User Manual

FluxY: *iMS2Flux.pl*

(v.7.1a)

Revision: July 1, 2013

Table of Contents

1. Overview

iMS2Flux was developed to facilitate high throughput isotopic labelling experiments quantified by mass spectrometry (MS) techniques emphasizing ^{13}C metabolic flux analysis (MFA).

> *However the software may be used to verify and correct any MS data from any isotopic labelling experiments, regardless of the final application.*

iMS2Flux is the core tool in the FluxY toolset[1] and as such was written to be platform independent (running on any platform with Perl support). This program combines tools to automatically verify and correct mass spectrometry (MS) data from isotopic labelling experiments. The corrected data can then be output to a generic format for general use or output to a format ready for direct inclusion with a variety of 3rd party ^{13}C-MFA software, specifically 13CFLUX, OpenFLUX and 13CFLUX2.

The software is written in Perl and has been tested on Windows, Mac OS, Linux and Unix systems using Perl 5. The software is provided uncompiled, meaning have full and direct access to the source code and may modify or extend it to suit your needs. It also means that beyond the tested systems it should run on any computer system supporting at least Perl 5[2].

To quickly get started with the software please see the getting started guides, including how to download and install the software (and, if necessary, Perl), which are provided for Windows (XP, Vista and 7), Mac OSX, Linux and Unix. Also see the included example for processing a set of amino acid measurements.

This software is run from a command line, but includes an optional GUI for Windows. For ease of use the majority of configuration options are provided via a configuration file.

- The GUI is available as a separate download from the same web site.
- The configuration file is completely presented in Appendix C, and each of its parts are explained in detail in the corresponding sections (2-5).

Program Summary

This program accepts input data from one or more files, each of which contains one or more data sets, chromatograms or samples respectively, of MS measurements. Each data set contains information on mass isotopomer measurements for the given compounds, and depending on the format, may contain other information as well. If multiple data sets represent biological or technical replicates they may be treated as individuals or averaged.

- For maximum flexibility within the software the main program treats operates on fragments (as opposed to compounds which may or may not be fragmented), i.e. compounds or techniques that produce no fragmentation are processed as a compound with one fragment.

1. The FluxY toolset is a group of tools for automating steady state, dynamic and kinetic MFA.
2. If not already installed you must first install Perl on the system to be used.

The software currently has built-in support for six compound classes, including derivatization method(s) where applicable. These compound classes are summarized in Table 1.

- Details on supported compounds can be found in their compound data file, **`DataClass_XX.pm`** - where XX refers to the two letter compound class identifier, located in the **`FluxY_Lib/MS`** sub-directory.

A summary of this data can be generated for each of the compound classes, see Section 6.

After data input comes data verification, where the MS data is subjected to a variety of data checks (see Section 3 for details). If any data does not pass any of the checks a flag file is generated flagging the data that failed. At this point the program stops to allow the problematic data and/or the chromatogram to be reviewed, revised or even removed, and the data check may be optionally repeated before continuing.

- This flag file is located in the same directory as the data file, and is formatted in a simple tab separated values text file which can be easily imported into a spreadsheet.

Table 1: Compound Classes

Compound Class Identifier	Compound	Supported Derivatizing Agents	Compounds / Fragments
AA	Amino Acids	TMBDS[1]	multiple (20) / multiple (7)
CW	Cell Wall Compounds	--	multiple (34) / single (1)
FA	Fatty Acid Methyl Esthers	--	multiple (10) / multiple (4[2])
GL	Glucose	Meox[3]+TMS[4] Meox[3]+Ac2O[5]	single (1) / multiple (6) single (1) / multiple (4)
GY	Glycerol	TMS[4] TFAA[6]	single (1) / multiple (3) single (1) / multiple (3)
SM	Soluble Metabolites	--[7]	multiple (81) / single (1)[7]

1. t-butyl-trimethylsilyl
2. Full molecular ion and the McLafferty fragments
3. methoxyamine
4. trimethylsilyl
5. 2,3,4,5,6-Penta-O-acytl-D-glucose-O-methyloxime
6. Acetic acid, trifluoro-,1,2,3-propanetriyl ester
7. The soluable metabolite class includes compounds derivatized with TMS and/or Meox+TMS, and it also includes specific fragments where available. However the purpose of this class is to process any group of these soluable metabolites despite not having all of the same properties. As such this class was written to treat each soluable metabolite as a single compound without generalizing for derivatization or fragmentation.

- The flag file is formatted to match the default data input/output format so as to facilitate overlay of data for easy analysis.

Once the data has been accepted, the software normalizes the data and optionally performs one or more correction steps (see Section 4 for details).

Lastly the processed data may be saved in a variety of formats, as individual or averaged data, or may be used to generate the specific measurement data for the inclusion in 13C MFA model files. See Section 5 for a complete description and list of supported files.

The configuration file is divided into four parts, Input, Data Checking, Data Correction and Output. Each part of the configuration file is detailed in the next four sections, with the last section describing the command line syntax, or how to run the software from the command line.

- A sample configuration file is provided with the software.
- Getting started guides including how to download and install the software are provided for Windows, Mac OSX, Linux and Unix. These are located in the `Getting_Started` directory.
- Also available in the `Getting_Started` directory is an example of a complete workflow using the automated chromatogram extraction tools provided with Waters QuanLynx™ software.

What's New in Version 7

Version 7.1 of this software contains two new experimental features in addition to a number of small bug and consistency corrections. As indicated by the minor revision number there is no change to the existing configuration file, with the new features currently supported by command line switches. These features are (1) support to print the standard deviations associated with the average carbon labelling option and (2) support for performing the correction for natural abundance omitting the backbone carbon atoms.

In addition, the actual correction subroutines were alterred to be more generally applicable. Specifically any hard coded reference to carbon as the labelled atom was removed from the natural abundance and original biomass calculations. These are now optionally specified in the appropriate function call, and default to carbon if not specified.

An indexing error was corrected in the correction for original biomass and specific data checks were added to the values in the OBM data file. In addition the data check error messages were generalized to identify data sources other than the configuration file.

Some of the numbered error/warning messages have been renumbered to enforce a consisteny among different programming sections.

The changes have been recorded in a change log, a new text file change.log, located in the same folder as the main program. Please see the log for a complete list of changes, and the affected perl perl modules.

The two new experimental features are accessible via command line only, and not through the Windows based GUI (which has not been updated).

2. Input

There are three distinct input phases:

1. **Command Line Arguments** (processed by `ParseCommandLine`).

Command line arguments are used to provide quick access to program options (such as help) or to over-ride regular operation (such as zeroing, rather than flagging missing data for input format QL).

- For a detailed explanation of the command line arguments see Section 6.

2. **The Configuration File** (processed by `ParseConfigFile`).

The configuration file provides settings for various program options and the names of additional files containing pertinent information. By default this file is named “config.txt” and is in ***the same directory*** as the data file(s).

- A sample configuration file is provided with the software.

3. **Input Data Files** (processed by `GetInputData`).

The primary input data file(s) contains the MS data. Additional data files may be optionally provided for a variety of information, such as original bio-mass data, the names of data sets, externally provided retention times and model file(s).

Each data set is composed of the same group of Mass Isotopomer fragments. The first mass of these fragments is assumed to be the M+0 mass isotopomer, and used to identify the fragment (by comparison to the fragment’s calculated mass).

