

d2ome uses heavy water labeling and LC-MS data to estimate protein turnover rates. It runs via a graphical user interface (GUI) or from the command line.

1. Preparing input to d2ome

First inputs should be generated for the program. The program uses mzML and mzIdentML formatted files for spectral and database search results, respectively. Assume there are six time points of heavy water labeling. At each time point, it is assumed there are two experiments (replicates or fractions). There is no limit on the number of replicates/fractions per time point. Assume the following raw files will be used:

A0day_1.raw A0day_2.raw; A1day_1.raw A3day_2.raw; A3day_1.raw A3day_2.raw;
A8day_1.raw A8day_2.raw; A15day_1.raw A15day_2.raw; A21day_0.raw
A21day_1.raw

The first step is to generate mzML (mass spectra) files from raw files. MSConvert tool of Proteowizard converts the raw data into mzML formatted files. The Proteowizard version 3.0.10702 or later versions should be used. Earlier versions do not handle indexing of high mass accuracy spectra correctly. Parameter settings required for MSConvert are shown in **Figure 1 (centroid MS1 data)**. In particular, “**Write Index**” should be **checked**, and “**Use Zlib compression**” should be **unchecked**. Output format should be set to “mzML”. Note, new developments, in particular, implementations of mass isotopomer dynamics[1] to improve rate constant estimations are carried for centroid MS1 data type. Therefore, using **centroid data is preferable**.

The second step is to do database searches to identify peptide sequences and proteins from the MS/MS data in the mzML file. If you are using Mascot's "Mascot Daemon," you will need to specify that the input file format is in mzML. This is specified in the Mascot's parameter file. An example of a parameter file setup is shown in **Figure 2**. To export the database search results in mzIdentML format, using a setting similar to the one in **Figure 3 A – B** in the "Auto-export..." option of the Mascot Daemon. In the filtering options section of the "Auto-export.." "**Group Proteins**" should be **unchecked**, "**Require bold red**" should be **checked**, **Figure 3 A**. In the Protein Hit Information section, **Figure 3 B**, "**Description**" and "**Length in residues**", both should be set to "**check**". **Uncheck** the "**Include query level information**", **Figure 3 C**.

Currently, d2ome supports mass spectral data in either **centroid** or **profile modes** (in MS1) to quantify mass isotopomers. The **centroid mode** is **preferable**. In this mode, the processing is faster, and it uses new developments on mass isotopomer dynamics[1].

