QUANTIFY 1.1.1

Manual

Matthias S. Klein
Institute of Functional Genomics
University of Regensburg
Josef-Engert-Str. 9
D-93053 Regensburg, Germany

matthias.klein@klinik.uni-regensburg.de

Table of Contents

1 License Agreement 1
2 Introduction 2
3 Installation 3
3.1 System Requirements 3
3.2 Installation 3
4 Manual 4
4.1 User Files 4
4.2 Step-by-Step Guide 8
4.3 Quick Start 15
5 Troubleshooting 17
6 Creating Peak Integrals Using AMIX 18
6.1 AMIX Peak Integrals from 1D Spectra 18
6.2 AMIX Peak Integrals from 2D HSQC Spectra 20

1 License Agreement

This program comes with no warranty, you use it at your own risk. You may use and distribute this program freely, as long as credit is given to the original author.

All product, brand and company names in this document may be trademarks of their respective owners.

2 Introduction

QUANTIFY 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 not be visible in other samples. This makes exact quantification of selected compounds a hard task, as each peak of the compound may or may not be overlapped with other compounds. This is not only true for 1D spectra, but also for 2D spectra such as ^{1}H-^{13}C HSQC spectra.

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

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

The features of QUANTIFY include:

- Multiplication of each integral with an individual calibration factor
- Checking for obligatory peaks before quantification
- Automatic outlier removal
- Reliability checking for compounds where only few peaks are visible in the spectrum
- Checking lower limits of quantification
- Automatic normalization to single compounds, e.g. creatinine
- Automatic correction for different dilutions
- Calculation of means and technical errors for replicate measurements
- Logging of the used parameters and configuration to allow reproducible results

3 Installation

3.1 System Requirements

Quantify.exe may be run using the following operating systems:

- Windows XP (Service Pack 1 or 2)
- Windows 2000 (Service Pack 3 or 4)
- Windows Server 2003 (Service Pack 1 or 2)
- Windows Vista

On other operating systems such as Linux or Mac OS X, you may run the source code file *quantify.m* directly. For this, it is required to have installed the mathematical software Matlab (The Mathworks, Natick, MA, USA). Copy the files *quantify.m* and *quantify.fig* from the folder *QUANTIFY* to your Matlab working directory, start Matlab and type *quantify* to start QUANTIFY. QUANTIFY is backward compatible at least until Matlab version 7.1.0.246 (R14) Service Pack 3.

3.2 Installation

Open the file *quantify.zip* with unzipping software (e.g. Winzip) and extract the files.

Run the file *MCRInstallerR2007b.exe*. This installs the required Matlab run-time routines and should be available from the same source as the file *quantify.zip*, for example from the software section on http://genomics.uni-regensburg.de/.

To start QUANTIFY, run the unzipped file *quantify.exe*.

4 Manual

4.1 User Files

4.1.1 Required Files

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

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

title: *Spectrum 1*	
Compound A	
Peak 1	*Integral value**
Peak 2	*Integral value**
Peak 3	*Integral value**
Compound B	
Peak 1	*Integral value**
Peak 2	*Integral value**
title: *Spectrum 2*	
Compound A	
Peak 1	*Integral value**
Peak 2	*Integral value**
...	...

* The integral values may be any positive or negative number including zero and should have been divided by the number of nuclei contributing to the peak.

Each spectrum title has to start with '*title: '*. Please make sure to include a blank line after each spectrum title and after the last peak of each compound. The number and order of compounds does not have to be the same for all samples, although it is recommended to have the same compounds in each sample to get consistent results. 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. For example files see the folder *QUANTIFY\Examples*.

The *.txt* files have to contain exactly two columns that are separated by a tabulator.

For using AMIX *.txt* files please see Chapter 6 for further information.

In case you want to use individual *dilution factors* (see Section 4.2 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.

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: Spectrum 1 **df4**	
...	...
title: Spectrum 2 **df13.7**	
...	...
title: Spectrum 3 **df200**	
...	...

In case you measure single samples repeatedly, QUANTIFY can calculate means and technical errors (see Section 4.2). Replicate spectra have to be titled *a, *b, *c and so on in the peak integral file, where * may be any string, e.g. sample1_a, sample1_b, sample1_c.

4.1.2 Optional Files

Optionally, a file containing peak information and a file containing lower limits of quantification may be used.

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

	Obligatory peak	Number of nuclei	Calibration factor
Compound A			
Peak 1	0	3	1.39
Peak 2	1	1	1.12
Peak 3	0	1	0
Compound B			
Peak 1	0	1	1.23
Peak 2	0	2	1.07
...	...	...	...

The column *Obligatory peak* can be used to mark single peaks as obligatory. This may be 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. 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.

In the column *Number of nuclei*, the number of nuclei contributing to the peak has to be entered. This number is used for a reliability check in case less peaks than the *peak threshold* are available for one compound.

In the column *Calibration factor*, calibration factors for each peak have to be entered. The calibration factors can be determined experimentally using a dilution series of the pure compound. The measured peak integrals have to plotted versus the concentration values and fitted with a regression line going through the point of origin, see Figure 4.1 for an example. The slope of the regression line is the calibration factor. The calibration factors may be any positive or negative number. In case a peak shall not be used for quantification, e.g. due to overlap with other peaks, enter 0 as calibration factor.

The *.txt* file has to contain exactly four columns that are separated by a tab stop.