For a ^{13}C labelling experiment, each fragment has a predetermined number of Carbon atoms (referred to here as numC) that may or may not be labelled. The software determines the end of a fragment by an out of order mass. The last mass of a fragment is M+numC. The software does not require the full set of mass isotopomers, however you may not drop a measurement in the middle of the fragment, only at the end. An information message is printed if less than the full set of mass isotopomers are found for a given fragment.

The program can also extract boundary data around each mass fragment. If included it can extract ’n’ values preceding the M+0 data point, i.e. M-n, and/or following the last carbon, i.e. the M+numC+n data points which may be monitored for their relative value. In this operation all fragments must have the same set of extra data points.

> *Warning: if boundary data is indicated in the configuration, but not actually present in the data, actual measurement data will be stripped off instead. This may leave the fragment unusable or generate excessive warnings about incomplete fragments.*

Supported MS File Formats

The program supports multiple input formats for the MS input file, however the options available for processing the data depend on the MS data provided.

- Details on supported input formats can be found in the appropriate parsing module '`Parse---.pm`' located in the **`FluxY_Lib/MS`** sub-directory.

Currently supported formats are:

TSV - Tab Separated Values (`ParseTSV.pm`)

This is a generic data format, used for both input and output. The basic TSV data file only contains one type of measurement value, and neither retention time[1] nor flags, so some options are not available. The TSV data table can be either horizontally or vertically oriented. The default is the vertical orientation, in which the first column has compound names, the second column has compound masses, with each column after this containing MS data from one chromatogram or sample.

- See Example_AA.txt, from the example in the Example_AA directory.

QL - QuanLynx (`ParseQL.pm`)

The program was originally designed to process high throughput MS data automatically extracted by Waters QuanLynx™ software and exported using their default report format. This format provides two data values, the **AREA** and **RESPONSE**, either of which may be used by the program, along with the retention time and detector flags.

> *When processing multiple measurement sets in Quanlynx, to be processed by this program, it is important to group together into sample lists data that represent experimental replicates, and to organize them such that replicates are provided in a consecutive order within the sample list.*

- For more detailed information on using this format, the complete workflow using the automated chromatogram extraction tools provided with Waters QuanLynx software is described in the included document; `MSto13C with QuanLynx and FluxY.pdf`, available in the `Getting_Started` directory.
- In particular see the appendix for compatibility requirements.

MAM - M. Agilent Macros (`ParseMAM.pm`)

Added to provide support for data exported directly to Excel from an Agilent MS using Agilent based macros (custom macros used at IPK-Gatersleben). This data must first be saved as a text file in TSV format (exported from Excel).

- This provides an alternative template for importing data where missing data is automatically filled in with zero values, even when in the middle of a fragment.

Configuration File: Part 1. Input

The first part of the configuration file consists of ten entries related to the input and setup of the MS data. There is one entry per line. for complete information on the configuration file syntax see Appendix C.

Line 1 contains a space separated list of MS data file names.

- At least one data file must be specified.
- If more than one data file is specified they must all be consistent, i.e. they must contain the same type of compounds, the same format, etc...

1. In this format retention time data may be supplied by an auxiliary file in the same format, TSV, where each value is a retention time instead of a measurement.

- If multiple files are specified, they are input in the order provided and the data is combined in a single matrix, no different than if the data came from a single file.
    - This is often a more convenient way to concatenate data from different experiments, than to do so manually.

Line 2 specifies the file type or format of the MS data file(s) specified in line 1.

- Each MS file type has an associated abbreviation, enter the abbreviation.
- Currently three formats are available, TSV, QL and MAM, as specified above.

Line 3 specifies the compound class associated with the MS data file(s).

- Each MS compound class has an associated abbreviation, enter the abbreviation.
- Currently six compound classes are available, see Table 1.

Line 4 specifies the derivatizing agent if used when processing the samples.

- This entry will be ignored if the compound class is not associated with a derivatizing agent (e.g. CW). In such a case it may be filled in or you may specify "None" to explicitly indicate no derivatizing agent.
- Each derivatizing agent has an associated abbreviation, enter the abbreviation.
- Currently 4 derivatizing agents are supported for three of the compounds, see Table 1.
    - Not all derivatizing agents can be used with all compounds, if an incorrect pair is specified the program will generate an appropriate error message.
- The current implementation of the software only supports specifying one derivatizing agent, however it is possible to utilize multiple derivatizing agents where some are always used, such as the use of Meox with glucose (GL, see Table 1) or in the case of fatty acids (FA) which are augmented by a methyl ester as opposed to be derivatized.

Line 5 is an option for the QL report format. In this format two measurement values are provided, the Response (R) and the Area (A).

- The response may be peak height or other defined measurement metric.
- If Line 2 is not QL, enter R here.

The next three configuration entries deal with experimental replicates. This information is used when calculating averages and standard deviations. If there are no experimental replicates, or you do not want the data averaged just enter the number 1.

Line 6 specifies the number of replicates in your MS data sets. There are two ways in which this may be specified:

- If the number of replicates is the same for each group of data sets, enter that as a single number.

    For example if you have 4 groups of 4 replicates for a total of 16 chromatograms/samples, enter 4.

- If the number of replicas are variable, then enter a space separated list of numbers that add up to the total number of replicates.

    For example if you have 3 groups of 3 replicates followed by one group of 5 replicates for a total of 14 data sets, enter: 3 3 3 5.

- The replicates will be grouped in the order of presentation of data in the MS data file.

  For the 3 3 3 5 example, the program assumes the first three data sets are one set of replicated experimental data, then the second set of three, the third set of three and the last five.

- If the data sets in the MS data file(s) are not in order either use a ‘1’ here (no replicates), or if they must be aggregated, then first rearrange the data sets so that they are presented in order.

When using a number of replicates other than 1, it is often desirable to associate an group name. The next two lines are used to optionally specify group names.

Line 7 is a simple Yes/No (Y,N) asking if you are going to provide an optional file with experimental names.

Line 8 specifies the name of this file.

- The file of replicate/experiment column headers is line oriented with one header per line. The number of headers must match the number of replicates, i.e. 1 header per set of replicates.
- This line is skipped if line 7 is N.

The last two configuration lines in this section specify any additional measurements taken outside of the number of mass isotopomers for each compound.

- All data to be processed must be consistent in the number of extra measurements.

Line 9 specifies the number of measurements preceding M.

- This line is labelled with M-1 in the configuration file, however you can specify any number preceding the M+0 measurement.
- Enter 0 for no extra measurements.

Line 10 specifies the number of measurements following M+numC.

- This line is labelled with M+numC+1 in the configuration file, however you can specify any number following the M+0 measurement.
- Enter 0 for no extra measurements.

3. Data Checking

Before performing any correction the program performs one or more data checks.

Data checks are integral to any form of automation and high throughput. Once processed it may not be possible to determine if there was poor data, translational mistakes or other errors. Further extending the automation to data extraction from the chromatograms adds additional distance between the user and the raw data, and with high throughput experiments it is not desirable to verify each data point. Experience with Waters QuanLynx^{TM} led to the development of a data check targeting errors in automated extraction through retention time analysis. The design of this software has left it easily extendable to additional error checking techniques that arise.

