral library from LC-MS/MS raw files in the Spectronaut software [13], using the Pulsar search engine (version 15.2.210819.50606).

1. Go to the Library Perspective in Spectronaut and click on "Generate Library from Pulsar / Search Archives..." under the “Spectral Library” tab " in the bottom left corner.
2. Choose an experiment name and click on "Add Runs from File…" or "Add Runs from Folder…" and select the runs from which you want to create the library. Click “Next" to continue.
3. Select the protein database(s) you want to use by clicking "Fasta File...". Fasta file(s) can be assigned per run or for the whole experiment. Click “Next”.
4. Choose search settings by clicking “Search Settings”. Select either the default schema which can be modified or a previously saved setting schema. Schemas can be assigned at either the experiment or run level. Only one setting schema can be set per run. When nothing is selected for a run, default settings will be applied.
5. In the settings, define the enzyme and enzyme digestion type. For a standard LiP-MS experiment, trypsin/P and semi-specific digest type are used as PK can generate unspecific cleavage sites. The remaining parameters remain unchanged. Click “OK” to continue.
6. (Optional) Add previously searched data to further enrich the depth of your library. Click “Next” to proceed.
7. (Optional) Specify gene annotations for this experiment. To use gene annotation, import an organism specific annotation file (GO Annotations EBI) in the Database perspective. Click “Next” to continue.
8. Choose settings for library generation. In a standard LiP experiment, use the BGS Factory Settings (default). Click “Next” to continue.
9. Start the library generation by clicking “Finish”.

## Performing a DIA library-based analysis in Spectronaut

In this section, we describe how to perform a label-free quantification analysis on the peptide level using the Spectronaut software [13].

1. Go to the Analysis Perspective and click on “Set up a DIA Analysis from File...” in the bottom left corner. This will let you navigate to your LC-MS/MS run files or folders that are to be included in this experiment and click on "Open". Provide a name to this experiment on top of the window.
2. Assign a spectral library to each file by clicking on “Assign Spectral Library...”. You can select the library from the Recently Used list, From File, or From Library Perspective. Click on “Load” to proceed. Click on “Next” to continue.
3. Specify search and extraction settings for DIA analysis by selecting the BGS Factory Settings (default) schema.
4. In the Quantification settings, set "Minor (Peptide) Grouping" to "by Modified sequence" and make sure "Data Filtering" is set to "Qvalue sparse" (see ***Note 13***).
5. Go to the "Post Analysis" node and set "Differential Abundance Grouping" to "Minor Group (Quantification Settings)" to analyze your data at a peptide level and "Smallest Quantitative Unit" to "All Ions" for a LiP-Quant analysis. For single-dose experiments, see ***Note 14***. When done, click "Next" to continue.
6. Select a protein database you want to include in this DIA search. The protein database(s) that were used during library generation are already preselected. If you want to include a different protein database, select individually.
7. Specify conditions to perform statistical tests during post analysis. Define the conditions in the "Condition" column and specify which condition is the reference (typically the vehicle control samples). Click “Next” to proceed.
8. (Optional) Specify gene annotations for this experiment. To use gene annotation, import an organism specific annotation file (GO Annotations EBI) in the Database perspective. Click “Next” to continue.
9. The settings for this experiment will be displayed in a summary window. Click "Finish" to start the analysis.

## Post-analysis Processing

Here, we advise how data quality can be assessed and how the *Candidates list* and *Report* can be exported once the DIA library-based analysis is done.

1. Go to the Post Analysis Perspective in Spectronaut. It shows summary information about identification, quantification, and results from the differential abundance test, hierarchical clustering, principal component analysis and GO terms enrichment and clustering.
2. (Optional) Take a look at summaries, tables, and plots available to examine the data quality. Follow the Spectronaut manual for details.
3. Under Differential Abundance, go to the Candidates node to visualize a table with results of differentially regulated peptides. The table is, by default, filtered by a *q*-value (multiple testing corrected *p*-value) of 0.05 and an absolute log2 ratio of 0.58. Remove these default filters by selecting "No filter".
4. Export the *Candidates list* by clicking "Export table...".
5. Go to the Report Perspective and select a report schema in a Normal Report (long format) node. For a LiP-Quant analysis, the following columns are required: R.Condition, R.FileName, R.Replicate, PG.Genes, PG.ProteinAccessions, PG.ProteinNames, PG.Coverage, FG.Quantity (alternatively FG.MS2Quantity or FG.NormalizedMS2PeakArea), EG.ModifiedSequence, EG.PrecursorId, EG.StrippedSequence, EG.Qvalue. We also provide our pre-defined schema “LiP\_Scheme.rs” that can be obtained from <https://github.com/RolandBruderer/MiMB-LiP-Quant> and imported in Spectronaut.
6. Click on "Export Report..." to save the report.

