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].

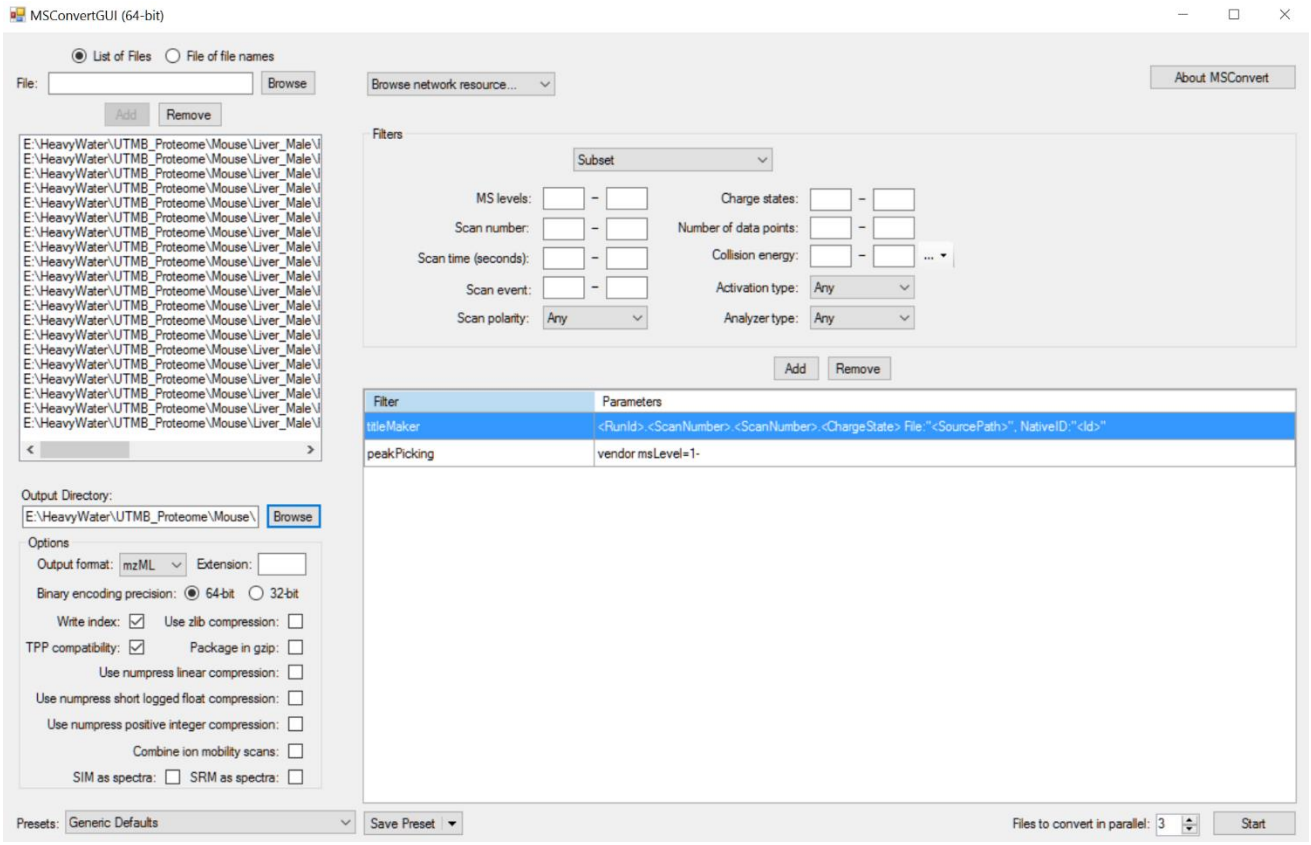


Figure 1. Input parameter set-up for MSConvert to generate mzML (**centroid MS1**) file from a (for example) raw file. Note that with the latest version of d2ome, generating MS1 scans in centroid mode is **preferred**.