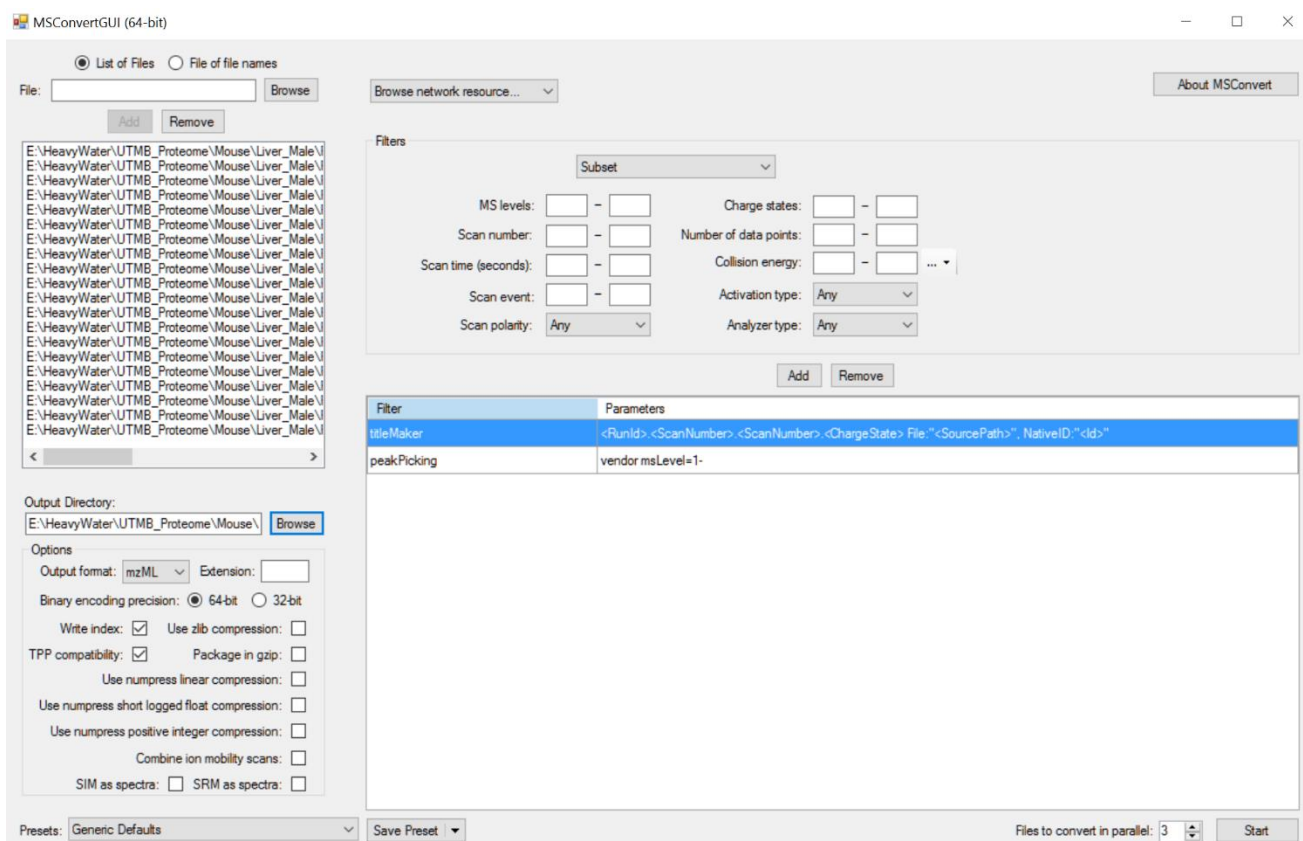


Figure 1. Input parameter set-up for MSConvert to generate mzML (**centroid MS1**) file from a (for example) raw file. Note that with the latest version of d2ome, generating MS1 scans in centroid mode is **preferred**.