## LiP-Quant Analysis

The LiP-Quant pipeline was developed by Piazza and Beaton et al. to analyze LiP-MS experiments in which cellular lysates are treated with a small molecule at multiple concentrations [11]. This approach serves to prioritize true protein targets of a small molecule using the machine learning-based LiP-Quant score (***see Note 15)***. For more details on performing a single-dose analysis, see ***Note 14***.

Here, we describe how to implement LiP-Quant, and provide a revised version of the R script from Piazza and Beaton et al. [11], now compatible with Microsoft Windows and macOS as well as with the latest version of Spectronaut 15, that enables calculation of the individual peptide subscores that make up the LiP-Quant score for each peptide. Briefly, peptides are first filtered from the Spectronaut *Candidates list* based on differential abundance (an absolute log2 fold-change > 0.46) and statistical significance (*q*-value < 0.01, one sample two-sided t-test with Storey correction for multiple testing) as described by Piazza and Beaton et al. [11]. Subsequently, multiple comparisons of metabolite-treated to vehicle control samples should be selected to obtain the statistical significance level (*q*-value) per peptide from the Spectronaut report; this contributes to one subscore of the LiP-Quant score (see ***Note 15***). We recommend selection of samples treated with metabolite concentrations above the reported physiological concentration, compared to vehicle control, since this is where the strongest effects are expected. After this step, dose-response correlation analysis (using the drc package) is performed on all pre-filtered peptides across the entire concentration range of a metabolite to compute a correlation coefficient (R) for a sigmoidal fit to the data [15]. A peptide with a high R (close to 1) has a high correlation. Additional subscores are computed based on, first, whether a protein is present in a list of contaminant proteins previously determined from positive control experiments (called “crapome” or “*Protein-Frequency-Library*” (PFL)), and second, the number of altered peptides from a single protein found in the top ten percent of all peptides (ranked by *q*-value). Finally, the LiP-Quant score is calculated for each peptide and protein from the weighted subscores and reported in the resulting tables. The resulting tables also provide information on half maximum effective concentration (EC50) values of the metabolite for each peptide, derived from the sigmoidal fit of the metabolite dose-response curve. Thus, relative binding affinities of proteins for the metabolite of interest can be further used to rank multiple identified targets.

The following input files are required for a LiP-Quant analysis using our provided script:

* *Candidates list*. The results table of differentially regulated peptides from pairwise comparison of conditions with no default filters applied.
* *Report.* The Spectronaut report generated on the entire dose response experiment.
* *Protein-Frequency-Library (PFL) table* (optional). The list of identified contaminant proteins from HeLa cells named “LiP-PFL.txt” containing two columns: PFL.FG (e.g., UniProtKB entry) and PLFrequency (fraction of occurrences in ground truth experiments; 0 means that a protein is not found as potential contaminant in any experiments). This is only applicable to experiments with cultured mammalian cells.
* *Table of known targets* (required for model training; otherwise optional). A .txt file table having two columns: ProteinNames (e.g., UniProtKB entry name) and Target (Yes or No).
* *Concentration table* (optional). Can be generated automatically based on the experimental conditions.

## Running LiP-Quant Analysis

Here, we provide a workflow for a LiP-Quant analysis using an R script accessible on <https://github.com/RolandBruderer/MiMB-LiP-Quant>. Make sure that you have installed the recommended (or the newest) version of R (<https://www.r-project.org>) and RStudio (<https://www.rstudio.com>).