A check for missing data is always performed. If data is missing the program generates a list of the missing data, and requires the user to correct this problem before continuing.

- With *QL* data, detection flags are used to differentiate between missing data and manually deleted peaks.
- With *MAM* data, the full fragment is generated with zero's in place of missing data.

The remaining data checks are optional and may be specified in the configuration file.

There are two classes of data checks, pre- and post-correction checks. If any data fails to pass a pre-correction data check the software generates a list of affected data and allows the user to review the data before continuing. The post correction data check generates an analysis file, but does not specifically flag any data. The user may optionally revise the data, even go back to re-evaluate the chromatogram. Alternatively if the user decides that the data is useable, they may simply disable the check.

Pre Correction Checks

Pre-correction data checks are those that can be applied directly to the measurement data. There are currently three optional data checks for boundary limits, and two checks (one required, one optional) specifically targeted to evaluating automated data extraction.

- Pre-correction data checks can be found in **`OptionalChecks.pm`**, located in the **`FluxY_Lib/MS`** sub-directory.

Missing Data

The check for missing data flags missing MS data measurements, not zero measurements but measurements with no entries. This includes both cases of error in data acquisition and those where no discernable peak is found. While this check is targeted at automated data acquisition, it is always performed as it can also catch copy/ paste errors or other data handling errors.

Readings above the Detector Limit (upper bound)

All MS devices have a linear region of operation, in terms of quantifying the ions. Measurements above this level of quantification is no longer comparable. A data check can be set to find any single measurements outside of this operating range.

- Although it affects the usability of the entire fragment, only the individual measurement is flagged for review.

Poor Peak Quality (lower bounds)

Some peaks can be so small as to be indistinguishable from noise, or in fact there is no peak, just noise. The poor peak quality check compares the sum of an entire fragments' mass isotopomers to a threshold value.

- If below the threshold, the entire fragment is flagged for review.

Data Value Threshold (lower bound)

This data check provides a mechanism for applying a minimum threshold for individual mass isotopomers within a fragment. This threshold is a percentage of the sum of the entire fragments' MS measurements. Any individual mass isotopomer making up that sum that is less than the specified percentage is set to zero.

Retention Time Analysis

Automate data acquisition from raw chromatograms can introduce its' own source of errors. One particular error is the selection of the wrong peak. This may be caused, for example, by peaks that are too close or overlapping, or by an error in the software.

For larger numbers of samples, as in high throughput or even a few sample groups with four to five replicates and/or multiple fragments, it is possible to perform meaningful statistical analyses on the extracted data.

If retention time data is available, either directly integrated with the MS data (as with the QuanLynx report format) or specified as a separate data file, then it may be analysed on a per compound basis. Depending on the software you may get individual retention times for each fragment or even individually for each measurement within each fragment.

The average retention time for each compound may then be calculated and the standard deviation used as a measure of accuracy for each individual measurement. If any measurement is more than a specified number of standard deviations from the mean, then that measurement is flagged for review.

- For a complete example (including illustrations) of using the retention time analysis with Waters QuanLynx™ software see Appendix A.
- This check combines the same compounds over all data sets, so it is important to only analyze grouped data that are expected to have the same retention time.

Post Correction Checks

Post-correction data checks are those that are typically applied to the data post processing (although they may also be applied without selecting any processing). There is currently one post-correction data check which is targeted at identifying contamination and coeluting substances. Contamination may occur during processing of the samples in the lab or during the MS analysis. Errors due to coelution is when two or more substances elute from a chromatographic column at the same time, making it difficult or impossible to identify the desired compound/fragment.

- Post-correction data checks can be found in **`OptionalChecks2.pm`**, located in the **`FluxY_Lib/MS`** sub-directory.

Average Carbon Labelling

The average carbon labelling (ACL) is used in general to determine the percentage of labelling yielded by a given biological sample. However when combined with an unlabelled control sample, it can also be used to find contamination and/or coelution in measurement data.

A calculation of the total average carbon labelling on a per fragment basis and for the entire chromatogram. This calculation depends on the number of carbon atoms in the fragment and both the raw fragment measurements and the processed (corrected/ normalized) fragment.

- This calculation is performed post correction, however it may be run with or without any correction having taken place, for example allowing for the analysis of previously corrected data.

In the ideal case when an unlabelled set of measurement data is corrected it should be 100% unlabelled (M+0). In general with noise and experimental error, we set a threshold such as 1% ^{13}C-labelled to indicate a clean unlabelled measurement.

- For a complete example (including real data table) of applying ACL see Appendix B.
- Note that the unlabelled sample is only valid for analysing the labelling sample which were processed concurrently.

Configuration File: Part 2. Data Checking

The second part of the configuration file consists of seven entries related to data verification. There is one entry per line. for complete information on the configuration file syntax see Appendix C.

Line 11 is the MS detector limit threshold (upper limit).

- Enter the linear detector threshold for your MS.
- This check is skipped if the value is 0.

Line 12 is used to detect when a chromatogram peak has poor quality.

- Enter the minimum acceptable quantification, for example to distinguish from background noise (dependent on s/n of MS).
- This check is skipped if the value is 0.

Line 13 is used to perform a retention time analysis.

- Enter the number of standard deviations (acceptable range) for the retention time variation (fractional values are allowed). Typical values are in the range of 1-2 but varies with peak spread and interference.
- This check is skipped if the value is 0.

The next two lines are used to specify an externally provided retention time file, in TSV format.

- Retention time data is included in the QL report format and therefore it is not necessary to provide a dedicated retention time data file.

Line 14 is a simple Yes/No (Y,N) asking if you are going to provide an optional file with retention time data.

Line 15 specifies the name of this file.

- This line is skipped if line 14 is N.

Line 16 specifies the data threshold for zeroing data.

- Enter the percentage (in decimal form) of the total peak quantification below which the measurement value will be set to zero.
- This check is skipped if the value is 0.

Line 17 is a simple Yes/No (Y,N) asking if you want to perform the average carbon labelling analysis.

- If the answer is yes (Y) then an additional file will be generated with one average labelling value per fragment and a total average labelling value per data set, for each experimental data set.
- If you have specified the use of replicate groups in line 6 (other than 1), the values will be calculated one per fragment for each replicate set.
- In general the correction for natural abundance (line 20) should be selected when performing this analysis.
- This check is skipped if the value is N.

4. Data Correction

Before using MS measurement data from isotopic labelling experiments it is desirable to perform one or more correction steps. For labelling experiments the most common type of correction is for naturally occurring stable isotopes, or natural abundance (NA). In addition to NA, it may also be desirable to correct for the natural abundance in the original bio-mass (OBM), although this depends more on the type and length of the experiment, i.e. does the OBM make a significant contribution to the final mass. Finally there are corrections that are targeted to specific MS techniques or compounds, such as proton loss and proton gain.