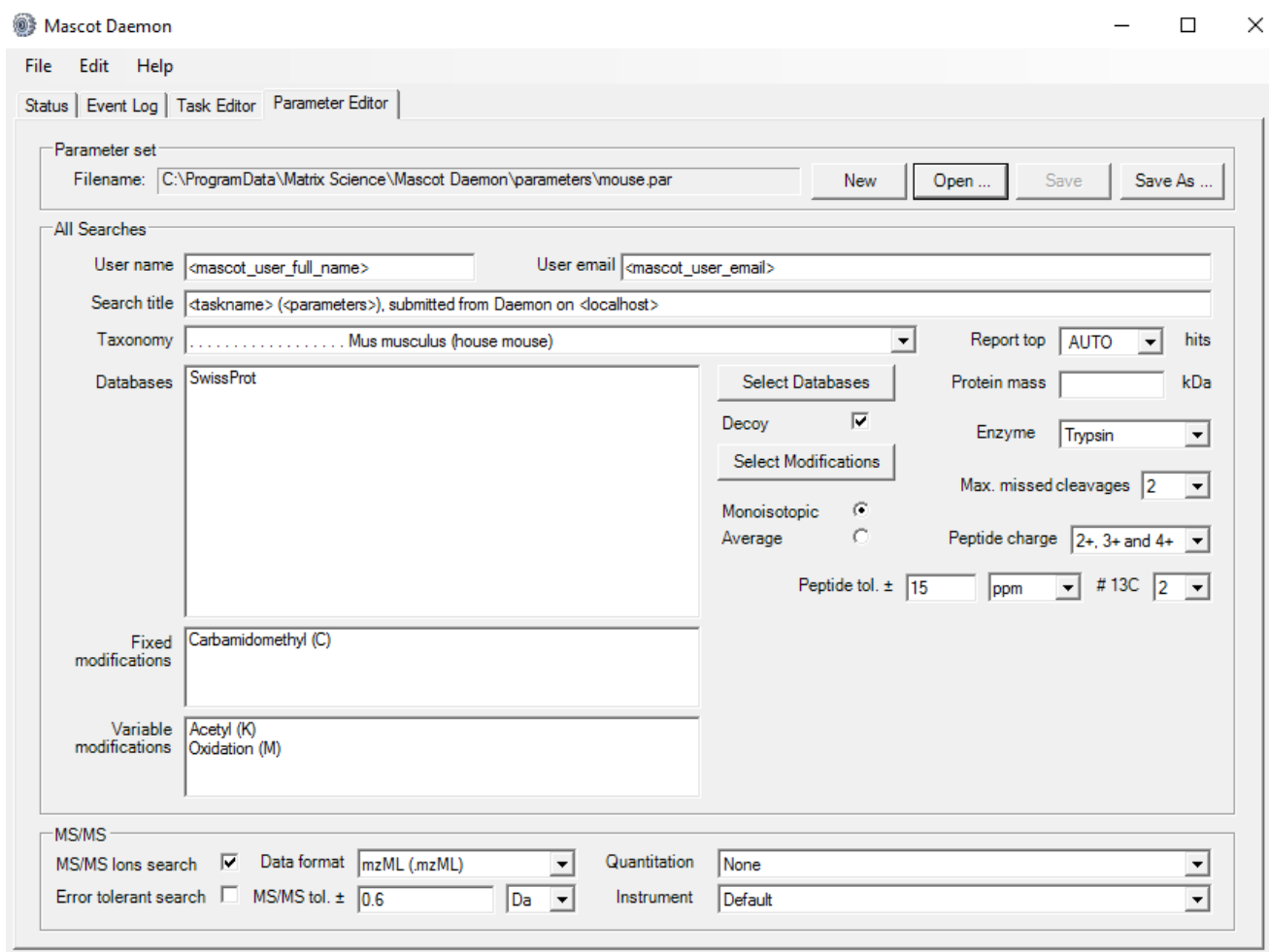


Figure 2. An example of a parameter setting using Mascot’s “Mascot Daemon” interface.

It is important for d2ome to set the input file format to “mzML”.

Choose first

☐ PMF ☒ MS/MS

Choose second

mzIdentML

Format

Significance threshold p<	0.05	<input type="radio"/> Identity <input checked="" type="radio"/> Homology
Target FDR (percentage)	5	
Ions score cut-off	0	
Max. number of hits	AUTO	
Protein scoring	<input type="radio"/> Standard <input checked="" type="radio"/> MudPIT	
Include same-set proteins	<input type="checkbox"/>	
Include sub-set proteins	0	
Group proteins into families	<input type="checkbox"/>	
Require bold red	<input checked="" type="checkbox"/>	
Unigene index	None	
Use Percolator	<input type="checkbox"/>	
Preferred taxonomy	All entries	

Include search information ☐

Figure 3 A. The filtering section of the Mascot Daemon’s “Auto-export...” options. “Group Proteins” should be unchecked, “Require bold red” should be checked.

Include protein hit information ☒

Protein hit information

Score ☐

Significance threshold ☐

Expectation value ☐

Description ☒

Mass (Da) ☐

Number of queries matched ☐

Percent coverage ☐

Length in residues ☒

pI ☐

Taxonomy ☒

Taxonomy ID ☒

Protein sequence ☐

emPAI ☐

Protein quantitation ☐

Figure 3 B. Protein Hit Information section of Mascot Daemon's "Auto-export..." options. The shown are the settings that are required for d2ome. The "Description" and "Length in residues", both should be checked.

Include peptide match information ☒

Peptide match information

Experimental Mr (Da) ☐

Experimental charge ☐

Calculated Mr (Da) ☐

Mass error (Da) ☐

Start ☐

End ☐

Number of missed cleavages ☐

Score ☐

Homology threshold ☐

Identity threshold ☐

Expectation value ☐

Sequence ☐

Frame number ☐

Variable Modifications ☐

Number of fragment ion matches ☐

Query title ☐

Peptide quantitation ☐

Unassigned queries ☐

Show duplicate peptides ☒

Include query level information ☐

Query level information

Query title ☐

seq(), comp(), tag(), etc. ☐

Query level search parameters ☐

MS/MS Peak lists ☐

Matched Fragment Ions ☐

Export data for all queries ☐

OK Cancel

Figure 3 C. Uncheck the “Include query level information”.

2. Running d2ome using GUI.

Download all binaries into a single folder. These files should be as shown in **Figure 4**.

