PTMdecoder User Manual

(Last updated: June 10, 2025)

Content

1 Introduction

1.1 Software Overview

PTMdecoder is tool for **discriminating and quantifying isobaric modification peptidoforms (IMPs)** in data-dependent acquisition (DDA) bottom-up proteomics. PTMdecoder imposes **no restrictions** on the type or number of peptidoforms in a multiplexed spectrum, and requires no prior spectral library for individual peptidoforms. It delivers both the peptidoform- and site-level results, offering novel insights into protein modification states that are often overlooked by current proteomic data analysis workflows.

1.2 Installation

1 Hardware and OS requirements:

A computer with a 64-bit version Windows 10 (and above) is required to run PTMdecoder, and 16GB of RAM or more.

2 Environment setup:

Download the **MATLAB Runtime R2022a (9.12)** installer from the MathWorks website. Run the installer and follow the on-screen instructions.

3 Install PTMdecoder:

Download the installation package from http://fugroup.amss.ac.cn/software/PTMdecoder/PTMdecoder.html and launch the installer.

2 Running PTMdecoder Using the Graphical User Interface (GUI)

Launch PTMdecoder by double-clicking “PTMdecoder.exe”. The GUI of PTMdecoder (**Figure 1**) provides three independent workflows.

1. Single run deconvolution
2. Single run re-quantification

3. Pairwise comparison

Key Features:

- Model execution: Each workflow consists of multiple self-contained analysis modules. Enable or disable individual analyses using the checkbox at the top of each panel of analysis. Disabled analyses will be skipped during execution.
- Parameter reusability: Click the “Load parameters from previous task parameter file” button in any analysis step to rapid configurate using .param files from prior tasks.
- Batch processing: For datasets requiring identical parameters (e.g., replicates), each analysis module includes a dedicated “Tasks” panel for batch operations.

To run an analysis in PTMdecoder, select a workflow, fill in the parameters, and click the 'Save Parameter & Run' button in the Save & Run section; the execution status will update in real time in the Output panel below. This manual provides step-by-step guidance for configuring parameters and interpreting results for each workflow.

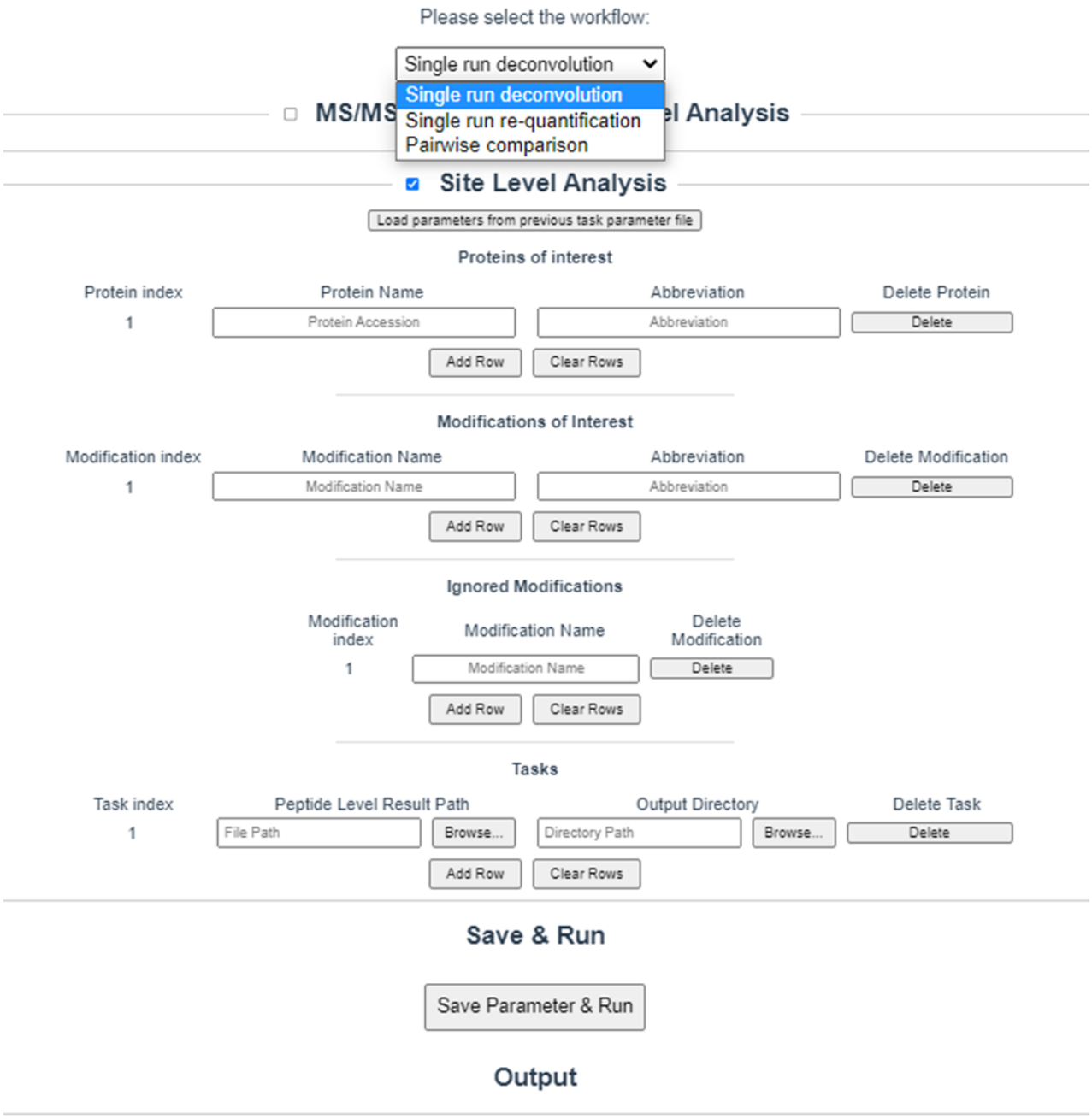

Figure 1. The GUI of PTMdecoder

2.1 Single run deconvolution

This workflow implements a two-stage analytical pipeline centered on individual mass spectra identifications. The sequential stages provide:

- **MS/MS & Peptidoform-Level Analysis**: Deconvolutes MS/MS spectra and extracted ion chromatogram (XIC) peaks to resolve the abundance of component IMPs in the samples.

- **Site-Level Analysis**: Computes site-level quantifications by aggregating abundances of all peptidoforms containing user-specified modification sites.

1 MS/MS & Peptidoform-Level Analysis

This module performs high-resolution deconvolution of MS/MS spectra and XIC peaks to discriminate and quantify co-eluting IMPs (**Figure 2**).