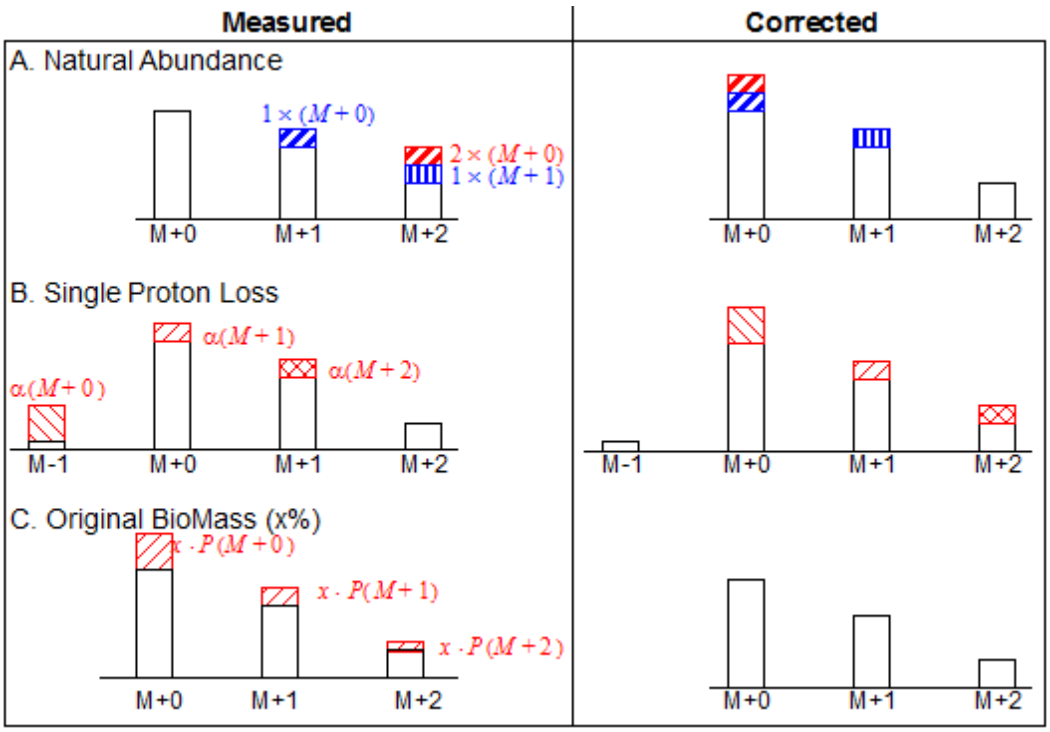


Figure 1: Overview of MS data correction. On the left is a representation of a compound with two carbon atoms, i.e. 3 mass measurements, and on the right is the corrected form of the same MS data, after the specified correction (for the proton loss correction an additional mass measurement is required prior to the M+0 mass). The coloured parts are used to illustrate the portion of each measurement that is reassigned after correction.

Correction

Each fragment is corrected separately over it's set of mass isotopomers, and the resulting data is scaled to 100% (per fragment).

> *Because each fragment is scaled relative to itself it is important to consider the use of value thresholding and the Poor Peak Quality data checks. After scaling the relative measured quantities are lost, and all fragment information* ***appear*** *to have the same weighting.*

If no correction mechanisms are specified, the uncorrected data is scaled and may be used as is.

Each data correction is described below, they are given in the order they would be applied if more than one correction is selected.

- Data correction functions can be found in the **`CorrectFragment.pm`**, located in the **`FluxY_Lib/MS`** sub-directory.

Correcting For Proton Loss/Gain

This correction is useful when a given compound or analytical technique results in a fraction of the compound losing or gaining a proton, such as documented with ion-trap MS used together with gas chromatography (yielding a percentage proton loss) and the McLafferty ion of fatty acid methyl esters (shown to have a percentage proton gain). If the entire compound lost or gained a proton this would not be required, instead the chemical formula could simply be adjusted.

The premise of the correction is that there is an equal probability of this loss or gain across all molecules of the compound, regardless of the number of labelled atoms. Based on this premise a single Proton gain or loss can be described by a scaling factor (Figure 1).

To perform the correction for proton loss each compound/fragment requires one additional measurement prior to the M+0 mass, i.e. mass M-1. For the proton gain one additional measurement posterior to the M+numC mass, is required i.e. mass M+numC+1.

This correction requires solving for a scaling factor using a non-linear set of equations. The iterative method for solution may determine that there is no valid solution. In this case the original measurement set is left uncorrected.

- Important: it is only possible to perform one or the other and not both of these corrections. on a set of data. Multiple simultaneous proton loss/gain corrections are not supported at this time.
- These corrections are not for general use, but should only be used for very specific cases, such as those described above.

Correcting For Natural Abundance

Natural abundance correction is performed for the specified set of atoms that make up the (derivatized) compound/fragment. Standard labelling probabilities of the first three isotopes of these atoms are considered in performing this correction, i.e. P(M+0), P(M+1) and P(M+2). A correction matrix is generated based on the number of each type of atom being considered and their natural isotope probabilities. The matrix size itself is related to the number of mass isotopomers in the fragment being corrected.

The isotopic compositions are derived from the *2011 NIST Atomic Weights and Isotopic Compositions for All Elements*. However users may easily substitute regional or local isotopic percentages by editing the atomInfo structure in the **`atomInfo.pm`** module located in the **`FluxY_Lib/Chemistry`** sub-directory.

Although targeting ^{13}C MFA, there is no artificial restraint to limit the atoms used in the correction matrix. The only restraint is a mathematical one, which requires the first atomic isotope to be for the atomic mass that matches the atomic mass (rounded to the nearest integer), for example the first isotope of carbon is for mass 12, which matches the atomic

mass (12.0107). This allows not only for different compounds, but also for different derivatizing agents (e.g. Si in TBDMS) or alternative analytical techniques.

- The correction for natural abundance should be applied to all measurements from isotopic labelling experiments.

Correcting For Original Bio-Mass

Depending on the underlying experiment the original, unlabelled, biomass may make up a significant portion of the final biomass. In this case the labelling profile of each metabolite is skewed to represent a combination of the original unlabelled (i.e. naturally abundant labelled) biomass and the experimentally labelled biomass.

This correction subtracts the original bio-mass (OBM) of each fragment (as a percentage of total OBM), from the measured fragment (also scaled). At this time only the natural abundance of the Carbon atoms is used to distribute the OBM over each measured mass isotopomer in a fragment.

- This correction requires an additional data file containing the percentage of OBM for each data set.
- The OBM file has one value (the percentage in decimal form) per line, and the order of the values must correspond to samples/columns in the MS data file.

Configuration File: Part 3. Data Correction

The third part of the configuration file consists of five entries related to the data correction (processing) of the MS data. There is one entry per line. for complete information on the configuration file syntax see Appendix C.

Line 18 is a simple Yes/No (Y,N) asking if you want to perform the proton loss correction.

- Only one of proton loss and proton gain may be selected.

Line 19 is a simple Yes/No (Y,N) asking if you want to perform the proton gain correction.

- Only one of proton loss and proton gain may be selected.

Line 20 is a simple Yes/No (Y,N) asking if you want to perform the natural abundance correction.

- This is the default correction for isotope labelling data.
- In general this correction should be selected when performing the average carbon labelling analysis (line 17).

Line 21 is a simple Yes/No (Y,N) asking if you want to perform the original biomass correction.

- If yes (Y) than an additional file must be provided containing the original biomass data.

Line 22 specifies the name of this file.

- This line is skipped if line 14 is N.

5. Output

The last step is to use the data to generate output. In fact the software may skip all of the verification and correction steps and simply operate as a translator from input to output. The main format is the text based *tab separated values* (TSV), for ease of use and viewing in a spreadsheet and as a general input and exchange format for other programs.