1. Download the GitHub folder using the link above.
2. Install R 4.1.1 and RStudio 1.4.1717.
3. Open RStudio and load the LiP-Quant script by clicking on “File > Open File...” from the downloaded folder.
4. (Optional) Adjust a significance level cutoff (Qvalue, default: 0.01, line 108 of the script), an absolute fold-change cutoff (FCcutoff, default: 0.46, line 110 of the script) and the size of a training set (ldaBGsize, default 400, line 112 of the script) if desired. We recommend using the default settings.
5. Execute the script by clicking on the “Source” button. If packages required for LiP-Quant are not installed, they will be installed automatically upon sourcing. If you encounter an error with a package installation, try to restart R and repeat this step, or install manually.
6. In the pop-up window, define a working directory where you saved your files to be analyzed. We recommend using the same folder where the files from GitHub are located.
7. Depending on whether you want to use adjusted score weights determined using your own training datasets (see ***Note 16***), or our pre-calculated score weights (see ***Note 17***), follow Procedure 1 (section 3.10.1) or Procedure 2 (section 3.10.2), respectively.

### Procedure 1 – Machine learning-based determination of adjusted score weights

In this procedure, training datasets provided by the user can be used to determine adjusted score weighting for the four classifiers established by Piazza and Beaton et al. [11].

1. If adjusted score weighting should be determined and applied, select “YES” (1). Default is “NO”.
2. Select which correlation method should be used (default: Pearson).
3. Select what intensity value transformation should be applied (default: normal) and select if dose response correlation testing should be performed (default: YES).
4. Define, if you want to use a protein frequency library (PFL). Default is Automatic (“YES” if "LiP-PFL.txt" file is present in the working directory, else “NO”). See ***Note 18*** for further information.
5. Select the unfiltered *Candidates list(s)* from the working directory. To train your own classifiers, we recommend using more than one training dataset (see ***Note 16***). This process might take a few minutes.
6. Select condition comparisons from which differential peptides should be used for the training. For example, the control sample versus treated samples (with concentrations of the metabolite above its reported physiological concentration). If more training datasets are available, indicate which condition comparisons should be used for model training from each dataset.
7. Select a table with known target proteins.
8. Select the *Spectronaut report*.
9. The next step will be step 3 of the following section 3.10.2 in which the dataset to be analyzed should be specified.

### Procedure 2 – LiP-Quant analysis of new targets

In this procedure, we provide instructions on how to run a LiP-Quant analysis without training new classifiers. In this case, pre-calculated score weights based on 3 training datasets will be used (see ***Note 17***).

1. If score weighing is not desired, click on “NO” in the popup menu.
2. Select which correlation method should be used (default: Pearson).
3. Import the unfiltered *Candidates list* from the working directory.
4. Select sample pairs to be statistically compared in further analysis. We recommend selecting samples treated with metabolite concentrations above its reported physiological concentration, compared to the vehicle control.
5. Select if a list of known targets is to be used. If “YES”, import a table with the known target(s), else click on “NO” to continue. See ***Note 19*** for details.
6. Import the Spectronaut report to perform ML-based ranking.
7. Select if a concentration table should be imported or not. If “YES”, import a concentration table from the working directory, else a standard concentration table is generated from the report.
8. Select which concentration unit is used in the concentration table. Default is nM. Once selected, a folder named “DRCplot” is generated in the working directory and dose response analysis start. DRC plots for all modified sequences are saved in the “DRCplots” folder.
9. Upon dose response analysis, output tables with ranked peptides/proteins based on calculated LiP scores are saved and additional plots (e.g., for the LiP score distributions) are generated in the working directory. See ***Note 20*** for guidelines on interpreting LiP-Quant results.

