MetaboQuant 1.3

A software tool for quantification of small molecules from NMR spectra

User Manual

Institute of Functional Genomics
University of Regensburg
Josef-Engert-Str. 9
93053 Regensburg, Germany

matthias.klein@ur.de

Table of Contents

License Agreement

Introduction

MetaboQuant is an aid for quantifying small molecules like amino acids in 1D and 2D NMR spectra of biologic fluid samples. Biofluids such as urine, blood plasma or milk and other biologic fluid samples such as tissue extracts are very complex in their composition, and compounds may be present at high concentrations in some samples, while they may be lowly concentrated and thus not all peaks might be visible above noise level in other samples. Additionally, each peak of a compound may be overlapped with other compounds, thus changing the peak integrals and corrupting quantification results. This is not only true for 1D spectra, but also (to a lesser extent) for 2D spectra such as ^{1}H-^{13}C HSQC spectra. Other reasons may also lead to corrupted peak integrals, such as errors in peak picking and integration routines. Corrupted peak integrals will present as outliers as compared to other peak integrals of the same compound.

MetaboQuant employs several algorithms to detect and exclude such outliers. This leads to reliably accurate and precise quantification results.

MetaboQuant takes peak integrals derived from NMR spectra as input and returns absolute concentration values of the observed compounds as output.

What MetaboQuant can do:

- Exact quantification of small molecules in complex mixtures

What MetaboQuant cannot do:

- Peak picking of NMR spectra
- Peak integration of NMR spectra
- Peak fitting in NMR spectra

The features of MetaboQuant:

- Dividing peak integrals by the number of contributing nuclei
- Scaling to a reference substance
- Multiplication of each peak integral by an individual calibration factor
- Checking for the presence of obligatory peaks before quantification
- Removal of compounds where too few peaks were found in the spectrum
- Reliability checking for compounds where only few peaks are visible in the spectrum
- Automatic outlier removal
- Checking lower limits of quantification
- Automatic normalization to single compounds, e.g. creatinine
- Automatic correction for different individual dilutions
- Calculation of means and technical errors for replicate measurements
- Logging of parameters and configuration to allow reproducible results

System Requirements

MetaboQuant may be run as stand-alone application using the following operating systems:

- Windows (Windows 2000 or newer)

On other operating systems such as Mac OS, you may run MetaboQuant from within Matlab (The Mathworks, Natick, MA, USA). For this you need:

- Matlab version 7.1.0.246 (R14) Service Pack 3 or newer

Installation

Stand-Alone Application (Windows)

Visit http://genomics.uni-regensburg.de/site/institute/software/metaboquant and download the following files:

- *MCRInstallerR2007b.exe*
- *MetaboQuant.zip*

Run the file *MCRInstallerR2007b.exe* to install the required Matlab run-time routines. Extract the file *MetaboQuant.zip* to a new folder. To run MetaboQuant, double-click on the file *MetaboQuant.exe*. Starting MetaboQuant may take some minutes, as the Matlab run-time environment takes some time to be loaded.

Running MetaboQuant from Within Matlab

Visit http://genomics.uni-regensburg.de/site/institute/software/metaboquant and download the following file:

- *MetaboQuant.zip*

Extract the file to a new folder.

Start Matlab and change the working directory to the MetaboQuant installation folder, *e.g.* using the command *cd*. To run MetaboQuant, type *MetaboQuant* in the Matlab command line.

1 Manual

1.1 Getting Started

After starting MetaboQuant, you will be asked whether to use *Basic Mode* or *Advanced Mode* (see Figure 1.1).

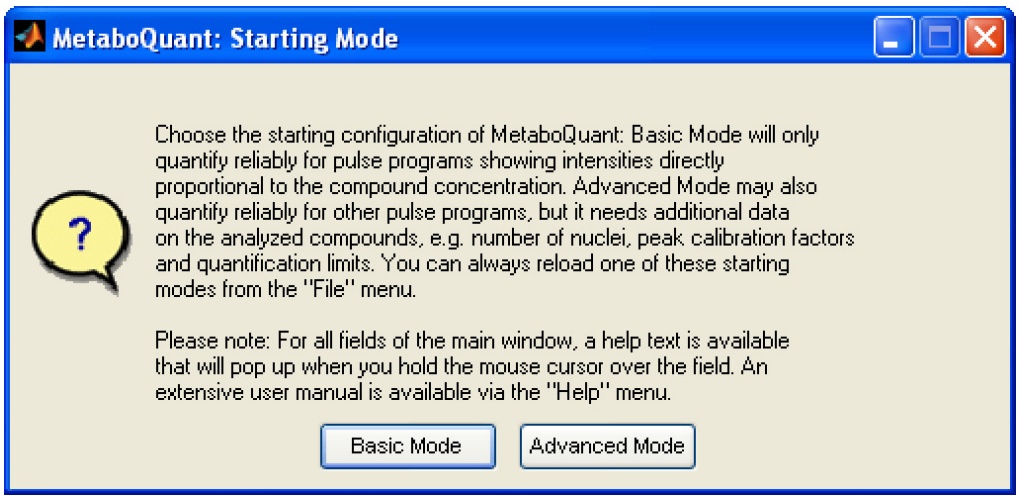

Figure 1.1: Choice of start mode.

Basic Mode is intended to get a quick overview of your data. *Basic Mode* will only yield reliable results in case a pulse program was used that yields intensities directly proportional to the compound concentration. *Advanced Mode* is configured to yield reliable results for all kinds of pulse programs. You will need additional data on the compounds in question to use *Advanced Mode*.

Please note that the different modes only affect the initial parameter settings and that all parameters can be changed in the main program window. You can reload the pre-defined parameter sets for *Basic Mode* and *Advanced Mode* at all times from the *File* menu.

For all fields of the main window, a help text is available that will pop up when you hold the mouse cursor over the field.

1.2 Basic Mode

When choosing *Basic Mode* in the MetaboQuant start window (Figure 1.1), MetaboQuant is being pre-configured to allow a quick start. In this mode, many implemented features are disabled. Anyway, this mode may allow a quick overview of your data. Please note that quantification results from this mode may be only semi-quantitative. For more accurate quantification, please refer to the *Advanced Mode* described in Section 1.3. The main features active in *Basic Mode* are automatic outlier removal based on deviations from the median and exclusion of compounds where too few peaks were found in the spectrum.

In *Basic Mode*, the main window looks as shown in Figure 1.2.

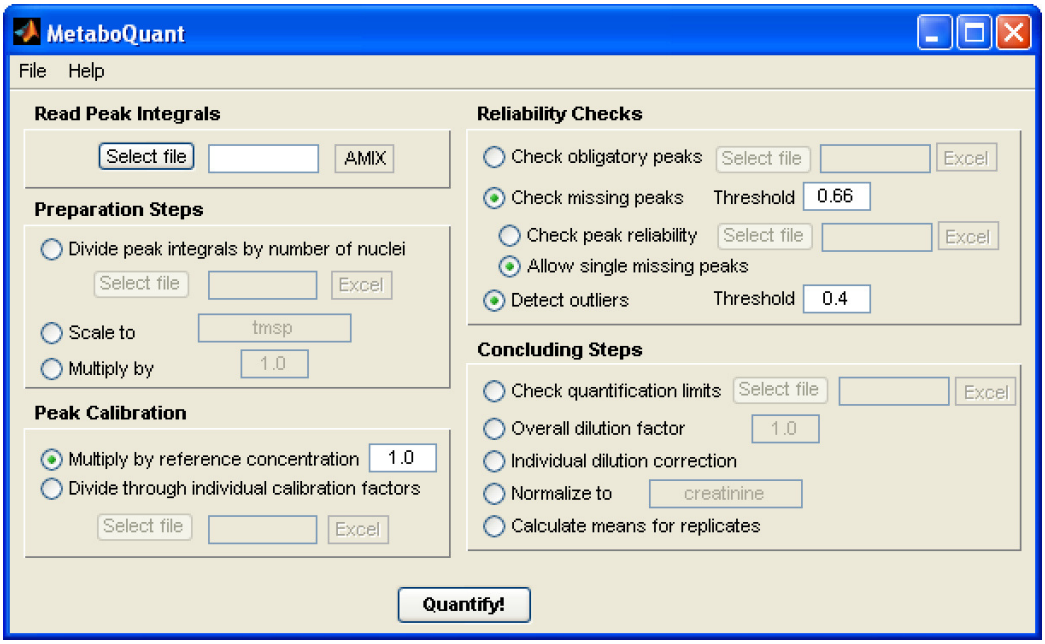

Figure 1.2: The main window of MetaboQuant in *basic mode* with default parameters