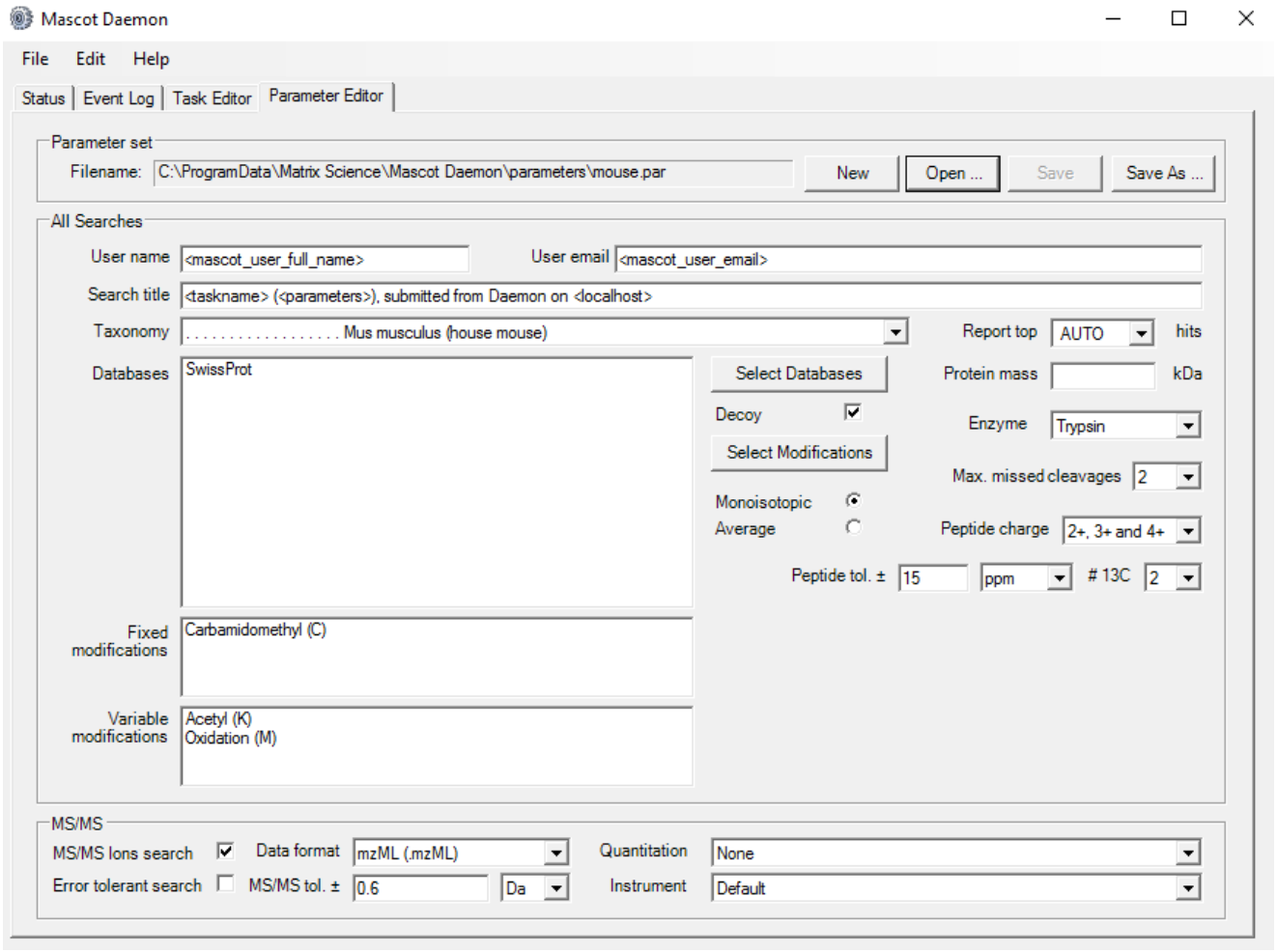


Figure 2. An example of a parameter setting using Mascot's "Mascot Daemon" interface.

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

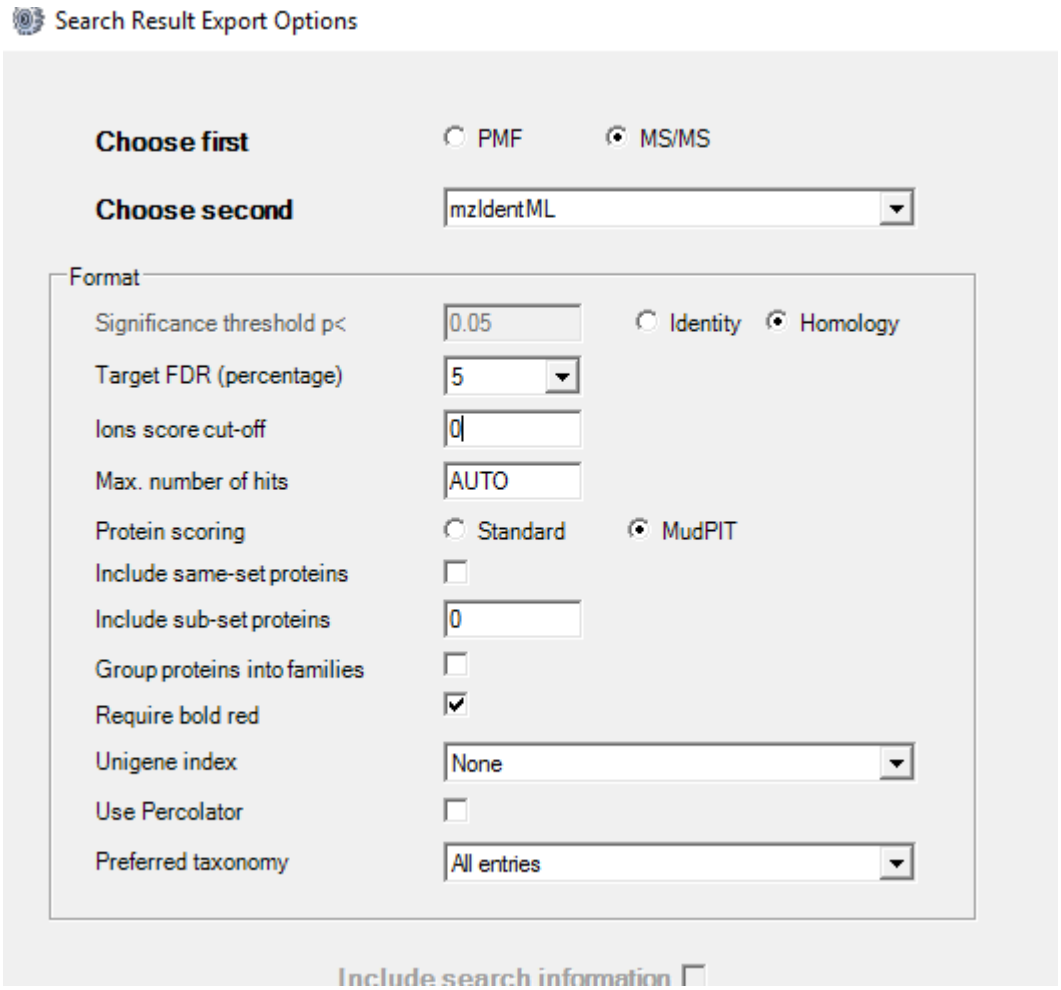


Figure 3 A. The filtering section of the Mascot Daemon's "Auto-export…" options. "Group Proteins" should be unchecked, "Require bold red" should be checked.

