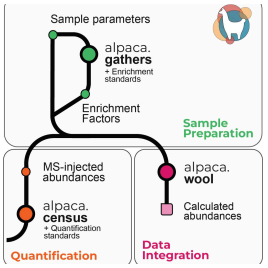
Mar 21, 2025

Absolute Quantification of Proteome Abundances with the Alpaca Pipeline



DOI
https://dx.doi.org/10.17504/protocols.io.5jyl8dx57g2w/v1

Borja Ferrero Bordera[1,2], Sandra Maaß[2]

[1]LMU Munich; [2]Universität Greifswald

B Borja Ferrero Bordera
LMU Munixh

OPEN ACCESS

DOI: https://dx.doi.org/10.17504/protocols.io.5jyl8dx57g2w/v1

Protocol Citation: Borja Ferrero Bordera, Sandra Maaß 2025. Absolute Quantification of Proteome Abundances with the Alpaca Pipeline. **protocols.io** https://dx.doi.org/10.17504/protocols.io.5jyl8dx57g2w/v1

Protocol status: Working
We use this protocol and it's working

Created: February 16, 2025

Last Modified: March 21, 2025

Protocol Integer ID: 120476

Keywords: proteomics, bioinformatics, python, absolute proteome quantification, data analysis, Mass-Spectrometry, alpaca, absolute quantification of proteome abundance, various proteomics data format, proteomics pipeline, analysis of absolute protein quantification data, absolute protein quantification data, protein abundances in complex biological sample, proteome abundance, absolute protein quantification, alpaca pipeline the quantification module, protein abundance, mass spectrometry data, processing of mass spectrometry data, quantification module, complex biological sample, alpaca pipeline, alpaca, proteomic, mass spectrometry

Funders Acknowledgements:
People Programme (Marie Skłodowska-Curie Actions) of the European Union's Horizon 2020 Programme
Grant ID: 813979

Abstract

The quantification module of the ALPACA (Absolute Protein Quantification) proteomics pipeline is designed to facilitate the analysis of absolute protein quantification data. This Python-based tool streamlines the processing of mass spectrometry data, enabling researchers to accurately determine protein abundances in complex biological samples. By integrating with various proteomics data formats and employing robust statistical methods, the module ensures precise quantification, thereby enhancing the reliability of downstream analyses.

Getting Started

1 Install Alpaca library in case it has not been installed before and import the package.

```
from alpaca_proteomics import alpaca
```

2 Make sure that your data is processed as explained in the protocol for Data Processing with Alpaca.

Protocol

NAME

Pre-processing Proteomics Data with the Alpaca Pipeline

CREATED BY

Borja Ferrero Bordera

Preview

3 Absolute proteome quantification through Alpaca has been optimized for the use of intact proteins as standards due to the advantages listed below.

In case other approaches like stable-isotope-labeled (SIL) spiked-in peptide standards were used, a previous step should be performed to calculate the amounts of native reference proteins (e.g. using Skyline) and provide those as anchor proteins as described in this protocol.

Note

Compared to labelled peptides or concatemers, anchor proteins present several advantages:

Enhanced Quantitative Accuracy: intact protein standards closely mimic the native proteins in a sample, ensuring that they undergo similar digestion and ionization processes during mass spectrometry analysis. This similarity leads to more accurate quantification compared to peptide-based standards, which may not fully replicate the behavior of the intact protein.

Consistency in Sample Preparation: Using intact proteins as standards ensures that they are subjected to the same sample preparation steps as the target proteins, including denaturation, reduction, alkylation, and digestion. This consistency minimizes variability and potential biases introduced during sample processing.

Applicability to Complex Samples: Intact protein standards are particularly beneficial when analyzing complex biological samples, such as tissue extracts or serum, where protein-protein interactions and matrix effects can influence quantification. Their use helps account for these complexities, leading to more reliable results.

Facilitation of Method Validation and Standardization: Incorporating intact protein standards into quantitative proteomics workflows aids in method validation and standardization across different laboratories and studies. This practice enhances the reproducibility and comparability of quantitative data.

4 Prepare your anchor protein standards file providing the protein abundances in molar amounts.

Important: For proper recognition, column headers for ProteinID should be “Accession” and the column for molar amounts should contain “fmol”. Ensure that the given IDs are the same as the Accession column in the analyzed dataset.

	Accession	MW (kDa)	Amount (fmol)
	O75475	48.2	75
	P02768	10.1	50
	P05067	32.5	25
	Q00653	20.9	30
	Q9Y6K9	65.8	100

Quantification

5 Import the standards file as a dataframe, for example, as described below.

```
standards_file = 'UPS2.xlsx'
st_proteins = alpaca.eats(standards_file)
```

5.1 **Optional:** If the standards were only added in given samples, those should be specified in a list. Make sure that the sample names are exactly the same as in the formatted dataset.

```
spiked_samples = ['iBAQ Before_Induction_01', 'iBAQ Control_01',
'iBAQ Diamide_01']
```

6 The function *alpaca.census()* allows for the transformation of MS-intensities into molar abundances based on a set of anchor proteins (provided in the standards_file; e.g. UPS2 standards) by fitting log2-transformed molar abundances to log2-transformed intensities.

The function takes 2 dataframes:

- Clean dataframe (described in protocol)
- Anchor protein dataframe (see details above)

Additionally, the intensity column should be specified.

```
quant_df, id_standards, coef, inter, r2 = alpaca.census(clean_df,
st_proteins,
                                                        lfq_col='iBAQ',
                                                        filter_col =
'Sample',
                                                        added_samples =
spiked_samples)
```

Quantification is performed based on fitting known anchor protein amounts to the measured label-free intensities. An R^2 >= 0.90 is recommended for intact protein

standards, while R^2 >= 0.95 would be advised when using SIL peptide standards.

The resulting dataframe *quant_df* contains a column with the calculated molar abundances. Additionally, *id_standards* dataframe contains the identified anchor proteins in the measured samples. Slope, intercept and R2 are also returned by this function.

Optional arguments can be specified to adjust quantification parameters:
- In case only a few samples were spiked with the given standards, specify the filter column name ("*filter_col*") and the spiked sample ("*added_samples*").

More information on the function arguments can be found at **https://borfebor.github.io/alpaca_proteomics/quantification/**

At this point, ***quant_df*** contains the protein abundances measured in the MS analysis. For biological
insights further analysis could be performed.