# Notes

1. We recommend storing the stock solution of PK in thin-walled PCR tube strips, allowing the use of multichannel pipettes for LiP experiments.
2. When TCEP-HCl is dissolved in water, the resulting pH is approximately 2.5. This acidic pH can cause partial precipitation of DOC. Therefore, we recommend dissolving TCEP-HCl in 1 M HEPES pH 7.4.
3. We recommend using approximately 4 pellet volumes of LiP buffer for the resuspension of the cell pellet. It is important that the total protein concentration of the lysate is higher than 2 µg/µL prior to final dilution. More detailed instructions on how to grow model cellular systems, such as *E. coli*, *S. cerevisiae* or HeLa cells can be found in Cappelletti et al. [16] and Piazza and Beaton et al. [11].
4. We advise removal of endogenous metabolites and cofactors (e.g., NADH, ATP) from cell extracts by gel filtration prior to LiP-MS analyses, to reduce the overall activity of the lysate. Endogenous metabolites could be converted to other compounds by catalytically active enzymes in the lysate, which may lead to unwanted indirect effects on protein structures. In addition, removal of endogenous metabolites could render more accessible binding sites that are typically occupied or saturated. Thus, the signal-to-noise ratio of both single- and multi-dose experiments is improved.
5. We recommend diluting the PK stock solution down to 0.2 µg/µL to mitigate possible volumetric errors when pipetting smaller volumes (< 5 µL). Also avoid adding larger volumes (> 5 µL) to the sample, because the PK solution at 4 °C can alter sample temperature and thus affect protease activity.
6. A LiP-MS experiment can be used to identify metabolite-protein interactions in either a single-dose (LiP-SMap) or a multiple-dose (LiP-Quant) format. In LiP-SMap, interactions were probed in *E. coli* lysates using metabolite concentrations estimated from previous metabolite measurements under physiological conditions. We recommend testing concentrations up to 5x higher than measurements of *in vivo* levels of the metabolite in a LiP-SMap experiment. In a LiP-Quant analysis, selection of a concentration range is one of the most important steps. We recommend using a 10-fold dilution series consisting of three concentration steps below and three concentration steps above a reported physiological concentration, including the physiological concentration itself, plus a vehicle control sample (untreated). Exposing the cell extract to a range of metabolite concentrations enables capturing interactions of various affinities. Concentrations substantially higher than the maximal reported *in vivo* values should be avoided as they can cause non-specific effects. It is important that the volume of metabolite solution added to the cell extract remains as low as possible to avoid sample dilution. We recommend preparing a dilution series that allows addition of 1 µL metabolite solution.
7. PK activity can be optimized if needed by altering enzyme-to-substrate ratio and incubation time. We recommend using our standard conditions of 5 minutes at 1:100 enzyme-to-substrate ratio. If a lower proteolytic activity is desired, samples can be exposed to PK for less than five minutes, but we do not recommend incubation times lower than 1 minute since this can substantially affect sample reproducibility. In such cases, the enzyme-to-substrate ratio should instead be adjusted to decrease the PK activity.
8. Inactivation of PK by heat (i.e., 99 ºC) for 5 minutes is a key step of the LiP workflow. Ensure that the temperature of 99 ºC in the samples is reached.
9. Samples are cooled down to avoid opening of tube lids due to vaporization in the samples.
10. Iodoacetamide is light-sensitive and degraded upon irradiation, resulting in an incomplete and inefficient alkylation reaction.
11. To obtain improved sensitivity and proteome coverage, fractionation techniques are recommended. More details on how to fractionate peptide samples are described by Piazza and Beaton et al. [11] or Rappsilber et al. [12].
12. Acquire DIA-MS spectra using the defined variable windows as described in Table 1 below.
13. In Spectronaut 15, "Data Filtering" is set to "Qvalue sparse" as default. This filtering step involves an imputation strategy of missing values. More details about data filtering and imputation can be found in the Spectronaut manual. If the user does not intend to use any imputation strategy, “Data filtering” should be adjusted to “Qvalue”. In general, we recommend using the global imputation strategy for both LiP-SMap and LiP-Quant experiments when using Spectronaut 15.
14. Here, we describe how to perform a label-free peptide quantification on samples treated with a single metabolite concentration (LiP-SMap) using a DIA library-based analysis in the Spectronaut software. Follow the steps as described above for LiP-Quant experiments with a minor modification. In the "Post Analysis" node, set both "Differential Abundance Grouping" and "Smallest Quantitative Unit" to "Minor Group (Quantification Settings)" and make sure that the “Use All MS-Level Quantities” option is unticked as quantities of peptides in DIA-MS are derived from MS2 level only. These settings will automatically aggregate peptide quantities of all charge states detected for a modified peptide sequence. When done, proceed with the step 6 described above in the “Performing a DIA library-based analysis in Spectronaut” section. When finished, summary information about identification, quantification, and results from the differential abundance test, hierarchical clustering, principal component analysis and GO terms enrichment (if selected in the analysis) and clustering are shown in different nodes. To identify peptides that change in abundance between the defined conditions, go to the Candidates node to visualize a table with results. The table is, by default, filtered by a *q*-value (multiple testing corrected *p*-value) of 0.05 an absolute log2 ratio of 0.58. The significance and absolute log2 ratio thresholds can be changed by the user directly on top of the table if desired. The *Candidates* *list* table can be exported by clicking "Export table...". To obtain quantities of individual peptides, go to the Report Perspective and select a report schema in a Normal Report (long format) node that can be further processed and analyzed by the user if desired.