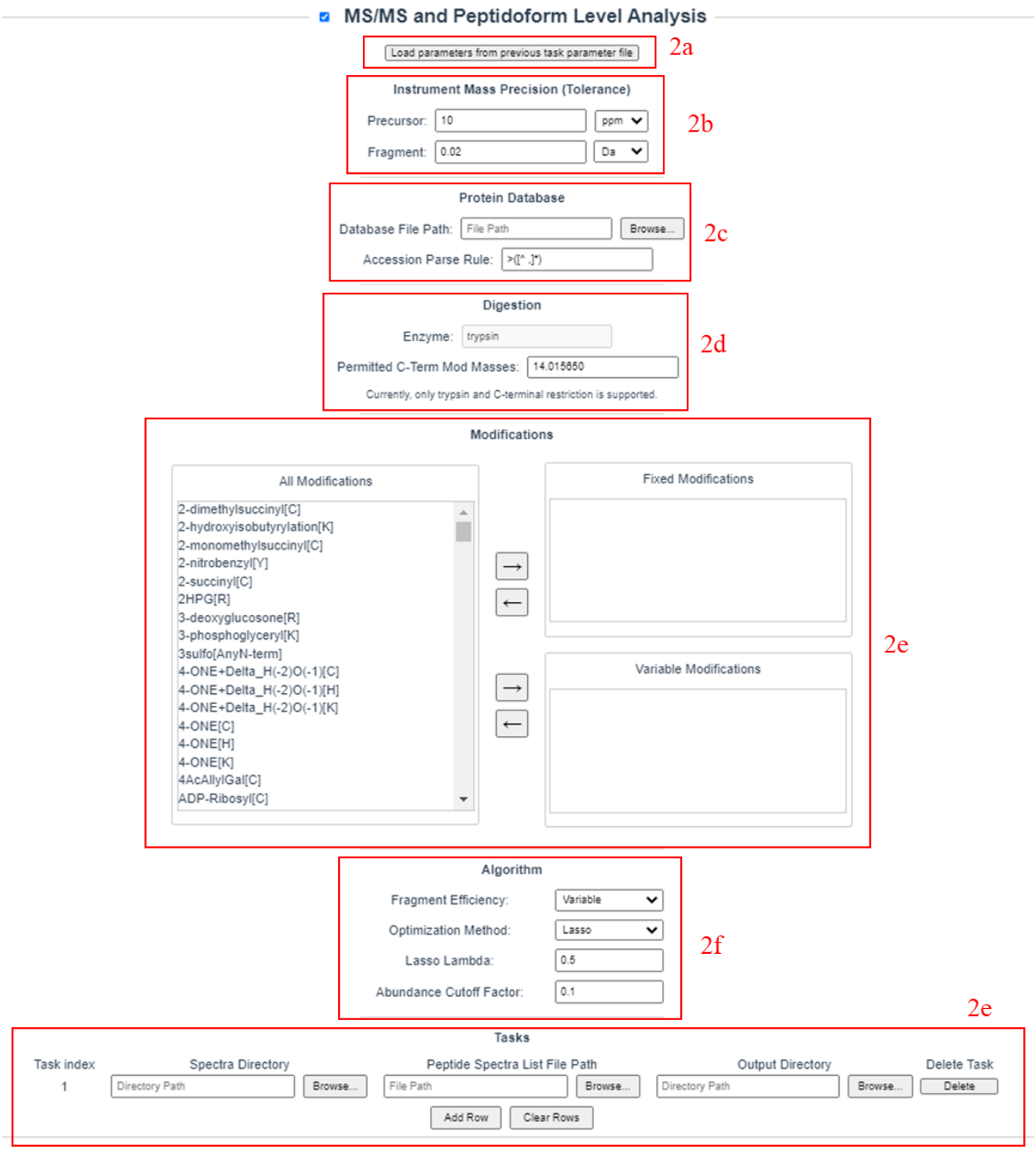

Figure 2 MS/MS & Peptidoform-Level Analysis

i. Input Requirements

Required inputs:

a) Spectra directory path (folder containing all MS/MS spectrum files (mgf, ms1, ms2 pre-converted via **pParse v2.2.1**, usage guidelines are provided in Appendix 1) used in the

corresponding search engine results)

b) Protein sequence database (.fasta)

c) Search engine results (currently supports only a custom “peptide spectra list” format; a conversion script is provided for Mascot .dat files, refer to Appendix 2)

- Structure of “peptide spectra list” format:

1. Peptide sequence line:
   - A single line containing uppercase letters representing a peptide sequence (e.g., GKGGKGLGK).
   - Indicates that the subsequent spectrum lines correspond to modified forms of this peptide (unmodified peptides are invalid).
2. Spectrum lines:
   - One or more lines following each peptide sequence line.
   - Each line specifies a spectrum that identifies the modified peptide above.
   - Format: [MGF filename][TAB][.dta spectrum identifier]
     - Fields are separated by a tab character.
   - Example:

SampleX_2025_HCDFT.mgf SampleX_2025.1234.1234.2.0.dta

- Requirements:
  - Each peptide sequence **MUST** have at least one spectrum line.
  - All .dta files referenced in spectrum lines must exist in the Spectra directory.
- Example:

KSAPSTGGVKK SampleX_2025_HCDFT.mgf SampleX_2025.1234.1234.2.0.dta SampleX_2025_HCDFT.mgf SampleX_2025.5678.5678.2.0.dta GKGGKGLGK SampleY_2025_HCDFT.mgf SampleY_2025.9012.9012.2.0.dta

ii. Parameters

a) “Load parameters from previous task parameter file” button (**Figure 2a**). Imports saved parameter files (.param) to automatically configure all parameters in this module.

b) Tolerance settings (**Figure 2b**)

- “Precursor”: Mass tolerance for precursors (supports ppm and Da)
- “Fragment”: Mass tolerance for fragment ions (Da-only)

c) Protein database (**Figure 2c**):

- Database file path. Click “Browse...” to select a protein sequence database file (*.fasta).
- Accession Parse Rule. Specify a regular expression pattern to extract protein accessions from fasta headers. Example for UniProt header : “>sp|P12345|TEST Example1 and test”, “>([^ ,]*)” to match “>sp|P12345|TEST”, and “>.*\|(.*)\|” to match “P12345”.

d) Enzymatic constraints (**Figure 2d**). The software currently only supports trypsin. The “Permitted C-Term Mod Masses” field specifies masses of allowed C-terminal modification (semicolon-separated). Must exactly match the masses declared in search engine. For example, “14.015650” (covers methylation) or “6.020129;10.008269” (covers “Label_13C(6)” and “Label_13C(6)15N(4)”). Critical: isotope labels in peptide C-term should be explicit declared here.

e) Modifications (**Figure 2e**).

- All (fixed/variable) modifications used in the database search must be declared.
- “All Modifications” section loads modifications from the predefined file “PTMdecoder/public/modify.ini”. To add custom modifications not present in the default list, manually edit “modify.ini” following the syntax guidelines detailed in Appendix 3.

f) Algorithm parameters (**Figure 2f**)

- “Fragment Efficiency”. This parameter configures the algorithm used to account for fragmentation efficiency during MS/MS spectrum deconvolution. Three options are available: “Variable” (recommended, as PTMdecoder's novel implementation provides optimal handling of fragmentation efficiency in different classes), “Constant” (assumes fixed efficiency for different classes), and “Equal” (assigns uniform weights to all fragments).

- "Optimization Method". This parameter selects the regression algorithm for deconvolution calculations. Two options are available: OLS (Ordinary Least Squares) for standard linear regression without regularization, and Lasso (L_1-regularized) which incorporates L_1 penalty to automatically identifies and keeps only the important IMPs while filtering out weak or unreliable signals (particularly useful when working with complex samples or low-quality spectra). When Lasso is selected, the regularization intensity can be tuned via the Lasso Lambda parameter below.
- "Lasso Lambda". L1 regularization factor (visible only when Lasso is selected in "Optimization Method").
- "Abundance Cutoff Factor". Specifies the intensity threshold for excluding low-abundance peaks during XIC deconvolution. The final cutoff is calculated by multiplying this factor by the abundance of most abundant IMP in each XIC peak.

g) Batch processing configuration (**Figure 2g**). This section allows users to configure and run multiple independent analysis tasks simultaneously. Each task uses the parameter settings defined in the previous sections.