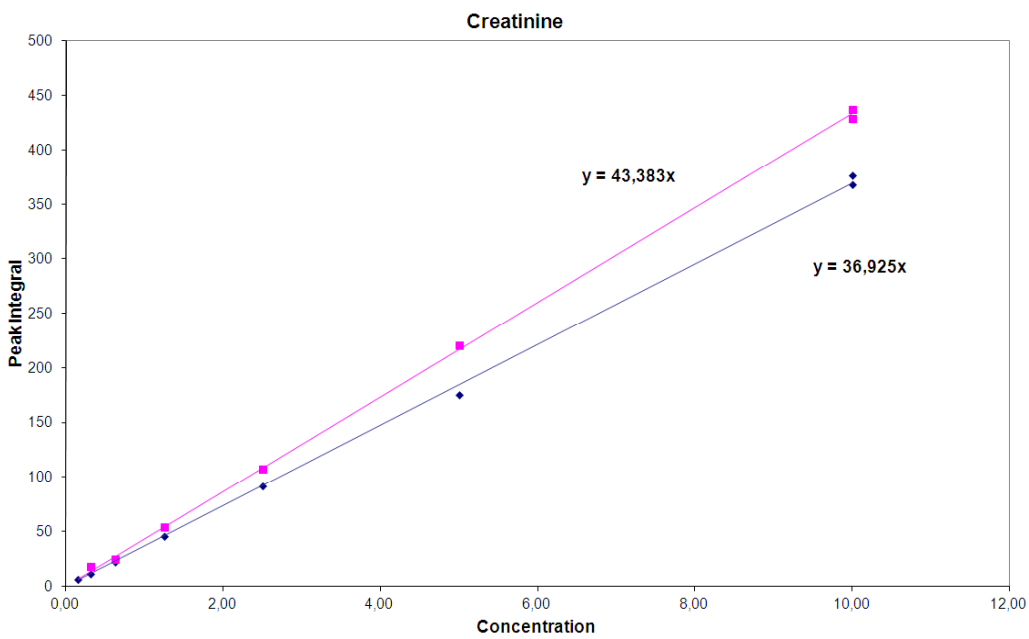

Figure 4.1: Dilution series for determination of peak calibration factors

Lower 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 LLOQ's may be any value and have to be present as follows in the file:

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

The tab stop separated *.txt* file has to contain exactly two columns that are separated by one tab stop.

4.2 Step-by-Step Guide

In this section, the usage of QUANTIFY will be described step by step. For each step an example is given (painted in blue). The main window of QUANTIFY (Figure 4.2) is divided into six panels that are intended to be filled out in their numerical order.

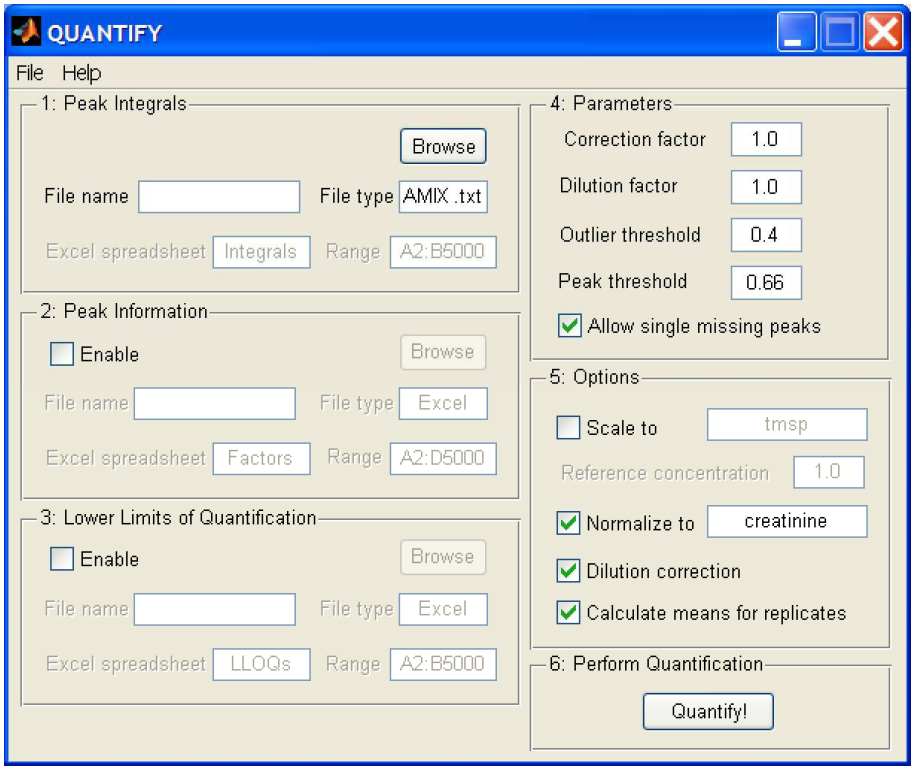

Figure 4.2. The main window with default parameters

For all fields of the main window, a help text is available that will pop up when you hold the mouse curser over the field. When starting QUANTIFY, default parameters will be filled out in most fields. In the following paragraphs, all items of QUANTIFY are described.

Menu Bar

File

- *Load default parameters* will load the default parameters.
- *Load custom parameters* will load previously saved parameters.
- *Save custom parameters* will save the current parameters except for the peak integral file.
- *Exit* will close QUANTIFY.