15. The LiP-Quant score, built into a classifier using linear discriminant analysis (LDA), is a combined score calculated from weighted subscores of features that contributed the most to the identification of true targets in positive control experiments [11]. The LiP-Quant score consists of the following features (subscores): (1) dose-response correlation analysis, (2) the presence of a protein with altered peptides in a list of proteins called "crapome" or “*Protein-Frequency-Library*" (PFL), (3) the number of peptides from a protein appearing in the top ten percent of all peptides ranked by *q*-value in the *Candidates list* from Spectronaut, and (4) the statistical significance (*q*-value) calculated from relative peptide abundances between metabolite-treated and vehicle sample. The PFL was built by Piazza and Beaton et al. for proteins that appeared in 9 out of 11 positive control experiments in HeLa cells, as non-specific targets (not known to bind a specific drug) [11].
16. To establish the criteria that contribute to the identification of true targets and train new classifiers, the user can follow the instructions in Piazza and Beaton et al. [11]. In general, we recommend a retraining, if the experimental setup changes significantly, i.e., concentration layout, protease, *Protein-Frequency-Library* (PFL), etc. Importantly, the selection of positive control experiments must be done with care and high-quality training data for a retraining are essential.
17. These datasets are based on treatment of a HeLa lysate with rapamycin, calyculin A and staurosporin. All datasets are available via the PRIDE partner repository with the dataset identifiers PXD018204 and PXD015446. The pre-calculated adjusted score weights are: (1) 1.317320 for the statistical significance (*q*-value) calculated from relative peptide abundances between metabolite-treated and vehicle sample, (2) 9.191420 for dose-response correlation analysis, (3) 1.332628 for the number of peptides from a protein appearing in the top ten percent of all peptides ranked by *q*-value in the *Candidates list* from Spectronaut, and (4) 1.452496 for the presence of a protein with altered peptides in the PFL. As shown by Piazza and Beaton et al. [11], the most dominant component of the LiP-Quant score is correlation to a sigmoidal fit of the small molecule dose-response curve. Hence, peptides with a sigmoidal correlation coefficient R close to 1 will be scored higher.
18. The PFL can be used only for experiments where cultured mammalian cells (e.g., HeLa) are under investigation. In order to study other experimental models, the user must build a new list of potential protein contaminants. If the PFL is not available, this subscore will not be incorporated into the final LiP-Quant score.
19. When the list of known targets is provided in the LiP-Quant analysis of new targets, the information is used for plotting purposes. For example, distribution of LiP-Quant scores of the known targets will be highlighted in the resulting plots.
20. We provide guidelines for interpreting LiP-Quant results. The purpose of the LiP-Quant score is to provide a user with a list of ranked target peptides/proteins. The higher the LiP-Quant score, the more likely it is that this peptide/protein is a true metabolite binder. However, note that orthogonal validation experiments are required to confirm that identified metabolite-protein interactions are not due to indirect effects. The LiP-Quant score follows an absolute scale between 0 and 6, where 6 represents the maximum possible likelihood of being a genuine target. As described by Piazza and Beaton et al. [11], this absolute scale helps to make a direct comparison between experiments. The determination of a LiP-Quant score threshold was based on aggregating results from five positive control experiments, in which drugs with known targets (rapamycin, calyculin A, selumentinib, FK506 and fostreicin) were added to HeLa lysate. Upon model training, the distribution of peptide LiP-Quant score appeared bimodal, displaying a clear enrichment of peptides from known target proteins at LiP-Quant score > 1.5. The value of this threshold was derived from the median score obtained in the experiments plus three standard deviations. Therefore, we recommend applying a LiP-Quant score threshold of 1.5 for a confident identification of metabolite binding proteins. Furthermore, every peptide scored by LiP-Quant has an EC50 value assigned to it, representing the concentration of a metabolite at half-maximum of a peptide response (those peptides for which the computed EC50 value is not within the concentration range used for the titration experiment should be excluded). Relative binding affinities of proteins (based on the computed EC50 values of peptides) to the metabolite of interest can accordingly be used to rank multiple identified targets. Finally, it is important to bear in mind that identified target peptides in putative metabolite binding proteins may represent binding sites, but that conformational changes in other parts of the protein due to metabolite binding may also be identified by LiP-MS approaches. This will be affected both by whether metabolite binding triggers larger structural changes in the protein and also by how well the protein sequence is covered in the MS experiment.