- "Spectra Directory": Folder path for MS/MS spectra files.
- "Peptide Spectra List File Path": Path to the custom "peptide spectra list" file (contains search engine identification results).
- "Output Directory": Output folder path for saving results.
- Three buttons: 'Add Row' (creates a new task), 'Clear Rows' (removes all tasks), and 'Delete' (removes the corresponding task).

iii. Output Files

Filename	Description
report_msms.txt	MS/MS-level deconvolution results
report_peptide_all.txt	Peptide-level quantification
report_spectra_may_FP.txt	Lists MS/MS spectra where site-discriminating fragment ions exhibit multicollinearity

(File contents explained in the next chapter.)

2 Site Level Analysis

This module performs quantification at the site level. It aggregates peptidoform-level quantifications for user-specified modification sites of interest (**Figure 3**).

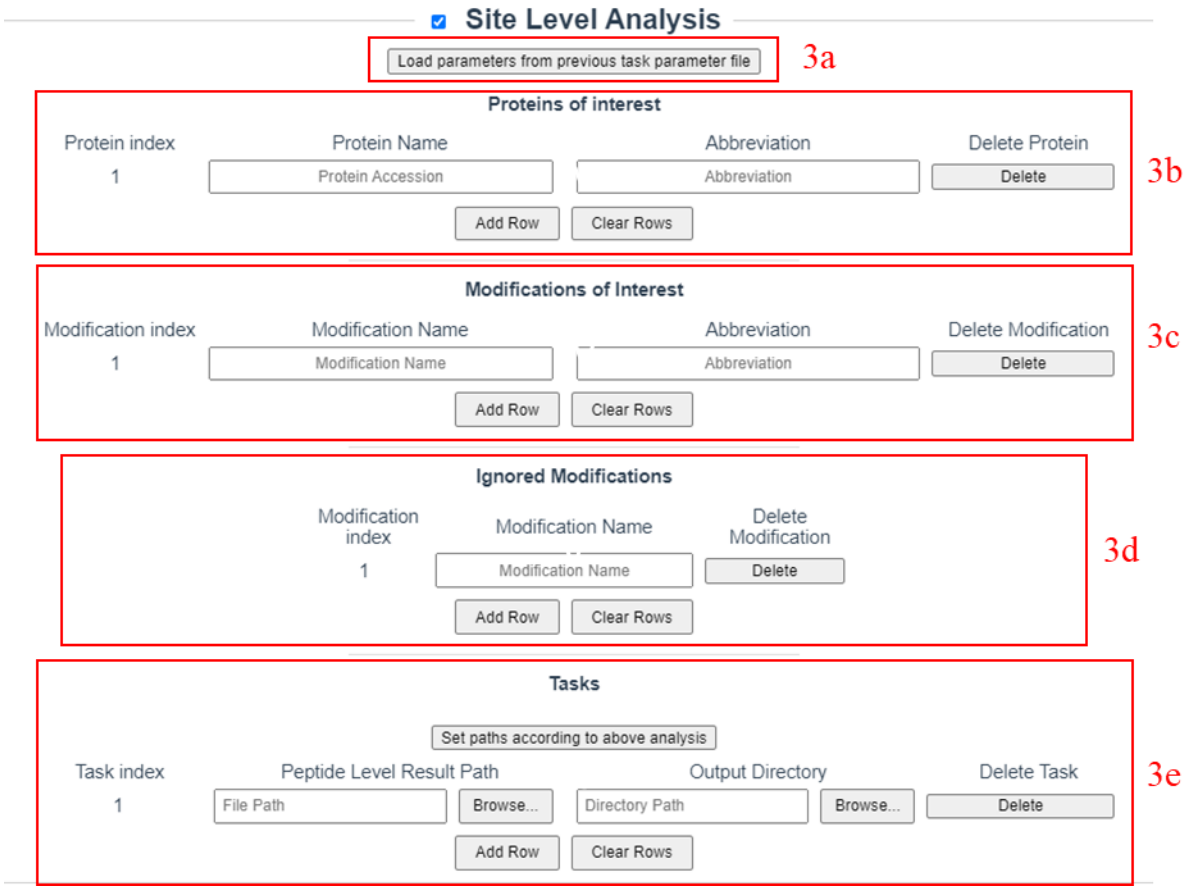

Figure 3 Site-level analysis

i. Input Requirements

Prerequisite: This module requires results from the MS/MS & Peptidoform-Level Analysis. Ensure either:

- The previous module has been executed, or
- The current run includes the MS/MS & Peptidoform-Level Analysis

Required input file:

- peptidoform-level deconvolution results (result_peptide_all.txt)

ii. Parameters

a) “Load parameters from previous task parameter file” button (**Figure 3a**). Imports saved parameter files (.param) to automatically configure all parameters in this module, except for the “Tasks” section.

b) “Proteins of interest” (**Figure 3b**). Protein(s) of interest and corresponding abbreviation(s)

c) “Modifications of Interest” (**Figure 3c**). Modification(s) of interest and corresponding abbreviation(s)

d) “Ignored Modifications” (**Figure 3d**). Modifications to be ignored (semicolon-separated) in the site level summary. For example, “Label_13C(6)” or “Label_13C(6)15N(4)”. Used to summarize the same modification in isotopically labeled samples. At this point, isotopic labeling needs to be separated from the modification of interest and ignored in the peptidoform strings. For example, “Label:13C(6)+Methyl” should be separated into “Label:13C(6)” and “Methyl”. Then “Label:13C(6)” are specified as fixed modifications and “Methyl” are specified as variable modifications. The modification of interest is “Methyl” and the ignored modification is “Label:13C(6)”.

e) Batch processing configuration (**Figure 3e**). This section allows users to configure and run multiple independent analysis tasks simultaneously. Each task uses the parameter settings defined in the previous sections.

- “Peptide Level Result Path”. Path to report_peptide_all.txt from MS/MS & Peptidoform-Level Analysis
- “Output Directory”. Output folder path for saving results.
- Three buttons: 'Add Row' (creates a new task), 'Clear Rows' (removes all tasks), and 'Delete' (removes the corresponding task).

iii. Output Files

Filename	Description
report_site.txt	Site-level quantification results
report_peptide_uninterested.txt	Peptides excluded from site-level aggregation

(File contents explained in the next chapter.)

2.2 Single run re-quantification

This workflow implements a two-stage analytical pipeline centered on verified peptidoform-level result. The sequential stages provide:

- **Peptidoform level re-quantification**: Deconvolutes the selected XIC peaks from aligned results to re-quantify the component IMPs in the samples.
- **Site-Level Analysis**: Computes site-level quantifications by aggregating abundances of all peptidoforms containing user-specified modification sites.

1 Peptidoform level re-quantification

This module performs re-quantification of peptidoforms using user-provided checked XIC peak result (e.g., aligned XIC peaks across replicates) (**Figure 4**).