If replicates are used, the corrected/processed data sets may be averaged.

- The current implementation requires that the replicates must be presented in consecutive order in the data file, i.e. it will not reorder the data sets.
- Regardless of the use of data replicates a standard deviation is calculated (as required by some model formats), even when not statistically accurate (too few replicates).
  - It is up to the user to determine if sufficient replicates are available to have meaningful statistics.
- To more accurately reflect the (MS) machine accuracy a constant additive offset may be specified for the standard deviations.
- In the case of only one replicate, the standard deviation is returned as 5% + the constant offset (if specified).

Supported output formats

All data can be saved in TSV formatted files, however this program also allows for the generation of model specific formats for some popular 3rd party ^{13}C MFA software. In the latter case only averaged data is used (all of these programs make use of statistical replicates).

The model specific data may be generated independently (each in an individual file), or may be directly integrated into an existing model of the specified format. In the latter case one or more model files may be specified. If so, then the measurements may be directly integrated into the models, either replacing the corresponding measurements and errors, or appending to the existing data set.

TSV - Tab Separated Values

TSV is the main input and output format used by this software. It is a generic exchange format which can easily be imported into spreadsheet software as well as being used directly as input to other programs.

Since the TSV is a flat (two dimensional) format, multiple TSV files are required to represent multiple data values, such as both a measurement and a retention time, or an average and a standard deviation, with one TSV file per data type.

MS and LM - 13CFLUX Network Configuration File

13CFLUX uses a simple text based, column oriented network configuration file called the Flux TaBLe or FTBL file (with extension .ftbl). There are two different sections for entering MS labelling data, the MS specific **M**ass **S**pectrometry or the more generic **L**abel **M**easurement sections.

- http://www.13cflux.net/13cflux/index.jsp

OF - OpenFlux Network Configuration File

Openflux uses a simple text based column oriented network configuration file for easy handling in spreadsheets (extension .csv). MS statistical data is represented in a single format stored in separate sections, the measurement and error sections.

- http://openflux.sourceforge.net/

C2 - 13CFLUX2 Network Configuration File

13CFLUX2 uses an XML based network configuration file called FluxML (with extension .fml). It directly supports a variety of measurement specific formats, as well as a generic one. This software generates the model (labelling measurements) and data sections.

- http://www.13cflux.net/13cflux2/

Configuration File: Part 4. Output

The fourth part of the configuration file consists of ten entries related to the output and model generation. There is one entry per line. for complete information on the configuration file syntax see Appendix C.

Line 23 is a simple Yes/No (Y,N) asking if you want to print the retention time values.

- This is used only when (a) retention time values are available and (b) they are not already in TSV format.

Line 24 is a simple Yes/No (Y,N) asking if you want to print the matrix of MS measurement values.

- This is used when the MS data file is not already in TSV format.
- All raw data is output including additional MS readings, i.e. M-n and M+numC+n.

Line 25 is a simple Yes/No (Y,N) asking if you want to print the processed MS data values.

- This prints out each data set individually, not averaged (see line 26).
- The processed data does ***not*** include additional MS readings.
- This will print out one set of normalised data after all selected processing operations have completed.
  - If you want to see the effect of individual processing steps you must select them one at a time.
  - The name of the output file is modified to include the processing steps applied.
- If no processing steps are selected, this will produce a set of normalized raw data without additional measurements.

Line 26 is a simple Yes/No (Y,N) asking if you want to print the average of the processed data calculated over the specified replicates.

- Exactly like the previous operation, but with averages.

Line 27 is a simple Yes/No (Y,N) asking if you want to print the standard deviations.

- The standard deviations as associated with the averaged data (see line 26).

The next five options deal with generating model specific data. This data may be generated separately or integrated directly into a model file.

Line 28 is a simple Yes/No (Y,N) asking if you want to generate model data.

- If yes (Y), one file will be printed for each set of experimental replicate data.
- The actual format of the file depends on the type of model it is being generate for (see line 29).

Line 29 species the format of the model file to generate.

- MS: 13CFLUX - MASS_SPECTROMETRY section.
- LM: 13CFLUX - LABEL_MEASUREMENTS section.
- OF: OpenFlux - measurement and error sections.
- C2: 13CFLUX2 - model (labelling measurements) and data sections.

Line 30 is a simple Yes/No (Y,N) asking if you want generate the data directly in a model file.

Line 31 is a simple Yes/No (Y,N) asking if you want to append to (rather than overwrite) any existing data in that section.

- This is useful if you are adding different types of data that are processed separately, such as amino acids and glucose.
- This line is skipped if line 30 is N.

Line 32 contains a space separated list of model file names.

- If line 30 is Y then at least one model file must be entered and it must match the format specified in line 29.
- For each model file, one new model file is generated for each set of averaged data.
- This line is skipped if line 30 is N.

Line 33 specifies a constant additive offset for the standard deviation.

- This may be used when the calculated standards deviations are lower than the accuracy of the measurement devices used.

6. How to Run *iMS2Flux*

This section explains how to run the software from the command prompt. If you are a Windows user and prefer the graphical user interface please see the GUI getting started guide accompanying the GUI application.

To install the software please see the appropriate getting started guide for your platform (Windows, Mac OSX, Linux or Unix).

- The getting started guides are located in the `Getting_Started` directory.
- The following examples assume `iMS2Flux.pl` has been installed following the "Run From Anywhere" instructions in the getting started guides. If not you must specify the path to the program when calling it.

General Usage

The basic software operation is quite simple. First set up a configuration file in the same directory as your MS data file. The easiest way to do this is to copy the configuration file provided with the example and simply edit it as necessary.

Next open a command prompt and change directories to the directory containing your data and configuration file and type:

```
PROMPT> iMS2Flux.pl <ENTER>
```

That's it. If there are any problems with the configuration file or the MS data file you will be notified by error message. Some non-syntactic error conditions are also flagged as warnings. When the program encounters an error it stops, however with warnings it will continue, even if the result is invalid.

- Any output that is generated goes into the same directory as the MS data file.
- If other optional files are being used, such as original biomass data, they must also be located in the same directory.

Alternate General Usage

Alternatively the configuration file may have a different (more meaningful) name. In this case the program will accept a single command line argument that specifies the name of the configuration file:

```
PROMPT> iMS2Flux.pl [configuration file name] <ENTER>
```

Printing Compound Information

There is a built in print command to generate a file with the complete listing of supported compounds for a given compound class (and derivatizing agent if applicable). This file will be located in the current directory.

```
PROMPT> iMS2Flux.pl -P1 XX YYY <ENTER>
```

- Where XX is the compound type and YYY is the derivatizing agent (if applicable), for a complete list see Table 1.

Getting Help

It is also possible to generate a help file on demand, which provides an overview of the programs operation, detailed operational syntax and a concise description of the configuration file:

```
PROMPT> iMS2Flux.pl -help <ENTER>
```

iMS2Flux Syntax

```
iMS2Flux.pl [CONFIGURATION_FILE_NAME] [-Z] [-P2] [-BB #]
iMS2Flux.pl -P1 XX YYY
iMS2Flux.pl -help
```

[CONFIGURATION_FILE_NAME] is an optional parameter.
By default CONFIGURATION_FILE_NAME is **config.txt** and is not required as a parameter to this program. Optionally you may provide the configuration data in a file with a different name, and through this parameter provide the name of the configuration file to the program.