Prerequisites for using *Basic Mode* without changing any parameters are:

- Use of a pulse program that yields integrals directly proportional to the compound concentration, *e.g.* 1D inverse-gated spectra or time-zero HSQC spectra. When using other spectra, you should consider using individual calibration factors in the panel *Peak Calibration*. For this, you will need a *Peak Information File* (see Sections 2.2 and 1.5 for details).
- The peak integrals should have been divided by the number of contributing nuclei. In case this has not been done, activate the button *Divide peak integrals by number of nuclei* in the panel *Preparation Steps*. For this, you will need a *Peak Information File* (see Sections 2.2 and 1.5 for details).
- The peak integrals should have been divided by the integral of the reference substance, *e.g.* TSP. If this has not yet been done, activate the button *Scale to* in the panel *Preparation Steps* and enter the name of the reference compound in the matching field.

In the following paragraphs, a brief description of the required user inputs is given for the case that all prerequisites are fulfilled. For a description of all available parameters see Section 1.3.

First, you have to select the file where the peak integrals of the compounds of interest are stored. In the panel *Read Peak Integrals* click on *Select File* and choose the corresponding file. For details on the format of the file see Section 2.1. On Linux and other operating systems (other than Windows), there might be difficulties when reading Excel files, therefore it is recommended to use text files in these cases.

In the panel *Peak Calibration*, enter the concentration of your reference substance in the corresponding field.

You do not need to change any of the pre-filled values in the other panels.

In the end, click on the button *Quantify!*. This will open a dialog where you can choose the output file. Please do not use dots (".") in the file name. On Linux and other operating systems (other than Windows), there might be difficulties when writing Excel files, therefore it is recommended to use text files in these cases. After choosing the file name, quantification is performed and the result file opened.

For details on the result file see Section 1.4.

Example: Click *Select File* in the panel *Peak Integral File*. A file selection window will pop up. Change the file type to *Tab stop separated .txt* and choose the file *urine_control.txt* in the folder *Examples\Input_Files\Peak_Integral_Files.*

1.3 Advanced Mode

In this section, the usage of the *Advanced Mode* of MetaboQuant will be described step by step.

The Main Window

The main window of MetaboQuant in *Advanced Mode* is shown in Figure 1.3.

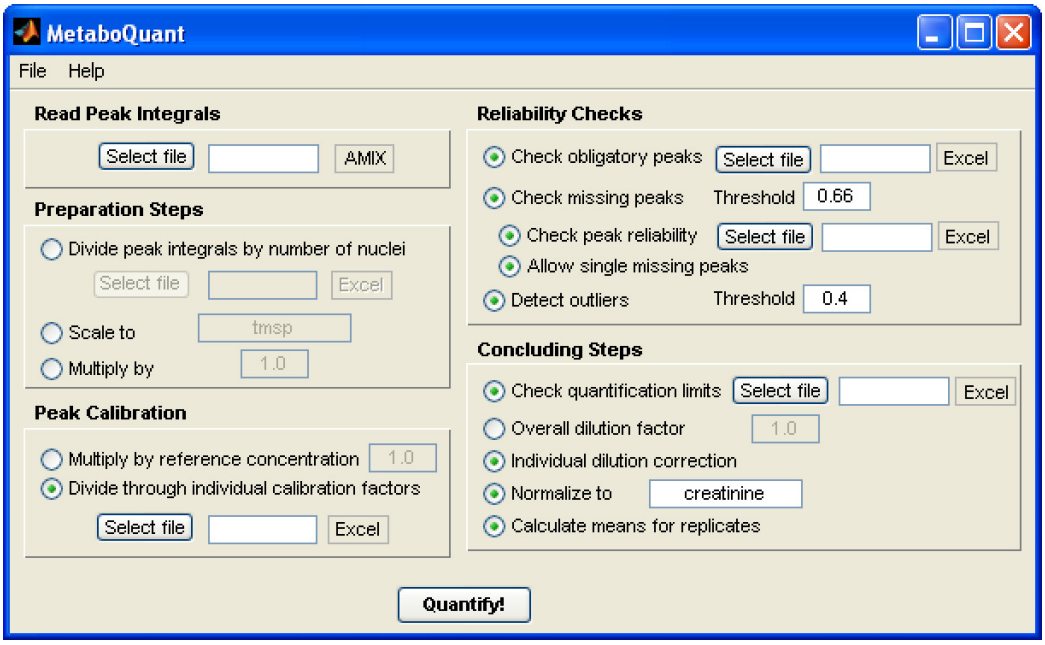

Figure 1.3: The main window with parameters for *advanced mode*

In the following paragraphs, all items of MetaboQuant are described.

The Menu Bar

- ***File* Menu**
  - **Load "Basic Mode" configuration**
    This will load the default parameters of *Basic Mode*, enabling a quick start without using additional files.
  - **Load "Advanced Mode" configuration (needs additional files)**
    This will load the parameter set for *Advanced Mode*.
  - **Load custom configuration**
    This will load a parameter set previously saved by the user.
    #
  - **Save custom configuration**
    This will save the current parameter set in a parameter file. Please note that the peak integral file name is not stored in the parameter file.
  - **Exit**
    This will close MetaboQuant.

- ***Help* Menu**
  - **Manual**
    This will open the manual (this document).
  - **About MetaboQuant**
    This will show the program version and additional information.

Read Peak Integrals

In this panel you have to choose the file where the integral values are stored. This file may be either an Excel *.xls* file, a tab stop separated *.txt* file or an Amix Analytic Profiler *.txt* file. For information on the data format of the file please see Section 2.1. Click on *Select File* and choose the matching file type and the peak integral file. After returning to the main window, the file name including the path of the chosen file is shown in the field right of the *Select file* button. Holding the mouse over this field will also show the file name as a pop-up window. The file type of the chosen file is shown in the rightmost field.

Please note that Microsoft Excel is required to read Excel files. If Excel is not installed on your computer, please use *.txt* files instead. For Linux and other operating systems (other than Windows), there might be difficulties when reading Excel files, therefore it is recommended to use text files as format in these cases.

MetaboQuant can also read files created by the Bruker Amix Analytic Profiler. For details on the parameters to use in Amix Analytic Profiler see Chapter 4.

Example: Click *Select file.* A file selection window will pop up. Change the file type to *Tab stop separated .txt* and choose the file *urine_control.txt* in the folder *Examples\Input_Files\Peak_Integral_Files.*

Preparation Steps

- **Divide peak integral by number of nuclei**
  Activate this button to divide each peak integral by the number of nuclei contributing to this peak. More than one nucleus may contribute to a certain peak, e.g. for CH_3-groups usually all three protons yield a signal at the same spectral position and, thus, all three signal sum up to a single peak. Therefore, it is necessary to divide the peak integral by the number of nuclei contributing to it. In case your peak integrals have already been divided by the number of nuclei, as in Amix output files (Bruker BioSpin, Rheinstetten, Germany), this check box does not need to be enabled. MetaboQuant will read the numbers of nuclei contributing to one peak from the *Peak Information File* (see Sections 2.2 and 1.5 for details). You will have to specify this file by clicking *Select file*. The file may be either an Excel *.xls* file or a tab stop separated *.txt* text file. Choose the matching file type and the file name and press *Open*. The file name including the path of the chosen file is shown in the field to the right. Holding the mouse over this field will also show the file name as a pop-up window. The file type is shown in the rightmost field.