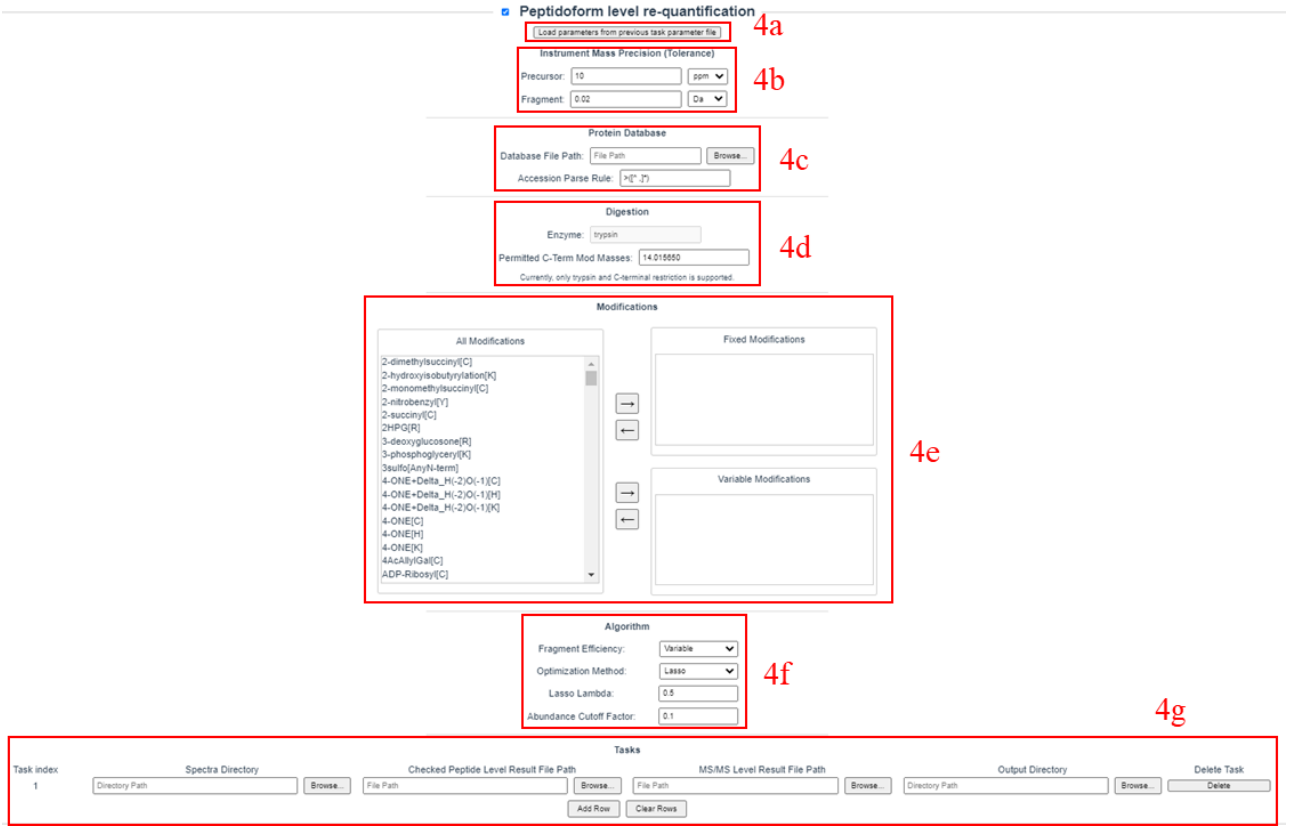

Figure 4 Peptidoform level re-quantification

i. Input Requirements

Prerequisite: The standard “MS/MS & Peptidoform-Level Analysis” must be completed first to generate required input files.

Required inputs:

a) Spectra directory path (folder containing all MS/MS spectrum files used in the corresponding search engine results)

b) Checked Peptide Level Result File (report_peptide_all.txt from PTMdecoder's previous peptide-level that has undergone XIC verification)

c) MS/MS Level Result File (report_msms.txt from PTMdecoder's previous MS/MS-level analysis)

ii. Parameters

a) "Load parameters from previous task parameter file" button (**Figure 4a**). Imports saved parameter files (.param) to automatically configure all parameters, except for some file paths in the “Tasks” section.

b) Shared Parameter Sections. All parameters from these sections function identically to the MS/MS & Peptidoform-Level Analysis:

- Instrument Mass Precision (Tolerance) (**Figure 4b**)
- Protein Database (**Figure 4c**)
- Enzymatic Constraints (Digestion) (**Figure 4d**)
- Modifications (**Figure 4e**)
- Algorithm Parameters (**Figure 4f**)

c) Batch Processing Configuration (**Figure 4g**). This section allows users to configure and run multiple independent analysis tasks simultaneously. Each task uses the parameter settings defined in the previous sections.

- “Spectra Directory”: Folder path for MS/MS spectra files.
- “Checked Peptide Level Result File Path”: modified report_peptide_all.txt from PTMdecoder's previous peptide-level that has undergone XIC verification.
- “MS/MS Level Result File Path”: Path to report_msms.txt from standard MS/MS &

Peptidoform-Level Analysis.

- “Output Directory”: Output folder path.
- Three buttons: 'Add Row' (creates a new task), 'Clear Rows' (removes all tasks), and 'Delete' (removes the corresponding task).

iii. Output Files

Filename	Description
report_peptide_all_requant.txt	Re-quantified peptide-level results

(File contents explained in the next chapter.)

2 Site Level Analysis

This module performs quantification at the site level as described previously, but uses peptidoform-level results generated from the re-quantification process as input. Specify the “report_peptide_all_requant.txt” file in “Peptide Level Result Path”. All other parameters can be configured identically to the standard site-level analysis. (**Figure 3**).

Output files:

Filename	Description
report_site_requant.txt	Site-level re-quantification results
report_peptide_uninterested_requant.txt	Peptides excluded from site-level re-quantification aggregation

(File contents explained in the next chapter.)

2.3 Pairwise comparison

This workflow implements a two-stage analytical pipeline centered on pairwise comparison of site-level result file. The sequential stages provide:

- **Pairwise Ratio of Changes Calculation**: Computes of pairwise ratio of changes based on site-level analysis result files, providing comparisons at both peptidoform and site levels, organized by modification sites.
- **Aggregating Ratios of Changes into a Consolidated File**: Combines ratios of changes from

multiple pairwise comparisons into a single output file.

1 Pairwise Ratio of Changes Calculation

This module calculates pairwise ratio of changes based on site-level analysis result files (**Figure 5**).

