The program, d2ome, uses heavy water labeling and LC-MS data to estimate protein degradation rate constants. The following is the workflow of the process.

The program can be run from the command line (with no limit on the number of experiments - both time points and replicates) or using a GUI (with up to 15 LC-MS runs). The GUI prepares files.txt and quant.state files that are used as mass spectral inputs and the parameters for the computational analysis procedures, respectively.

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 1A (**centroid MS1 data**)** and **Figure 1B** (profile 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**. The processing is faster, and it uses new developments on mass isotopomer dynamics[1].

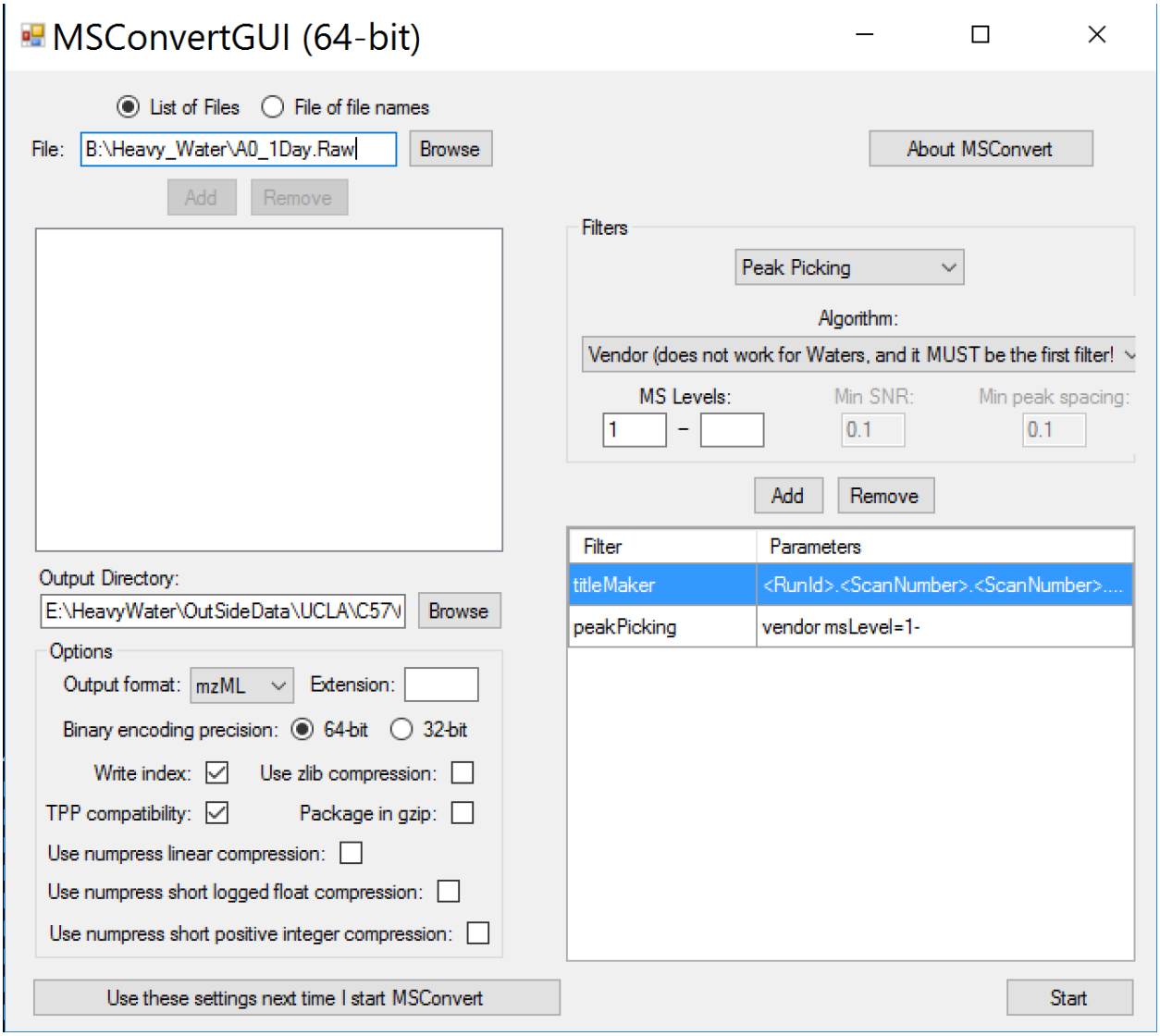


Figure 1 A. Input parameter set-up for generating mzML (**centroid MS1**) file from a (for example) raw file. Note that with the latest version of d2ome, generating MS1 scans in centroid mode is **preferred**.