-Z With QL data type only - Overrides default behaviour for missing data - prints missing data table, replaces any missing data with zero, and continues processing.

-BB Specify whether to correct the fragments with a specified percentage of the carbon bacbone. The valid range for this value (#), given in decimal, form is from 0 to 1 inclusive.
- This option only affects the natural abundance correction.
- If not specified the default is to correct with backbone carbon.
- To exclude the backbone carbon set the value to 0.
- while a fractional value is supported, the usage is currently experimental and not normally used.

-P1 Prints the fragment/compound mass tables and associated data related to supported fragments/compounds for the specified compund type (XX), and derivatizing agent (YYYY) if necessary.

-P2 Prints the standard deviations associated with the average carbon labelling data check. This option is ignored if the optional data check is not enabled in the configuration file.

-help Prints the full help information including the format of the configuration file.

Appendix A: Retention Time Analysis

Data acquisition begins with the harvesting of cells from labelling experiments for MS evaluation. This includes the use of built in tools in both the actual hardware (Mass Spectrometry device) and software for setting up and acquiring raw MS data for multiple experiments.

We take this automation further using both MS and custom tools to (1) perform bulk extraction of the desired compounds from the raw MS data file(s), and (2) to verify the extracted data to allow for full automated data processing.

In this example we illustrate the use of a novel data verification called retention time analysis. With individual retention times being automatically provided by various automated chromatogram extraction tools for each fragment of each compound it becomes possible to perform statistical analyses on this data to identify mistakes in peak selection by the automated extraction software. The other pre-correction data checks are also discussed in this illustration.

In Figure 2A we see an example of raw MS data represented in chromatograms. This series of chromatograms illustrate four GCMS measurements about the peak for Valine 264. The four measurements include a standard, two experimental replicates for the Ler mutant and one for the CVI mutant. These four samples, and more, were processed using the Waters QuanLynx^{TM} software to automatically identify and quantify the desired amino acid mass isotopomers. In all but the first case (the standard) the software correctly identifies the desired peak.

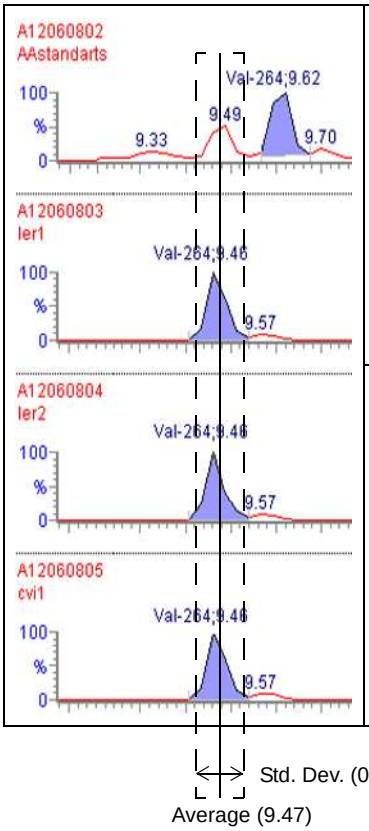
The QuanLynx software allows for the identified measurement peaks to be automatically quantified into a text based report. This report can then be imported into our software (using input data format QL).

In order to facilitate automatic data extraction we require confidence that the extracted data is correct. Although this is fairly obvious for a person to detect, any high throughput workflow requires that there is the least amount of user interaction. For that reason there is a series of data checks performed on the automatically extracted data before further processing. These data checks include, but are not limited to:

- a detector threshold limit - to flag integrated results above the linear limit of the MS device,
- a poor peak quantity - to flag compounds whose cumulative peaks are indistinguishable from noise,
- a missing data check - to flag missed peaks (which may include both cases of error in data acquisition and those where no peak is actually present), and
- a retention time check - based on the number of standard deviation of each computed retention time from the average retention time (calculated over all experiments and over all fragments of a given compound).

The first two of these checks are already widely used in software for processing/ correcting MS data, but with the inclusion of automated data extraction it is necessary to introduce additional checks such as the missing data (not a problem with manual extraction) and retention time data checks. The design of this software readily allows for

A.



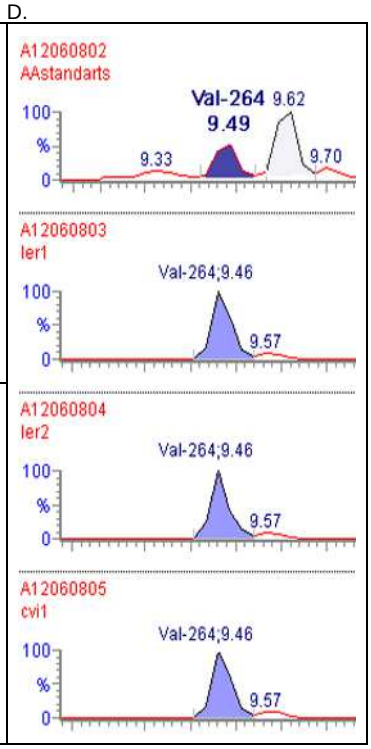
B.

	A	B	C	D	E	F	G	H
1			Std 1	Std 2	Ler 1	Ler 2	CVI 1	CVI 2
14	Val	260	9.50	9.46	9.46	9.46	9.46	9.46
15		261	9.50	9.46	9.46	9.46	9.46	9.46
16		262	9.49	9.46	9.46	9.46	9.46	9.46
17		263	9.50	9.49	9.46	9.46	9.46	9.46
18		264	9.50	9.62	9.46	9.46	9.46	9.46
19		288	9.50	9.49	9.46	9.46	9.46	9.46
20		289	9.50	9.49	9.49	9.46	9.49	9.46
21		290	9.50	9.49	9.49	9.46	9.49	9.46
22		291	9.49	9.49	9.49	9.46	9.49	9.46
23		292	9.49	9.49	9.49	9.46		9.46
24		293	9.50	9.49	9.49	9.49	9.49	9.46
25		302	9.50	9.49	9.49	9.49	9.49	9.49

C.

	A	B	C	D	E	F	G	H
1			Std 1	Std 2	Ler 1	Ler 2	CVI 1	CVI 2
14	Val	260	RT	0	0	0	0	0
15		261	RT	0	0	0	0	0
16		262	0	0	0	0	0	0
17		263	RT	0	0	0	0	0
18		264	RT	RT	0	0	0	0
19		288	RT	0	0	0	0	0
20		289	RT	0	0	0	0	0
21		290	RT	0	0	0	0	0
22		291	0	0	0	0	0	0
23		292	0	0	0	0	MD	0
24		293	RT	0	0	0	0	0
25		302	RT	0	0	0	PPQ	0

D.



the inclusion of additional data checks be they general or specific to some type of MS data.

Figure 2: This figure illustrates the MS data representations (as chromatograms) and work flow throughout the process of data extraction and checking. (A) Illustrates the chromatograms for the same fragment (Valine 264) from four different experiments. (B) shows a portion of the retention time matrix focusing on Valine, with flagged data highlighted in yellow (1-2 S.D.) and pink (>2 S.D.). (C) shows the same segment of the general data checks matrix indicating cells which failed one (or more) data checks. The outlined values in (B) and (C) correspond to the chromatograms shown in (A). Finally (D) shows the chromatogram after manually correcting the peak selection. RT - flagged for retention time, MD - missing data and PPQ - poor peak quality.