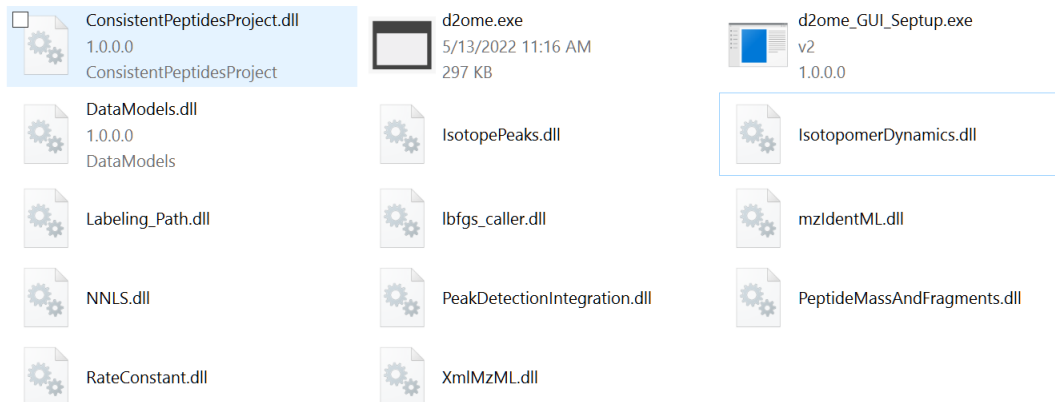


Figure 4. The binaries to run d2ome from either command line or using a GUI.

The GUI and a Visualization Tool are started by d2ome_SetUp_GUI.exe. The application form is shown in **Figure 5 A**.

The GUI automates filling of the mzML and mzid files from a folder. mzML and mzid file names should match, e.g., SomeFile.mzML, SomeFile.mzid. It is also possible to enter the files manually as in the previous GUI version. The autofill mode starts with “Browse” button. A user can copy and paste a folder path directly into the box. The GUI will sort the files into matching pairs (mzML and mzid). A Tab Controller like the one shown in **Figure 5 B** should appear. The user will fill the labeling duration (Time) and body water enrichment (BWE) cells for each experiment. Shown in **Figure 5 B** is an analysis consisting of 18 LC-MS experiments. The labeling time points are: (0, 0, 1, 1, 2, 2, 3, 3, 4, 4, 5, 5, 6, 6, 14, 14, 21, 21). The user will specify the labeling time units (Days or Hours). There are nine labeling timepoints in this example. For each labeling timepoint, there are two biological replicates. The corresponding total body water enrichments are (0, 0.0, 0.0304, 0.0235, 0.0325, 0.0322, 0.0309, 0.0281, 0.0259, 0.0257, 0.0359, 0.0287, 0.0359, 0.0265, etc).

Alternatively, the user can use the Manual Input button. In this mode, each of the mzML and mzid files and the corresponding labeling duration time and body water

enrichment are entered separately. Click on the “Browse” button for the mzML file to find an mzML file. Enter the labeling duration (Time box) and body water enrichment (BWE box), and press the “Add” button to add the data to the list.

The “Sort” button will sort the input data and order it in a sequence with increasing labeling duration.

Users can use the “Clear All” button to remove all entered files. The “Delete” button will remove selected files only.

“BWE” designates the body water enrichment with D₂O that corresponds to the specific labeling experiment. For example, if the body water enrichment is 7%, enter 0.07 into the box under “BWE”. “Rate Constant” method currently allows one options: one-parameter (determine the degradation rate constant).

Peptide consistency (Figure 5 B) is four. It means that only peptides that have been identified (passed the FDR threshold) and quantified in at least four different timepoints of labeling will be used in the estimation of rate constants for proteins.

Peptide score (Figure 5 B) is the threshold peptide score (Mascot Ion Score) in peptide-spectrum matches. Currently, the software uses results in the mzid format generated by Mascot. The threshold score is the ion score of Mascot. The program will check that the peptides passed the FDR threshold used in Mascot (using a reversed sequence database).

The mass accuracy to be used in the peak detection.

The GUI creates the *files.txt* and *quant.state* files (described below), input and parameter files, respectively.

To select the output directory click on the “Browse” button at the bottom of the GUI. One can also copy and paste the folder path to the box next to the button. After the data have been entered, the output directory has been chosen; the quantification is started by the “Start” button, which is next to the Output Directory box, **Figure 5 A**.

Note that the software expects that there are **NO white spaces** in the folder names.