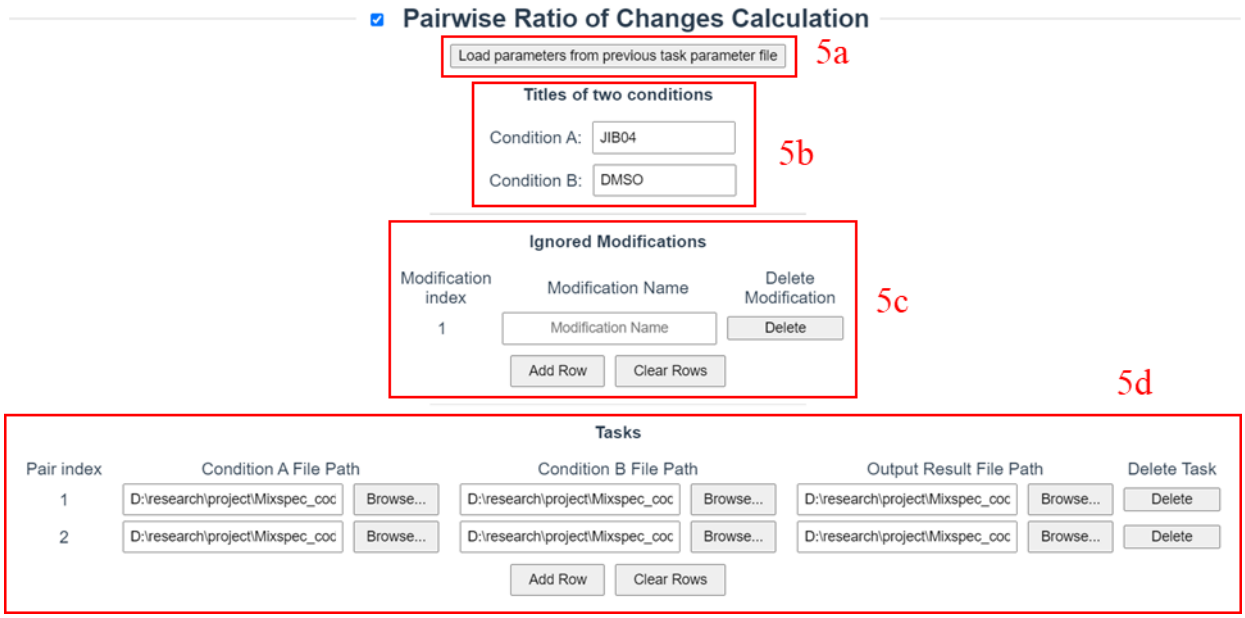

Figure 5 Pairwise ratio of changes calculation

i. Input Requirements

a) This module requires results from the Site-Level Analysis.

b) Required input file:

- Site-level deconvolution results (result_site.txt or result_site_requant.txt)

ii. Parameters

a) “Load parameters from previous task parameter file” button (**Figure 5a**). Imports saved parameter files (.param) to automatically configure all parameters.

b) “Titles of two conditions” (**Figure 5b**). Set names for the two conditions to be compared. These will appear as column headers in the result file.

c) “Ignored Modifications” (**Figure 5c**). Specify modifications to exclude during pairwise comparison (semicolon-separated list). Uses the same logic as the "Ignored Modifications"

parameter in the site-level analysis. For example, “Label_13C(6)” or “Label_13C(6)15N(4)”.

d) Batch Processing Configuration (**Figure 5d**). This section allows users to configure and run multiple independent analysis tasks simultaneously. Each task uses the parameter settings defined in the previous sections.

- “Condition A File Path”: Specify the site-level analysis result file path for Condition A (e.g., result_site.txt files of treatment group samples). The file should follow the standard output structure from site-level analysis.
- “Condition B File Path”: Specify the site-level analysis result file path for Condition B (e.g., result_site.txt files of control group samples). Must maintain identical data structure to Condition A for valid comparison.
- “Output Result File Path”: Path for pairwise comparison results. The generated file will follow the structure described below.
- The output file provides site-centric pairwise comparisons, reporting ratio of changes in quantification for each modification sites along with all contributing peptidoform-level quantification in pairs. This structured format enables manual calculation of peptidoform ratios in Excel, where users may further normalize the ratios using quantification ratio of normalization peptides.

iii. Output files.

The analysis generates a user-named structured file organized by modification sites. Each entry begins with a site identifier, followed by all corresponding peptidoforms and their quantitative comparisons between Condition A and B. For every site, the file reports both the measured quantifications and their calculated ratios. The file is easy manually processed in Excel to calculate normalized ration using the quantification ratio of normalization peptides.

2 Aggregating Ratios of Changes into a Consolidated File

This module combines ratios of changes from multiple pairwise comparisons into a single output file. (**Figure 6**)

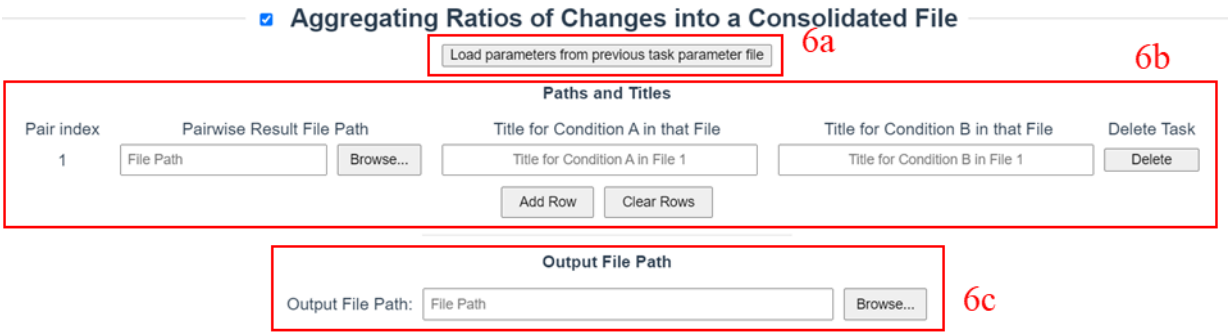

Figure 6 Aggregating Ratios of Changes into a Consolidated File

i. Input Requirements

This module requires results from the "Pairwise Ratio of Changes Calculation" module.

ii. Parameters

a) "Load parameters from previous task parameter file" button (**Figure 6Figure 5a**). Imports saved parameter files (.param) to automatically configure all parameters.

b) "Paths and Titles" (**Figure 6Figure 5b**). Manage the source files and their corresponding condition titles for aggregation.

- "Pairwise Result File Path". Specify a pairwise comparison result file that will be incorporated into the final output.
- "Title for Condition A in that File". Specify how the first condition from this pairwise comparison will be labeled in the output's column headers.
- "Title for Condition B in that File". Specify the column header for the second condition.
- Three buttons: 'Add Row' (creates a new entry), 'Clear Rows' (removes all entries), and 'Delete' (removes the corresponding entry).

c) "Output File Path". Specify the output destination for the aggregated result. Click the "Browse..." button to select the desired location and filename for the final output file.

iii. Output files

The analysis generates a user-named structured file organized by modification sites. Each entry begins with a site identifier, followed by all corresponding peptidoforms and their quantitative comparisons across user-specified entries. For every site, the file reports both the measured quantifications and their calculated ratios. The file is easy manually processed in Excel to calculate

normalized ration using the quantification ratio of normalization peptides.

3 Running PTMdecoder from the Command Line

The PTMdecoder Command Line Interface (CLI) provides advanced users with scriptable access to IMP analysis capabilities.

Prerequisite: Ensure MATLAB Runtime R2022a (9.12) is installed before proceeding

3.1 Parameter File Structure

Configuration files (.param) use simple key-value syntax:

```
# Comment lines start with #
parameter_name = parameter_value


# Example:
msms_peptide_level_on = 1
output_dir_path = ./results/
```

3.2 Module Control Switches

There are some core switches to control workflow execution:

Module Switch	Values	Description	Notes
msms_peptide_level_on	0/1	Activates MS/MS and peptidoform analysis	Mutually exclusive with peptide_requant_on
peptide_requant_on	0/1	Activates peptide re-quantification	Mutually exclusive with msms_peptide_level_on
site_level_on	0/1	Activates site-level analysis	
merge_to_pair_level_on	0/1	Activates pairwise comparison	
merge_pairs_level_on	0/1	Activates comparison integration	

Parameter files can only include enabled modules and their corresponding parameters. Unused modules can be omitted from the configuration file.