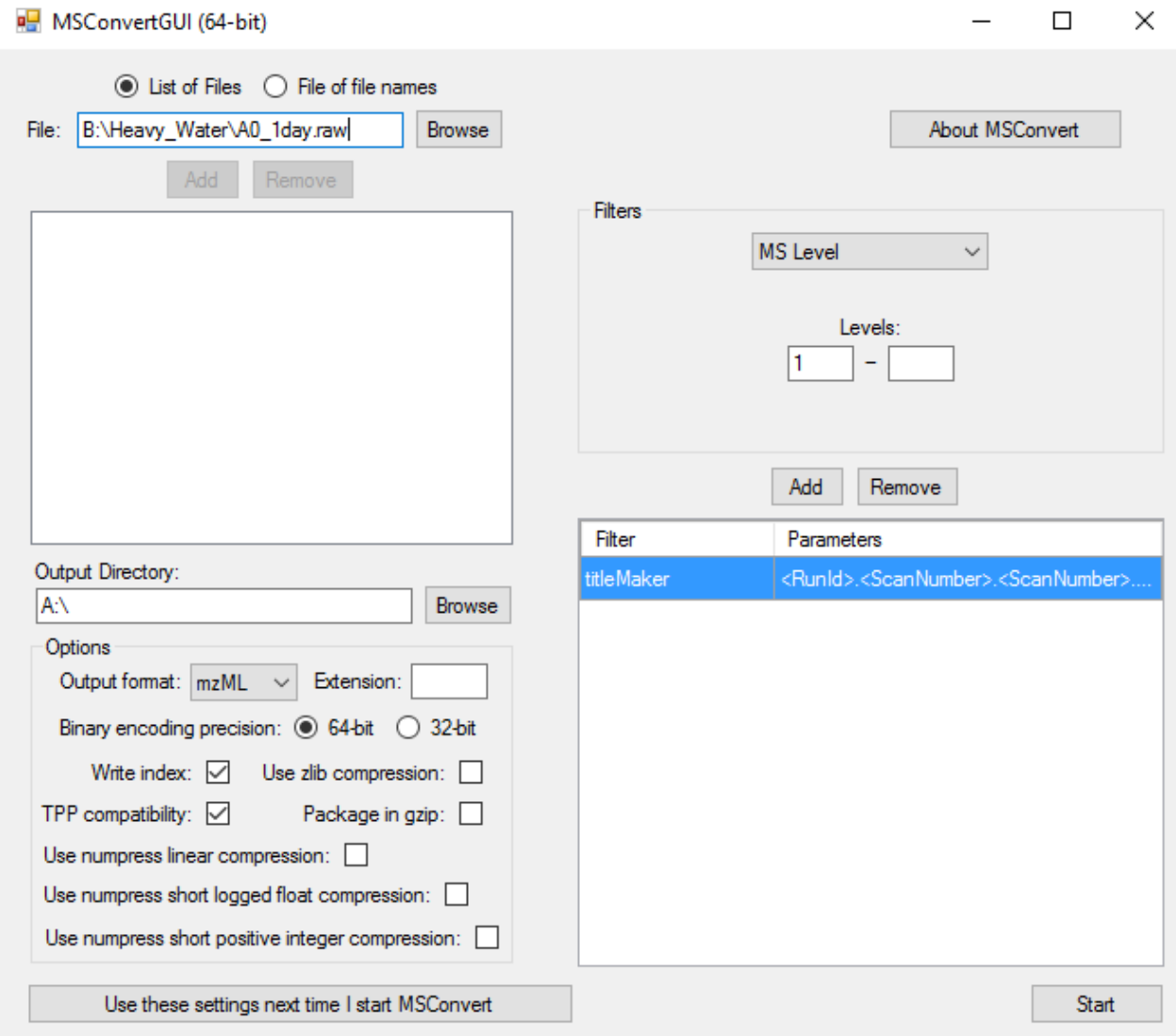


Figure 1 B. Input parameter set-up for generating mzML (**profile MS1**) file from a (for example) raw file.