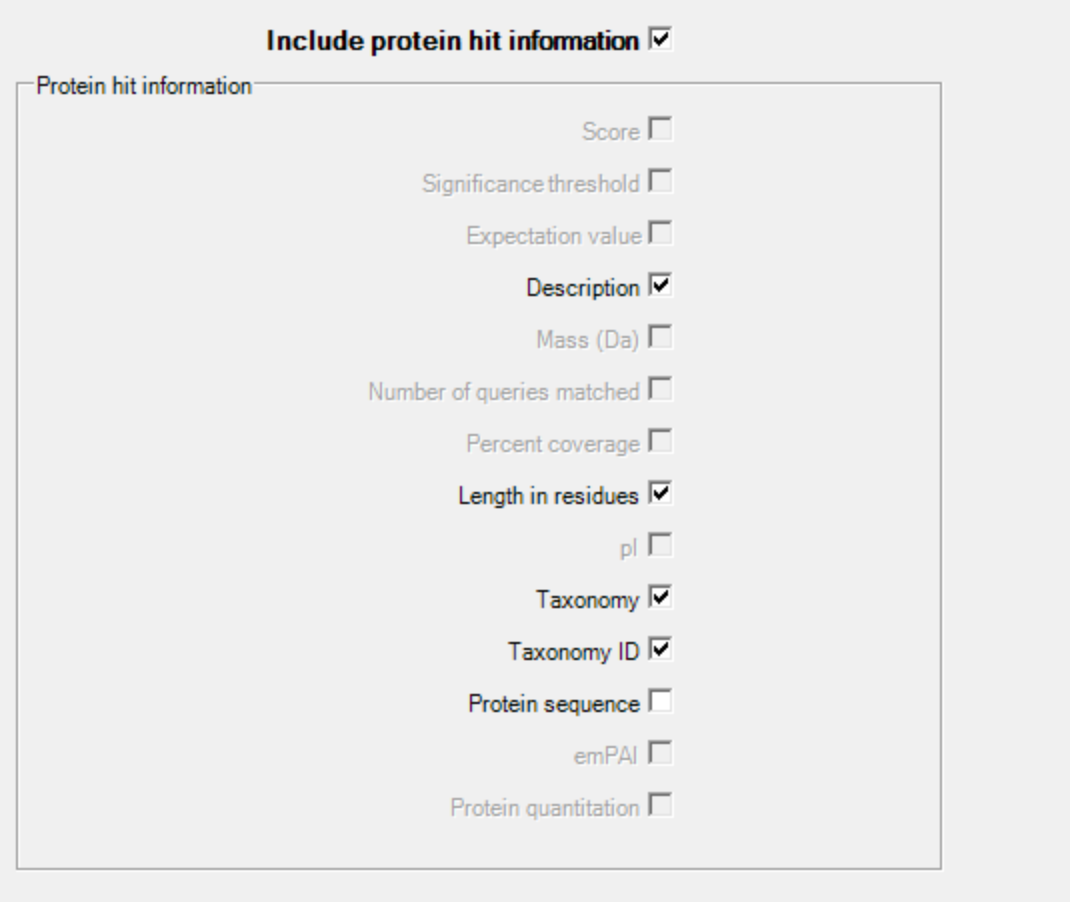


Figure 3 B. Protein Hit Information section of Mascot Daemon's "Auto-export…" options. The shown are the settings that are required for d2ome. The "Description" and "Length in residues", both should be checked.

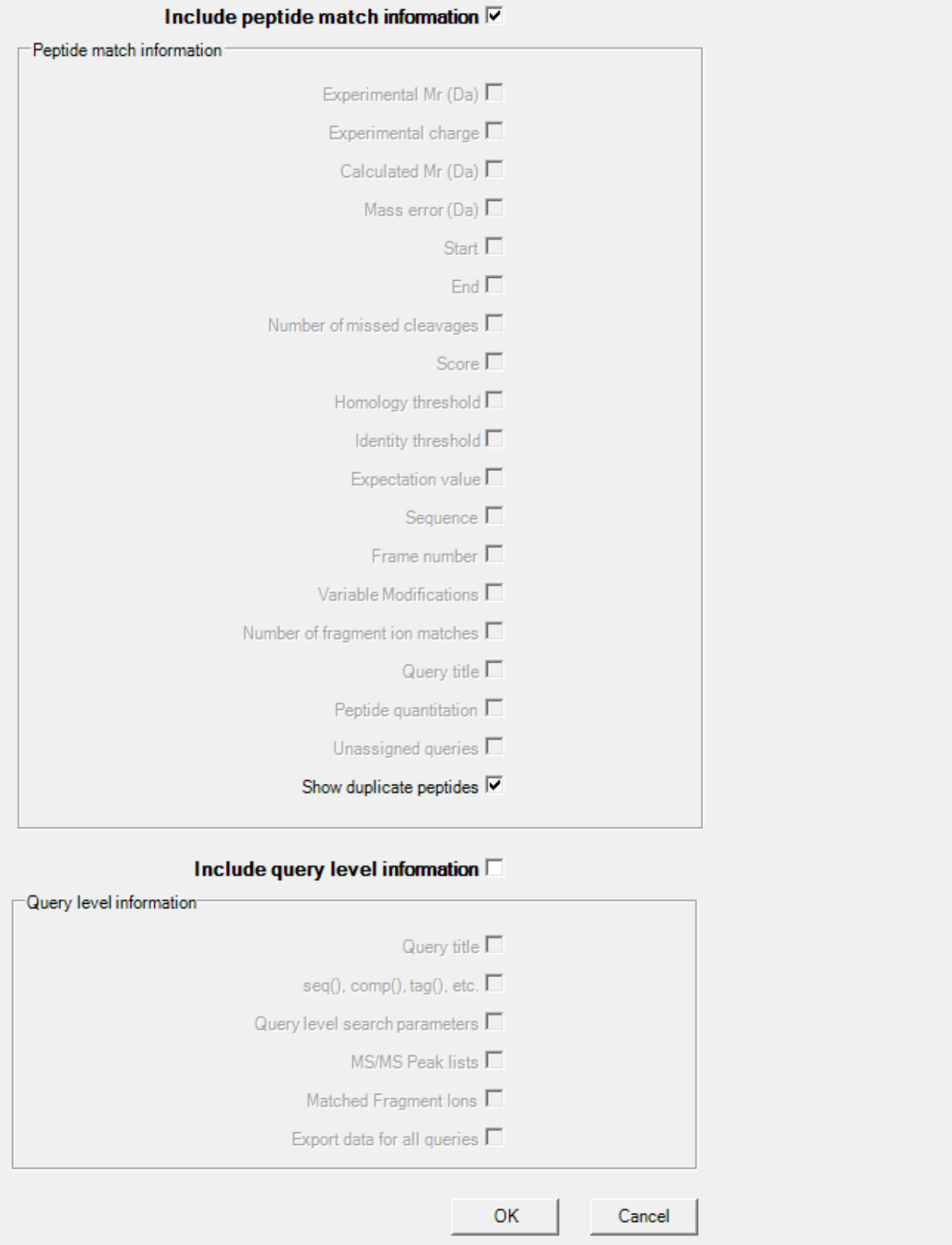


Figure 3 C. Uncheck the "**Include query level information**".

