DynaMet DOCUMENTATION

## General

DynaMet provides a fully automated workflow for liquid chromatography mass spectrometry (LC-MS) raw data analyses allowing for metabolome-wide investigations of dynamic isotope labeling experiments. DynaMet enables untargeted extraction of labeling profiles by grouping metabolite features in different samples with isotopic patterns changing over time. Moreover, integrated tools for expressive data visualization enhance result inspection. DynaMet was developed for Python based LC-MS data analysis framework eMZed 2.

## Installation

DynaMet requires eMZed 2.3.1 or higher. If you have not yet installed eMZed you can find latest emzed2 version including installation instructions [here](http://emzed.ethz.ch/installation.html).

After eMZed installation you have to install two additional packages ‘hires’ and ‘pacer’. To install the packages start eMZed and type into IPython console:

 press enter

 press enter

Dynamet is distributed as eMZed extension called package or apps. To install the package type emzed.project.install\_wheel() into eMZed IPython console.



After installation open a new IPython console and DynaMet can be started by entering the command emzed.app.dynamet().

## Requirements

***Samples and Experimental Design:*** Data analysis needs a series of samples from a time course labeling experiment where the carbon source is shifted from natural labeled carbon source to 13C labeled carbon source. Details about the design of such experiments for different cell types can be found in the literature.

***LC-MS Data:*** The workflow was developed for high mass resolution spectra. Since peak grouping is based on element specific m/z value differences of isotopic peaks it requires high mass resolution spectra with mass accuracy of at least 0.008 for those delta values. Quality of isotope analysis will increase with increasing separation of mass isotopic peaks mass and thus with mass resolution. We tested the approach on an Orbitrap instrument acquiring at R = 60000. However the tool comprises a routine called ‘suitability test’ to evaluate whether the data set is suited for the workflow. In general, best results will be obtained when measuring the complete sample set in a single batch on a new column. DynaMet accepts LC-MS data in mzXML and mzML format (peakmaps). If you do not have a converter for MS data of your instrument you can find more information [here](http://en.wikipedia.org/wiki/Mass_spectrometry_data_format#cite_note-autogenerated1-26). **Before starting DynaMet create a new project folder and paste the peakmaps of your data set into the folder.**

# Running DynaMet with eMZed 2



Start eMZed and type emzed.app.dynamet() into command line of the iPython shell and press *Enter* .



The DynaMet main window will pop up.



First, always choose the folder containing your LC-MS data (peakmaps) by pressing on the folder button (top, right) and select your project folder.

## Configure DynaMet Parameters:





***Feature detection***

*Change\_pipeline\_config -> feature\_detection*



Feature detection configuration comprises for different parts. Only selected parts are displayed for parameter modification.

1. **peakmap\_processing :**
   1. ignore\_blanks: all sample files (peakmaps) containing label blank, 'Blank' or 'BLANK' in filename (\_label\_) or at the end of filename (\_label.) are ignored.
   2. orbitrap\_data: Allows removing shoulder artefact peaks around intense peaks. Intense peaks in mass spectra not only show the detected mass but a number of so-called side-lobes (A. Kaufmann, P. Butcher, K. Maden, S. Walker, M. Widmer. Analyst 2011, 126, 1898.) or shoulder peaks. Those peaks significantly hamper analysis. If you choose the Orbitrap data options those peaks are autamically removed from spectra.
2. **Alignment:**
   1. ***retention time:*** uses emzed rtAlign module providing clustering based retention time alignment algorithm developed by Lange (Lange E, et al. A geometric approach for the alignment of liquid chromatography-mass spectrometry data. Bioinformatics. 2007;23:I273–I281)



***Parameters:***

*rt\_alignment*: Retention time alignment will be performed if values is set to on 'on'. You can switch it off by selecting 'off'.

*maxMzDifferencePairfinder*: max allowed difference in m/z values for pair finding.

*mz\_diff*: max allowed difference in m/z values for super imposer.

*rt\_diff*: max allowed difference in rt values for searching matching features.

* 1. ***mass alignment:*** Performs affine linear mz-correction for a feature table. The approach is based on spiking mass calibrants into all samples to correct for mass drifts occurring overtime. This table needs columns ``mz\_hypot`` for the m/z value calculated from the mass of the isotope, ``rtmin``, ``rtmax`` for the retention time window where the peak is expected to elute from the column in order to restrict the match of the table against the ``mz\_reference\_table`.