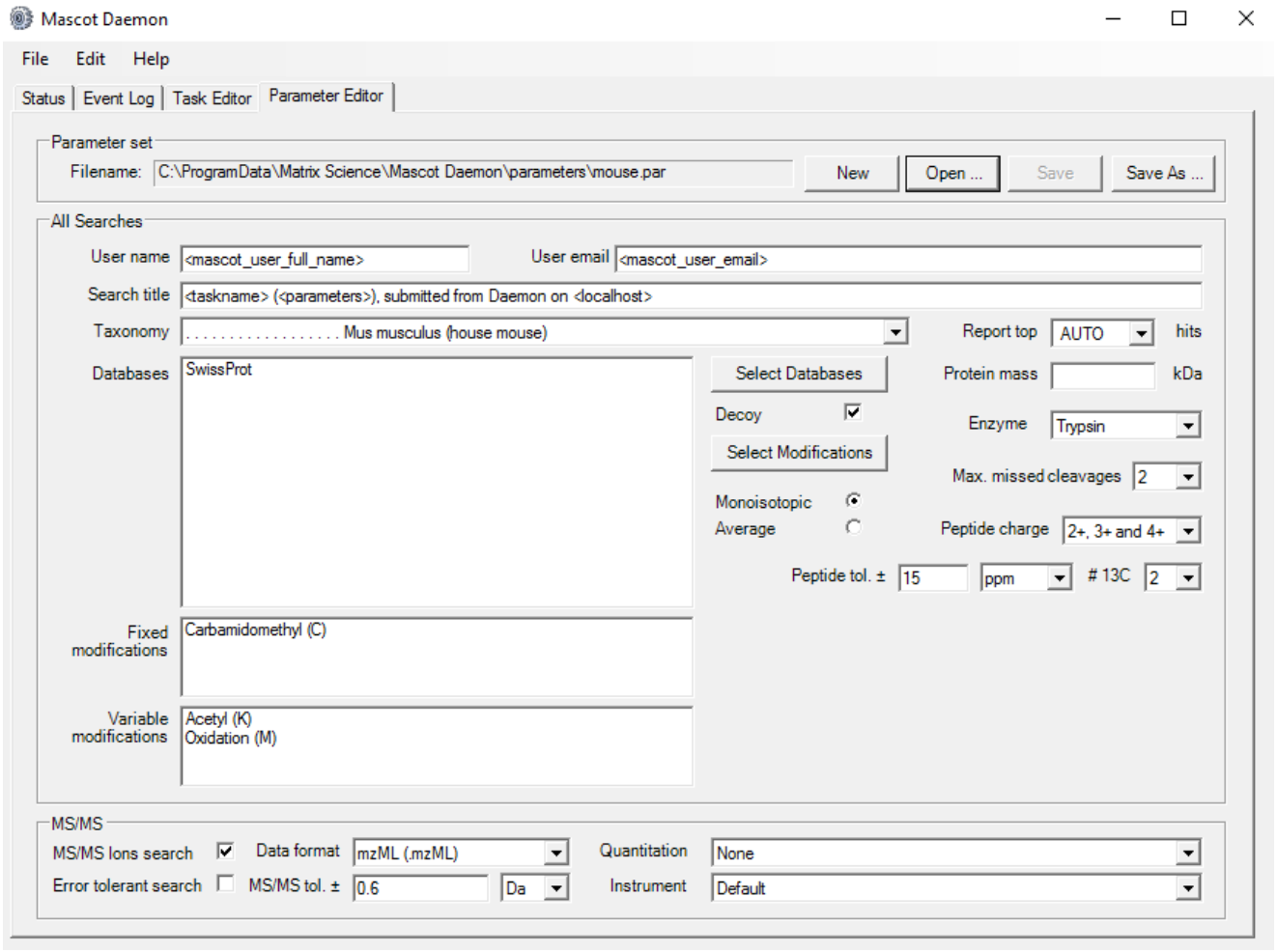


Figure 2. An example of a parameter setting using Mascot's "Mascot Daemon" interface.

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

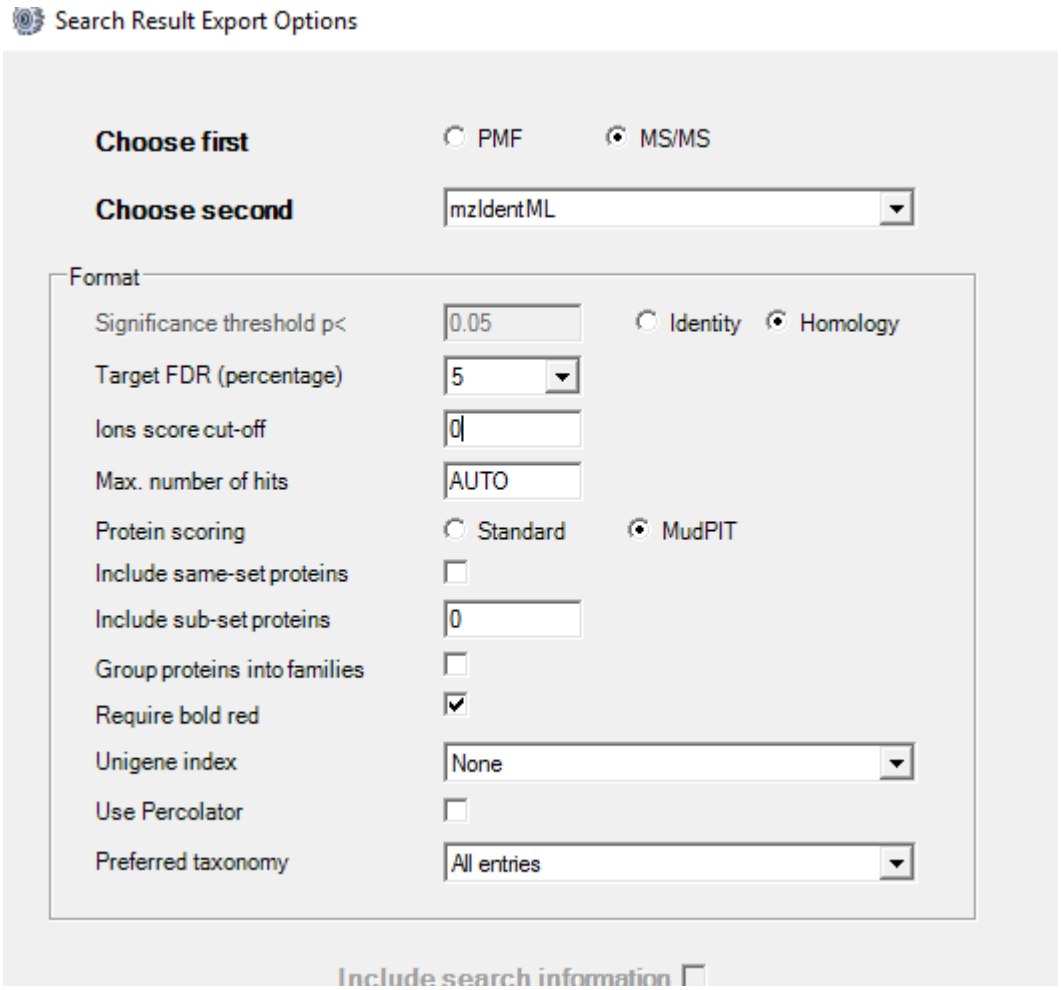


Figure 3 A. The filtering section of the Mascot Daemon's "Auto-export…" options. "Group Proteins" should be unchecked, "Require bold red" should be checked.