2. Running d2ome using GUI.

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

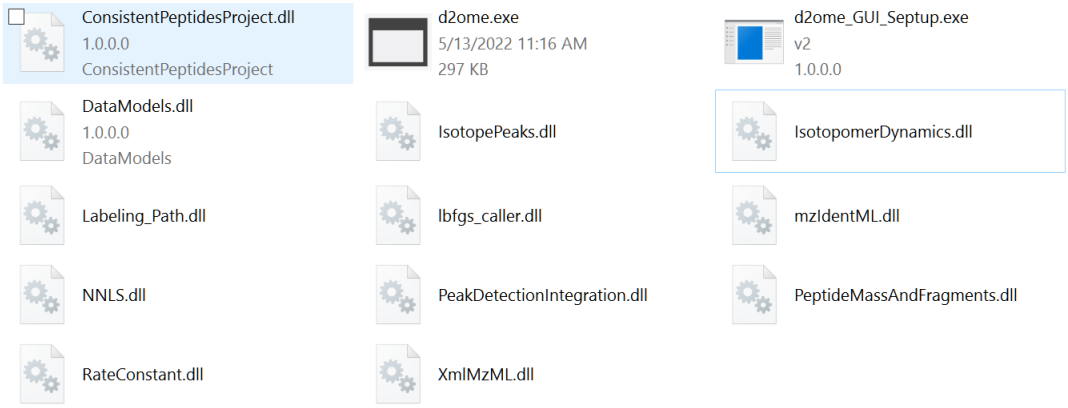


Figure 4. The binaries to run d2ome from either command line or using a GUI.

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 D_2O 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.

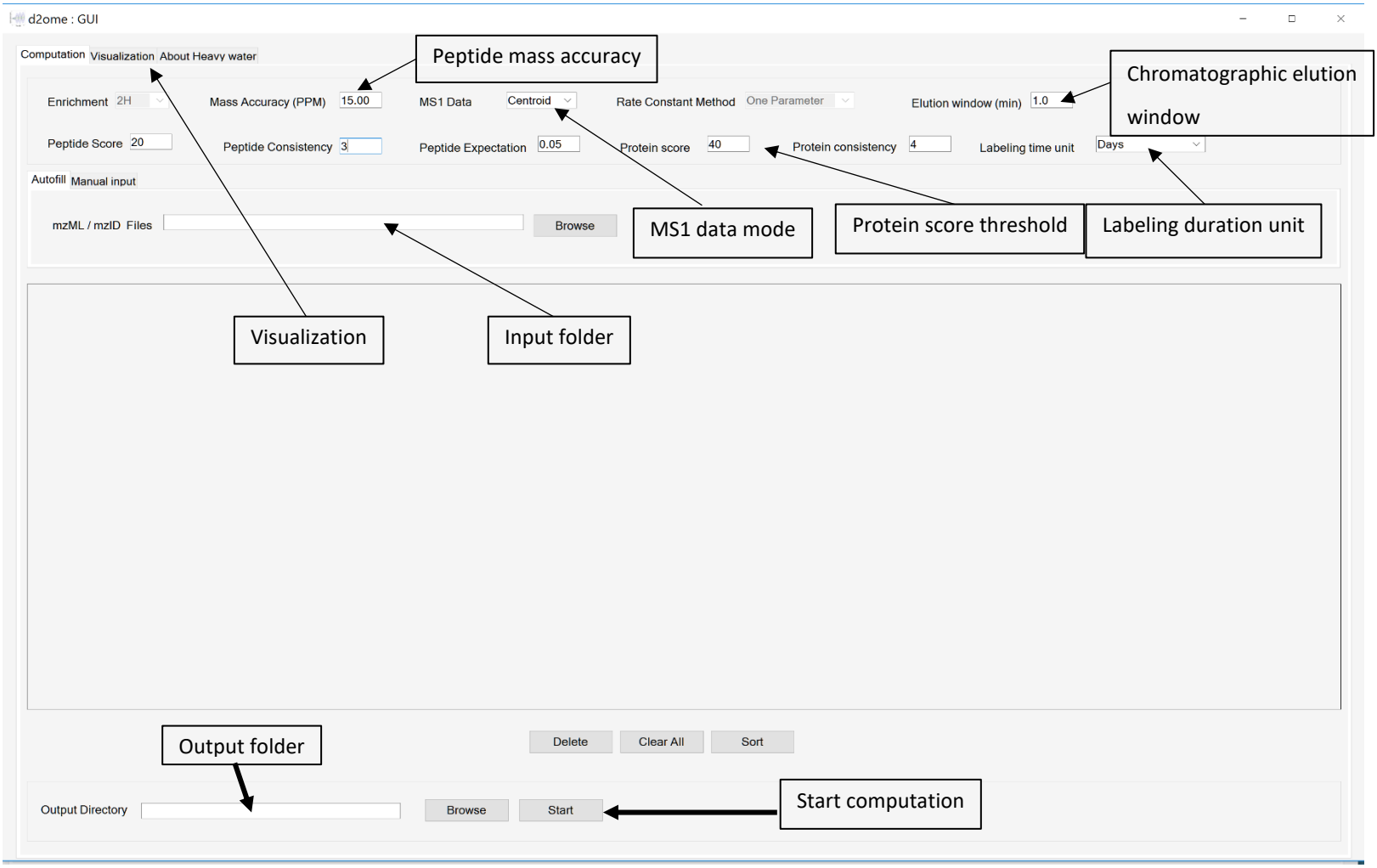


Figure 5 A. d2ome GUI before data initialization.