Help

- *Manual* will open the manual (this document).
- *About QUANTIFY* will show the program version.

Panel 1: Peak Integrals

As a first step, 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 used files please see Section 4.1.1. Select *Browse*, choose the matching file type and the peak integral file and hit enter. After returning to the main window, the file name including the path of the chosen file is shown in the field *File name*. The file type of the chosen file is shown in the field *File type*.

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 Microsoft Excel files, you will have to enter the name of the spreadsheet containing the data (default is *Integrals*) in the field *Sheet* and the range where the values are stored (default is *A2:B5000*) in the field *Range*. Make sure not to include headlines in the given range.

QUANTIFY can also read data from Bruker AMIX Analytic Profiler *mprofile.txt* output files. For details on the parameters to use in AMIX Analytic Profiler see Chapter 6.

Example: Click *Browse.* 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 *QUANTIFY\Examples\Input_Files\Peak_Integral_Files.*

Panel 2: Peak Information

In this panel you have to choose whether you want to use additional peak information. This information is necessary for individual peak calibration and additional peak reliability checks. To use this additional information, check the *Enable* check mark.

In this case, you will have to specify a file where the information is stored by selecting *Browse*. The file may be either an Excel *.xls* file or a tab stop separated *.txt* file. For the data format of the used file please see Section 4.1.2. After choosing the file type and the file in the file browser, the file name including the path of the chosen file is shown in the field *File name*. The file type is shown in the field *File type*.

Please note that Microsoft Excel is required to read Excel files. If Excel is not installed on your computer, please use *.txt* files instead.

When using Excel files, you will have to specify the spreadsheet name and the range where the values are stored in the fields *Sheet* and *Range*. Please make sure not to include any headlines in the area defined in the *Range* field.

Example: Click on the check box *Enable*. Click *Browse*. A file selection window will pop up. Change the file type to *Tab stop separated .txt*, go to the folder *QUANTIFY\Examples\Input_Files\Peak_Information_and_LLOQ_Files and* choose the file *urine_peakinfo.txt.*

Panel 3: Lower Limits of Quantification

Here you have to choose whether you want to check if the calculated concentrations are above individually defined lower limits of quantification (LLOQ's). Check the *Enable* check box to activate this function.

You will have to specify a file where the LLOQ's are stored by selecting *Browse*. 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 4.1.2. After choosing a file, the file name including the path of the chosen file is shown in the field *File name*. The file type is shown in the field *File type*.

Please note that Microsoft Excel is required to read Excel files. If Excel is not installed on your computer, please use *.txt* files instead.

When using Excel files, you will have to specify the spreadsheet name and the range where the values are stored in the fields *Sheet* and *Range*. Please make sure not to include any headlines in the area defined in the *Range* field.

Example: Click on the check box *Enable*. Click *Browse*. A file selection window will pop up. Change the file type to *Tab stop separated .txt*, go to the folder *QUANTIFY\Examples\Input_Files\Peak_Information_and_LLOQ_Files* and choose the file *urine_LLOQs.txt.*

Panel 4: Parameters

These parameters control the outlier detection and reliability checking of the results, among others. The parameters are pre-filled with defaults that should yield good results. Anyway, you may change these parameters in order to optimize your results.

Correction Factor

Each concentration is multiplied by this factor before checking if the concentration is above the LLOQ. The value may be any positive or negative number, default value is 1 (no correction).

This factor 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 *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.

Dilution Factor

Each concentration is multiplied by this factor after checking if the concentration is above the LLOQ. The factor may be any positive or negative number, default value is 1 (no correction).

The dilution factor may be used if all samples are diluted in the same way. The difference to the parameter *correction factor* named 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* in Panel 5.

Outlier Threshold

This is a threshold for outlier detection. For compounds having more than one peak, the median of all peaks is calculated. Values that differ by more than the *outlier threshold* from the median relatively are excluded as outliers. The outlier threshold has to be positive (>0), default value is 0.4.

Peak Threshold

This is a threshold for controlling the reliability of a result by checking how many of the peaks available for a molecule are present in a spectrum. The ratio of found peaks to total peaks has to exceed the *peak threshold*. The peak threshold has to be in the range of 0 to 1, default is 0.66.

Allow Single Missing Peaks

This parameter controls whether single missing peaks of one compound shall not lead to the exclusion of this compound even when it drops below the *peak threshold* then. As single peaks might always be missed by the peak-picking routine, activating this option may yield concentration values that might have been excluded otherwise. In default setting, this option is activated.

Example: Leave all parameters on their default settings.

Panel 5: Options

This panel controls additional functions of QUANTIFY.

Scale To

If this check box is enabled, all integral values of one spectrum will be divided by the value of the first valid peak 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, this check box does not need to be enabled.

Reference Concentration

Here you may enter the concentration of the references substance specified in the *Scale to* field. 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 50 µL of D_2O containing 10 mmol/L tmsp as reference substance , you have to enter *1.25* as reference concentration, as 10 mmol/L * (50 µL / 400 µL) = 1.25 mmol/L.

Normalize To

If this check box is enabled, 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 here.

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. For details see Section 4.1.1.

Calculate Means for Replicates

If this check box is activated, QUANTIFY will search for replicate spectra of the same sample, indicated by the spectrum title. For details see Section 4.1.1.

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.

Example: Enable the check box *Normalize to* and enter *creatinine* in the corresponding text field. Disable the check boxes *Scale to*, *Dilution correction* and *Calculate means for replicates*.