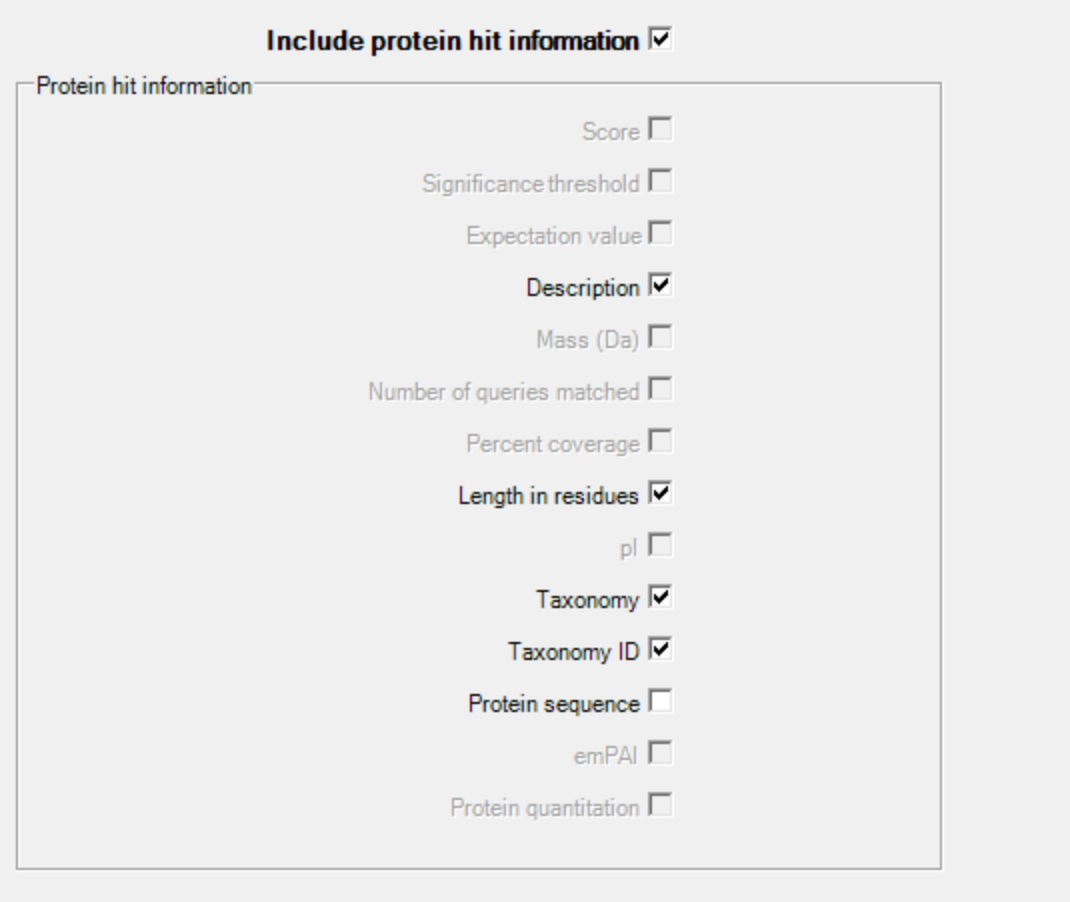


Figure 3 B. Protein Hit Information section of Mascot Daemon's "Auto-export…" options. The shown are the settings that are required for d2ome. The "Description" and "Length in residues", both should be checked.