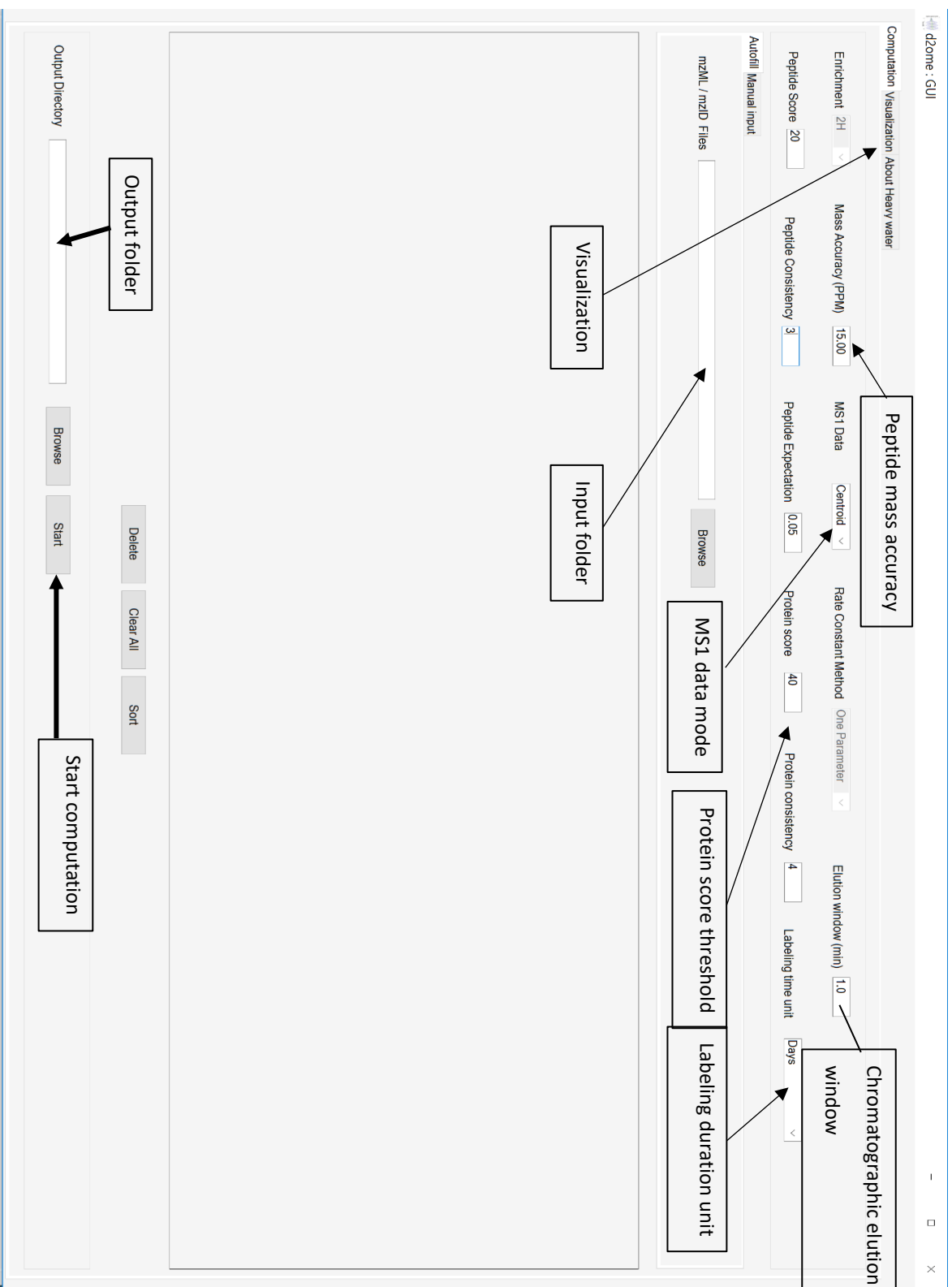


Figure 5 A. d2ome GUI before data initialization.

Computation Visualization About Heavy water

Enrichment 2H

Mass Accuracy (PPM) 20.0

MS1 Data

Centroid

Rate Constant Method

One Parameter

Elution window (min) 1.0

Peptide Score 20

Peptide Consistency 4

Peptide Expectation 0.05

Protein score 40

Protein consistency 4

Labeling time unit Days

Autofill Manual input

mzML / mzID Files 321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML Browse

