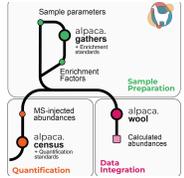


Mar 24, 2025

Integration of Sample Preparation and Subcellular Fraction Enrichment in Alpaca

DOI

<https://dx.doi.org/10.17504/protocols.io.j8nlk9z46v5r/v1>



Borja Ferrero Bordera^{1,2}, Sandra Maaß²

¹LMU Munich; ²Universität Greifswald



Borja Ferrero Bordera

LMU Munich

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.j8nlk9z46v5r/v1>

Protocol Citation: Borja Ferrero Bordera, Sandra Maaß 2025. Integration of Sample Preparation and Subcellular Fraction Enrichment in Alpaca. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.j8nlk9z46v5r/v1>

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: February 25, 2025

Last Modified: March 24, 2025

Protocol Integer ID: 123349

Keywords: proteomics, mass-spectrometry, data analysis, proteome, python, alpaca, protein abundances, subcellular, data integration, alpaca proteomics data integration module, subcellular fraction enrichment in alpaca, based absolute protein quantification result, absolute protein quantification result, mass spectrometric abundance information, accurate quantification of subcellular fraction, absolute protein quantification, subcellular fraction enrichment, interpretability of mass spectrometry, mass spectrometry, protein amount, subcellular fraction, alpaca, integration of enrichment procedure, protein, enrichment procedure, sample preparation, integration of sample preparation, proteomic

Funders Acknowledgements:

People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme

Grant ID: 813979

Abstract

The Alpaca proteomics data integration module focuses on the final stages of data analysis for absolute protein quantification, transforming the mass spectrometric abundance information to absolute protein amounts in the cell(s) based on the experimental preparation. Moreover, this module allows for the integration of enrichment procedures during sample preparation, allowing for accurate quantification of subcellular fractions. Altogether, this module enhances the robustness and interpretability of mass spectrometry-based absolute protein quantification results.

Troubleshooting

Getting Started

- 1 This tutorial describes modules that integrate sample preparation details to calculate absolute protein abundances in femtomole from mass spectrometry (MS) data. The results allow for accurate biological interpretation directly calculating protein concentrations derived from basic sample preparation workflows but also meet the requirements of specialized experimental setups (e.g., employing proteome enrichment).

Protocol



NAME

Absolute Quantification of Proteome Abundances with the Alpaca Pipeline

CREATED BY
Borja Ferrero Bordera

Preview

- 2 **Integration of Sample Preparation Details in Alpaca**
Alpaca supports the integration of two types of sample preparation workflows, depending on whether a specific fraction of the proteome has been enriched or not. The two corresponding workflows are:
 - **Integration of basic sample preparation workflows** (see Step 5)
 - **Integration of sample preparation workflows with protein enrichment** (see Step 6)

To ensure proper integration of sample preparation steps into downstream analysis, the following input tables must be prepared and loaded as pandas DataFrames:

 - **Sample preparation table** – Required for both workflows. Includes sample-specific metadata (see Table 1; Step 3).
 - **Enrichment standards table** – Required **only** for enriched samples. Contains protein standards used to estimate enrichment factors (see Table 2; Step 4).
- 3 To integrate sample preparation details into the protocol, you must provide a table (pandas DataFrame) containing the relevant experimental information, as shown in **Table 1**. This table includes the sample volumes, protein concentrations, enrichment settings, and optional details for standards used during subcellular fractionation or proteome enrichment. The table should be imported following the format described below.

Table 1. Sample preparation details input supported by alpaca proteomics.

	Condition	SampleVolume	ProteinConcentration	AmountMS	CellsPerML	TotalCultureVolume	ProteinSRM	fmoISRM	Enrichment	EnrichmentMode	Std
	Cond1_t0	2.31	2.99	9.67	4.54	7.54	TNAMLN	4.44	FALSE		3.9
	Cond2_t1	2.5	0.2	4.1	5.13	2.62	AJFVYC	4.85	TRUE	Concentration	2.4
	Cond3_t2	7.38	6.56	2.77	3.66	3.8	BYEKSC	9.71	TRUE	Enrichment	5.7