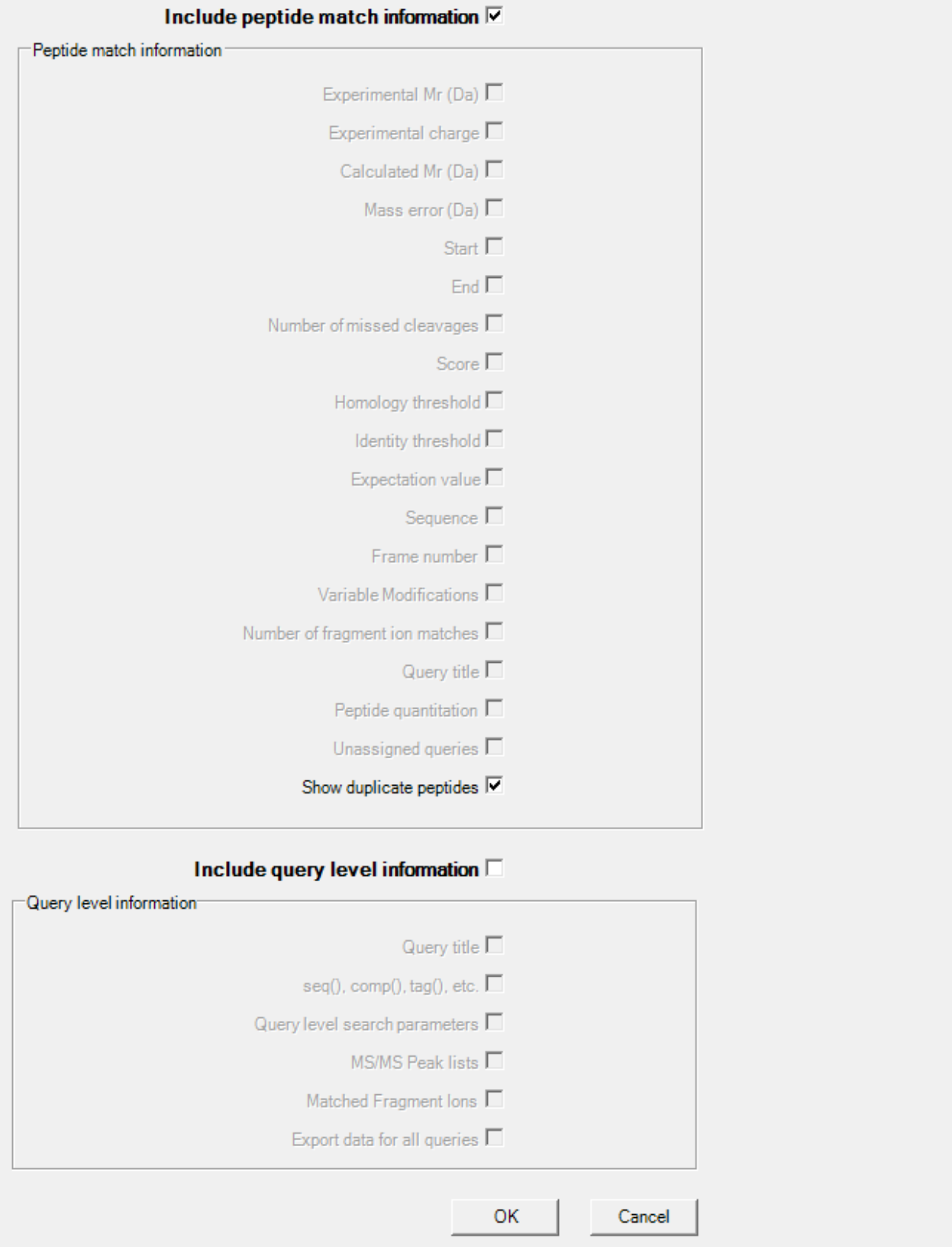


Figure 3 C. Uncheck the "**Include query level information**".

2. Preparing files.txt file

The location of the mzML and mzID files that were prepared in the first step plus information regarding the labeling time 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.

3. 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. Values are 1, 2, 3.

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

Three models have been integrated in the software and can be used by choosing a value for "NParam_RateConst_Fit". 1 stands for one-parameter model in which only the decay rate constant is undetermined, 2 stands for two-parameter model, and 3 stands for three-parameter model. 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.

4. 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**.

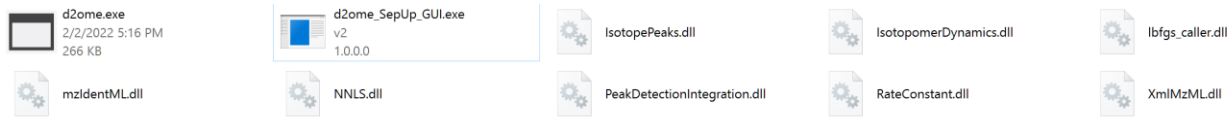


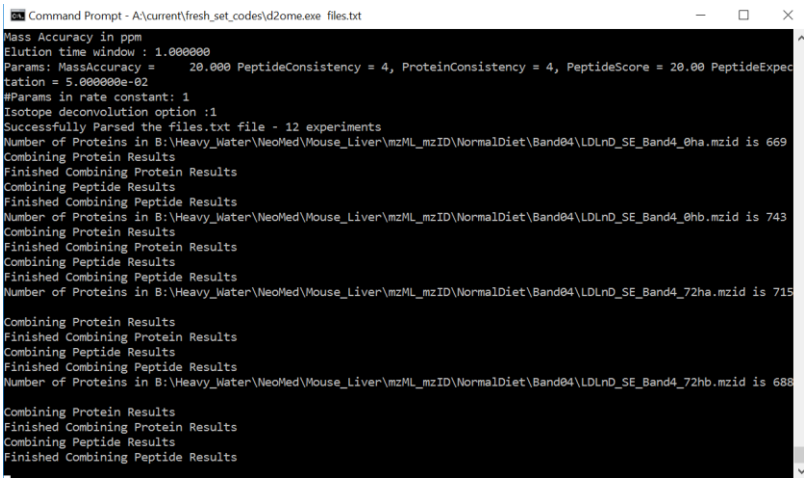
Figure 4. The list of binaries that are necessary to run d2ome from either command line or using a GUI.

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.

---

5. Running the program using GUI.

We have newly created a new GUI and a Visualization Tool for the results. It is started by a click of a mouse on the d2ome_SetUp_GUI.exe. The application form (Tab Controller) shown below should appear, **Figure 5 A**.

The new GUI allows automated 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 are is in the previous GUI version. For using autofill mode, click on the “Browser” button following “mzML / mzID files”. A user can also insert a folder path directly into the box. The GUI will sort the files into matching pairs. 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; **Figure 5 C**. Shown in Figure 5 C is an analysis consisting of 20 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) days. There are nine labeling timepoints. 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, **Figure 5 D**. In this mode, each of the mzML and mzid files and the corresponding labeling duration time and body water enrichment is entered separately. Click on the “Browse” button for the mzML file to find an mzML file. Click “Open” to place it on the Box. Do the same for the corresponding mzid 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. In the Manual Input mode, the file names for mzML and mzid files do not need to match, e.g., OneFIle.mzML can be matched with SecondFile.mzid.