mzML_Filename	MzID_Filename	Time	BWFE
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_01.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_01.mzid	0	0
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_Rovshan_DO2_mouselive_wv_28Feb2022_274.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_Rovshan_DO2_mouselive_wv_28Feb2022_274.mzid	0	0
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_14.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_14.mzid	1	0.0304
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_05.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_05.mzid	1	0.0235
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_19.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_19.mzid	2	0.0322
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_13.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_13.mzid	2	0.0325
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_18.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_18.mzid	3	0.0281
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_08.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_08.mzid	3	0.0309
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_04.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_04.mzid	4	0.0259
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_06.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_06.mzid	5	0.0359
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_07.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_07.mzid	5	0.0287
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_02.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_02.mzid	6	0.0265
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_16.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_16.mzid	6	0.0359
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_12.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_12.mzid	14	0.0485
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_09.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_09.mzid	14	0.0473
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_15.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_15.mzid	21	0.0374
\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_17.mzML	\\UT321319\Shared_Folder\UTMB_male_mice_liver\Long_Grad_mzML_220min_OT60k_26Aug21_RSL_17.mzid	21	0.0399

Delete Clear All Sort

Output Directory C:\Users\hndobem\UTMB\USERS\H\Desktop Browse Start

Figure 5 B. d2ome GUI in the Autofill mode.

Preparing files.txt file

The following is an examples of the text file(called files.txt in our example but can have any name):

0 B:\Heavy_Water\A0day_1.mzML B:\Heavy_Water\A0day_1.mzid 0
0 B:\Heavy_Water\A0day_2.mzML B:\Heavy_Water\A0day_2.mzid 0
1 B:\Heavy_Water\A1day_1.mzML B:\Heavy_Water\A1day_1.mzid 0.05
1 B:\Heavy_Water\A1day_2.mzML B:\Heavy_Water\A1day_2.mzid 0.05
3 B:\Heavy_Water\A3day_1.mzML B:\Heavy_Water\A3day_1.mzid 0.05
3 B:\Heavy_Water\A3day_2.mzML B:\Heavy_Water\A3day_2.mzid 0.05
5 B:\Heavy_Water\A5day_1.mzML B:\Heavy_Water\A5day_1.mzid 0.05
5 B:\Heavy_Water\A5day_2.mzML B:\Heavy_Water\A5day_2.mzid 0.05
7 B:\Heavy_Water\A7day_1.mzML B:\Heavy_Water\A7day_1.mzid 0.05
7 B:\Heavy_Water\A7day_2.mzML B:\Heavy_Water\A7day_2.mzid 0.05
14 B:\Heavy_Water\A14day_1.mzML B:\Heavy_Water\A14day_1.mzid 0.05
14 B:\Heavy_Water\A14day_2.mzML B:\Heavy_Water\A14day_2.mzid 0.05
21 B:\Heavy_Water\A21day_1.mzML B:\Heavy_Water\A21day_1.mzid 0.05
21 B:\Heavy_Water\A21day_2.mzML B:\Heavy_Water\A21day_2.mzid 0.05
30 B:\Heavy_Water\A30day_1.mzML B:\Heavy_Water\A30day_1.mzid 0.05
30 B:\Heavy_Water\A30day_2.mzML B:\Heavy_Water\A30day_2.mzid 0.05

The results will be in the folder where the program was run. In this case, the folder, E:\GUI_RUN.

6. Output.

All outputs are reported in csv formatted files. For each protein that passed the thresholds specified in quant.state file, two main output files will be created: ProteinAccession.RateConst.csv and ProteinAccession.Quant.csv.

The ***.Quant.csv** file contains comprehensive information about each peptide (which passed the thresholds specified in quant.state file) of a protein. Each peptide entry is a row of information. The information are amino acid sequence, distinctness of the sequence, charge state of the precursor, theoretical m/z ratio of the precursor, theoretical isotope abundances, total labeling (theoretical, before the start of the labeling), precursor m/z value (measured), the highest Mascot Ion score, Mascot expectation, mass accuracy (in ppm), scan number, the integrated abundance of the mass isotopomers (six), elution start and end times that were used to calculate the isotopomer abundances, peak width of the monoisotope in the mass-to-charge domain, total labeling from experimental isotopes. The information is repeated for each experiment. Ion Score of 0 indicates that the peptide was not observed in that particular experiment. The total labeling (molecular percent enrichment) is by default calculated only for the 1st heavy isotope. If the entries are blank for an experiment, it means that the peptide was not fragment in that LC-MS.