Please note that this feature uses the same *Peak Information File* as the features *division through calibration factors*, *obligatory peak checking* and *peak reliability checking*. Selecting a file here will thus also change the file used for the other mentioned features.

- **Scale to**
  If this check box is enabled, all integral values of one spectrum will be divided by the integral of the reference substance specified in the corresponding text field. In case your peak integrals are already scaled to the reference substance signal, as in Amix output files (Bruker BioSpin), this check box does not need to be enabled. Please note that in case the references substance has more than one peak, the integral value of the first peak of the reference substance is used.

- **Multiply by**
  In case this check box is activated, each peak integral is multiplied by a correction factor before calculating the concentrations. The factor has to be specified in the corresponding filed and may be any positive or negative numeric value, default value is 1.0. This option is useful for example if your reference substance concentration is different in some of your spectrum sets, for example due to a change of supplier. The difference to the parameter *overall dilution factor* described below is that the correction factor is applied *before* checking for the LLOQ's, as a change in the reference substance concentration will result in false absolute concentrations in the first hand.

Peak Calibration

In this panel, you have to choose the peak calibration mode, either using individual calibration factors or using a reference concentration for scaling.

- **Multiply by reference concentration**
  If this check box is ticked, all peak integrals will be multiplied by the specified reference concentration to yield the final concentration. This makes sense only in case the peak integrals have been scaled to the integral of the reference substance. In case this has not been performed beforehand, you can achieve this by activating the option *Scale to* in Panel *Preparation Steps*. The reference concentration has to be positive (>0). It is important to correct the reference concentration to your sample amount.
  For example, if you add 400 µL of sample to 200 µL buffer and 50 µL of D_2O containing 10 mmol/L tmsp as reference substance, you have to enter *1.25* as reference concentration, as follows from the calculation:
  10 mmol/L * (50 µL / 400 µL) = 1.25 mmol/L.

- **Divide through individual calibration factors**
  In case this option is activated, each peak integral will be multiplied by an individual calibration factor. These factors are taken from the *Peak Information File* that you will have to specify. Individual peak calibration shall compensate for uneven excitation pulses, differences in magnetization transfer efficacy and other effects. This is especially necessary in

multidimensional experiments. You will find instructions on how to determine the calibration factors in Section 2.2.
You will have to specify a file where the information is stored by clicking *Select file*. The file may be either an Excel *.xls* file or a tab stop separated *.txt* text file. For the data format of the used file please see Section 2.2. Choose the matching file type and the file name and press *Open*. The file name including the path of the chosen file is shown in the field to the right. Holding the mouse over this field will also show the file name as a pop-up window. The file type is shown in the rightmost field.
Please note that this feature uses the same *Peak Information File* as the features *division through number of nuclei*, *obligatory peak checking* and *peak reliability checking*. Selecting a file here will thus also change the file used for the other mentioned features.

Please note that Microsoft Excel is required to read Excel files. If Excel is not installed on your computer, please use *.txt* files instead. For Linux and other operating systems (other than Windows), there might be difficulties when reading and writing Excel files, therefore it is recommended to use text files as format in these cases.

Example: Click *Select file*. A file selection window will pop up. Change the file type to *Tab stop separated .txt* and choose the file *urine_peakinfo.txt* in the folder *Examples\Input_Files\Peak_Information_and_LLOQ_Files*.

Reliability Checks

These parameters control the outlier detection and reliability checking of the results. The parameters are pre-filled with defaults that yield good results on ^{1}H-^{13}C-HSQC spectra of biofluids. Anyway, you may change these parameters in order to optimize your results.

- **Check obligatory peaks**
  When this button is activated, for each peak an *obligatory peak value* (taken from the *Peak Information File*) is evaluated. If a peak is marked as obligatory, but is not found in a spectrum, the compound is not quantified for this sample. This is useful for molecules with both high-intensity and low-intensity peaks, where one strong peak is often overlapped by other signals. By marking a weaker peak of the molecule as obligatory, this compound will only be quantified if the weaker peak is present. This is only meaningful if one of the weaker peaks is seldom overlapped. This will reduce the number of incorrect quantification results. When using this option, a *Peak Information File* has to be specified by clicking *Select file*. The file may be either an Excel *.xls* file or a tab stop separated *.txt* text file. For the data format of the used file please see Section 2.2. Choose the matching file type and the file name and press *Open*. The file name including the path of the chosen file is shown in the field to the right. Holding the mouse over this field will also show the file name as a pop-up window. The file type is shown in the rightmost field.

Please note that this feature uses the same *Peak Information File* as the features *division through number of nuclei*, *division through calibration factors* and *peak reliability checking*. Selecting a file here will thus also change the file used for the other mentioned features.

- **Check missing peaks**
  When activation this check box, a reliability check using the number of found peaks is performed. The threshold controls how many of the peaks available for a molecule have to be present in a spectrum. The ratio of found peaks to total peaks has to exceed the threshold. The threshold has to be a numeric value in the range of 0 to 1. Default is 0.66, this means that at least two of three peaks of one compound have to be present to quantify a compound.

- **Check peak reliability**
  This check shall refine the check for missing peaks (see above) by reducing the number of peaks that are falsely excluded due to being below the threshold for missing peaks. For this, the entry *Intensity* in the *Peak Information File* is used. In cases where for one compound less peaks are available than required by the threshold for missing peaks, the number of nuclei contributing to the found peaks is compared to the number of nuclei contributing to the peaks not found in the spectrum. If the maximal number of nuclei contributing to one of the found peaks is higher than the maximum number contributing to the not-found peaks, the peak is deemed to be reliable. This is based on the assumption that peaks arising from many nuclei will have higher intensities. In cases of low abundance, it is to be expected that only peaks with many nuclei will be visible above noise level. These peaks will thus allow a reliable quantification in these cases. When using this option, a *Peak Information File* has to be specified by clicking *Select file*. The file may be either an Excel *.xls* file or a tab stop separated *.txt* text file. For the data format of the used file please see Section 2.2. Choose the matching file type and the file name and press *Open*. The file name including the path of the chosen file is shown in the field to the right. Holding the mouse over this field will also show the file name as a pop-up window. The file type is shown in the rightmost field.
  Please note that this feature uses the same *Peak Information File* as the features *division through number of nuclei, division through calibration factors* and *obligatory peak checking*. Selecting a file here will thus also change the file used for the other mentioned features.

- **Allow single missing peaks**
  This parameter controls whether single missing peaks of one compound shall not lead to the exclusion of this compound even in case the number of found peaks would drop below the threshold for missing peaks then. As single peaks might always be missed by a peak-picking routine, activating this option may yield valid concentration values that might have been excluded otherwise. In default setting, this option is activated.

- **Detect outliers**
  This button activates an internal outlier detection and removal routine. After dividing all peak integrals by the individual calibration factor or multiplying by the reference concentration (depending on the option selected in the panel *Peak Calibration*), all peaks of one compound are analyzed. For compounds with more than one peak, the median of all peaks is calculated, and peaks differing from the median relatively by more than the specified threshold will be treated as outliers and excluded from analysis. The threshold has to be a positive numeric value (>0), default value is 0.4. In case only two peaks are available for a compound, and that these two peaks differ by more than the threshold from their median, the lower peak integral is chosen, as peak overlap usually results in too high peak integrals.

Concluding Steps

- **Check quantification limits**
  Here you have to choose whether you want to check if the calculated concentrations are above individually defined lower limits of quantification (LLOQ's). Tick the check box *Enable* to activate this function. In this case, you will have to select a file containing the LLOQ's by clicking on *Select file*. The file may be either an Excel .xls file or a tab stop separated .txt file. For the data format of the used files please see Section 2.3. After choosing a file, the file name including the path of the chosen file is shown in the field to the right. Holding the mouse over this field will also show the file name as a pop-up window. The file type is shown in the rightmost field.