The “Sort” button will sort the input data and order it in a sequence of 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 isotope labeling durations. For example, if the body water enrichment is 7%, enter 0.07 into the box under “BWE”. “Rate Constant” method allows two options: one-parameter (determine the degradation rate constant) and two-parameter (determines the degradation rate constant and the asymptotic enrichment).

Peptide consistency (**Figure 5 C**) is four. It means that only peptides that have been identified and quantified in at least four out of the seven timepoints of label exposure (in this example) will be used in the estimation of rate constants for proteins.

Peptide score (**Figure 5 C**) 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 is 20 PPM.

The GUI creates the files.txt and quant.state files, 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, **Figure 5 C** (in this case E:\GUI_test). After the data have been entered, the output directory has been chosen; the quantification is started by clicking on 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.

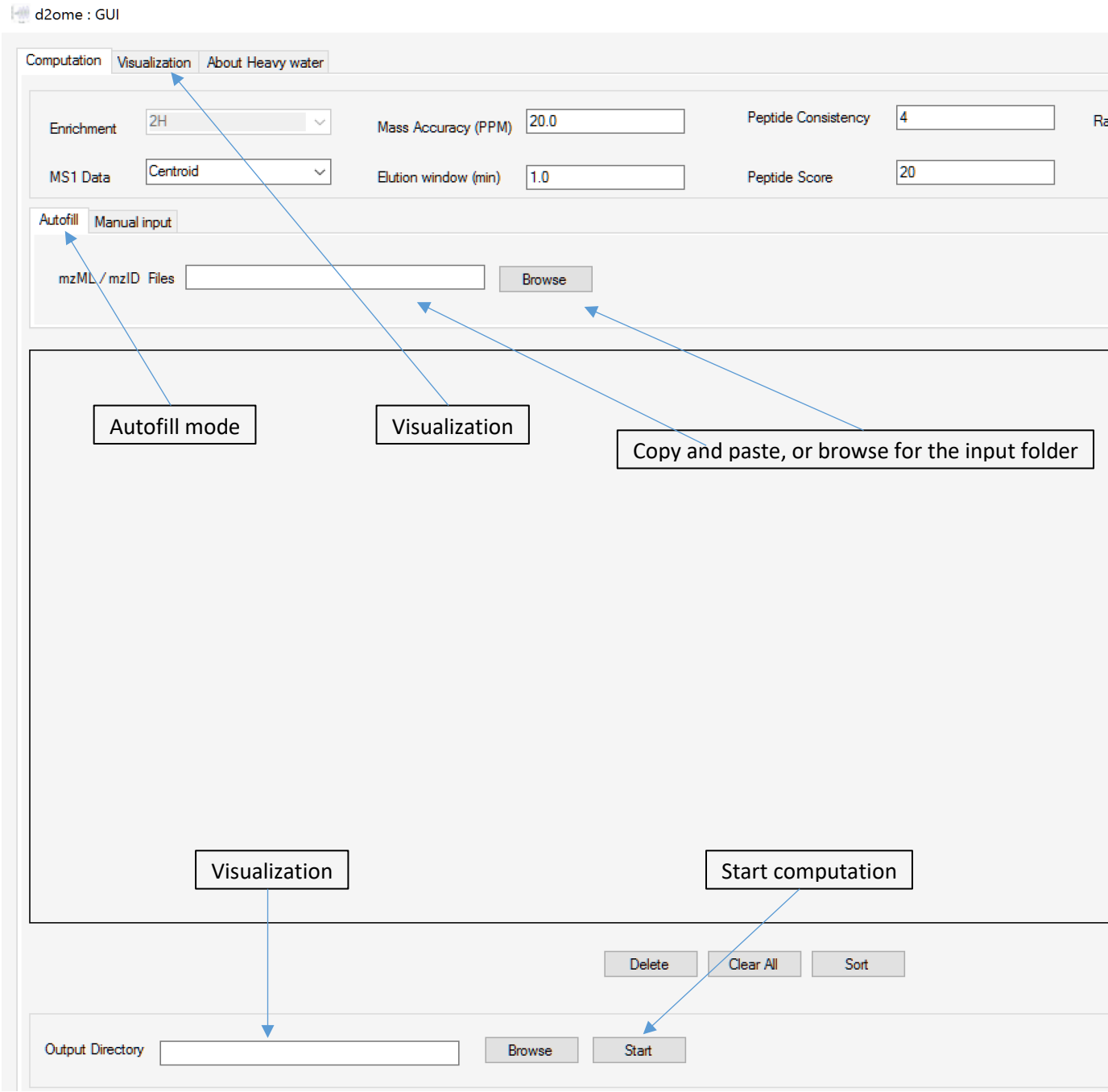


Figure 5 A. d2ome GUI before data initialization.