Parameters:

*mz\_alignment*: Mz alignment will be performed if values is set to on 'on'. You can switch it off by selecting 'off'.

*process\_calibration\_table*: For mz calibration, a table for targeted extraction is needed. You can choose between 'default', 'load different', 'inspect / modify', and 'build new'.

* Default is the calibration table of current config. If you run the tool for the first time, default table is as shown below.

****

**Figure:** Default mass calibration table. A mass calibration table needs columns ``mz\_hypot`` for the mz value calculated from the mass of the isotope, ``rtmin``, ``rtmax`` for the retention time window where the peak is expected to elute from the column in order to restrict the match of the table against the ``mz\_reference\_table`.



* 'Load different' allows loading any table. When selecting a dialog window occurs which allows selecting a mass calibration table. By default, an empty table is provided.

When pressing the ' help' button requirements to mass calibration tables are specified:



If you choose the empty table, the tool continues with 'build\_new'. Else it continues with the inspect / modify part. This avoids choosing calibration tables, which contains not the compounds spiked into the samples.

* 'inspect / modify': This allows you adapting retention time windows of the calibration table to samples of your data set. Select a peakmap to check for retention time windows by pressing the right bottom.





After choosing a peakmap, an information window pops up specifying how to adapt retention time windows within Tables.



To continue -> press ok.



**Figure:** Calibration table with selected peakmap.

After pressing ok table explorer opens calibration table with selected peakmap and you can inspect and modify retention times. We recommend removing all peaks from calibration table where no match is found. In addition, it is possible to delete rows of the table or to add compounds by cloning rows. To this end right click on button at the beginning of the row of interest and a window pops up. Choose the 'Clone row' or 'Delete row'. The right click also allows undoing manual manipulations.

'build new': This option allows building a new calibration table from scratch. If you build a new calibration table we recommend to choose the compounds in a way that it covers the complete m/z range measured and that number of calibrants fulfill the minPoint criterion, e.g. if minPoint is 5 you have to define at least 5 peaks in your calibration table. When selecting a window pops up afterwards remembering minimal requirements to the calibration table.



To continue -> press ok.



An empty Table pops up where you can enter the compound values manually. If the number of provided rows is not sufficient, you can add additional rows by right clicking the button at the beginning of each row. To continue simply close the table (click on x in upper right corner).

Remark: rtmin and rtmax values are calculated in seconds, but are automatically displayed in minute format. Thus if you enter 60 in a rtmin cell and press enter, the value 1.00m is displayed. You can enter the same value in minutes directly by typing 1.0m.

*To continue -> close the table.*

Further steps are now as describes under 'modify / inspect' (see above).

*mztol:* maximal tolerated mass difference to match a peak with calibrant (units)

*minR2*: Stop criterion when removing outlier points. Values are in Range [0 - 1].

minPoints: Minimal number of points for calibration curve fitting. You need at least as many peaks in your calibration table!

*maxtol*: Maximal tolerated mass deviation after calibration. Stop criterion when removing outlier points.

*interactive*: if 'True' manual inspection and data point removal is enabled. For automatic data processing this is not recommended!

1. **Peak\_detection:** Extracted ion chromatogram peaks (EIC-peaks) are detected using openMS feature finder metabo (Kenart E., et al. Molecular & Cellular Proteomics. 2014; 13: 348-359.)



**TIP:** You obtain direct help for a selected parameter when remaining with mouse pointer on parameter for several seconds.

* 1. Default setting display a reduced number of parameter settings that allow adapting the finder to your LC-MS system



*common\_noise\_threshold\_int*: intensity threshold below which peaks are regarded

as noise. For given example all peaks with max intensity below 1000 counts are ignored.

*common\_chrom\_peak\_snr*: minimum signal-to-noise a mass trace should have. For instance the definintion of LOD is snr>3.

*common\_chrom\_fwhm*: typical peak width (full width at half maximum).

*mtd\_mass\_error\_ppm:* Allowed mass deviation. The value should not be confused

with mass\_accuracy. It corresponds to the boundaries of mz values (mzmin detected, mz mass detected) observed for exctracted ion chromatographic peaks.