- **Condition:** Identifier for the condition or timepoint in which the parameters were applied.
- **SampleVolume:** Volume (in μL) of the protein extract used for digestion.
- **ProteinConcentration:** Measured protein concentration (in $\mu\text{g}/\mu\text{L}$) of the sample.
- **AmountMS:** Sample amount (in μg) injected into the mass spectrometer.
- **CellsPerML:** Determined number of cells per mL in the original culture.
- **TotalCultureVolume:** Total culture volume harvested (in μL).
- **ProteinSRM (Optional):** Accession of a protein used in SRM-based enrichment quantification.
- **fmoISRM (Optional):** Femtomoles of the SRM-targeted protein used to estimate enrichment.
- **Enrichment (Optional):** Boolean flag indicating whether a sample has been enriched (TRUE or FALSE).
- **EnrichmentMode (Optional):** Type of enrichment – either Enrichment or Concentration. See **Figure 3** for details.
- **StdDilution (Optional):** Dilution factor applied to the stock solution of enrichment standards before addition. Use if undiluted. Relevant only when enrichment is calculated using `alpaca.gathers()`
- **StdVolume (Optional):** Volume (in μL) of enrichment standards added to the sample. Also used only with `alpaca.gathers()`

The **enrichment type** should be specified in the **Sample Preparation table** (Table 1) based on the strategy used to quantify the proteome fraction of interest (**Figure 3**).

- **Enrichment** refers to protocols that selectively isolate specific protein subsets based on their properties (e.g., enrichment of membrane proteins based on hydrophobicity).
- **Concentration** applies to workflows where a proteome fraction is concentrated to allow quantification of total abundance (e.g., concentration of extracellular proteins from highly diluted supernatants).

This distinction ensures that the correct calculation method is applied during data processing.

Enrichment approaches

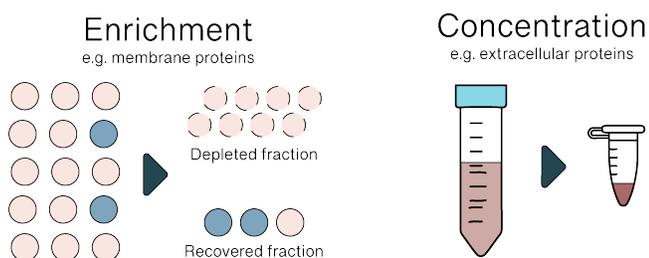


Figure 3. Types of enrichment supported by alpaca pipeline.

- The enrichment standards file should be set-up as described in Table 2. The columns **Accession**, **MW (kDa)** (molecular weight in kilodaltons), and **StdConcentration** (standard concentration in the stock solution, in $\mu\text{g}/\mu\text{L}$) are required for correct enrichment factor calculation. The Protein column was included here for readability and improved interpretation but is not required for the pipeline to run.

Table 2. Example of an enrichment standards file

Protein	Accession	MW (kDa)	StdConcentration ($\mu\text{g}/\mu\text{l}$)
Lysozyme	P00698	14.3	17.78032
Alcohol Dehydrogenase	P00330	36.8	45.75635
Soybean Trypsin Inhibitor	P01071	20	24.86758

Integration of sample preparation workflows

- This protocol allows the integration of sample preparation details for experiments in which protein concentrations were not altered prior to quantification (e.g., direct measurement of raw lysates).

Quantification of non-enriched proteins

e.g., raw lysate proteins

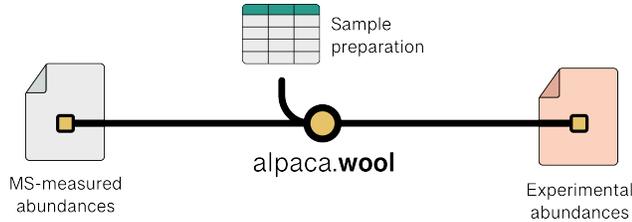


Figure 1. Workflow for Sample Preparation Integration in Alpaca

This diagram illustrates the input and output structure for integrating sample preparation details into the proteomics quantification pipeline.

- **Grey Document (left):** Input file containing fmol-based protein quantifications (typically obtained from mass spectrometry analysis).
- **Central Table:** Sample preparation metadata table (see **Table 1**), including experimental details such as sample volume, protein concentration, enrichment type, and standard volume.
- **Gold Circle:** Indicates the integration step where data from the quantification file and sample prep table are merged.
- **Pink Document (right):** Output file containing adjusted protein abundances, now corrected for sample preparation parameters (e.g. molecules per cell).