The rows of ***.RateConst.csv** file of a protein contain: peptide uniqueness (distinct or shared sequence with other peptides) of the sequence, rate constant (in unites reciprocal to the time units used in the files.txt file and specified in quant.state), correlation between the fit and experimental data, root mean square error, and absolute deviation between

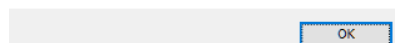
the theoretical and experimental isotope profiles (before the start of labeling), peptide charge, sequence mass-to-charge ratio, number of accessible hydrogens (NEH), number of data points (NDP), R^2 of the theoretical fit, and averaged abundance of the monoisotope.

Proteins.csv file contains the list of proteins and their Mascot scores for proteins that passed the specified (in quant.state) thresholds.

7. Visualization.

Once the quantification is finished, there will be a message on the screen. Press the

Finished... Please check the for the results in E:\GUI_test folder.



“Ok” button. Then press “Visualization” button on the Tab Controller. An output like the one in **Figure 6** should appear on the screen. The program automatically reads the results from the output directory, sorts the proteins by name, and shows the first protein, its rate constant, standard deviation, and its peptides, their characteristics (charge state, R^2 of the fit, computed rates, etc.). To look at the results of theoretical fit, one can choose any of the peptides on the screen by “click” or “up”, “down” arrows. The figures for all peptides of all proteins can be saved (in jpeg format) by clicking on the button “Export all proteins”. Alternatively, the protein on the screen can be exported by clicking the “Export Protein_Name” button. The visualization tool can also be used to view previously quantified data.

