BIOINFORMATICS APPLICATIONS NOTE Vol. 22 no. 23 2006, pages 2962–2965
doi:10.1093/bioinformatics/btl526

Systems biology

MetaQuant: a tool for the automatic quantification of GC/MS-based metabolome data

Boyke Bunk[1], Martin Kucklick[1], Rochus Jonas[2], Richard Münch[1], Max Schobert[1], Dieter Jahn[1,*] and Karsten Hiller[1]

[1]Institute of Microbiology, Technical University of Braunschweig, Spielmannstr. 7, D-38106 Braunschweig, Germany and [2]Institut für Bioverfahrenstechnik, Technische Universität Braunschweig, Gaußstr. 17, D-38106 Braunschweig, Germany

Received on May 16, 2006; revised on October 9, 2006; accepted on October 10, 2006
Advance Access publication October 17, 2006
Associate Editor: Nikolaus Rajewsky

ABSTRACT

Summary: MetaQuant is a Java-based program for the automatic and accurate quantification of GC/MS-based metabolome data. In contrast to other programs MetaQuant is able to quantify hundreds of substances simultaneously with minimal manual intervention. The integration of a self-acting calibration function allows the parallel and fast calibration for several metabolites simultaneously. Finally, MetaQuant is able to import GC/MS data in the common NetCDF format and to export the results of the quantification into Systems Biology Markup Language (SBML), Comma Separated Values (CSV) or Microsoft Excel (XLS) format.

Availability: MetaQuant is written in Java and is available under an open source license. Precompiled packages for the installation on Windows or Linux operating systems are freely available for download. The source code as well as the installation packages are available at http://bioinformatics.org/metaquant

Contact: d.jahn@tu-bs.de

1 INTRODUCTION

During the last few years more than 200 genomes have been completely sequenced. Obtained data processed via diverse methods of bioinformatics provide first insights into the genetic and metabolic capacity of these organisms. The identification of thousands of molecules that exist within a living cell opens a window to the understanding of complex biological systems. Currently, besides transcriptomics and proteomics, high throughput metabolomics is one of the most challenging techniques in the context of systems biology (Oliver *et al*., 1998; Fiehn, 2002). The aim of metabolomics is the identification of the metabolites of living cells and via comparison of these metabolites present under different growth conditions to better understand metabolic adaptation strategies. Gas chromatography in combination with subsequent mass spectrometry (GC/MS) allows the detection of hundreds of derivated metabolites simultaneously (Strelkov *et al*., 2004). However, current GC/MS-based methods usually only provide semi-quantitative information on the concentration of the detected substances. Although this information is characteristic for the investigated organism and growth condition, it does not provide the necessary information for the understanding and modeling of metabolic kinetics or other concentration-related studies (Nielsen and Oliver, 2005). Currently, most of the GC/MS-related softwares are able to analyze GC/MS data in order to identify substances (Stein, 1999; Katz *et al*., 2004) or to compare whole chromatograms. In this context, the program MetAlign allows a comparative analysis of different chromatograms with the elimination of noise and the detection of local shifts for the purpose of filtering and normalizing of the recorded raw data (America *et al*., 2006). Many quantification softwares that are provided by the manufacturer of commonly used GC/MS instruments such as ChemStation or Xcalibur are able to perform only a limited number of calibration runs. Chemstation allows for 20 calibration levels or calibration points per substance, XCalibur for 50 and Shimadzu GC/MS-solution for 64. These programs are able to use several reference mass fragments (ions) or even the total spectrum for peak detection. However, they examine either simply the intensity of one mass fragment or the summed intensity of all mass fragments (TIC – total ion chromatogram) for quantification. This approach is not sufficient for quantitative analysis of complex chromatograms containing several 'fused peaks', 'shoulder peaks', 'rider peaks' or 'hidden peaks'. Moreover, the quantification functions of these programs are only useful for the quantification of a few substances. The automated simultaneous quantification of hundreds of metabolites with concentration ranges of several orders of magnitude is not possible. These quantification processes are essential for quantitative metabolomics. Additionally, these programs need frequent manual intervention and are therefore time-consuming. To our knowledge there is no software currently available which performs an accurate and nearly automatic quantitative determination of a large number of metabolites in terms of nmol. Therefore, we developed MetaQuant, an open source, Java-based software for the simultaneous and automatic identification and quantification of a high number of metabolites.

2 FEATURES

MetaQuant provides an interactive and easy-to-use graphical user interface. The program was implemented in Java to take advantage of the platform independence of this object-orientated programming language. Thus, MetaQuant can be used with nearly all currently available systems (e.g. Windows XP, Linux). To further simplify the installation process we offer precompiled installation packages

*To whom correspondence should be addressed.