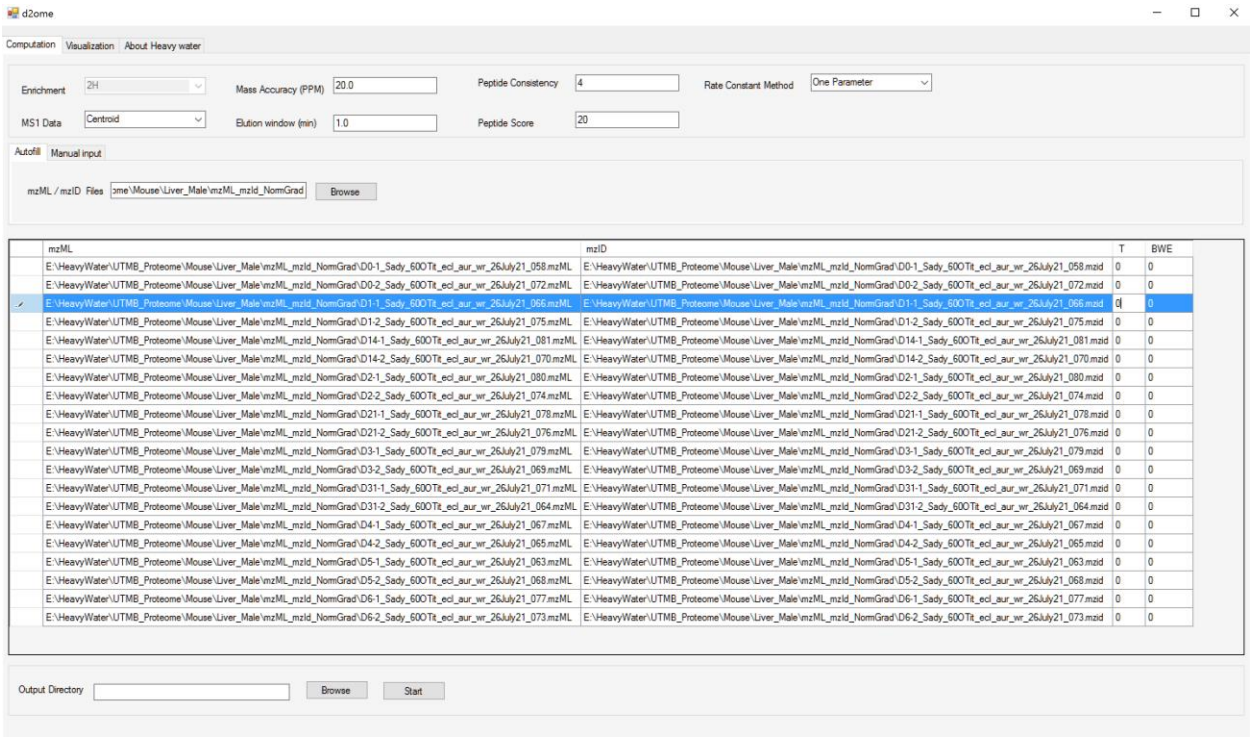


Figure 5 B. d2ome GUI in the Autofill mode.