  Please note that Microsoft Excel is required to read Excel files. If Excel is not installed on your computer, please use .txt files instead. For Linux and other operating systems (other than Windows), there might be difficulties when reading Excel files, therefore it is recommended to use text files as format for in these cases.

  Example: Click *Select file*. A file selection window will pop up. Change the file type to *Tab stop separated .txt* and choose the file *urine_LLOQs.txt* in the folder *Examples\Input_Files\Peak_Information_and_LLOQ_Files.*

- **Overall dilution factor**
  If this option is activated, each concentration value is multiplied by this factor (after checking if the concentration is above the LLOQ, in case *check limits of quantification* was chosen). The factor may be any positive or negative numeric value, default value is 1.0. The overall dilution factor may be used if all samples are diluted in the same way. The difference to the parameter *correction factor* mentioned above is that diluting the sample will not result in false absolute concentrations. Therefore, the dilution factor is applied *after* checking the LLOQ's. If each sample is diluted in a different way, you may use the *Individual dilution correction* explained below.

- **Individual dilution correction**
  If this check box is activated, individual dilution correction will be performed on all spectra. This may be used in case different sample amounts were used for the different spectra, or in case of tissue extracts, where different tissue amounts were employed. The *individual dilution factor* for each spectrum has to be provided in the *Peak Integral File* as part of the spectrum title. Please be aware of that using samples with very different dilutions may cause changes in chemical shift and thus hamper quantification. For details see Section 2.1.

- **Normalize to**
  If this check box is ticked, all concentrations will be divided by the concentration of the compound entered in the corresponding text field. For urine, this is usually done with the creatinine concentration, but you may choose any compound.

- **Calculate means for replicates**
  If this check box is activated, MetaboQuant will search for replicate spectra of the same sample. Replicates have to be indicated in the spectrum title, for details see Section 2.1. Means and technical errors (TE's) will be calculated in case replicates are found. For TE calculation only the first two replicate spectra are used.

Quantify!

When the button *Quantify!* is clicked, you are requested to choose an output file. This may either be a tab stop separated *.txt* file or a Microsoft Excel file. Please do not use dots (".") in the file name, as this may cause missing file extensions.

Please note that Microsoft Excel is required to write Excel files. If Excel is not installed on your computer, please use *.txt* output files instead. For Linux and other operating systems (other than Windows), there might be difficulties when writing Excel files, therefore it is recommended to use text files as format in these cases. For Excel files, it will automatically be determined which Excel version is installed on the computer and the matching file format and file extension (*.xls* or *.xlsx*) will be chosen.

The status of the calculations is shown as a progress bar during the quantification. After successful quantification, the result file will be automatically opened.

Example: Click *Quantify!* A *save as*-window will pop up. Select *Save.*

1.4 Result Files

An example for an Excel result file is shown in Figure 1.5. Additional example result files can be found in the folder *MetaboQuant\Examples\Output_Files*.

The compound concentration values will be stored in the spreadsheet *Results*. For each compound the mean, the standard deviation and the minimal and maximal values over all spectra will be shown at the lower end of the table. Additionally, the number of samples containing this compound is shown. Please note that in case one or more of the features *individual dilution correction*, *normalize to* or *calculate means for replicates* were selected, the spreadsheet containing the original results will be named *Uncorrected Results* instead of *Results*.

All used parameters, file and folder names, additional information such as date and version numbers, user data, warnings and error messages will be stored in the spreadsheet *Configuration*. All factors will have a single quotation mark (‘) as prefix to circumvent Excel to convert the variable type in an unwanted way.

In case compounds were excluded because the number of found peaks was below the threshold for the check for missing peaks, these compounds and the respective peak intensities are stored in the spreadsheet *Too Few Peaks*. If for an excluded compound the *reliability check* was successful, the values of this compound are stored in the spreadsheet *Accept after Reliability Check*.

If MetaboQuant identified peaks as outliers, all original values of this compound are stored in the spreadsheet *Outliers*.

	A	B	C	D	E	F	G
1		tmsp	alanine	creatinine	hippuricacid	taurine	citricacid
2	M02a	6.01106035		5.11151127		0.5425398	0.6672607
3	M03a	6.01106035	0.31333764	11.3110268	4.52818979	0.40361721	1.37600122
4	M05a	6.01106035	0.38812946	11.9302793	0.88494328	0.56647469	1.07826411
5	M06a	6.01106035		4.83247691	5.6322102		
6	M07a	6.01106035	0.64464121	8.96948977	0.53190112	0.32959717	1.18644781
7	W01a	6.01106035	0.53265229	4.75414713	1.23225415		2.1372904
8	W02a	6.01106035	0.10385545	4.42004131	0.91935238		1.60446228
9	W03a	6.01106035	0.26222326	17.0044899	4.10001793		2.16415063
10	W05a	6.01106035	0.32378463	13.0993454	17.0690314	0.61037674	4.56921526
11	W06a	6.01106035		6.10641444	1.70205886	0.79236957	0.87220087
12	z1 Mean	6.01106035	0.36694628	8.75392222	4.06666212	0.5408292	1.73947703
13	z2 Standard (	9.3622E-16	0.17792423	4.39777696	5.21866019	0.16261373	1.18010724
14	z3 Minimum	6.01106035	0.10385545	4.42004131	0.53190112	0.32959717	0.6672607
15	z4 Maximum	6.01106035	0.64464121	17.0044899	17.0690314	0.79236957	4.56921526
16	z5 Number o	10	7	10	9	6	9
17							

Accept after Reliability Check | Outliers | Used Peaks | Configuration | Results

Figure 1.5: Example for an Excel result file. The different spreadsheets of the file can be seen on the lower margin of the figure.

Information about the number of used peaks for each compound of each spectrum is stored in the spreadsheet *Used Peaks*. For each metabolite, two entries are given, the left one containing the concentration from the *Results* spreadsheet, the right one indicating the ratio of used peaks to available peaks for this compound. For example, (6/6) would mean that all six peaks were used for the calculation of

the mean value, indicating a low variance in peak intensities, as indicated by the lack of excluded outliers. This implies a high reliability of the result. In contrast, a value of (1/6) would mean that only one out of six values was used, indicating either a low number of visible peaks or a high variance that led to outlier exclusions. In such cases, a manual checking of the other peaks, for example in the *Outliers* spreadsheet, might be necessary.

In case *Individual Dilution Correction* was chosen, the dilution corrected concentrations are stored in the spreadsheet *Individually Dilution Corrected* in the output file.

In case *Calculate Means for Replicates* was chosen, means and technical errors for replicate samples will be written to the spreadsheet *Mean* in the output file.

In case *Normalize to* was chosen, a spreadsheet named *Normalized to compound name* is created, e.g. *Normalized to creatinine*. This spreadsheet contains the normalized values.

If *Individual Dilution Correction, Calculate Means for Replicates* or *Normalize to* was chosen, the uncorrected results are written to the sheet *Uncorrected Results*.

All original values (before checking for limits of quantification and multiplying by calibration and dilution factors) will be stored in the spreadsheet *Original Values*.

If you choose *.txt* files as output, several *.txt* files will be generated containing the data. The files have a fixed number of columns that are separated by a tabulator. The file names are derived from the entered file name by adding extensions, for example *Filename_Results.txt*.

1.5 Step-by-Step Guide for Adding New Compounds

To perform an exact quantification for a compound, several specific values have to be added to a peak information file and a limit of quantification file. In the following paragraphs, the necessary steps will be described.

Peak Information File

In case you do not have a peak information file yet, create a new file, either a text .txt file or an Excel .xls file. In text files, the columns must be separated by tab stops. For details on the file format, see Section 2.2. The format should be as follows:

	Obligatory Peak	Intensity (Number of Nuclei)	Calibration Factor	Peak Used for Quantification?
Compound A				
Peak 1	0	3	1.39	1
Peak 2	1	1	1.12	1
Peak 3	0	1	1.37	0
Compound B				
Peak 1	0	1	1.23	1
Peak 2	0	2	1.07	1
...	...	...	...	...

Compound Name

Enter the compound name in the first field of the second column. Each compound has to have a unique name.