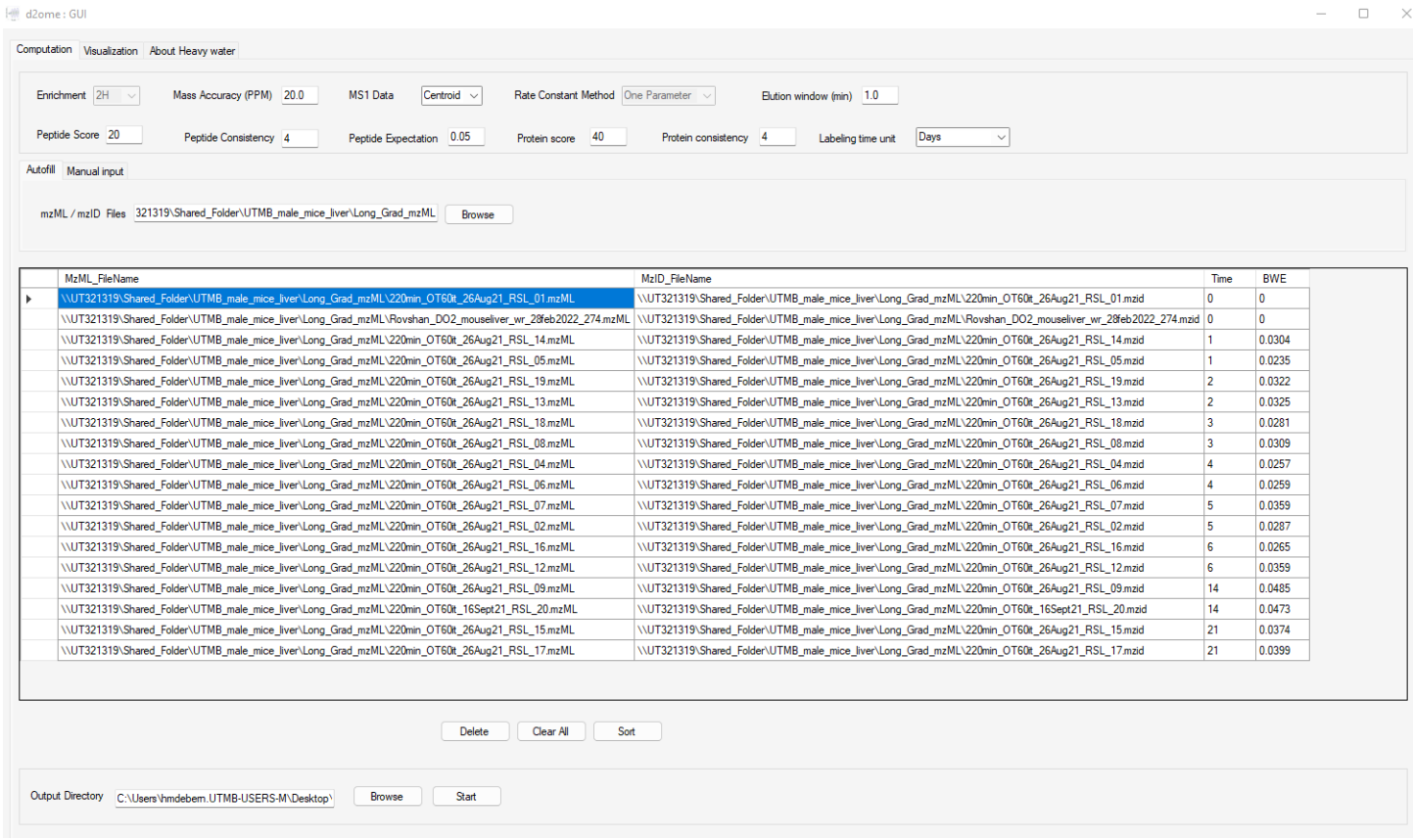


Figure 5 B. d2ome GUI in the Autofill mode.

3. Running d2ome from command line.

Preparing files.txt file

The location of the mzML and mzID files that were prepared in the first step plus information regarding the labeling duration and the enrichment level should be stored in a text file. The text file is then used when running the software.

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

>>>>>>>>>>>>>>

In every row, the first number is the number of labeling days, the second element is the address of the mzML file, the third element is the location of the mzid file that corresponds to the preceding mzML file, and the fourth element is the Body Water Enrichment level at that labeling day. For example, in the line, 7 B:\Heavy_Water\A7day_2.mzML B:\Heavy_Water\A7day_2.mzid 0.05 “7” is the number of labeling days for this sample, “B:\Heavy_Water\A7day_2.mzML” is the mzML file, “B:\Heavy_Water\A7day_2.mzid” is the mzid file. “0.05” is the body water enrichment level. Note that two repeating lines with labeling days of “7” indicate the repeats/replicates. There is no limit on the number of replicates in d2ome. In practice, we have used as many as 16 replicates per time point. Computed rate constants are in the reciprocal of the time units specified in this file.

4. Preparing quant.state file

To overwrite the default parameters of d2ome, one prepares another file: quant.state. Note that in this case the naming is important, and it has to be exactly the same as what is typed here. An example of **quant.state** file is:

<<<<<<<<<<<<<<

mass_accuracy = 15 ppm // mass accuracy: either in ppm or Da

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

protein_score = 50 //minimum protein score

peptide_score = 20 // minimum peptide score, ion score in Mascot, default is 30
peptide_expectation = 0.05 // maximum peptide expectation in Mascot

elutiontimewindow = 1 // time window (mins) to search for elution peak. From the time that highest scoring MS2 was triggered

protein_consistency = 4 // minimum number of experiments for protein consistency // default 4

peptide_consistency = 4 //mininum number of experiments for a peptide consistency // default 4

NParam_RateConst_Fit = 1 // The model for fitting rate constant. Currently, only 1.

Labeling_time_unit = Days // Days, Hours

>>>>>>>>>>>>>>

Note also that d2ome is sensitive to what comes before the equal sign and the sign itself. These should not be changed. Also, double slash stands for comment and whatever after that in a line is not read by the software. Therefore, one is only expected to change the numbers.