*mtd\_reestimate\_mt\_sd*: enables dynamic re-estimatation of m/z variance during mass trace collection state. It is recommended to use this parameter.

*epdet\_width\_filtering*: enable filtering of unlikely peaks width. If 'on' tool filters peaks with the 5% and 95% quantiles of the peak width distribution.

* 1. **This option is only suited for experts which are familiar with the peak detection approach.** Besides parameters mentioned under a further parameters are provided for modification.



*mtd\_trace\_terminaton\_criterion*: Termination criterion for the extension of mass traces. In `outlier` mode, trace extension cancels if a pre-defined number of consecutive outliers are found (see *trace\_termination\_outliers* parameter). In 'sample\_rate` mode, trace extension in both direction stops if ratio of found peaks versus visited spectra falls below `*min\_sample\_rate*` threshold.

*mtd\_trace\_termination\_ouliers*: mass trace extension in one direction cancels if set value of consecutive spectra without detected peaks is reached.

*mtd\_min\_sample\_rate*: Minimum fraction of scans along the mass trace that must contain a peak.

*mtd\_min\_trace\_length:* Minimum expected length of a mass trace (in seconds).

*mtd\_max\_trace\_length:* Maximum expected length of a mass trace (in seconds).

*epdet\_width\_filtering:* Same as described for default settings except thatin addition to 'auto' mode (called on in default) a fixes mode is also possible.

*epdet\_min\_fwhm:* Minimum full-width-at-half-maximum of chromatographic peak (in seconds). Ignored if parameter *epd\_width\_filtering* is off or auto.

*epdet\_min\_fwhm:* Maximum full-width-at-half-maximum of chromatographic peak (in seconds). Ignored if parameter *epd\_width\_filtering* is off or auto.

*epdet\_masstrace\_snr\_filtering:* apply post-filtering by signal-to-noise ratio after smoothing.

1. **Feature\_grouping:** a graph based feature grouper for high-resolution MS data. Feauture\_grouping is a core development of the pipeline since it is a key step to detect features with significant labeling incorporation.



*isolation width*: Maximal m/z difference between mass traces of different samples tolerated as same m/z value (value in U).

*Charge\_lower\_bond*: Minimal charge state z to consider.

*Charge\_upper\_bond*: Maximal charge state z to consider.

*max\_c\_gap*: Maximal allowed carbon gap width n. Here, n corresponds to a multitude of the mass difference of 13C and 12C times the charge state of the feature. Depending on chosen labeling strategy the maximal distance correspond to the number of carbon atoms of a metabolite. In case of one-carbon compound as sole carbon source max\_c\_gap is one for most core metabolites and corresponds to the size of precursors originating from central metabolism; e.g. biosynthesis of ATP requires ribose-5-phosphate which is fully labeled before significant label accumulates in ATP leading to a max\_c\_gap of 5 for ATP. However, the higher max\_v\_gap the higher the probability of mismatches.

*rel\_min\_area*: Lowest peak area relative to the base peak area accepted.

***5.Identification***

*Change\_pipeline\_config -> identification*





***Parameters***

*instr\_linear\_error*: Absolute instrumental linear error on area measurement 'e.g. LTQ-Orbitrap classic instrument 0.03 (3 %).

*idms\_sample*: Select whether data set contains an isotope dilution sample, which is a sample composed of 1 : 1 mixture of a natural labeled cell extract, and a cell extract from cells cultivated on [U-13C] labeled carbon source for at least 5 generations. Ideally, carbon source is 99% [U-13C] and thus all metabolites are labeled as carbon source. In addition, both cell extracts originate from the same strain or cell line cultivated applying same growth condition.

*c\_source\_labeling*: Enter value for labeled fraction of applied substrate for dynamic labeling experiment, e.g. 0.99. The correct value is important for the estimation of the number of carbon atoms.