Peak Names

Define unique names for each peak of the compound. These names have to be unique only within one compound, in other compounds the same peak names may be used! Enter each of the compound names in the first field of a line following the compound name.

Obligatory Peak

In the column *Obligatory peak*, you can mark single peaks as obligatory. A value of 0 means that the peak is not required to be present and 1 means that the peak must be present to quantify the compound. This is useful for molecules with both high-intensity and low-intensity peaks, where one strong peak is often overlapped by other signals. By marking a weaker peak of the molecule as obligatory, this compound will only be quantified if the weaker peak is present. This is only meaningful if one of the weaker peaks is seldom overlapped. This will reduce the number of incorrect quantification results. Usually, 0 should be chosen for all peaks.

Intensity (Number of Nuclei)

In this field, you have to enter the number of nuclei (usually protons) that give rise to this particular peak.

Calibration Factors

The calibration factors can be determined experimentally using a dilution series of the pure compound. Start by creating a solution with the maximal expected concentration, and dilute this solution by factor 2 until you reach either the minimum expected concentration or the average lower limit of quantification. Measure a sample of each dilution step using the same pulse program you will use for your biologic samples of interest. Calculate the peak integrals using the same software as you will use on your samples of interest.

Plot the measured peak integrals against the concentration values for each peak, e.g. using Excel. Fit the line with a regression line going through the point of origin as shown in Figure 2.1. The slope of the regression line is the calibration factor for this particular peak. The calibration factors may be any positive or negative number.

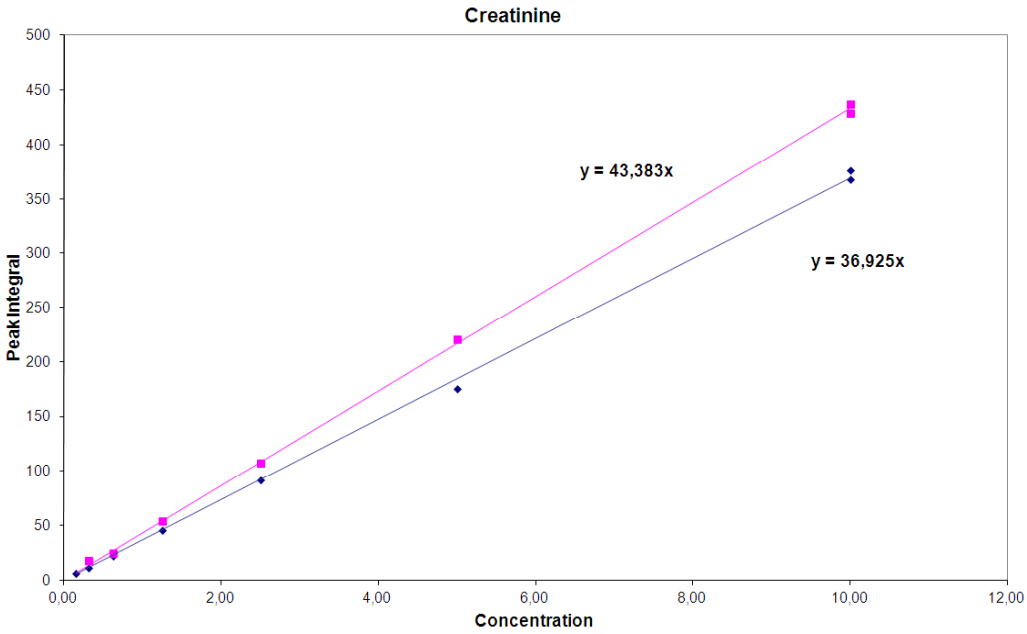

Figure 2.1: Dilution series for determination of peak calibration factors

Peak Used for Quantification?

In case a peak shall not be used for quantification, e.g. due to overlap with other peaks, enter 0 here. In all other cases, enter 1 here.

Limit of Quantification File

In case you do not have a limit of quantification file yet, create a new file, either a text .txt file or an Excel .xls file. In text files, the columns must be separated by tab stops. For details on the file format, see Section 2.3.
Enter the compound name in the first field of the second column. In case the file already exists, add a new line at the lower end of the file. The format should be as follows:

Compound	LLOQ [mmol/L]
Compound A	*0.15*
Compound B	*0.30*
...	...

To determine the lower limit of quantification, you have to create several solution of the pure compound with different dilutions. Measure each of the sample several times (e.g. six times) using the same pulse program as for your samples of interest, and integrate the peaks using the same procedure as for you samples of interest. For each peak, calculate the relative standard deviation of one dilution point. The lower limit of detection is the lowest concentration value, where the relative standard deviation for repeated measurements of one peak is below 20%. Enter this value in the second column for this compound.

2 File Formats

To use MetaboQuant, several files may be necessary. As a minimum, a file containing peak integral values has to be available. For more exact quantification results, files containing additional information on used peaks and compounds are necessary. These files are explained in the following sections:

2.1 Peak Integral File

A file containing the peak integral values from one spectrum or from several spectra is required to use MetaboQuant. Usually you should be able to create such a file using your spectrometer software, e.g. with Bruker Amix (Bruker BioSpin, Rheinstetten, Germany). This file may be a Microsoft Excel *.xls* file, a tab stop separated *.txt* file or an Amix Analytic Profiler *.txt* file.

The data have to be present as follows in the file:

title: *Sample 1*	
Compound name 1	
Peak name 1	*Integral value**
Peak name 2	*Integral value**
Peak name 3	*Integral value**
Compound name 2	
Peak name 1	*Integral value**
Peak name 2	*Integral value**
title: *Sample 2*	
Compound name 1	
Peak name 1	*Integral value**
Peak name 2	*Integral value**

* The integrals may be any positive or negative numeric values including zero.

Empty lines are ignored and may therefore be inserted at any position in the file. Each spectrum title has to start with '*title: '.* The number and order of compounds does not have to be the same for all samples. Peaks that shall not be used for quantification may be included in the list, in these cases the integral values has to be replaced by a string, for example *not used*. In case a peak was not found in the spectrum, its value should be zero ("0"). For example files see the folder *MetaboQuant\Examples*.

The *.txt* files have to contain exactly two columns that are separated by a tabulator. Make sure to use points (".") as decimal separators.

Excel files have to come in the format *.xls*. The data have to be stored in a spreadsheet named *Integrals*, starting in cell *A1*.

For creating Amix *.txt* files please see Section 4 for further information.

Individual Dilution Factors

In case you want to use *individual dilution factors* (see Section 1.3 for details), the individual dilution factor has to be written at the end of the spectrum title and has to start with *df*. All concentrations of this spectrum will then be multiplied by this factor. Please be aware of that using samples with very different dilutions may cause changes in chemical shift and thus hamper quantification.

Example: You have diluted sample 1 by factor 4, sample 2 by factor 13.7 and sample 3 by factor 200. The peak integral file then has to look as follows:

title: Sample 1 **df4**	
Peak name 1	*Integral value*
Peak name 2	*Integral value*
Peak name 3	*Integral value*
title: Sample 2 **df13.7**	
Peak name 1	*Integral value*
Peak name 2	*Integral value*
Peak name 3	*Integral value*
title: Sample 3 **df200**	
Peak name 1	*Integral value*
Peak name 2	*Integral value*
Peak name 3	*Integral value*

Replicate Samples

In case you measure single samples repeatedly, MetaboQuant can calculate means and technical errors (see Section 1.3). Replicate spectra have to be titled *a, *b, *c and so on in the peak integral file, where * may be any string, as shown below:

title: Sample 1**a**	
Compound name 1	
Peak name 1	*Integral value*
Peak name 2	*Integral value*
Peak name 3	*Integral value*
title: Sample 1**b**	
Compound name 1	
Peak name 1	*Integral value*
Peak name 2	*Integral value*
Peak name 3	*Integral value*
title: Sample 1**c**	
Compound name 1	
Peak name 1	*Integral value*
Peak name 2	*Integral value*
Peak name 3	*Integral value*

2.2 Peak Information File

This file contains additional information about the peaks of a compound to enable accurate quantification. This file may be either a Microsoft Excel *.xls* file or a tab stop separated *.txt* file.