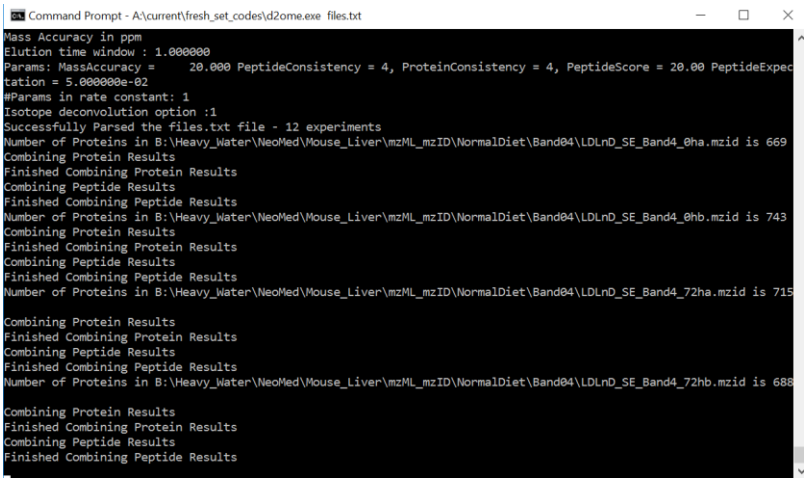
5. Running the program from the command line.

1. Download all binaries into a single folder. Assume, you have downloaded the files into the folder: C:\d2ome_exec. These files should be as shown in **Figure 4**.
2. From the directory where you have files.txt and quant.state (and possibly NEH.txt – to specify number of exchangeable hydrogens for each amino acid) files, use the command:

   E:\GUI_RUN>C:\d2ome_exec\d2ome.exe files.txt

   In the above line, it is assumed that the files.txt and quant.state files are in the E:\GUI_RUN folder.

   The program should start with a message on the output like this one:



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 1^{st} heavy isotope. If the entries are blank for an experiment, it means that the peptide was not fragment in that LC-MS.

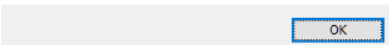
The rows of ***.RateConst.csv** file of a protein contain: peptide uniqueness (distinct or shared sequence with other peptides) of the sequence, rate constant (in unites reciprocal to the time units used in the files.txt file and specified in quant.state), correlation between the fit and experimental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



"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.

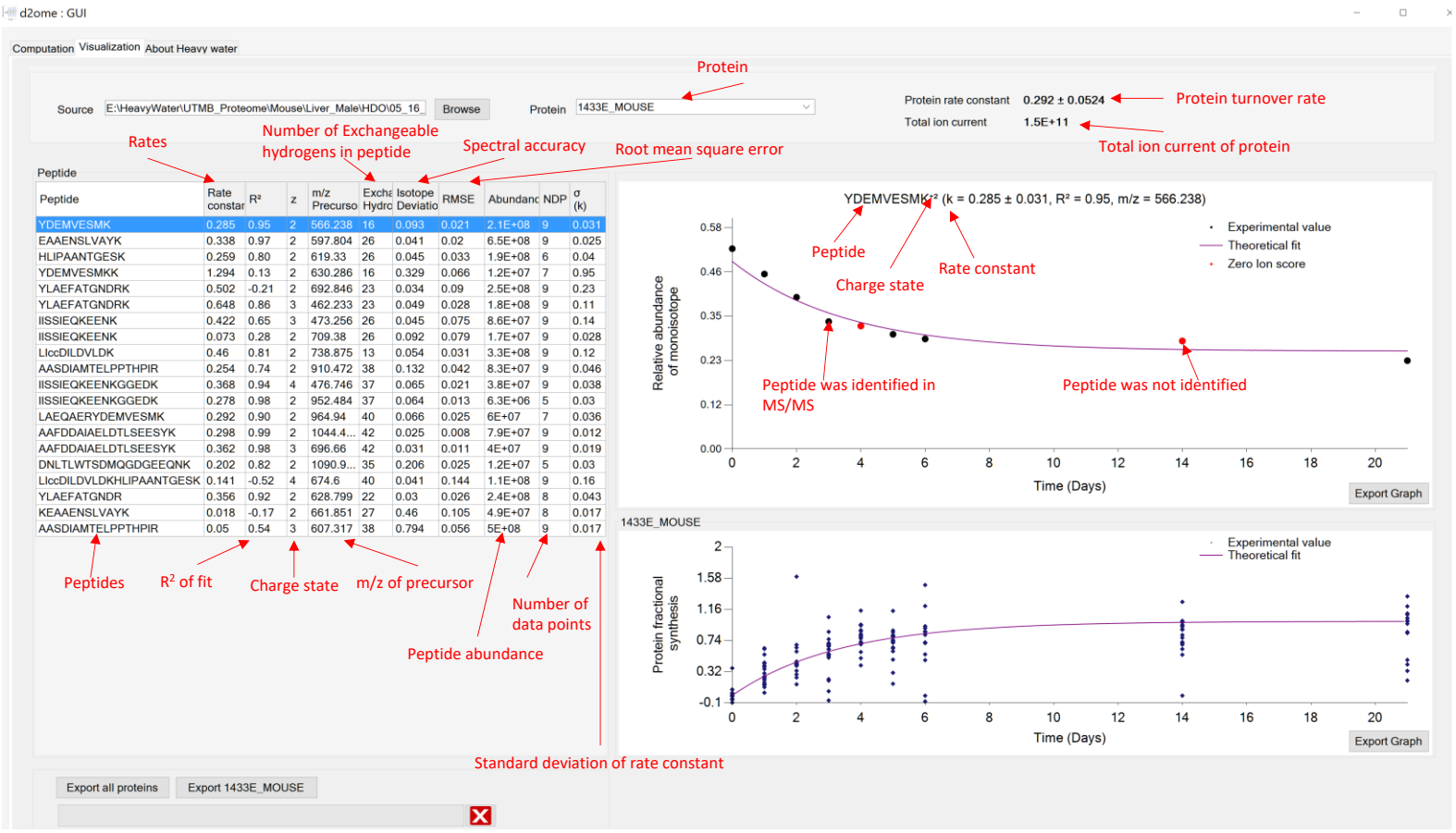


Figure 6. Output results for protein, 1433E_MOUSE.