# Acknowledgements

This work was supported by the European Research Council (grant agreement no. 866004), the EPIC-XS Consortium (grant agreement no. 823839), a Sinergia grant from the Swiss National Science Foundation (SNSF grant CRSII5\_177195), the National Center of Competence in Research AntiResist and the Promedica Stiftung, Chur.

# References

1. Chubukov V, Gerosa L, Kochanowski K, Sauer U (2014) Coordination of microbial metabolism. Nat. Rev. Microbiol. 12:327–340

2. Lindsley JE, Rutter J (2006) Whence cometh the allosterome? Proc. Natl. Acad. Sci. U. S. A. 103:10533–10535

3. Bennett BD, Kimball EH, Gao M, et al (2009) Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat Chem Biol 5:593–599. https://doi.org/10.1038/nchembio.186

4. Lomenick B, Hao R, Jonai N, et al (2009) Target identification using drug affinity responsive target stability (DARTS). Proc Natl Acad Sci U S A 106:21984–21989. https://doi.org/10.1073/pnas.0910040106

5. Gallego O, Betts MJ, Gvozdenovic-Jeremic J, et al (2010) A systematic screen for proteing-lipid interactions in Saccharomyces cerevisiae. Mol Syst Biol 6:. https://doi.org/10.1038/msb.2010.87

6. Savitski MM, Reinhard FBM, Franken H, et al (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. Science (80- ) 346:. https://doi.org/10.1126/science.1255784

7. Huber KVM, Olek KM, Müller AC, et al (2015) Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. Nat. Methods 12:1055–1057

8. Geer MA, Fitzgerald MC (2016) Characterization of the Saccharomyces cerevisiae ATP-Interactome using the iTRAQ-SPROX Technique. J Am Soc Mass Spectrom 27:233–243. https://doi.org/10.1007/s13361-015-1290-z

9. Piazza I, Kochanowski K, Cappelletti V, et al (2018) A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. Cell 172:358-372.e23. https://doi.org/10.1016/j.cell.2017.12.006

10. Diether M, Nikolaev Y, Allain FH, Sauer U (2019) Systematic mapping of protein‐metabolite interactions in central metabolism of Escherichia coli. Mol Syst Biol 15:. https://doi.org/10.15252/msb.20199008

11. Piazza I, Beaton N, Bruderer R, et al (2020) A machine learning-based chemoproteomic approach to identify drug targets and binding sites in complex proteomes. Nat Commun 11:. https://doi.org/10.1038/s41467-020-18071-x

12. Rappsilber J, Mann M, Ishihama Y (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat Protoc 2:1896–1906. https://doi.org/10.1038/nprot.2007.261

13. Bruderer R, Bernhardt OM, Gandhi T, et al (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol Cell Proteomics 14:1400–1410. https://doi.org/10.1074/mcp.M114.044305

14. Schopper S, Kahraman A, Leuenberger P, et al (2017) Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. Nat Protoc 12:2391–2410. https://doi.org/10.1038/nprot.2017.100