The *.txt* file has to contain exactly five columns that are separated by a tab stop. Make sure to use points (".") as decimal separators.

Excel files have to come in the format *.xls*. The data have to be stored in a spreadsheet named *Factors*, starting in cell *A1*.

The data have to be present in the file as in the following example:

	Obligatory peak	Intensity (number of nuclei)	Calibration factor	Peak used for quantification?
Compound A				
Peak 1	0	3	1.39	1
Peak 2	1	1	1.12	1
Peak 3	0	1	1.37	0
Compound B				
Peak 1	0	1	1.23	1
Peak 2	0	2	1.07	1

Obligatory Peak

The column *Obligatory peak* can be used to mark single peaks as obligatory. Possible values are 0 and 1. 0 means that the peak is not required to be present and 1 means that the peak must be present to quantify the compound. Usually, 0 should be chosen for all peaks. This column is only evaluated by MetaboQuant in case *Check obligatory peaks* is activated.

Intensity (Number of Nuclei)

In the column *Intensity (number of nuclei)*, the number of nuclei contributing to the peak has to be entered. This number is used for dividing peak integrals and for a reliability check in case only few peaks of one compound were found. This column is only evaluated by MetaboQuant in case *Divide integrals by number of nuclei* and/or *Reliability check for missing peaks* are activated.

Calibration Factor

In the column *Calibration factor*, calibration factors for each peak have to be entered, e.g. in the unit 1/(mmol/L). This column is only evaluated by MetaboQuant in case *Use individual calibration factors* is activated.

Peak Used for Quantification?

The entries of the column *Peak used for quantification?* indicate whether a peak shall be used for quantification or not. In case the peak shall be used for quantification, the entry should be 1. In case a 0 is entered here, the peak is not used for quantification. In this case, this peak is not used for counting the number of found or not-found peaks for checking missing peaks and peak reliability checking. This column is evaluated by MetaboQuant whenever one of the options *Check obligatory peaks*, *Divide integrals by number of nuclei*, *Reliability check for missing peaks* or *Use individual calibration factors* is activated.

2.3 Limits of Quantification File

In this file, the lower limits of quantification (LLOQ's) are stored for automated checking of the quantification results. This file may be either a Microsoft Excel *.xls* file or a tab stop separated *.txt* file.

The tab stop separated *.txt* file has to contain exactly two columns that are separated by one tab stop. Make sure to use points (".") as decimal separators.

Excel files have to come in the format *.xls*. The data have to be stored in a spreadsheet named *LLOQs*, starting in cell *A1*.

The LLOQ's may be any numeric values and have to be present as follows in the file:

Compound	LLOQ [mmol/L]
Compound A	*0.15*
Compound B	*0.30*

For an example, see the exemplary files in the *Examples* folder.

3 Troubleshooting

- **Starting MetaboQuant Fails**
  If MetaboQuant fails to open, please install the file *MCRInstallerR2007b.exe*. This file is available from the same source as MetaboQuant, for example from the software section on http://genomics.uni-regensburg.de/site/institute.

- **Opening Output Files Fails**
  In case the output file cannot be opened, check whether the file has the correct file extension (e.g. *.xls, .xlsx*). In case the file has no extension, add the matching extension. Missing file extensions are usually caused by filenames containing dots (“.”). This is due to a bug in Matlab.

- **Error: *Subscripted assignment dimension mismatch***
- **Error: *Attempted to access ...; index out of bounds***
  These error messages indicate problems while reading a peak integral value file. Make sure that the file matches the requirements (see Section 2.1 for Excel and text files or Section 4 for Amix files).

- **Warning: *Dilution factors were found only for x out of y samples!***
  This warning means that MetaboQuant could not recognize individual dilution factors for all samples. This might be because you did not specify dilution factors for all samples, or due to errors in the format of the dilution factors. Make sure that the dilution factor starts with *df* and that it is written at the end of the spectrum title without any following characters. Also make sure to use points (.) as decimal separators.

- **Error: *Data format not supported. No Amix file?***
  If you encounter this error message when reading peak integral values from a *.txt* file, most probably MetaboQuant tried to open a tab stop separated *.txt* file as an Amix *.txt* file. In case you use Amix *.txt* files, something seems to be wrong with the file format. Make sure you follow the steps in Section 4 when creating Amix *.txt* files.

  In case you use a tab stop separated *.txt* file, make sure you choose the entry *Tab stop separated text* from the file type menu when selecting the file using *Browse* in Panel 1 (see Figure 3.1).

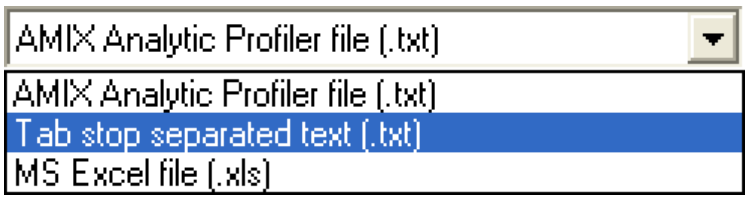

Figure 3.1: Choosing the integral value file type

4 Creating Peak Integrals Using Amix

In order to create peak integral files using Amix Analytic Profiler (Bruker BioSpin, Rheinstetten, Germany), you should follow the steps described below to ensure that the files are readable for MetaboQuant. The procedure is described for Amix version 3.9.3 and may differ for other versions. Parameters not explicitly named in this section should not affect the readability of the files and may therefore be chosen according to your preferences.

You have to activate both *identification* and *quantification* in Amix in order to get usable files.

The file containing the peak integral is called *mprofile.txt*.

4.1 Amix Peak Integrals from 1D Spectra

To quantify compounds from 1D spectra, the *.txt* file may not contain any 2D HSQC data. To achieve this, you have to deactivate the buttons for 2D/HSQC identification and quantification.