The block of code below describes how to proceed with the analysis.

```
# 1. Import Sample preparation details as described in Table 1.

sample_prep = pd.read_csv('params.csv', sep=',')

# 2. Integrate your quantified proteome (quant_df) with your
sample preparation (sample_prep)

results = alpaca.wool(quant_df, sample_prep)
```

Integration of simple sample preparation workflows with protein enrichment

- 6 This protocol also supports experiments where protein concentrations were modified during sample preparation (e.g., quantification of enriched membrane proteins).

Note

The pipeline is designed to use intact proteins as enrichment standards to quantify the enrichment or concentration steps. If isotopically labeled peptides (e.g., AQUA peptides) were used instead, enrichment factors should be calculated externally using dedicated SRM analysis software such as Skyline. These values can then be manually added to the sample preparation table, allowing the analysis to proceed using the standard protocol described above.

Quantification of enriched proteins

e.g., extracellular or membrane proteins

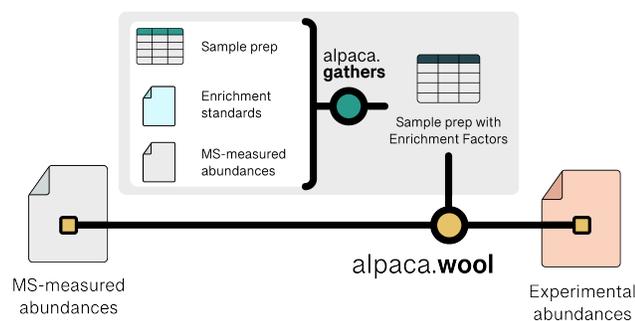


Figure 2. Workflow of Sample Preparation Integration Using alpaca.gathers

This schematic illustrates the data flow for calculating enrichment factors using the alpaca.gathers module:

- **Sample Prep Table:** Metadata file describing sample volumes, protein concentrations, and enrichment types (see Table 1).
- **Enrichment Standards:** File with known protein standards used to estimate enrichment efficiency (see Table 2).
- **MS-measured Abundances:** File containing the list of proteins quantified in fmol (obtained from alpaca.census function).

These three inputs are processed by alpaca.gathers, which calculates the enrichment factors for each sample. The output is a **sample prep table augmented with enrichment factors**, enabling further correction and normalization of proteomics data.

- The **grey file icon (left)** represents an input file containing proteins quantified in fmol (obtained from alpaca.census function).
- The **blue file icon** is the enrichment standards file (see Table 2).
- The **central teal circle** symbolizes the alpaca.gathers function.
- The **pink file icon (right)** is the downstream output used in further processing steps.

The enrichment standards file should be set-up as described in Table 2. The columns **Accession**, **MW (kDa)** (molecular weight in kilodaltons), and **StdConcentration** (standard concentration in the stock solution, in $\mu\text{g}/\mu\text{L}$) are required for correct enrichment factor calculation. The Protein column was included here for readability and improved interpretation but is not required for the pipeline to run.

Table 2. Example of an enrichment standards file

Protein	Accession	MW (kDa)	StdConcentration ($\mu\text{g}/\mu\text{l}$)
Lysozyme	P00698	14.3	17.78032
Alcohol Dehydrogenase	P00330	36.8	45.75635
Soybean Trypsin Inhibitor	P01071	20	24.86758

The block of code below describes how to proceed with the analysis.



```
# 1. Import Sample preparation details as described in Table 1.

sample_prep = pd.read_csv('params.csv', sep=',')

# 2. Import your enrichment standards file as described in Table
2.

enrichment_std = pd.read_excel('enrichment_std.xlsx')

# 3. Calculate the enrichment factors from the quantified
enrichment standards. The calculated
# enrichment factors are incorporated into the sample preparation
dataframe, which is updated
# and returned by the function

enrichment_std, sample_prep_updated = alpaca.gathers(quant_df,
enrichment_std, sample_prep)

# 4. Integrate your quantified proteome (quant_df) with your
sample preparation (sample_prep)

results = alpaca.wool(quant_df, sample_prep_updated)
```