3.3 Module-based parameters

This section details all module-specific parameters. Each module is activated by its corresponding switch parameter (<Module Switch> = 1). Only parameters for activated modules need to be included in the configuration file.

1 MS/MS and Peptidoform Level Analysis

Activation: “msms_peptide_level_on = 1”

Parameter	Description	Example Value
mod_file_path	Modification config file	modify.ini
fixed_mod	Fixed mods (semicolon-separated)	Carbamidomethyl (C)
variable_mod	Variable mods (semicolon-separated)	Acetyl (Protein N-term); Oxidation (M)
spec_dir_path	Raw spectra directory	.\spectra
ms1_tolerance_value	MS1 tolerance value	10
ms1_tolerance_type	MS1 tolerance unit (ppm/Da)	ppm
ms2_tolerance	MS2 tolerance value	0.05
alpha	Spectral noise-filtering threshold	0.01
fasta_file_path	Protein database file (*.fasta)	.\uniprot.fasta
parse_reg_exp	Protein name extraction pattern	>([^ ,]*)
pep_spec_file_path	Peptide-spectrum mapping file	.\pepSpecFile.txt
model	Model Selection (1/2/3 for Variable/Constant/Equal model)	1
method	Solving method (1 for OLS, 2 for Lasso)	2
lambda	Lasso lambda (lasso only)	0.5
result_filter_threshold	Peptidoform abundance Cutoff Factor	0.1
enzyme_name	Enzyme name (only trypsin is allowed for now)	trypsin
enzyme_limit_C_term_possible_mod	C-terminal modifications (semicolon-separated masses)	14.015650;42.010565
output_dir_path	Output directory	.\results

min_MSMS_num	Sets the minimum number of identified MS/MS spectra required for a precursor to be included in quantification. (Optional)	1

The “modify.ini” file containing PTM definations is located in the public folder of the software installation directory. For adding or modifying the PTMs, refer to Appendix 3. Both “fixed_mod” and “variable_mod” must strictly match the modification names specified in “modify.ini”.

2 Peptide Re-quantification Module

Activation: “peptide_requant_on = 1”

Parameter	Description	Example Value
mod_file_path	Modification config file	modify.ini
fixed_mod	Fixed mods (semicolon-separated)	Carbamidomethyl (C)
variable_mod	Variable mods (semicolon-separated)	Acetyl (Protein N-term); Oxidation (M)
spec_dir_path	Raw spectra directory	.\spectra
ms1_tolerance_value	MS1 tolerance value	10
ms1_tolerance_type	MS1 tolerance unit (ppm/Da)	ppm
ms2_tolerance	MS2 tolerance value	0.05
alpha	Spectral noise-filtering threshold	0.01
fasta_file_path	Protein database file (*.fasta)	.\uniprot.fasta
parse_reg_exp	Protein name extraction pattern	>([^ ,]*)
model	Model Selection (1/2/3 for Variable/Constant/Equal model)	1
method	Solving method (1 for OLS, 2 for Lasso)	2
lambda	Lasso lambda(lasso only)	0.5
result_filter_threshold	Peptidoform abundance Cutoff Factor	0.1
enzyme_name	Enzyme name (only trypsin is allowed for now)	trypsin

enzyme_limit_C_term_possible_mod	C-terminal modifications (semicolon-separated masses)	14.015650;42.010565
output_dir_path	Output directory	.\results
min_MSMS_num	Sets the minimum number of identified MS/MS spectra required for a precursor to be included in quantification. (Optional)	1
checked_peptides_res_path	Validated peptide quant file	.\report_peptides_checked.txt
msms_res_path	MSMS quant results	.\report_msms.txt

3 Site Level Module

Activation: “site_level_on = 1”

Parameter	Description	Example Value
protein_name_abbr_num	Number of proteins of interest	3
protein_name_abbr_X	Name & abbreviation of proteins of interest (X=1,2,...)	sp|P0C0S8|H2A1_HUMAN > H2A
mod_name_abbr_num	Number of modifications of interest	2
mod_name_abbr_X	Name & abbreviation of modifications of interest (X=1,2,...)	Acetyl > ac
pep_level_file_path	Peptide-level results	.\report_peptide_all.txt
output_intere_path	Site-level output	.\report_site.txt
output_unintere_path	Uninteresting peptides output in site-level analysis	.\report_peptide_uninterested.txt
ignore_strings_site_level	Modifications to ignore in site-level analysis (semicolon-separated)	"{Label:13C(6)}"

Syntax of “name and abbreviation” parts:

```
# Protein abbreviations
protein_name_abbr_num = 2
protein_name_abbr_1 = Uniprot_Accession > Abbreviation
protein_name_abbr_2 = sp|P62805|H4_HUMAN > H4


# Modification abbreviations
mod_name_abbr_num = 2
```

mod_name_abbr_1 = Modification_Name > Abbreviation

mod_name_abbr_2 = Methyl > me

Note: The value in *_name_abbr_num (e.g. protein_name_abbr_num, mod_name_abbr_num) **MUST exactly equal** the number of subsequent name-to-abbreviation entries.

4 Pairwise Comparison Module

Activation: “merge_to_pair_level_on = 1”

Parameter	Description	Example Value
left_right_out_num	Number of comparison groups	2
left_right_out_X	Input/output paths (X=1,2,...)	cond1.txt | cond2.txt > out.txt
left_name	Condition A name	Control
right_name	Condition B name	Treated
ignore_strings_pair_level	Modifications to ignore in pairwise comparison analysis	"{Label:13C(6)}"

Syntax of “left_right_out” parts:

File names of the paired files to be compared and their result output file names

left_right_out_num = 2

left_right_out_1 = [Left input] | [Right input] > [Output path]

left_right_out_2 = .\res\cond1.txt | .\res\cond2.txt > .\res\output.txt

5 Comparison Integration Module

Activation: “merge_pairs_level_on = 1”

Parameter	Description	Example Value
pair_num	Number of comparison results	3
pair_X	Comparison result file (X=1,2,...)	.\comparisons\pair_1.txt
left_right_name_X	Group names (X=1,2,...)	Control (rep1) | Treated (rep1)
final_output_path	Integrated results output	.\final_merged.txt

Example:

Aggregate Ratios of Changes into one File

merge_pairs_level_on = 1

```
# Pairs to be aggregated
pair_num = 3
pair_1 = .\pair_1.txt
left_right_name_1 = Treat(rep 1) | Control(rep 1)
pair_2 = .\pair_2.txt
left_right_name_2 = Treat(rep 2) | Control(rep 2)
pair_3 = .\pair_3.txt
left_right_name_3 = Treat(rep 3) | Control(rep 3)

# Path of the result file after aggregation
final_output_path = .\joint.txt
```

3.4 Execution Instructions

Save the parameter configure file and open command line prompt, run PTMdecoder as follow:

```
PTMdecoder_core.exe [parameter_file_1.param] [parameter_file_2.param] ...
```

PTMdecoder can process multiple parameter files sequentially in one command, and each parameter file runs in isolation.

4 Interpretation of Output

4.1 report_msms.txt

This file presents MS/MS identification results organized by peptides. The structure is as follows:

i. Peptide Entry (“P”-line)

a) Starts with “P” followed by the unmodified peptide sequence (e.g., PEPTIDEK).

b) Each “P” line acts as a header for its corresponding spectrum entries.

ii. Spectrum Entry (“S”-line)

a) Starts with “S” followed by:

- Dataset name: The dataset filename (e.g., Example_HCDFT.mgf).