15. Ritz C, Baty F, Streibig JC, Gerhard D (2015) Dose-response analysis using R. PLoS One 10:1–13. https://doi.org/10.1371/journal.pone.0146021

16. Cappelletti V, Hauser T, Piazza I, et al (2021) Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. Cell 184:545-559.e22. https://doi.org/10.1016/j.cell.2020.12.021

Figure 1: Schematic of LiP-MS workflow for identification of metabolite-protein interactions on a proteome-wide scale. A metabolite is incubated at a single concentration (LiP-SMap) or multiple concentrations (LiP-Quant) with a native-like cellular lysate and briefly exposed to a sequence-unspecific protease (proteinase K; PK) in the limited proteolysis step. Because binding of a metabolite to a target protein can result in altered protease accessibility, condition-specific (i.e., metabolite-specific) fragments of target proteins are generated. Protein fragments are further digested to peptides by trypsin and subjected to label-free quantitative MS analysis. Differentially abundant peptides between the metabolite-treated sample (metabolite-bound form) and the untreated sample (metabolite-free form) can identify protein targets of the metabolite. In LiP-SMap, structurally altered peptides can be identified using a Volcano plot, which displays log2-transformed fold-changes of peptide intensities between the metabolite-treated and untreated samples as a function of statistical significance (q-value). The LiP-Quant approach enables ranking of targets based on a machine learning-derived score (LiP-Quant score) calculated from weighted subscores of features determined to contribute the most to the identification of true targets, namely (1) dose-response correlation analysis (based on correlation to a sigmoidal fit of the metabolite dose-response curve), (2) the presence of a protein with altered peptides in a list of proteins called "crapome" or “Protein-Frequency-Library" (PFL), (3) the number of peptides from a protein appearing in the top ten percent of all peptides ranked by q-value in the Candidates list from Spectronaut, and (4) the statistical significance (q-value) calculated from relative peptide abundances between the metabolite-treated sample and the untreated sample.

Table 1: Definition of variable windows for DIA-MS measurements.

|  |  |  |  |
| --- | --- | --- | --- |
| **Window** | **Mass-to-charge ratio (*m/z*)** | **Charge (*z*)** | **Isolation window (*m/z*)** |
| 1 | 358 | 2 | 16 |
| 2 | 373 | 2 | 16 |
| 3 | 388 | 2 | 16 |
| 4 | 403 | 2 | 16 |
| 5 | 418 | 2 | 16 |
| 6 | 433 | 2 | 16 |
| 7 | 448 | 2 | 16 |
| 8 | 463 | 2 | 16 |
| 9 | 478 | 2 | 16 |
| 10 | 493 | 2 | 16 |
| 11 | 508 | 2 | 16 |
| 12 | 523 | 2 | 16 |
| 13 | 538 | 2 | 16 |
| 14 | 553 | 2 | 16 |
| 15 | 568 | 2 | 16 |
| 16 | 583 | 2 | 16 |
| 17 | 598 | 2 | 16 |
| 18 | 613 | 2 | 16 |
| 19 | 628 | 2 | 16 |
| 20 | 643 | 2 | 16 |
| 21 | 659 | 2 | 18 |
| 22 | 676 | 2 | 18 |
| 23 | 693 | 2 | 18 |
| 24 | 710 | 2 | 18 |
| 25 | 727 | 2 | 18 |
| 26 | 744 | 2 | 18 |
| 27 | 761 | 2 | 18 |
| 28 | 778 | 2 | 18 |
| 29 | 795 | 2 | 18 |
| 30 | 813 | 2 | 20 |
| 31 | 832 | 2 | 20 |
| 32 | 851 | 2 | 20 |
| 33 | 870 | 2 | 20 |
| 34 | 889 | 2 | 20 |
| 35 | 908 | 2 | 20 |
| 36 | 929.5 | 2 | 25 |
| 37 | 953.5 | 2 | 25 |
| 38 | 977.5 | 2 | 25 |
| 39 | 1006.5 | 2 | 35 |
| 40 | 1048 | 2 | 50 |
| 41 | 1111 | 2 | 78 |