8. Flexible Quality Control for Protein Turnover Rates.

This page presents a bioinformatics tool for determining protein turnover rates based on user customizable GOF measurements. The tool uses d2ome software quantification outputs to determine new protein turnover rates and their corresponding confidence intervals.

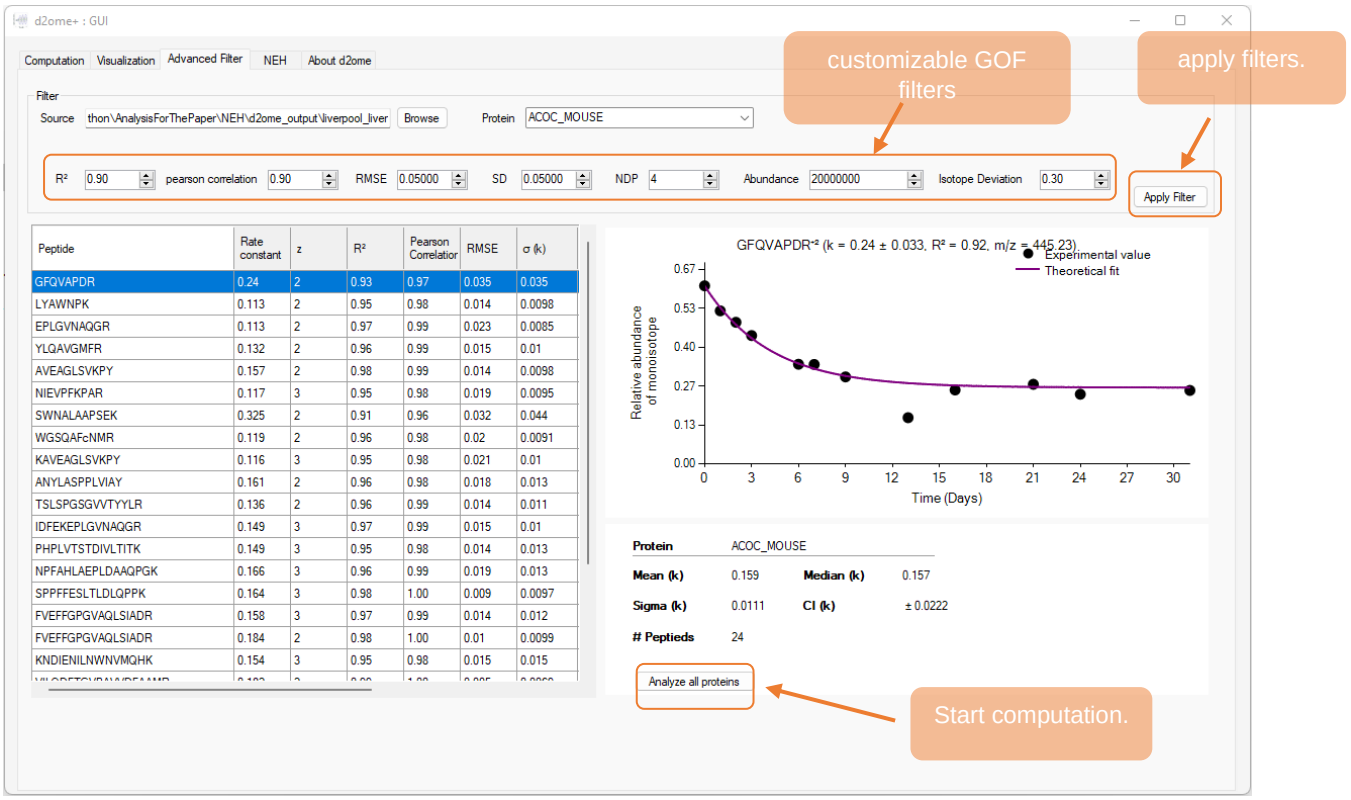


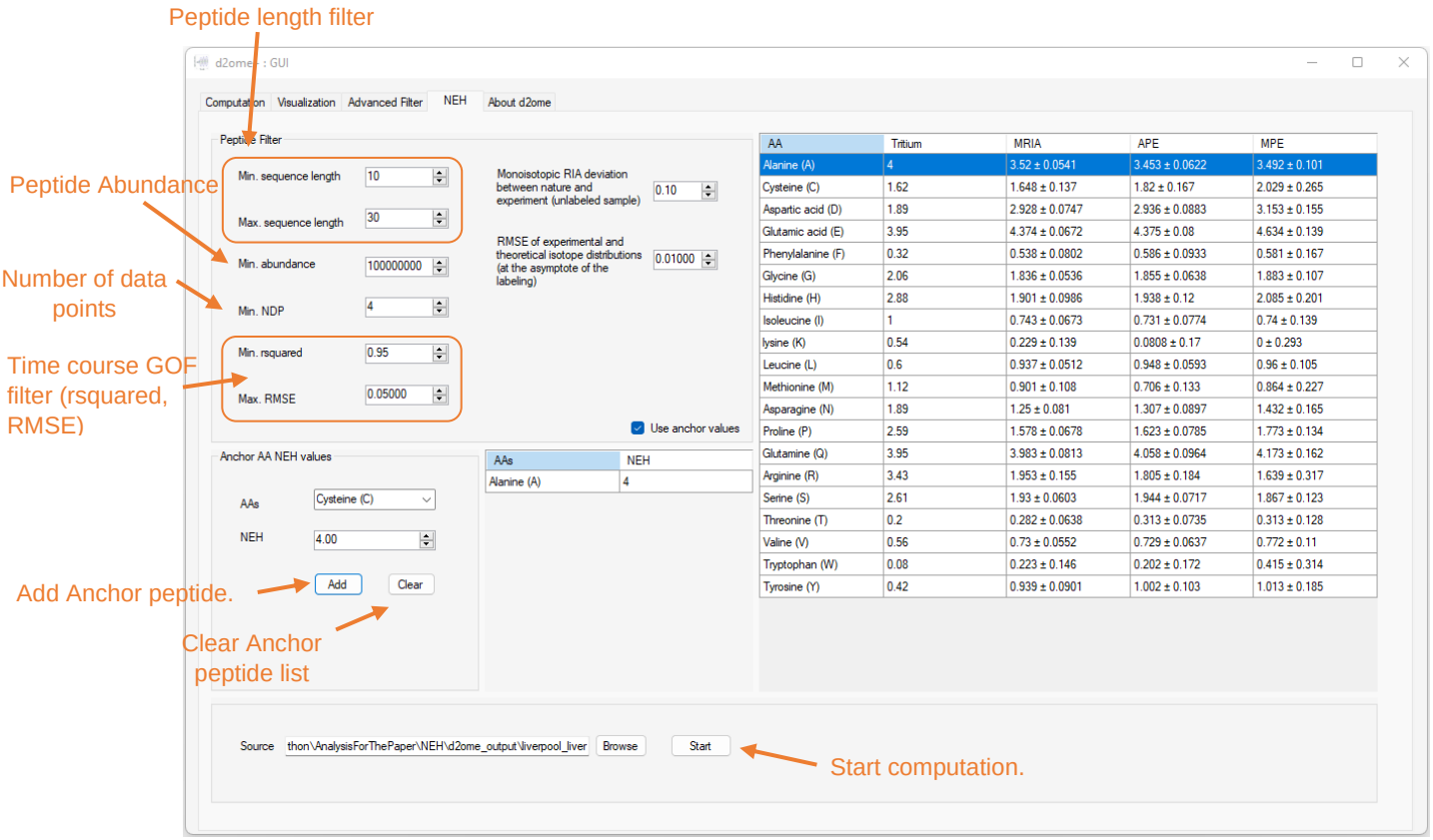
Figure 7. The graphical user interface (GUI) for advanced tool for protein peptide filtering.