Figure 2B illustrates a partial spreadsheet of retention times from the automatic evaluation of the GCMS data focusing on Valine and the experiments illustrated in the chromatograms in Figure 2A (the outlined values). This spreadsheet is coloured to provide visual feedback on the distance of each retention time from the average. The average retention time as taken over each of the experiments and all of the mass isotopomers for the fragments of Valine is 9.47 with a standard deviation of 0.02. The green cells have a retention time with less than one standard deviation (S.D.) from the average, while the yellow are between one and two S.D., and red is greater than two S.D.

Also of interest is a blank cell (experiment CVI1, fragment Val-292). This is an example of missing data. When the software does not automatically detect a peak, there is also no retention time associated with the peak.

The retention time analysis for Valine is overlaid on the chromatograms in part A to show the average (solid line) and one standard deviation (dotted box).

In Figure 2C we see the result of the data checks based on the options provided, a retention time (RT) spread of 1 S.D. and a poor peak quality (PPQ) of 500 and a detector limit of 10^7. The format is similar to the retention time data, this time coloured green for no problem, pink for failing one or more data checks, and red for missing data (MD).

Figure 2D illustrates the original four chromatograms from part A after manual correction. Notice the highlighted peak in the top chromatogram. After correcting the peaks identified by RTA the results were again automatically quantified with QuanLynx and saved to a text file in its native report format. After checking the other flagged peaks, and making manual corrections as required, the data check was run a second time resulted in no flags. Should additional flags be thrown, this process is iterated until all potential data problems have been dealt with. If flagged data is not changed, for example the retention times flagged for experiment Std 1 (see part C) after manual review, the criteria for flagging may be changed to accommodate false positives - in this case the retention time threshold value was set to 1.5 S.D.

Incorporating automated data checking with automated data extraction allows this workflow to minimize the number of chromatograms that the investigator is required to manually check (and if necessary correct).

Appendix B: Average Carbon Labelling

Average carbon labelling (ACL) can initially be used to verify the design of an isotope labelling experiment, and later as a standard data check for contaminated data.

When designing isotope labelling experiments it is desirable to have a quantification of the total label uptake in addition to the average label distribution. This allows the researcher to determine if the experimental protocol yields appropriate levels for analysis, while optimizing the experiment time course, media, or other properties. For example a steady state ^{13}C MFA experiments may target 10-20% labelling, but may also require analysis to be performed with an isotope-ratio MS (IR-MS) which is limited to a maximum of 20% ^{13}C labelling.

Once MS data has been produced ACL proves to be a powerful tool for identifying contaminated data. This includes both regular and known contamination, such as co-eluting substances during MS, and random contamination at any point in the processing. This method requires an unlabelled experiment to be performed in parallel. After correction for natural abundance the unlabelled data should have very little carbon labelling (ideally zero, but in practice there is always some noise). By analyzing the experimental data that has been processed and measured together with the unlabelled sample, ACL can be used to identify contamination in individual fragments.

Table 2 contains an example of the average carbon labelling from 5 experiments (with 4 replicates per set) of pea (*Pisum sativum)*. The first experimental set (unlabelled) consist of samples treated with an unlabelled-substrate (no ^{13}C-substrate), while the next 4 sets, experiments 1-4, come from a ^{13}C labelling experiment grown in a media containing uniformly labelled ^{13}C-glucose (in addition to unlabelled sucrose).

The 20 chromatograms yielded 63 common amino acid fragments, shown in Table 2. Each entry in the table represents the percentage labelling of each fragment for one experiment, calculated over the 4 replicates. This labelling is calculated after performing correction for natural abundance. At the bottom of the table is the labelling percentage calculated for the entire experiment.

Under ideal conditions the isotope labelling after correction for an unlabelled experiment should be 0. However, under real conditions this is not the case, small non-zero ^{13}C-labelling, less than 1%, is common. As can be seen in the table, the majority of amino acid fragments fall into this category, while 6 (in orange) fall into the possibly contaminated range (1-2%) and 5 into the definitely contaminated range (red). Some of these contaminations are due to co-eluting substances, such as the alanine and glycine M-159 fragments. Others may be due to low peak intensity with a low signal to noise ratio, such as the proline M302 measurement, or other processing/lab related causes.

The unlabelled experiment serves as a negative control. Overall it shows a reasonable level of isotope labelling, 0.75%. However the exclusion of fragments with 2% isotope labelling reduces the average labelling to 0.56% and with the further elimination of those fragments above 1% labelling, its final average carbon labelling is 0.41%. That is a reduction of 25% and 46% of the labelling respectively. Removing these same fragments from the labelled experiments results in an average reduction of only 2% and 4% respectively, i.e. still yielding good labelling information.