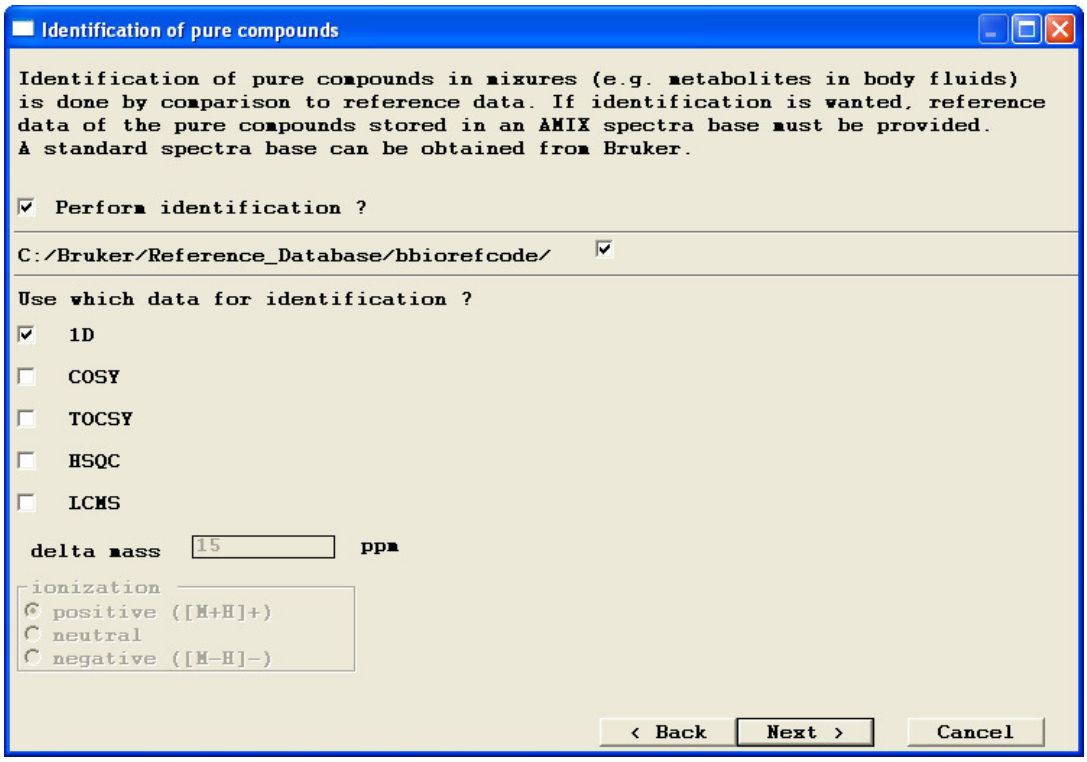

Figure 4.1: *Perform identification* and *1D* have to be enabled

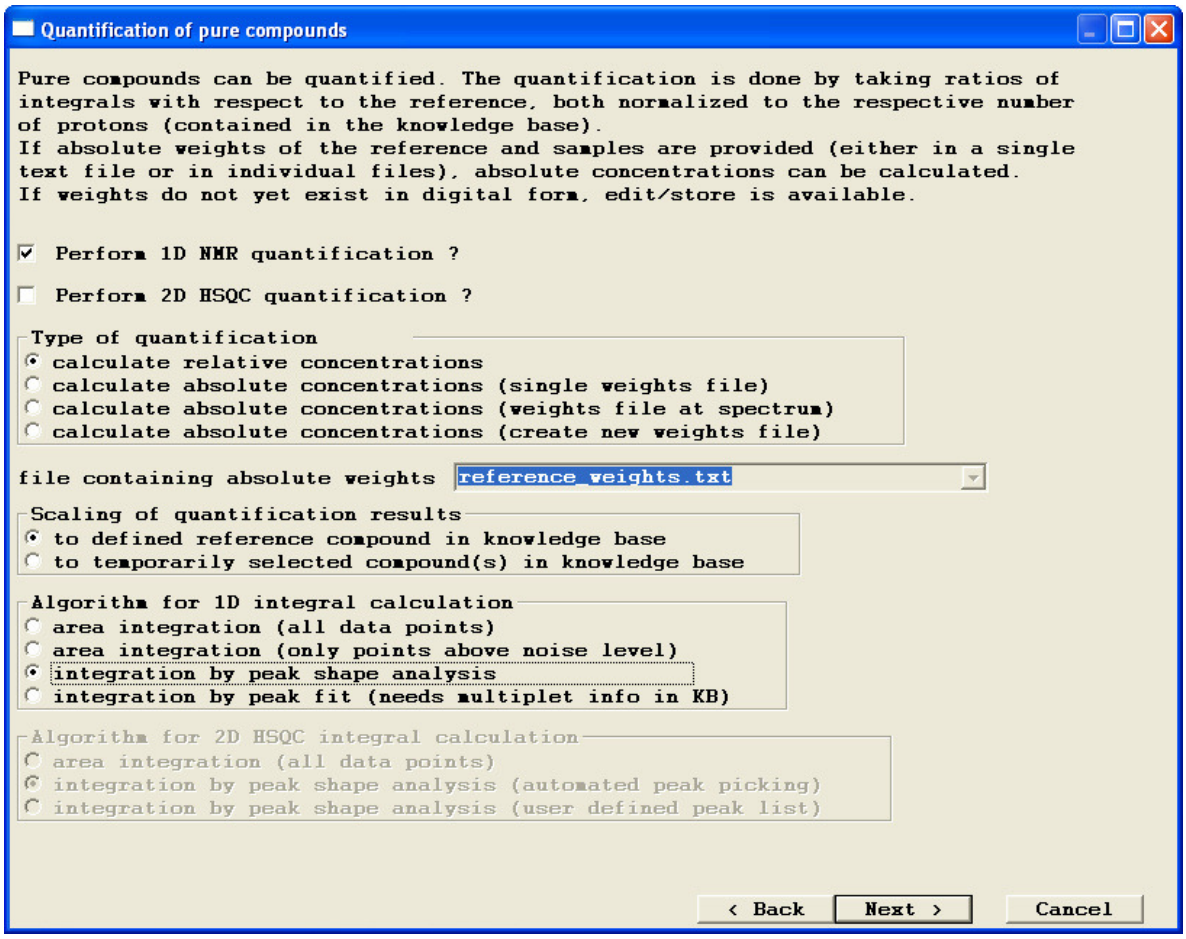

Figure 4.2: *Perform 1D quantification* has to be enabled

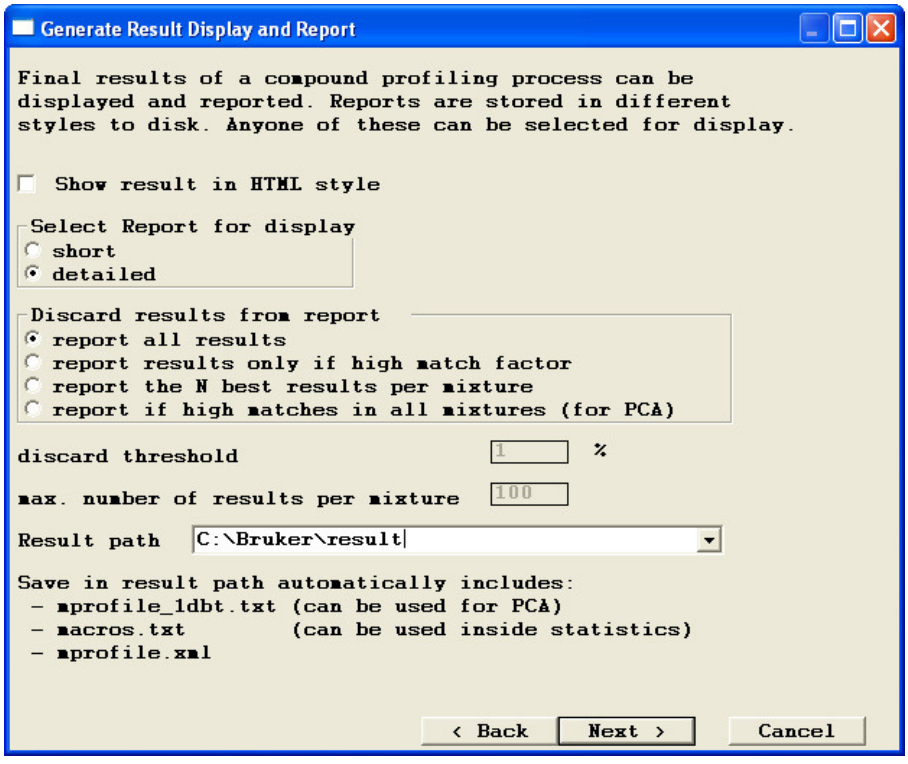

Figure 4.3: *Detailed* and *Report all results* have to be enabled

4.2 Amix Peak Integrals from 2D HSQC Spectra

When quantifying from 2D HSQC spectra, 1D spectra have to be quantified as well, although this data will not be used later on. MetaboQuant will look for 2D HSQC data in the file and use these for quantification.