*data base:* Choose an emzed integrated data base, or use your own data base. Assure that your data base is compatible with pipeline. Dynemet provides emzed integrates data bases KEGG, Pubchem, and human metabolome data base. In case you are choosing 'other' a Dialog opens which allows for loading your own data base:





Database formats table and csv are accepted. The column names 'mf' containg compounds molecular formula and 'm0' containing corresponding monoisotopic mass are mandatory. If you load one data base in csv format the file requires column name is in the first row of your data base. Column names have to be unique ( don't forget to switch file type filter to \*.csv). Currently, identification is only based on exact m/z values combined with possible adducts to assign a data base entry to a feature. In case features could be grouped by assigned adducts corresponding mass value is used for db assigning.

***Feature\_analyis***

*Change\_pipeline\_config -> data\_analysis*





Data analysis parameters define minimal quality criteria for feature selection.

***Parameters***

*min\_labeling*: threshold value for feature selection: Minimal labeled carbon fraction of one feature measured in at least one sample out of complete set of samples. If not fulfilled feature will be excluded

*feature\_frequency*: threshold value for feature selection: frequency of feature occurrence in sample set. Example 0.5 : feature is detected in at least every second sample. If not fulfilled, feature will be excluded

*max\_nrmse*: upper limit value for feature selection: Maximal normalized root-mean-square-error accepted for labeling profile fitting. If not fulfilled, no feature labeling curve analysis

***Suitablity\_test***

*Change\_pipeline\_config -> suitability\_test*

Configuration and application are desc

## Suitablility Test:

Dynamet's feature grouping algorithm relies on high mass accuracy data. Based on accurate m/z distances of mass isotopologues, feature grouper proposes possible elements (CHNOPS) that explain measured mz distances. Currently the grouper has only been tested in the negative ESI mode. Suitability test verifies for a set of compounds whether data fulfills requirements or not. The tool provides a default set of compounds (features) but data can also be tested for any user-defined compounds. Compounds must be natural labeled and at least two mass isotopologues for each compound should be measured in the sample. We recommend testing at least 10 different features distributed over complete m/z range measured.

**Configure suitability test**

To modify the setting click '*change\_pipline\_config*' -> '*suitablility\_test*'



The test takes into account the mass resolution of your instrument. Enter the mass Redolution R you applied for the measuring your samples. In case of Orbitrap instruments resolution R is calculated for each m/z value from unit Resolution. To do so the 'R at mz' parameter is required. For Instrument of the QExactive series R is defined at m/z 200 whereas in case of all other Orbitrap instruments it is defined at m/z 400. The parameter is ignored if field *'Orbitrap\_instrument*' is not selected.

**Run suitability test:**

### CHOOSE TEST\_TABLE

To run the test click '*suitability test*' in the main window



As test table you can use the current table of pipeline configuration (the default is build from compounds of mass calibration table), or load a table from folder. If you load your own table required tale columns are:

*id*: unique identifier; *name*: compound name (int);

*mf* : molecular formula (string);

*mf\_ion*: molecular formula corrected for adduct, example in case of M-H a H is substracted from formula (string);

*rtmin, rtmax*: bounderies of retention time window within peak elution is expected (float);

*adduct\_name*: name of adduct (for adduct assignment the Metlin nomenclature is applied: [M-H]- -> M-H) (string);

*polarity*: applied ESI mode: +/- (string);

z: charge state of ion (z>=1, int)



**Figure??**: Example of a table for suitability test

**Build a new test table from scratch**

Choose *build new* and press ok.



An empty Table pops up where you can enter the compound values manually. If the number of provided rows is not sufficient, you can add additional rows by right clicking the button at the beginning of each row. To continue simply close the table (click on x in upper right corner).

**Remark:** rtmin and rtmax values are calculated in seconds, but are automatically displayed in minute format. Thus if you enter *60* in a rtmin cell and press enter, the value *1.00m* is displayed. You can enter the same value in minutes directly by typing *1.0m.*

Next for each compound a window pops up that allows choosing adduct observed for compound. If you want to use more than adduct for the same compound you have to enter it twice in table with different id.



### Choose peakmap (sample) from data set



In the next step, all peaks specified in test-table are extracted from sample file. Choose a file containing all or at least most of compounds. When the *inspect peaks* is selected, a table with peak extraction results will pop up. We recommend to always check peak extraction since it is possible to adapt retention time windows manually. A pop up window will remind you that.