Table 2: Results of average carbon labelling analysis

Avg. C Labelling								
		Unlabeled	**Exp. 1**	**Exp. 2**	**Exp. 3**	**Exp. 4**		
ala	158	5.41%	17.98%	18.51%	18.02%	16.95%		Unlabelled >=2% labelling
	232	0.48%	14.30%	14.85%	14.35%	13.12%		Unlabelled < 2% & >= 1% labelling
	260	0.23%	13.96%	14.39%	13.91%	12.62%		
gly	144	3.13%	14.61%	14.52%	14.88%	14.34%		
	218	0.36%	12.77%	12.72%	13.22%	12.66%		
	246	0.71%	12.83%	12.85%	13.27%	12.74%		
	288	0.03%	11.86%	11.73%	12.18%	11.64%		
val	186	1.32%	13.98%	14.48%	14.03%	12.00%		
	260	0.28%	13.02%	13.58%	13.11%	11.00%		
	288	0.10%	12.69%	13.17%	12.73%	10.63%		
	302	0.07%	13.56%	14.06%	13.54%	11.39%		
leu	200	1.10%	12.50%	13.00%	12.90%	12.26%		
	274	0.37%	11.77%	12.34%	12.12%	11.60%		
	344	0.92%	11.27%	11.59%	11.05%	10.87%		
ile	200	0.61%	12.10%	12.22%	12.66%	11.67%		
	274	0.25%	11.38%	11.48%	11.96%	11.00%		
	344	0.20%	9.00%	8.97%	9.47%	8.47%		
pro	184	0.31%	9.95%	9.88%	10.08%	9.60%		
	258	0.17%	9.78%	9.61%	9.73%	9.27%		
	286	0.01%	8.48%	8.32%	8.62%	8.10%		
	302	10.26%	19.15%	13.20%	10.90%	11.02%		
	328	0.82%	8.61%	8.34%	8.24%	7.97%		
met	218	0.42%	11.91%	11.76%	12.57%	12.31%		
	292	30.60%	20.28%	20.63%	21.17%	39.08%		
	302	22.81%	29.31%	29.43%	29.87%	28.03%		
	320	0.48%	11.94%	11.80%	12.55%	12.35%		
	362	1.89%	10.27%	9.52%	11.28%	11.57%		
ser	288	0.62%	12.65%	12.45%	12.67%	12.19%		
	302	0.64%	12.56%	12.53%	12.60%	12.25%		
	362	0.52%	12.37%	12.35%	12.56%	12.04%		
	390	0.53%	12.42%	12.33%	12.58%	12.09%		
	432	0.05%	11.36%	11.19%	11.30%	10.89%		
thr	376	0.06%	10.59%	10.75%	11.64%	10.47%		
	404	0.10%	10.56%	10.64%	11.47%	10.44%		
	446	0.23%	10.17%	10.34%	11.00%	10.13%		
phe	234	0.19%	12.17%	12.93%	13.60%	12.13%		
	302	0.26%	13.00%	13.77%	14.24%	12.88%		
	308	0.49%	12.93%	13.90%	14.44%	12.92%		
	336	0.05%	12.12%	13.10%	13.66%	12.29%		
asp	302	0.51%	5.84%	6.15%	6.11%	5.80%		
	316	1.10%	6.32%	6.62%	6.44%	6.08%		
	390	0.15%	5.44%	5.69%	5.56%	5.17%		
	418	0.45%	5.59%	5.95%	5.87%	5.45%		
	460	0.14%	5.23%	5.41%	5.21%	4.88%		
glu	302	1.97%	9.81%	9.85%	10.01%	9.65%		
	330	0.49%	8.75%	8.97%	9.08%	8.58%		
	404	0.08%	8.12%	8.18%	8.26%	7.74%		
	432	0.95%	8.71%	8.88%	9.14%	8.64%		
	474	0.14%	7.75%	7.79%	7.91%	7.51%		
lys	302	0.44%	16.53%	16.81%	17.52%	16.61%		
	329	0.80%	11.90%	11.99%	12.31%	11.60%		
	431	0.23%	11.17%	11.34%	11.68%	10.94%		
	473	0.44%	11.26%	11.41%	11.76%	11.09%		
his3	302	0.21%	11.26%	12.13%	12.48%	11.73%		
	338	1.11%	12.05%	12.63%	12.91%	12.33%		
	412	0.12%	5.19%	6.77%	8.62%	7.86%		
	440	0.09%	11.35%	12.01%	12.24%	11.65%		
	482	0.57%	11.58%	12.24%	12.34%	11.72%		
tyr	302	0.30%	11.86%	11.97%	12.53%	12.14%		
	364	0.57%	11.62%	11.52%	11.69%	11.36%		
	438	0.24%	11.24%	11.15%	11.24%	10.95%		
	466	0.13%	11.12%	10.98%	11.14%	10.80%		
	508	0.48%	10.85%	10.87%	11.19%	10.86%		
Average Carbon Labbelling		0.75%	10.86%	11.27%	11.58%	10.68%		

Appendix C: Configuration File

This program requires a configuration file, by default named ***config.txt***. Optionally the configuration file may have a different name, one more meaningful. The format of this file is line oriented, with each line either blank, containing a comment or a single configuration entry. There are 33 configuration lines, all 33 lines are required, even though some of the data is optional.

The rules for each line are described below:

- Blank lines may be inserted anywhere.
- A line is a comment line if it starts with '//'.
- Comment lines and blank lines are ignored by the program.
- Configuration lines, as specified below, may contain either a single value or a space separated list of values.
- A colon is used to mark the end of descriptive text (to the left) and the start of configuration data (to the right.

Configuration file structure:

Line 1: The name of the data file(s) specified in a space separated list.

Line 2: **[TSV|QL|MAM]** - The data file format.
- TSV - for Tab Separated Values
- QL - for QuanLynx
- MAM - for M. Agilent Macros

Line 3: **[AA|CW|FA|GL|GY|SM]** - The type of MS data.
- AA - Amino Acids (multiple compounds with multiple fragments)
- CW - Cell Wall (multiple compound, no fragments)
- FA - Fatty Acids (Full and McLafferty Fragment)
- GL - Glucose (single compound with multiple fragments)
- GY - Glycerol (single compound with multiple fragments)
- SM - Soluble Metabolites (multiple compound, no fragments)

Line 4: **[NONE|TBDMS|TMS|TFAA|Ac2O]** - The derivatizing agent.
- TBDMS - t-butyl-trimethylsilyl (AA)
- TMS - trimethylsilyl (GL & GY)
- TFAA - Acetic acid, trifluoro-,1,2,3-propanetriyl ester (GY)
- Ac2O - 2,3,4,5,6-Penta-O-acytl-D-glucose-O-methyloxinme (GL)
- None - No derivatizing agent is specified.

Line 5: **[A|R]** - The type of data to process (A for area, R for response).
- Applies only to data type QL, otherwise either value is OK.

Line 6: **[number(s) >= 1]** - The number of consecutive replicates in the data file.
- A single number, all experiments have the same number of replicates.
- A space separated list specifying the number of replicates in the same order as presented in the data file(s)

Line 7: **[Y|N]** - Are you providing the replicates names/titles?

Line 8: The name of the file with the replicate names (skipped if line 7 is N).

Line 9: **[number >= 0]** - Number of measurement values preceding 'M+0'.

Line 10: **[number >= 0]** - Number of measurement values post 'M+numC'.

Line 11: **[number >= 0]** - The detector limit threshold (check skipped if 0).

Line 12: **[number >= 0]** - The poor peak quality threshold, sum of fragment (check skipped if 0).

Line 13: **[number >= 0]** - The retention time threshold, standard deviations (check skipped if 0).

Line 14: **[Y|N]** - Are you providing retention times in a separate file?

Line 15: The name of the file with the retention times (skipped if line 14 is N).

Line 16: **[number 0-1]** - The data value threshold (check skipped if 0).

Line 17: **[Y|N]** - Generate the Average Carbon Labelling summary.

Line 18: **[Y|N]** - Correct for proton loss, requires M-1 data measurements.

Line 19: **[Y|N]** - Correct for proton gain, requires M+numC+1 data measurements.

Line 20: **[Y|N]** - Correct for natural abundance.

Line 21: **[Y|N]** - Correct for original bio-mass.

Line 22: The name of the file with the original bio-mass percentages (skipped if line 21 is N).

Line 23: **[Y|N]** - Print the retention time matrix.

Line 24: **[Y|N]** - Print the unprocessed (uncorrected data).

- For QL input also supports **[A]** - area, **[R]** - response and **[Y|B]** - for both.

Line 25: **[Y|N]** - Print the processed (corrected/scaled) data.

Line 26: **[Y|N]** - Print the corrected data averaged over the number of replicates.

Line 27: **[Y|N]** - Print the standard deviation over the number of replicates.

Line 28: **[Y|N]** - Generate model data for supported MFA software.

Line 29: **[MS|LM|OF|C2]** - Which type of data should be generated?

- MS - 13CFLUX - MASS_SPECTROMETRY section.
- LM - 13CFLUX - LABEL_MEASUREMENTS section.
- OF - OpenFlux - measurement and error sections.
- C2 - 13CFLUX2 - model (labelling measurements) and data sections.

Line 30: **[Y|N]** - Combine the measurement data with one or more models?

Line 31: **[Y|N]** - Append the data to existing measurement data, default is to replace.

Line 32: The name of the model file(s) specified in a space separated list (skipped if line 31 is N).

Line 33: **[number >= 0]** - The additive offset for use when calculating std. dev. (skipped if line 31 is N).