- Spectrum name: The spectrum name or identifier (e.g., Example_HCDFT.123.123.2.0.dta).

b) Each “S” line is followed by one or more IMPs identified in that spectrum.

iii. IMPs and their abundances

a) Each IMP is listed with:

- Sequence: Composed of 20 standard amino acids (uppercase letters), underscores (“_”, denoting the N- or C-terminus of the peptide) and curly braces (“{}”, denoting modifications). For example, “_{propionyl}PEPT{phospho}IDEK{methyl}R_” represents the peptide “PEPTIDEKR” with:
  - N-terminal propionylation (“_{propionyl}”),
  - Phosphorylation at T (“T{phospho}”),
  - Methylation at K (“K{methyl}”).
- Relative Abundance: A number between 0-1 indicating the IMP’s relative abundance in the spectrum.

4.2 report_peptide_all.txt

This file presents peptide-level quantification results organized by the starting positions of peptides within protein sequences. The structure is as follows:

i. Protein Start Position Entry

a) One or more “<ProteinName>,<StartPosition>” pairs separated by “;”.

- Each pair contains protein names (e.g., sp|P84243|H33_HUMAN), a comma, then the start positions (e.g., 28). Example: “sp|P84243|H33_HUMAN,28;tr|B4DEB1|B4DEB1_HUMAN,28;”

b) Acts as a header for all IMPs mapped to these protein positions.

ii. IMP Entry (“*” line)

a) Starts with “*” followed by:

- IMP Name: Modified peptide sequence (e.g., “_K{Methyl}SAPSTGGVK{Methyl}K_”).
- Charge: Precursor charge state (e.g., “+2”).
- MS File: Dataset filename (e.g., “Example_HCDFT.mgf”).

- Monoisotopic Mass: Theoretical mass (e.g., 1234.5678).
- Mass Tolerance Bounds: Lower (1234.5608) and upper (1234.5748) bounds.
- XIC Area: Quantification value (e.g., 1500000.0).

Example: "* _K{Methyl}SAPSTGGVK{Methyl}K_ +2 MCF7_DMSO_1_HISTONE_0723_HCDFT.mgf 544.3271 544.321687 544.332574 5306206.625033"

b) Each "*" line is followed by one or more XIC Peak entries for this IMP.

iii. XIC Peak Entries

a) Each line starts with "@" followed by:

- Start RT: XIC peak start time, in minutes.
- End RT: XIC peak end time, in minutes.
- Selection Flag:
  - 0: The IMP was detected in this peak but not selected for quantification (low confidence or interference).
  - 1: IMP automatically selected for quantification.
  - 2 (user override): Manually chosen for quantification (overrides 1 if present).

Example: "@ 0.011456 0.598165 0.500000 1"

4.3 report_spectra_may_FP.txt

This file lists MS/MS spectra potentially containing false positive identifications resulting from multicollinearity among site-discriminating fragment ions. If there are many spectra shown in this file, it is strongly recommended to specify "Lasso" in "Method Selection" when conducting MS/MS and Peptidoform Level analysis. Each line follows the format "[Dataset Filename]\t[Spectrum Title]", where "[Dataset Filename]" refers to the acquisition .mgf file (e.g., Example_HCD.mgf) and "[Spectrum Title]" specifies the spectrum title (e.g., Example_HCD.1234.1234.2.dta).

4.4 report_site.txt / report_site_requant.txt

This file presents site-level quantification results organized by protein-site combinations. The structure is as follows:

i. Protein-Site Entry Line

- Format: “[SiteIdentifier]\t[TotalQuantification]”
  - [SiteIdentifier]: The site identifier follows a standardized naming convention that integrates three key components: the protein abbreviation, the modified residue position, and the modification abbreviation. For example, the identifier H2AK95me1 represents histone H2A lysine 95 monomethylation. Here, “H2A” denotes the protein (histone H2A), “K95” specifies the modified lysine residue at position 95 within the protein sequence, and “me1” indicates the type of modification (monomethylation).
  - [TotalQuantification]: Aggregated quantification value across all peptidoforms (e.g., 46784714644.35).
- Example: “H3K27me1 46784714644.351982”
- Followed by corresponding IMP entries

ii. IMP Entry Line (“*” line)

- Format: “*\t[IMP_Name]\t[Charge]\t[Dataset_File]\t[MonoMass]\t[MassLower]\t[MassUpper]\t[XIC_Area]”
  - IMP_Name: IMP sequence (e.g., _K{Methyl}SAPSTGGVK_).
  - Charge: Precursor charge state (e.g., +2).
  - Dataset_File: Source dataset filename (e.g., MCF7_HISTONE_0723.mgf).
  - MonoMass: Theoretical monoisotopic mass (e.g., 544.3271).
  - MassLower/Upper: Tolerance bounds, ±MS2 tolerance.
  - XIC_Area: Quantification value (e.g., 5306206.6).
- Example: “*\t_K{Methyl}SAPSTGGVK{Methyl}K_\t+2\tMCF7_HISTONE_0723.mgf\t544.3271\t544.3217\t544.3326\t5306206.6”

4.5 report_peptide_uninterested.txt / report_peptide_uninterested_requant.txt

This file contains peptide-level identification results excluded from site-level aggregation, sharing identical structure with report_peptide_all.txt except for the XIC peak entries. The structure is

as follows:

i. Protein Start Position Entry

a) One or more “<ProteinName>,<StartPosition>” pairs separated by “;”.

- Each pair contains protein names (e.g., sp|P84243|H33_HUMAN), a comma, then the start positions (e.g., 28). Example: “sp|P84243|H33_HUMAN,28;tr|B4DEB1|B4DEB1_HUMAN,28;”

b) Acts as a header for all IMPs mapped to these protein positions.

ii. IMP Entry (“*” line)

a) Starts with “*” followed by:

- IMP Name: Modified peptide sequence (e.g., “_K{Methyl}SAPSTGGVK{Methyl}K_”).
- Charge: Precursor charge state (e.g., “+2”).
- MS File: Dataset filename (e.g., “Example_HCDFT.mgf”).
- Monoisotopic Mass: Theoretical mass (e.g., 1234.5678).
- Mass Tolerance Bounds: Lower (1234.5608) and upper (1234.5748) bounds.
- XIC Area: Quantification value (e.g., 1500000.0).

Example: “* _K{Methyl}SAPSTGGVK{Methyl}K_ +2 MCF7_DMSO_1_HISTONE_0723_HCDFT.mgf 544.3271 544.321687 544.332574 5306206.625033”

4.6 report_peptide_all_requant.txt

This file contains re-quantified peptide-level results with strictly one XIC peak per IMP, following the same structure as report_peptide_all.txt but with the following constraints:

i. Protein Start Position Entry

a) One or more “<ProteinName>,<StartPosition>” pairs separated by “;”.

- Each pair contains protein names (e.g., sp|P84243|H33_HUMAN), a comma, then the start positions (e.g., 28). Example: “sp|P84243|H33_HUMAN,28;tr|B4DEB1|B4DEB1_HUMAN,28;”

b) Acts as a header for all IMPs mapped to these protein positions.

ii. IMP Entry (“*” line)

a) Starts with "*" followed by:

- IMP Name: Modified peptide sequence (e.g., "_K{Methyl}SAPSTGGVK{Methyl}K_").
- Charge: Precursor charge state (e.g., "+2").
- MS File: Dataset filename (e.g., "Example_HCDFT.mgf").
- Monoisotopic Mass: Theoretical mass (e.g., 1234.5678).
- Mass Tolerance Bounds: Lower (1234.5608) and upper (1234.5748) bounds.
- XIC Area: Quantification value (e.g., 1500000.0).