Panel 6: Perform Quantification

After filling in all fields, the main window of QUANTIFY will look like Figure 4.3.

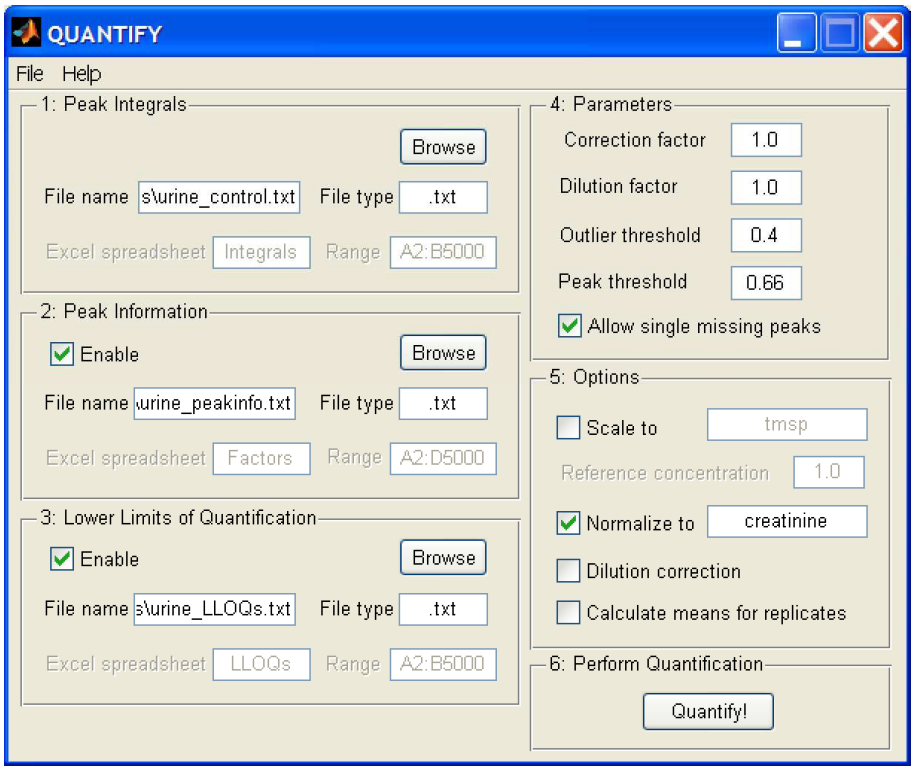

Figure 4.3. The main window of QUANTIFY after filling in all fields according to the examples in this text.

When the button *Quantify!* is selected, you are requested to choose an output file. This may either be a tab stop separated *.txt* file or a Microsoft Excel file. 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 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.

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

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

An example for an Excel result file is shown in Figure 4.4. Additional example result files can be found in the folder *QUANTIFY\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 spectra containing this compound is added.

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.

	A	B	C	D	E	F	G
1		tmsp	alanine	creatinine	hippuricacid	taurine	citricacid
2	M02a	1,17598496		1		0,10614078	0,13054079
3	M03a	0,53143366	0,02770196	1	0,40033411		0,12165131
4	M05a	0,50384909	0,03253314	1		0,0474821	0,09038046
5	M06a	1,24388807		1	1,16549138		
6	M07a	0,67016748	0,07187044	1		0,03674648	0,13227595
7	W01a	1,26438248	0,1120395	1	0,25919563		0,44956337
8	W02a	1,3599557	0,02349649	1	0,20799633		0,36299712
9	W03a	0,35349842	0,01542082	1	0,24111385		0,12726937
10	W05a	0,4588825	0,02471762	1	1,3030446	0,04659597	0,34881249
11	W06a	0,9843846		1	0,27873294	0,1297602	0,14283355

Outliers / Used Peaks / Configuration / Uncorrected Results / **Normalized to creatinine**

Figure 4.4. 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 will get 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 *Dilution Correction* was chosen in Panel 5, the dilution corrected concentrations are stored in the spreadsheet *Dilution Corrected* in the output file.

In case *Calculate Means for Replicates* was chosen in Panel 5, means and TE's will be written to the spreadsheet *Mean* in the output file.

In case *Normalize to* was chosen in Panel 5, a spreadsheet named *Normalized to compound name* is created, e.g. *Normalized to creatinine*. This spreadsheet contains the normalized values.

If *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 LLOQ's and multiplying by correction and dilution factors) will be stored in the spreadsheet *Original Values*.

If you choose *.txt* files as output, up to eight *.txt* files will be generated containing these data. The files will 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*.

4.3 Quick Start

QUANTIFY is pre-configured to allow a quick start. In this mode, most implemented features are disabled. Still, this mode may allow a quick view on your data. Please note that quantification results from this mode are only semi-quantitative. For accurate quantification, please refer to the step-by-step guide in Section 4.2. The main feature active in this mode is automatic outlier removal based on deviations from the median.

For quantification, you need a file that contains the NMR peak integral values obtained for the different compounds you are interested in. For details on the format of the file see Section 4.1.1.

Run the file *quantify.exe*. The main window will show up (Figure 4.5).