To modify the retention time window lef click on the boundery dot and keep the mouse botton pressed. You can now move the window boundary to the left or right. As soon as the window is placed correctly, press on the integrate button and rtmin and rtmax will be set to boundary values. All peaks that cannot be detected are removed automatically:



**Accurate peak selection is mandatory for reliable results !!**

**TIP:** you can also add or remove ions at that level by cloning or deleting rows. To add the new compound, enter mzmin, and mzmax values of the ion, select the peak with the integration bounderies, reintegrate it and, correct all other values ( adduct\_name, mf,….).

### Start test

The test will start automatically. A pop up window will inform you about the results



In addition details of the test result will be displayed in emzed iPython console.



## Run Analysis:

Remark: To get comfortable with DynaMet the package also contains a reduced example data set ‘Dynamet\_test\_example.zip’ comprising 9 mzXML files originating from a dynamic 13C labeling incorporation experiment of *Bacillus methanolicus* MGA3 grown on methanol.



Before starting analysis make sure that (i) the correct project path is selected, (ii) sample files are of type mzXML or mzML and belong to a time series of a dynamic labeling experiment, and (iii) all parameters are chosen correctly. If you are not sure whether your LC-MS data fulfill the minimal criteria, perform a suitability test first. To start the workflow, press the button 'run analysis' in the main window. When running the analysis for the first time workflow will request additional data.



From the data LC-MS data files (peakmaps) a list is creacted and you can select those samples belonging to the dynamic labeling experiment.

*To continue -> press ok*



Next, a window pops up. After pressing 'ok' a table pops up which allows defining sample order and sampling time points of each sample. Time points are calculated in seconds. Default time values are 0.0s, default sample order is derived from sorted sample file names (ascending)



For given example, time points are:



After finishing, close the table.

If you chose the option *Change\_pipeline\_config -> identification -> idms\_sample,* a further selection windows pops up which allows choosing an idms sample (mixture of natural labeled and uniformly 13C labeled cell extract).



To start DATA ANALYSIS press ok.

When analysis is done a window pops up informing about end of process.



**Remarks:** DynaMet uses 'pacer' a lightweight Python package for implementing distributed data processing workflows. In general, it manages, enhance and accelerate data analysis. For details see [here](https://pypi.python.org/pypi/pacer). Since all intermediate results from processing steps are cached, pacer is able to determine which steps are concerned by parameter change and re-running the analysis only executes those steps.

## Inspecting results



**Figure:** Presentation of results. All results are combined in a single table including plots, identification results or details of selected features.

Dynamet creates an explorable emzed Table object for all extracted features (see Figure). Plots and subtables can be opened by double clicking on the cell. Identification results and feature details are provided as feature-wise subtables. If you use implemented data base e.g. Kegg identification results contain column with direct link of of assigned compound to pubchem data base. Plots show mass isotopologue distribution heatmap and if possible a fit curve of labeling incorporation into metabolite pools.



To inspect results press the inspect\_result button.

A dialog box opens which allows you to select the latest result table (current) or to choose a result table from all result tables in the project RESULT folder (DynaMet automatically adds date and time to the result file name with while saving. Older results are not overwritten).





When Choosing **all** a box opens that allows you select result file out of all present in the result folder.



**Explaining the result table columns:**

***feature\_id***: Unique identifier number for each feature.

**adduct\_group**: Co-eluting features with m/z value differences which can be explained by different adducts of the same compound are grouped.

***z***: feature charge state.

***rt***: feature retention time in minutes.

***mz0****:* Mass of monoisotopic peak. Value is only assigned if feature was detected in natural labeled sample.

***min\_mz***: Lowest m/z value of one feature out of all samples .

***possible\_m0***: value calculated if adduct assignment was possible and m/z 0 was determined (see above).

***num\_c***: Number of estimated carbon atoms (for details see *details* subtable).

***flcluster\_id***: Features are clustered by fitting parameters t50 and std\_c\_13\_fraction . Grouped features have the same *flcuster\_id*

***label\_t50\_sec***: Time constant T50 (time required to reach 50 % of carbon atoms being labeled) resulting from parameters of first order fitting curves.