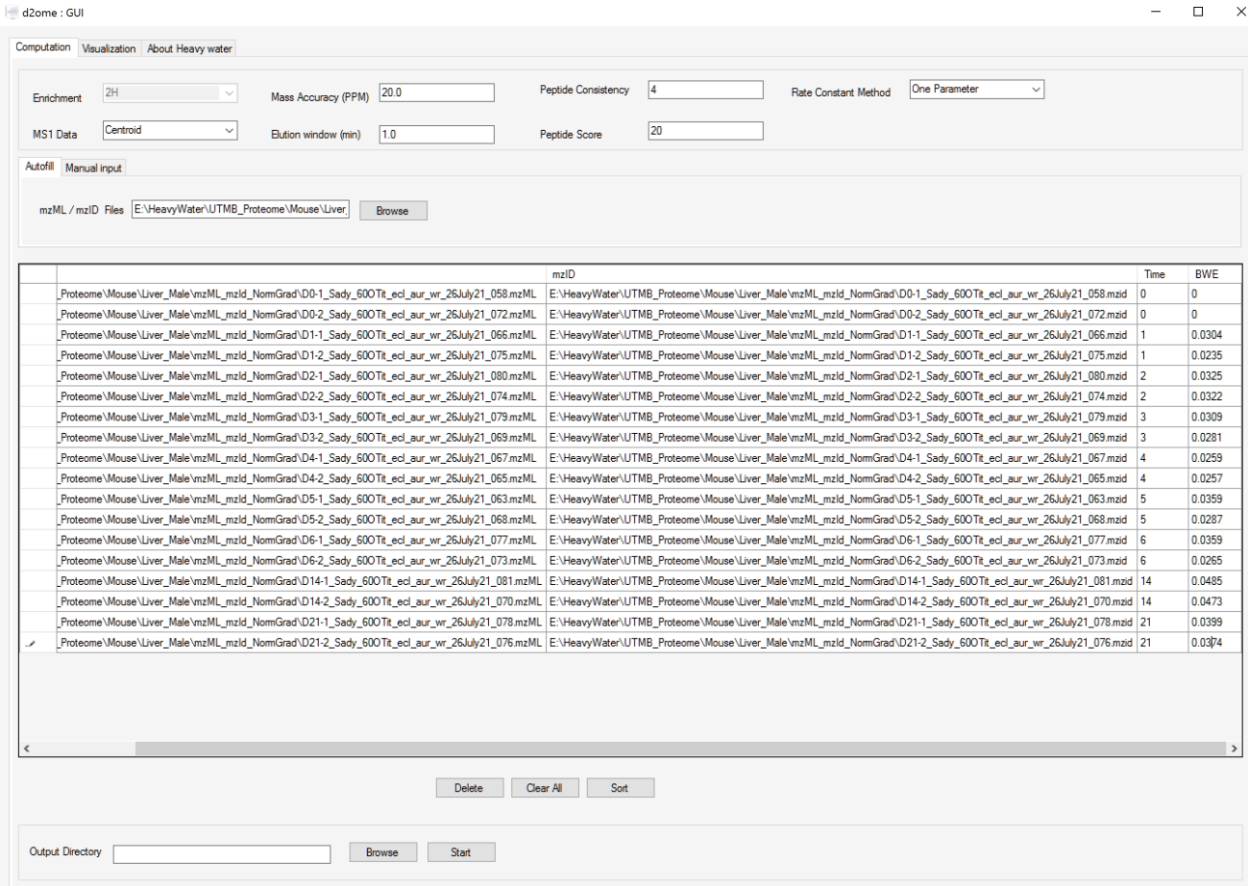


Figure 5 C. d2ome GUI with Autofill mode after filling labeling duration (Time) and body water enrichment (BWE) data for each experiment.

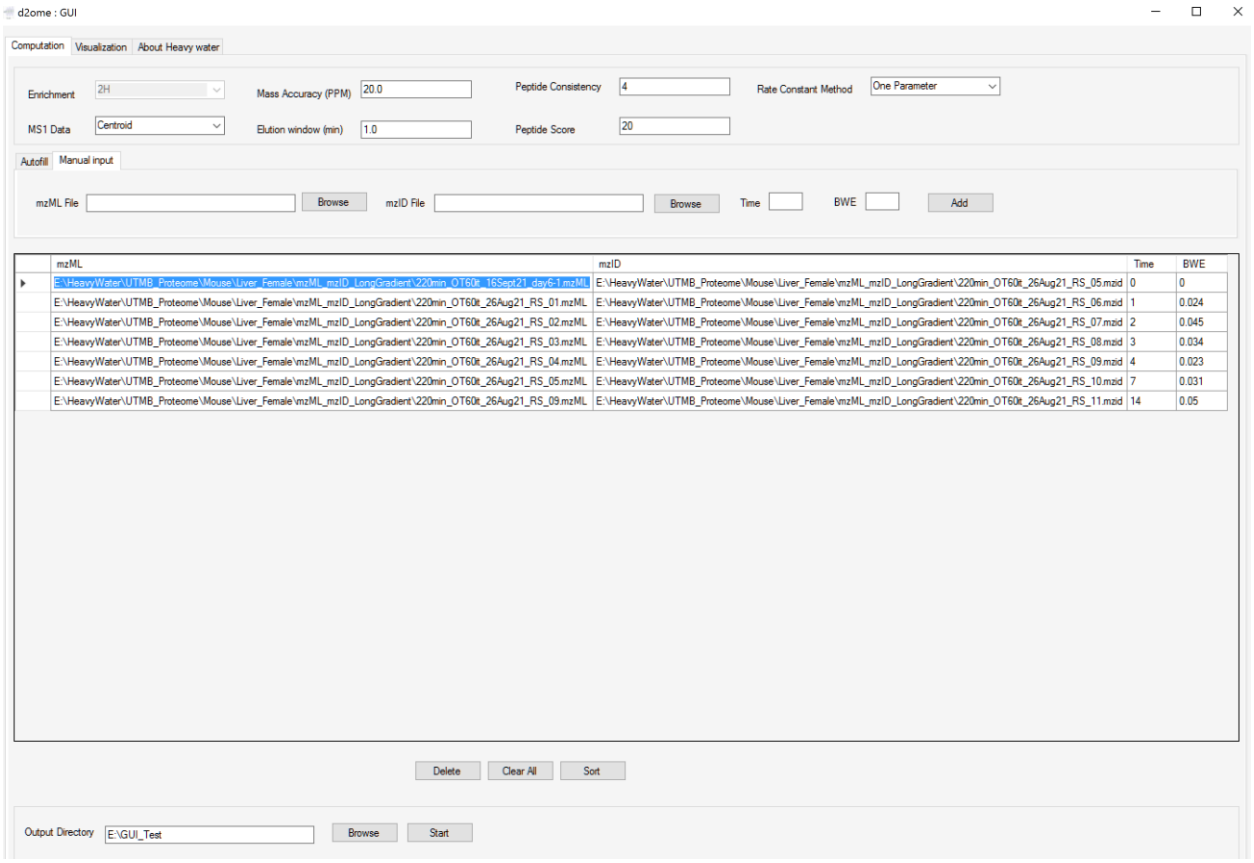


Figure 5 D. d2ome GUI with Manual input mode.

6. Output.

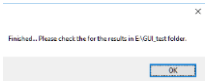
All outputs are reported in csv formatted files. For each protein that passed the thresholds specified in quant.state file, two main output files will be created: ProteinAccession.RateConst.csv and ProteinAccession.Quant.csv.

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), correlation between the fit and experimental data, root mean squared error, and absolute deviation between the theoretical and experimental isotope profiles (before the start of labeling).

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

7. Visualization.



The new GUI allows visualization of results. 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. A figure like the one in **Figure 6 A** 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 A. Visualization of results for protein, 1433B_MOUSE.

8. Citation

The data processing and modeling implemented in d2ome was described in[2]:

Sadygov RG**,** Avva J, Rahman M, Lee K, Sergei Ilchenko S, Kasumov and Borzou A., 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(11):3740-3748.

Mass isotopomer dynamics is presented in:

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.

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. Up to 15 experiments can be submitted for proteome dynamics study. There is no limit on the number of experiments when the program is called from the command line. 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

8. 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).

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.