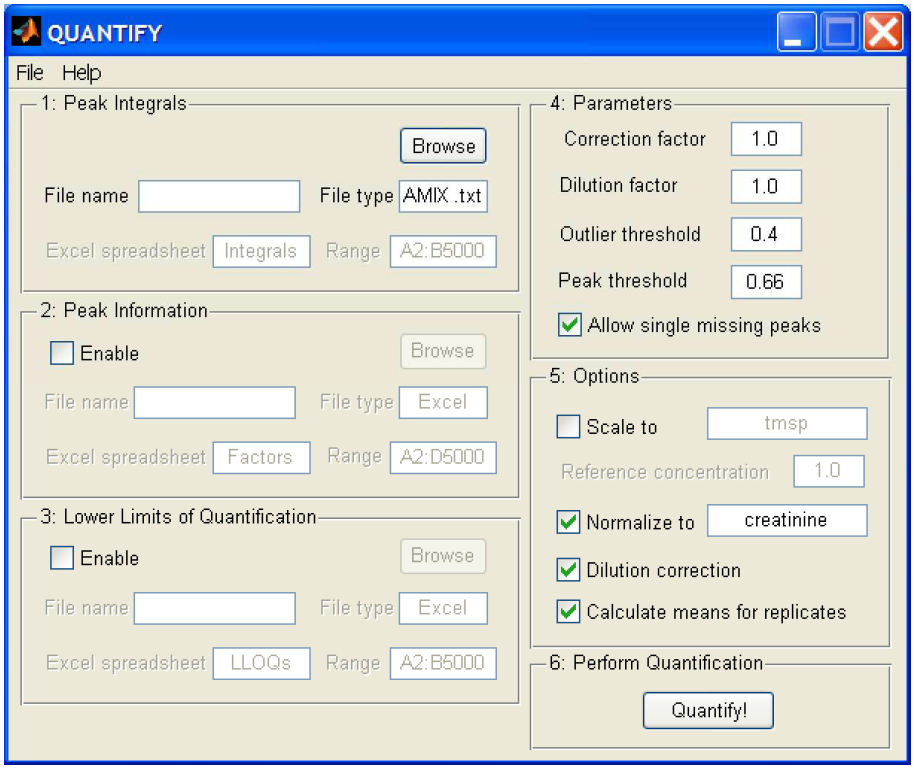

Figure 4.5. The main window of QUANTIFY in quick start mode with default parameters

Panel 1: Peak Integrals

Select *Browse* and choose the file containing the peak integral values.

Panels 2 - 4

You do not need to change the pre-filled default values in these panels.

Panel 5: Options

Activate the *Scale to* check box and enter the name of your reference substance. Enter the concentration of your reference substance in the field *Reference concentration*.

Panel 6: Perform Quantification

Select the button *Quantify!*. This will open a dialog where you can choose the output file name. Afterward, quantification is performed and the result file opened.

5 Troubleshooting

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

- **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 4.1.1 or Chapter 6 for AMIX files).

- **Warning: *Dilution factors were found only for x out of y samples!***
  This warning means that QUANTIFY 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 QUANTIFY 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 Chapter 6 when creating AMIX *.txt* files.

  In case you use a tab stop separated *.txt* file, make sure you choose the matching entry from the file type menu when selecting the file using *Browse* in Panel 1 (see Figure 5.1).

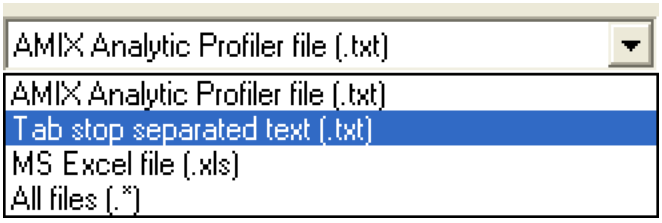

Figure 5.1. Choosing the integral value file type

6 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 QUANTIFY. 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.

6.1 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.