***std\_label\_t50\_sec***: Standard deviation of estimated time constant T50.

***std\_c13\_fraction\_calc***: gain k of first order fitting curve.

***pool\_t50\_sec***: Time constant T50 (half life time) of metabolite pool based on resulting from parameters of first order fitting curve of isotopologue M0 depletion.

***std\_pool\_t50\_sec***: Standard deviation of estimated pool half life T50.

***nrmse***: Normalized root mean square error of fitting curve.

***fit\_model***: Two first order fitting models are applied by workflow: logistic and pt1.

***dli\_label\_plots***: Dynamic labeling incorporation plots.



For each feature a heatmap is provided showing distribution of mass isotopologue abundances corrected for natural labeling. Mi correspondsd to ith mass isotopologue and Sj to the jth sample. If fitting was possible, a fitting curve is provided showing the number of incorporated labeled carbon atoms over time.

***M0\_dilution\_plots***: Monisotopic isotopologue dilution plots.



If fitting was possible for a feature a fitting curve is provided showing quaktiy of M0 depletion fit.

***feature\_clustering\_plot***:



Feature clustering plot provides an overview of grouped features.

***details***: Subtable with detailed information of one feature:



Details subtable provides feature-wise detailed results. The columns are:

*adduct\_group*: Co-eluting features with m/z value differences which can be explained by

different adducts of the same compound are grouped.

*possible\_adducts*: Assigned adduct e.g. M-H, M-2H

*feature\_id*: Unique feature identification number.

*mz*: Measured m/z value of mass peak.

*mzmin:* Lower boundary of feature peak mass trace .

*mzmax:* Upper boundary of feature peak mass trace

*z*: charge state

*possible\_m0*: value calculated if adduct assignment was possible and m/z 0 was determined

*rt*: feature retention time in minutes

*rtmin*: Lower retention time window boundary

*rtmax*: Upper retention time window boundary

*fwhm*: Full width half maximum of peak

*method*: Peak integration method; emzed provides a number of different integration

algorithms by default the tool applies exponential modified gauss EMG integrator (emg\_exact) and trapez.

*area*: Peak area counts \*s

*rmse*: root mean square error of peak area detect

*peakmap*: LC-MS underlying sample peapmap containing all measured spectra. Double

clicking opens peakmap explorer:



Details about how to work with peakmaps can be found [here](http://emzed.ethz.ch/tour.html#working-with-peak-maps).

*source*: File name of the peakmap.

*mz0*: m/z value of monoisotopic peak

*min\_mz*: Lowest m/z value of one feature out of all samples

*time*: Sampling time point. Tipp if you want to display an overlay of all feature peaks of one sample choose ‘expand selection by: time’ and click on one row to select all peaks of the same time.

*order*: Sample order as entered by the user

*num\_isotopes*: number i of mass isotopologue Mi

*num\_c*: Number of estimated carbon atoms of compound (for details see *q\_score*).

*min\_num\_c*: Lower boundary of carbon atoms estimation

*max\_num\_c*: Upper boundary of carbon atoms estimation

*q\_score*: Quality scoring value of carbon number estimation. Up to four different approaches are used and combined estimate the number of carbon atoms see below

*origin\_of\_c\_estimation*: (1) **by\_nl**: From mass isotopologue distribution of natural labeled compound. (2) **by\_idms:** If an idms sample is in the data set, the m/z distance between monoisotopic mass peak M0 and uniformly 13C labeled peak MUL is used (by\_idms). The fraction of 13C in labeled carbon source is taken into account. The approach is most reliable for substrates with uniformly 13C labeled fraction > 95%. (3) **by\_dli:** The number of highest Mi detected defines the lower limit *min\_num\_c*. (4) Finally, global carbon atoms bounderies are estimated from provided database by matching molecular formulas within a mass range of 25 units. If no database was provided Pubchem data base is used pubchem db

*mi\_fraction*: Area fraction of ith mass isotopologue of total feature area.

*mi\_frac\_corr*: mi\_fraction corrected for natural carbon isotope abundance.

*no\_C13*: Number of labeled carbon atoms detected in feature at time x.

*C13\_fraction*: Fraction of 13C in compound at time x.

***identification\_results***: A sub-table containing most feature information joined with data base. In case data bases provided by emzed were selected identification table provides urls to Pubchem DB. You can click on the link to open corresponding Pubchem entry.