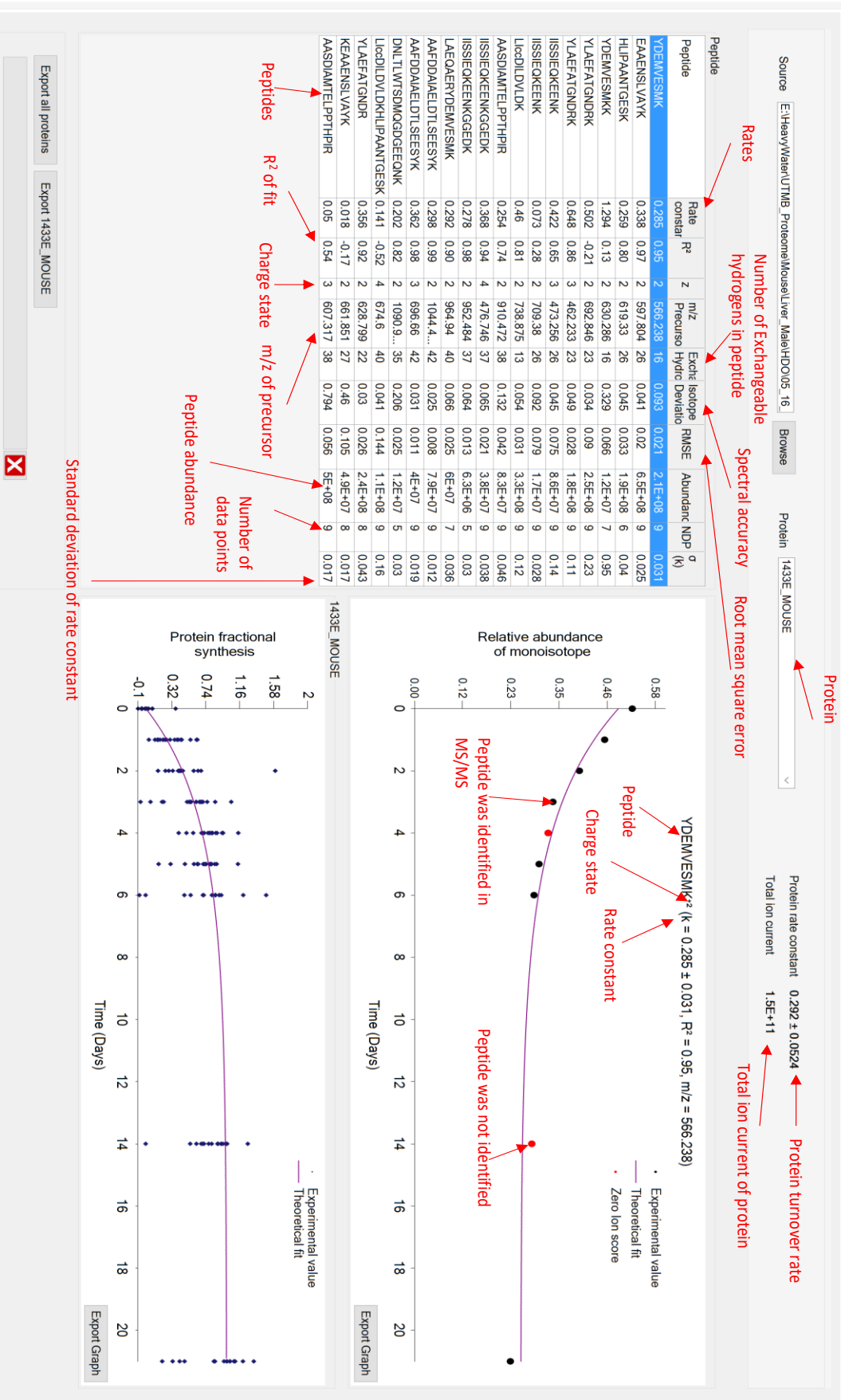


Figure 6. Output results for protein, 1433E_MOUSE.

8. Citation

The data processing in d2ome was described in [2]. The selection of mass isotopomers for label quantification is described in [1].

9. Additions/Changes.

11.08.2019

1. We have added support for user-defined values of the Number of Exchangeable Hydrogens in each amino acid. If the NEH.txt file is provided, the program will read the number of exchangeable Hydrogens information for each amino acid from the NEH.txt file and overwrite the default values. The change was made in RateConstant.dll. If the default values are used, the changes will have no effect on the results.

12.10.2019

2. We have added support for mzid files that were generated using MGF file input. The original support was for the mzid files generated using mzML files for peptide/protein identifications. Changes were made to mzIdentML.dll.

04.03.2020

3. GUI for d2ome has been created. It automates the creation of files.txt and quant.state files. All dll and exe files should be in the same folder, like in **Figure 4**.

Click on the d2ome_SetUp_GUI.exe file, and the GUI will start.

05.04.2021

4. Support has been added for mass spectral data acquired on Agilent instruments. The program will recognize Agilent mzML files and process them.

5. The number of experiments that can be entered in GUI has been increased.

11.05.2021

6. d2ome now works with centroid data in MS1. **Centroid mode for MS1 is preferable.** The processing is faster. The data processing in this mode also applies the recent development of mass isotopomer dynamics[1] to estimate deuterium enrichment.

The MS1 data mode is specified in the quant.state file using the following line:

```
MS1_Type = 1 // data type of MS1, 1 - centroid, 0 - profile
```

7. The program determines the number of mass isotopomers that can potentially be useful for estimating deuterium enrichment, given the body water enrichment. Peak detection and quantification will only quantify a peptide's needed number of mass isotopomers.

8. Currently, d2ome supports the following post-translational modifications (PTMs): carbamidomethylation of Cys, deamidations of Asn and Gln, oxidation of Met, and acetylation of Lys.

01.31.2022

9. New GUI, Visualization Tool, and the mass isotopomer dynamics[1] for peak detection have been added. With the new GUI, there is no limit on the number of input files (mzML and mzID pairs).

05.10.2022

10. We have added an option to quantify peptides that have not been identified in MS/MS. The RIA results from unidentified peptides are shown with red circles in the peptide RIA time course.

11. The output includes the estimations of standard deviations for peptides $\sigma(k)$ [3], average peptide abundance, the number of data points (NDP), peptide charge state, and m/z .
12. The peak estimation uses ranges of possible ion abundances (as discussed in [1]) to exclude from the quantification high mass isotopomers if their contribution is negligible.
13. There is no limit on the number of experiments that can be analyzed using GUI.
14. To speed up the data processing, the code is parallelized for creating a list of consistently identified proteins from all experiments.

References.

1. Sadygov RG. Partial Isotope Profiles Are Sufficient for Protein Turnover Analysis Using Closed-Form Equations of Mass Isotopomer Dynamics, *Anal Chem* 2020;92:14747-14753.
2. Sadygov RG, Avva J, Rahman M et al. d2ome, Software for in Vivo Protein Turnover Analysis Using Heavy Water Labeling and LC-MS, Reveals Alterations of Hepatic Proteome Dynamics in a Mouse Model of NAFLD, *J Proteome Res* 2018;17:3740-3748.
3. Sadygov RG. Protein turnover models for LC-MS data of heavy water metabolic labeling, *Brief Bioinform* 2022.