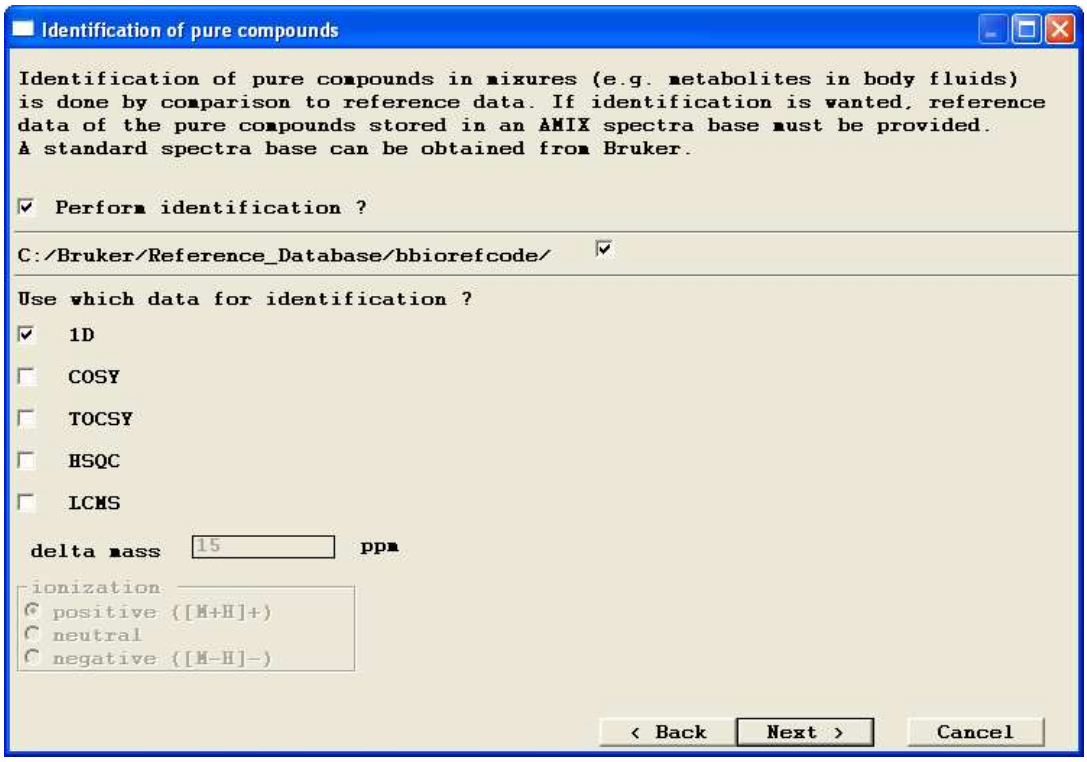

Figure 6.1. *Perform identification* and *1D* have to be enabled

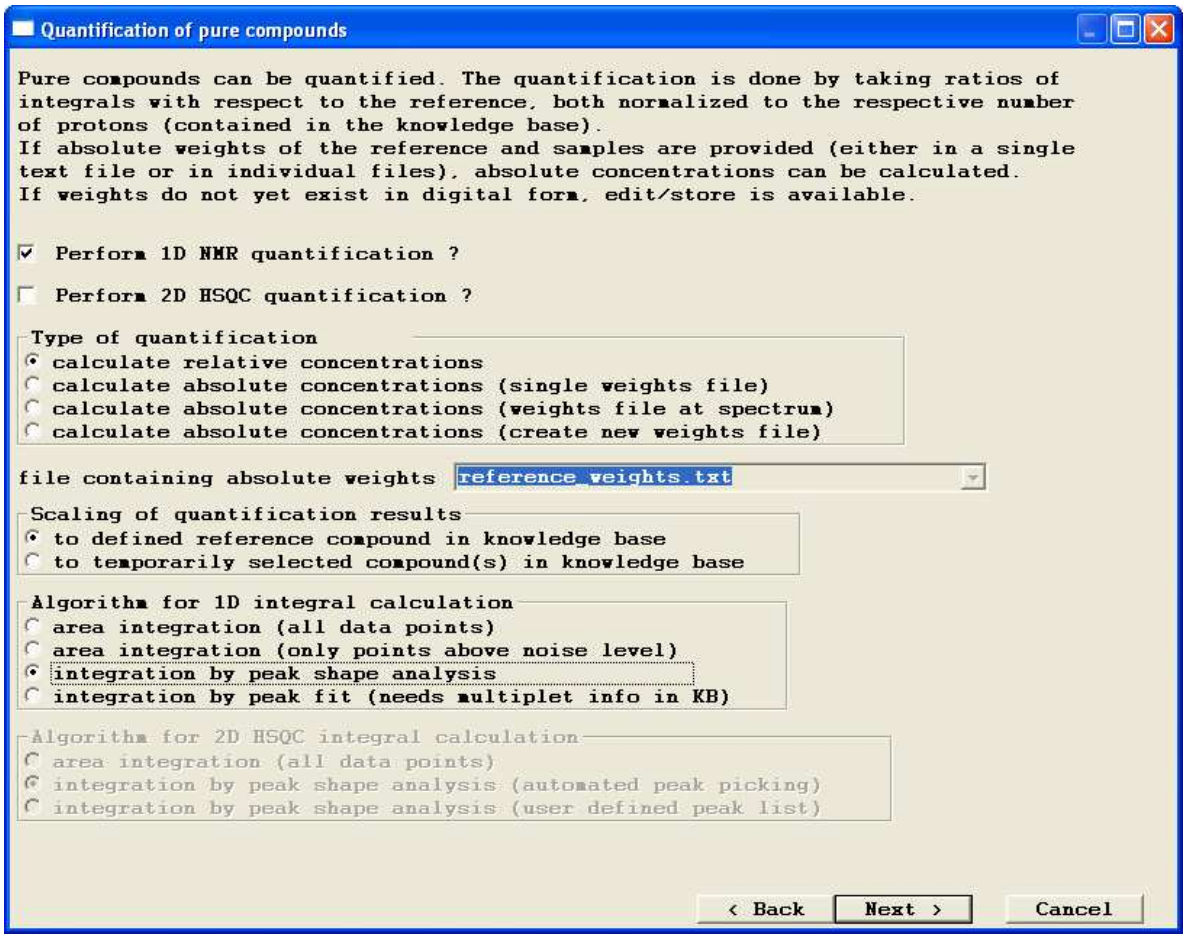

Figure 6.2. *Perform 1D quantification* has to be enabled

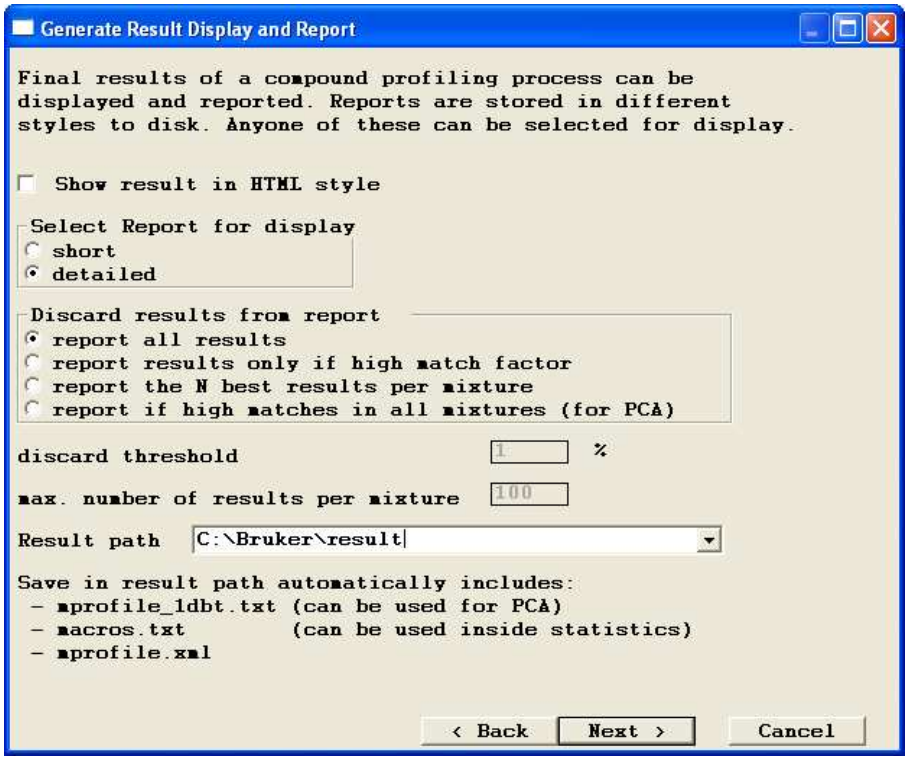

Figure 6.3. *Detailed* and *Report all results* have to be enabled

6.2 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. QUANTIFY will look for 2D HSQC data in the file and use these for quantification.