Example: "* _K{Methyl}SAPSTGGVK{Methyl}K_ +2 MCF7_DMSO_1_HISTONE_0723_HCDFT.mgf 544.3271 544.321687 544.332574 5306206.625033"

b) Each "*" line is followed by one or more XIC Peak entries for this IMP.

iii. XIC Peak Entries

a) Each line starts with "@" followed by:

- Start RT: XIC peak start time, in minutes.
- End RT: XIC peak end time, in minutes.
- Selection Flag: Always 1 (automatically selected during re-quantification).

Example: "@ 0.011456 0.598165 0.500000 1"

4.7 Pairwise comparison results

This file provides site-level quantitative comparisons between two experimental conditions, organized by modification sites. The structure consists of a header line followed by alternating site entries and their corresponding peptidoform (IMP) entries, organized as follows:

i. Header Line

The first line defines the comparison context with three fixed columns followed by two user-defined condition names: "Sites\tSequence\tCharge\t[Condition1]\t[Condition2]", in which [Condition1]/[Condition2] are user-specified names for compared conditions (e.g., SAHA and DMSO).

ii. Site-IMP Entries

The pairwise comparison results organize quantifications hierarchically by modification sites and their associated IMPs. Each modification site entry begins with a fully specified line containing the

site identifier, followed by one or more lines listing corresponding peptidoform details.

The first line of a site entry explicitly states the modification site identifier in the “Sites” column (e.g., H2AK5ac), followed by the IMP sequence, charge state (Charge), and quantifications for the two experimental conditions. Subsequent lines corresponding to the same site omit the Sites identifier, leaving the column blank to indicate inheritance from the preceding entry.

This structure allows direct comparison of quantification changes across conditions at both site and peptidoform levels. For example, a single site identifier like H2AK5ac may encompass multiple peptidoforms with varying charge states, each displaying condition-specific intensity values.

4.8 Aggregation results

This file consolidates pairwise comparisons across multiple replicates into a unified structure, extending the original pairwise format with two key enhancements:

i. Each modification site entry concludes with a dedicated row marked by the keyword “SUM” in the Sequence column. This row does not represent a specific IMP sequence but instead provides the summed quantification for all IMPs associated with the site.

ii. Expanded Columns for Ratios and Normalization: This file extends the original pairwise format by adding two new columns to the existing structure, enabling both quantitative comparisons and normalization workflows:

- Site-Level Ratio: A calculated ratio (e.g., SAHA/DMSO) derived from the SUM row values.
- Normalized Ratio: An empty column reserved for user-defined normalization.

Appendix 1 Usage of pParse for preparing the ms1, ms2, and mgf files for PTMdecoder

This manual has been validated with pParse v2.2.1. While other versions may function, ensure critical parameters remain unchanged if alternative versions must be used.

PTMdecoder requires mass spectra data in ms1, ms2, and mgf format for IMP analysis. This requires proper configuration of the upstream software pParse through either of these methods

A. Parameter Configure File (.cfg) Method

1. File generation

A .cfg file can be generated by using

```
pParse.exe
```

in command line, or modified from a previous .cfg file.

2. Critical parameters

Ensure the .cfg file containing these two critical directives:

```
delete_msn = 0
        # 0 Keep the MS1/MS2 files [recommended];
output_mgf = 1
        # 0 Do Not output MGF;
```

3. Execution command:

```
pParse.exe task.cfg
```

Note: Replace task.cfg with your actual configuration filename. Ensure the file resides in the working directory or provide its full path (e.g., “pParse.exe C:\demo\task.cfg”).

B. Command-line Execution

Users can run pParse through command line by

```
pParse.exe -b 0 -m 1 -D D:\mydata\First.raw
```

or

```
pParse.exe -b 0 -m 1 -D D:\mydata\
```

The “-b 0/-m 1” flags enforce the equivalent parameter configure file settings (delete_msn=0 and output_mgf=1), which guarantee the simultaneous generation of ms1, ms2, and mgf files.

For more details about usage of more options, please refer to the official user guide of pParse or use just “pParse.exe” to print the guide information.

Appendix 2 Preparing Peptide Spectra List File from Mascot Results *.dat file

PTMdecoder requires search engine results structured in a customized format as inputs. This appendix focuses specifically on preparing these files. If you are using results from the Mascot search engine (*.dat files), we provide MATLAB scripts to help you convert your data into the required format.

Note: Do **NOT** enable Mascot's built-in decoy generation. If you want to use FDR control module in the conversion script, generate a target-decoy protein sequence database prior to searching.

Conversion Scripts

The scripts named "preprocess_*.m" serve as entry points that invoke modular functions such as "ReadDatResultFolder" and "FilterWithChemPrior", where the suffix (*) corresponds to a specific experiment detailed in the paper [unpublished]. These scripts can be accessed at: https://github.com/YanFuGroup/PTMdecoder/tree/main/FDR_control_generate_pep_spec_list.

Workflow Overview

While the data preparation steps may vary slightly between experiments, the core workflow generally involves the following key functions:

1. "ReadDatResultFolder": Reads Mascot .dat files from a specified folder.
2. "FilterWithChemPrior": Applies chemical knowledge-based filters (e.g., removes unlikely modifications, or C-terminal modifications that may interfere with enzyme digestion).
3. "JudgeGroup" and "ComputeFDR": These functions jointly implement FDR control.
4. Experiment-specific filtering: Additional steps may be included to select PSMs based on criteria relevant to the specific dataset (e.g., selecting specific types of modifications).

You can use these scripts as templates for processing your own Mascot results and adapt the file paths, filtering criteria, and any experiment-specific logic according to your needs. The script outputs the required input files for PTMdecoder (e.g., pepSpecFile.txt). Additional files such as

report_msms_top1.txt are generated for analytical comparisons but are not essential for standard operation.

Appendix 3 Add custom modifications by editing “modify.ini”

The “modify.ini” is the configuration file for PTMdecoder’s post-translational modifications (PTMs). Users can manually edit this file to add new modifications or update existing parameters. The file is in the following format:

```
[modify]
total=1622
...
name45=Acetyl[K]
Acetyl[K]=K NORMAL 42.010565 42.036700 0
...
name<ID>=<mod_name>
<mod_name>=<amino_acid> <term_spec> <mono_mass> <average_mass> <neutral_loss>
...
```

The “total” field at the top of the file is an integer that specifies the total number of modifications listed in the file (e.g., “total=1622”).

Each modification is defined by two lines: the first uses a “name<ID>=” format (where <ID> is a sequential number) to declare the modification name (e.g., “Acetyl[K]”). The second line provides the modification details: <amino_acid> is the modified residue (using upper-letter codes), <term_spec> is the terminus specificity (supporting “NORMAL”, “PEP_N”, “PEP_C”, “PRO_N”, and “PRO_C”), <mono_mass> and <average_mass> are the monoisotopic mass and average mass (in Daltons). The <neutral_loss> field specifies the neutral loss, which can always be set to 0 since PTMdecoder currently does not use neutral loss information.