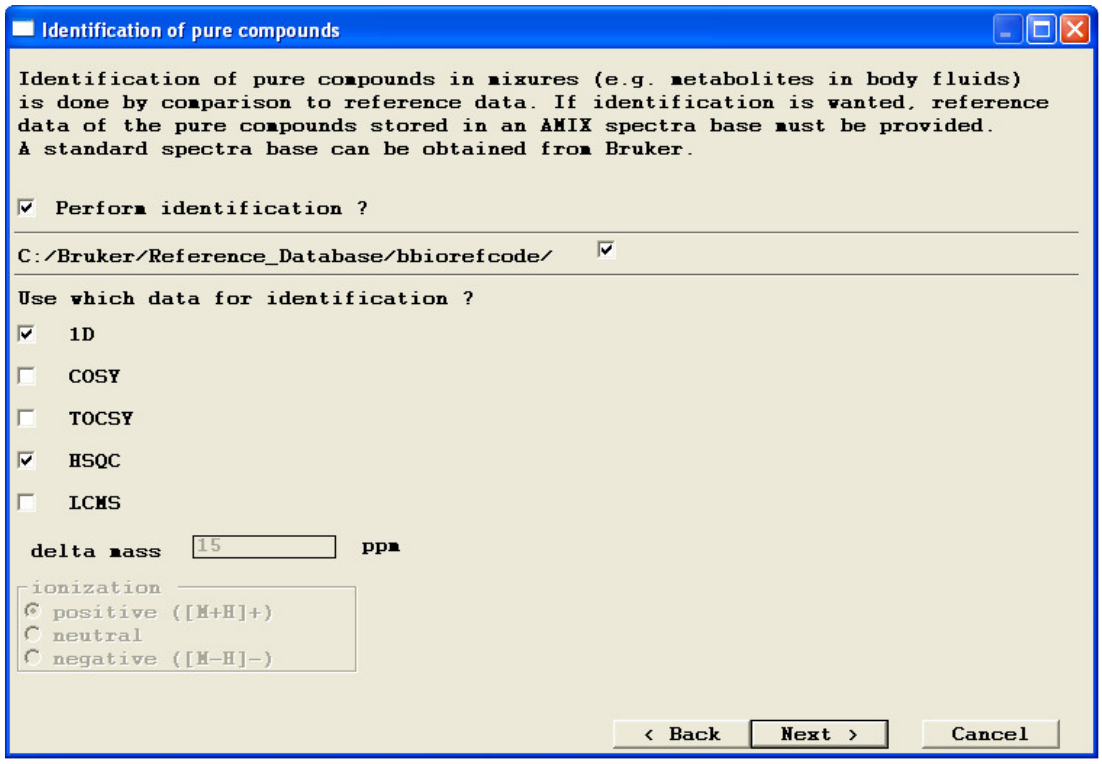

Figure 4.4: *Perform identification* has to be enabled both for *1D* and *HSQC*

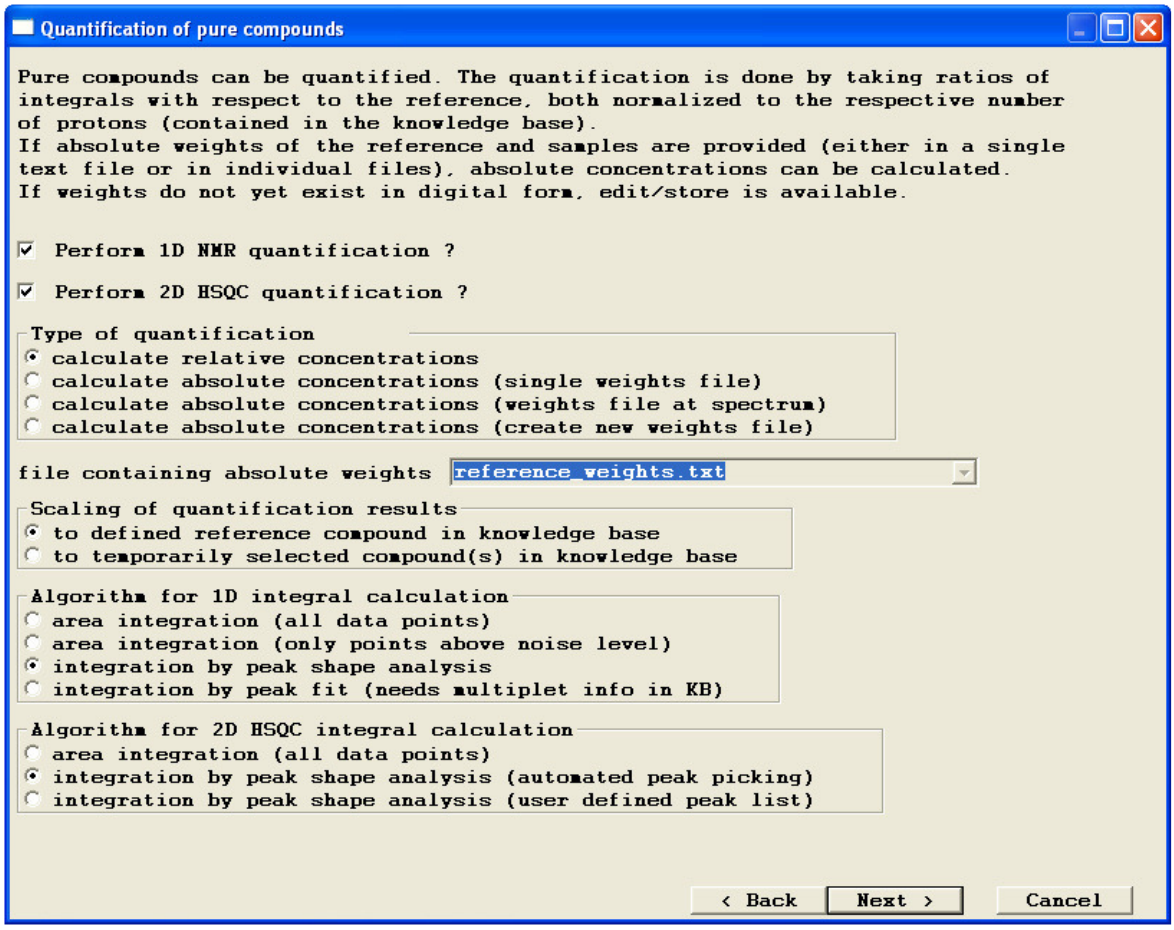

Figure 4.5: Both *Perform 1D quantification* and *Perform 2D quantification* have to be enabled

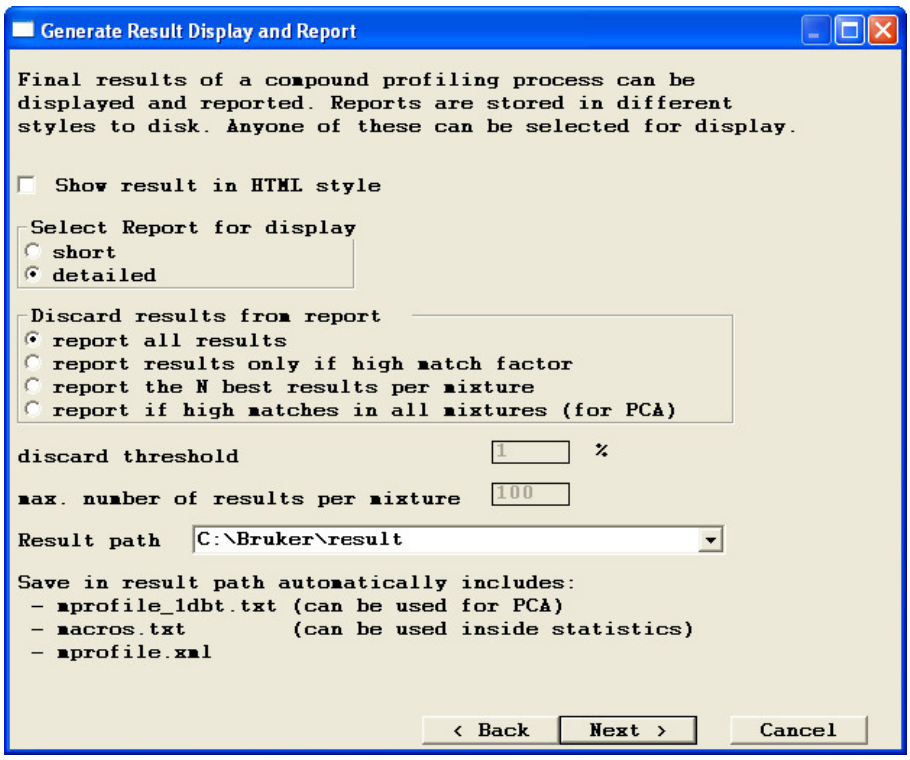

Figure 4.6: *Detailed* and *Report all results* have to be enabled