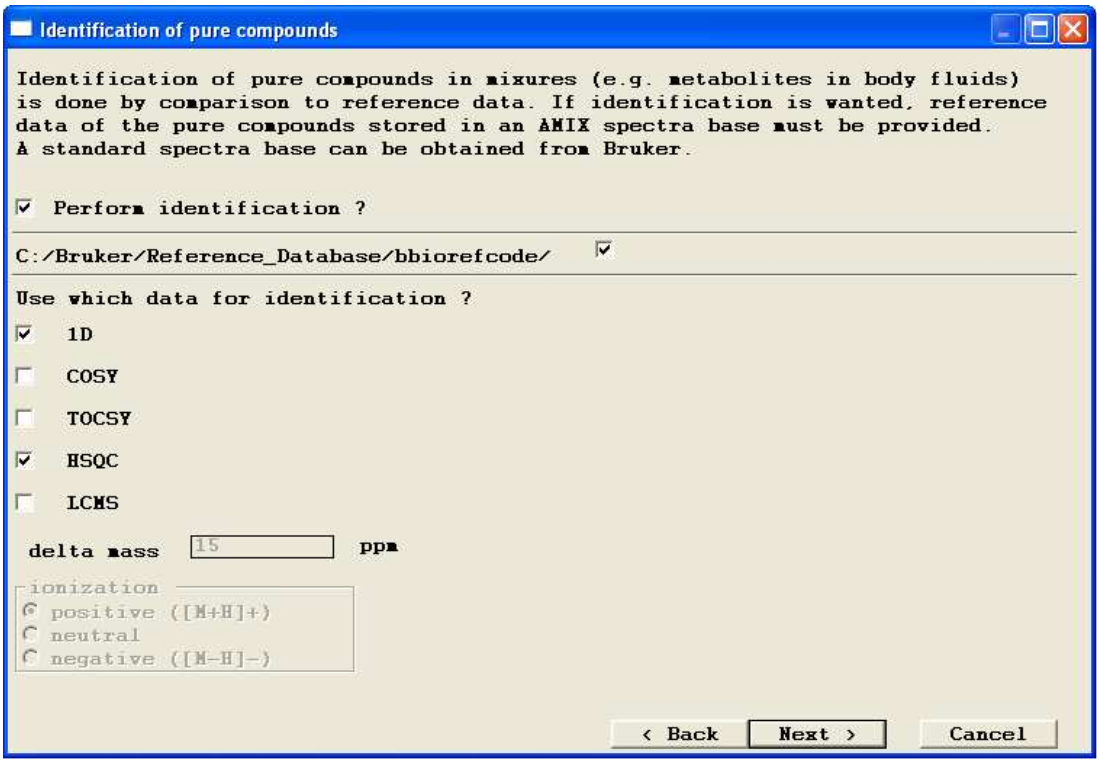

Figure 6.4. *Perform identification* has to be enabled both for *1D* and *HSQC*

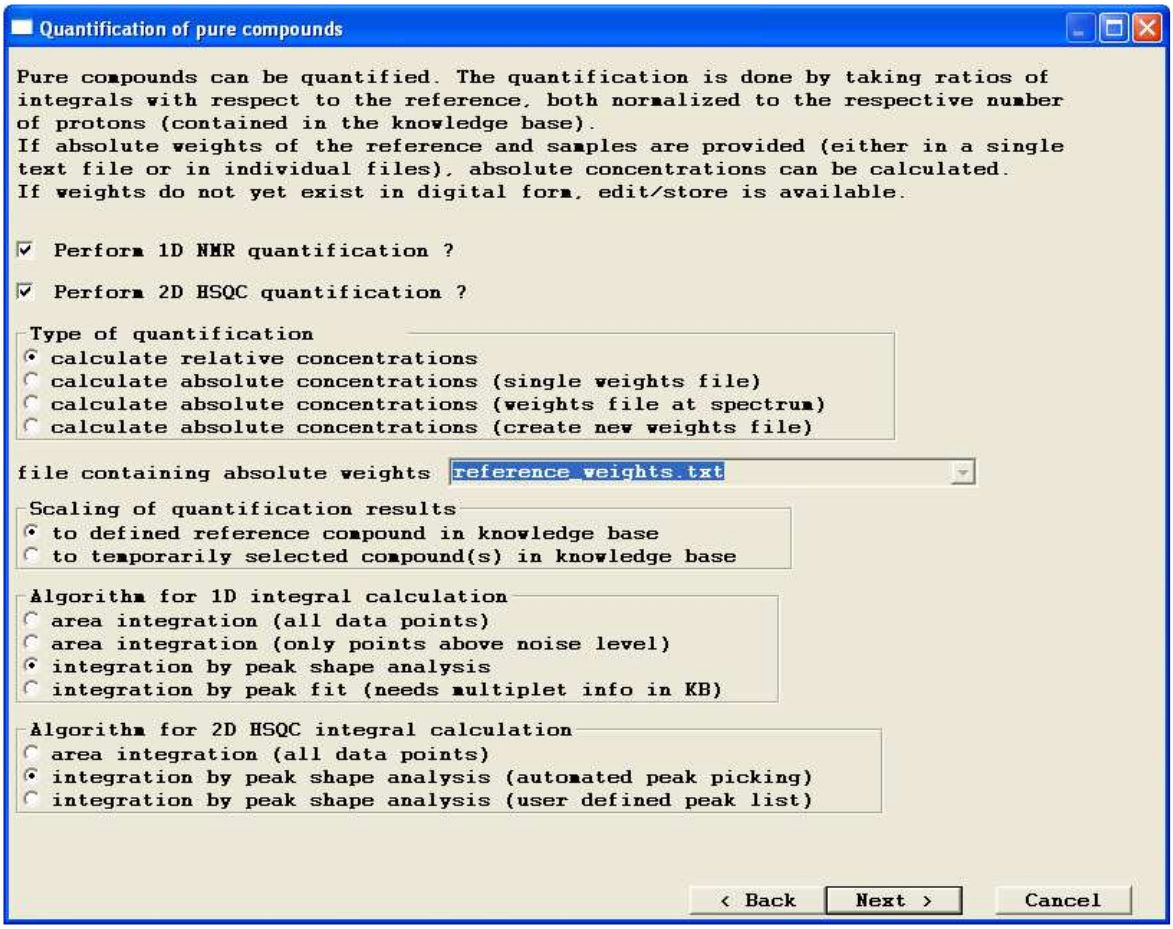

Figure 6.5. Both *Perform 1D quantification* and *Perform 2D quantification* have to be enabled

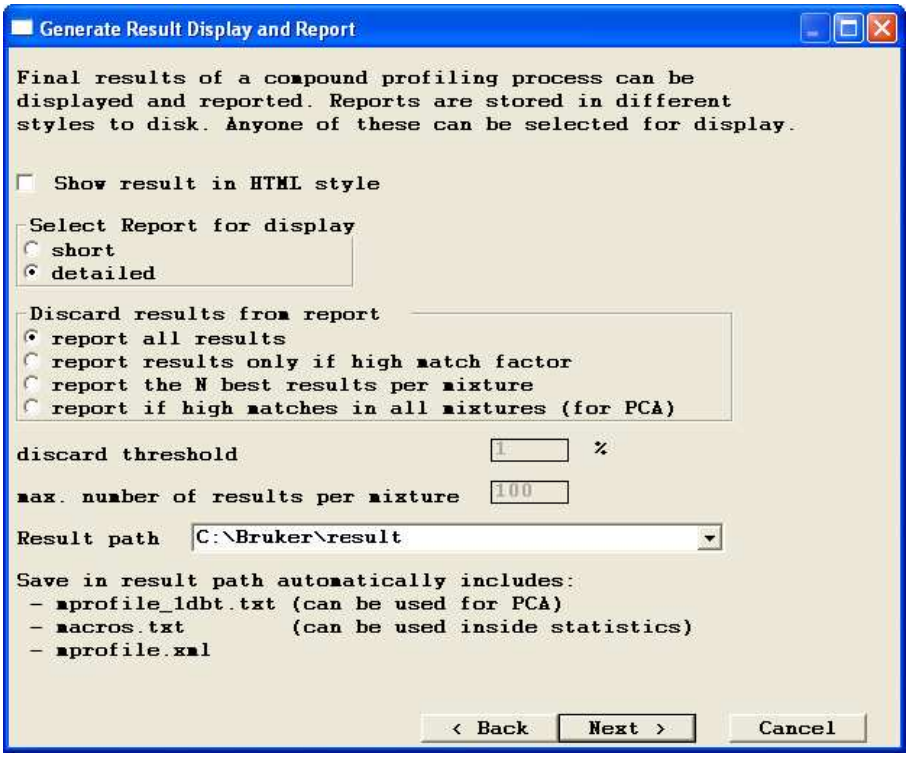

Figure 6.6. *Detailed* and *Report all results* have to be enabled