The filtering parameters incorporated in this tool are the coefficient of determination (R^2), the Pearson correlation coefficient (r), the root mean squared error (RMSE), the peptide abundance, the isotope deviation, and the number of experiments in which the peptide is identified and quantified. The tool enables users to visually inspect and validate the filtered peptides by providing the time-course plot of the experimental RIA values and

their comparison with the theoretical ones. Furthermore, it generates a comma-separated output file, called Analyzed_Proteins_and_Their_Rates.csv, which contains the newly computed turnover rate, the number of peptides used in the computation, the 95% confidence interval, and the standard deviation of the protein turnover rates quantified from the LC-MS experiment.

9. Determining numbers of exchangeable hydrogens from LC-MS data.

We have implemented data-driven analytical approaches to determine the number of exchangeable hydrogens for AAs using mass spectrometry data. A graphical user interface for inputting data to compute N_{EH} values for peptides and AAs is shown in **Figure** 8. The application implements the three analytical methods proposed in a recent study. The application has an option for the users to set the filtering criteria, including peptide abundance, sequence length, number of experiments peptide identified, and the goodness-of-fit measures (R^2, RMSE). In addition, the software allows users to incorporate an unlimited number of anchor amino acids NEH that are determined experimentally using GC-MS or other techniques.



AA	Tritium	MRIA	APE	MPE
Alanine (A)	4	3.52 ± 0.0541	3.453 ± 0.0622	3.492 ± 0.101
Cysteine (C)	1.62	1.648 ± 0.137	1.82 ± 0.167	2.029 ± 0.265
Aspartic acid (D)	1.89	2.928 ± 0.0747	2.936 ± 0.0883	3.153 ± 0.155
Glutamic acid (E)	3.95	4.374 ± 0.0672	4.375 ± 0.08	4.634 ± 0.139
Phenylalanine (F)	0.32	0.538 ± 0.0802	0.586 ± 0.0933	0.581 ± 0.167
Glycine (G)	2.06	1.836 ± 0.0536	1.855 ± 0.0638	1.883 ± 0.107
Histidine (H)	2.88	1.901 ± 0.0986	1.938 ± 0.12	2.085 ± 0.201
Isoleucine (I)	1	0.743 ± 0.0673	0.731 ± 0.0774	0.74 ± 0.139
lysine (K)	0.54	0.229 ± 0.139	0.0808 ± 0.17	0 ± 0.293
Leucine (L)	0.6	0.937 ± 0.0512	0.948 ± 0.0593	0.96 ± 0.105
Methionine (M)	1.12	0.901 ± 0.108	0.706 ± 0.133	0.864 ± 0.227
Asparagine (N)	1.89	1.25 ± 0.081	1.307 ± 0.0897	1.432 ± 0.165
Proline (P)	2.59	1.578 ± 0.0678	1.623 ± 0.0785	1.773 ± 0.134
Glutamine (Q)	3.95	3.983 ± 0.0813	4.058 ± 0.0964	4.173 ± 0.162
Arginine (R)	3.43	1.953 ± 0.155	1.805 ± 0.184	1.639 ± 0.317
Serine (S)	2.61	1.93 ± 0.0603	1.944 ± 0.0717	1.867 ± 0.123
Threonine (T)	0.2	0.282 ± 0.0638	0.313 ± 0.0735	0.313 ± 0.128
Valine (V)	0.56	0.73 ± 0.0552	0.729 ± 0.0637	0.772 ± 0.11
Tryptophan (W)	0.08	0.223 ± 0.146	0.202 ± 0.172	0.415 ± 0.314
Tyrosine (Y)	0.42	0.939 ± 0.0901	1.002 ± 0.103	1.013 ± 0.185

Figure 8. The graphical user interface (GUI) of d2ome for a number of exchangeable hydrogens estimation of amino acids.

10. Citation

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

11. 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 σ(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.

09.20.2023

15. Two parameter data modeling options for rate constant computation added

16. Tighter optimization parameter incorporated into LBFGS algorithm

17. The graphical user interface (GUI) for advanced tool for protein peptide filtering added

04.01.2024

18. Analytical approach for determining the number of exchangeable hydrogens for peptide and amino acids is incorporated to d2ome.

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.