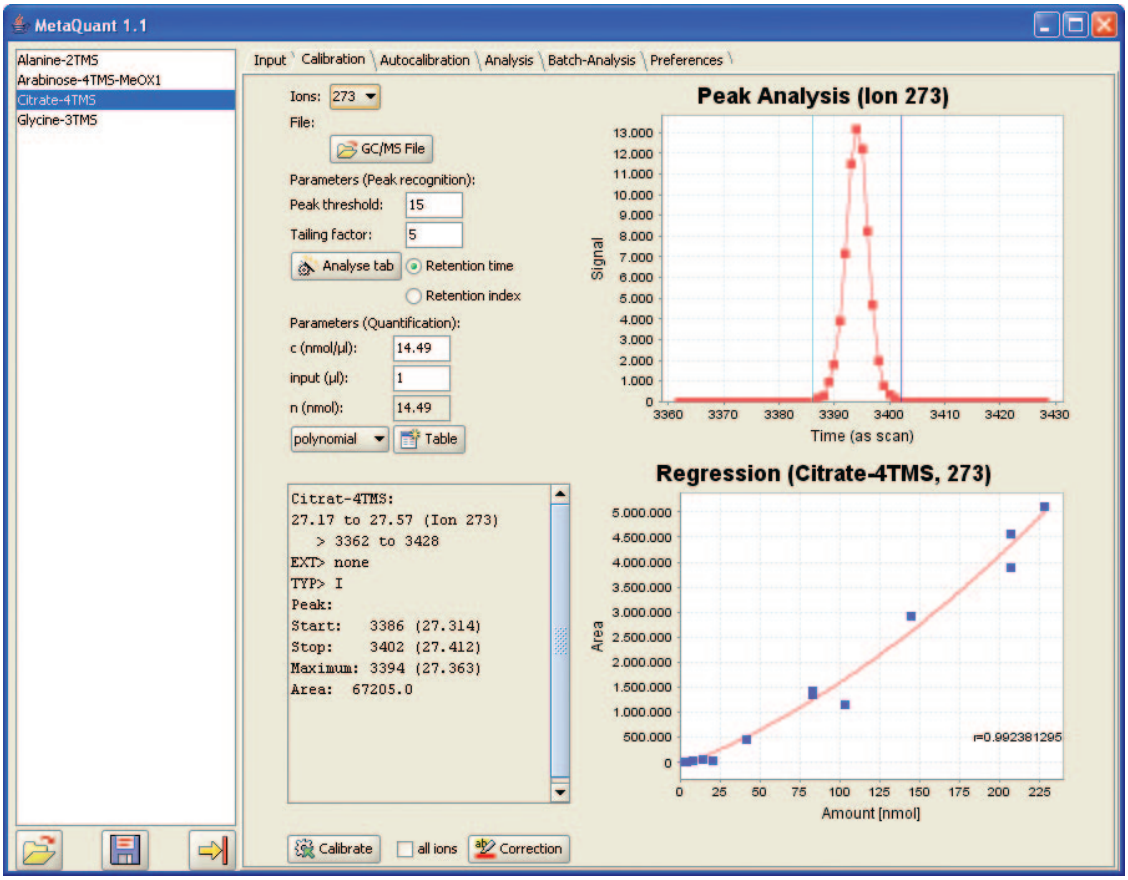

Fig. 1. Screenshot of the calibration dialog of MetaQuant: Peak detection and setting of calibration points for the regression analysis are performed here.

that can be started without any further requirements such as a Java Runtime Environment either on Windows or Linux operating systems. The main panel of the program provides access to the various functions of the software:

2.1 Function 1 - Metabolite definition

Metabolites are identified by the retention times or retention indices and by the detection of parts of their characteristic mass spectrum. Retention indices are retention times that are normalized by the retention times of n-alkanes as reference series (Castello, 1999). Therefore, retention indices are comparable between different GC instruments and various chromatographic conditions. Usually, either the area of a peak in the TIC or the area of a peak in only one specific ion chromatogram at the substance specific retention time is used for quantification even if more ions are used for metabolite identification. However, for most substances the quantification accuracy can be improved once parts of other specific ion chromatograms are used for integration. For this purpose, MetaQuant has the ability to define different sets of mass fragments for identification and quantification. Additionally, the user is able to define particular ions which are not derived from the target substance and therefore indicate a coeluting substance. The necessary metabolite specific parameters can be obtained by a deconvolution software like AMDIS. However, after this initial metabolite definition process no deconvolution software is needed anymore.

2.2 Function 2 - Calibration

To facilitate a fast and easy calibration process, we integrated an automatic calibration function into our software. After recording a GC/MS spectrum for a mixture of standard metabolites with known concentrations, MetaQuant is able to solve the integrals of the corresponding peaks in the spectrum. These peak areas in combination with the known substance concentrations are then used as calibration points for a regression analysis (Fig. 1). Although linear dependency between measured peak area and corresponding metabolite concentration is expected, in some cases polynomial, power or even exponential regression analysis of the data lead to more accurate results. Corresponding observations were made, in particular, if measurements of metabolite concentration were performed over several magnitudes. Consequently, MetaQuant provides the choice between different regression methods for the quantification process. The concentration range handled by MetaQuant is only limited by the used GC/MS instrument. The number of metabolites that can be used for calibration is only restricted by the available memory of the employed computer.

2.3 Function 3 – Quantification

The quantification process of MetaQuant consists of two steps. The first step is the detection of derivated metabolite specific peaks in the GC/MS chromatogram of interest. The algorithm used for peak recognition is failsafe and is able to correctly detect peaks even with

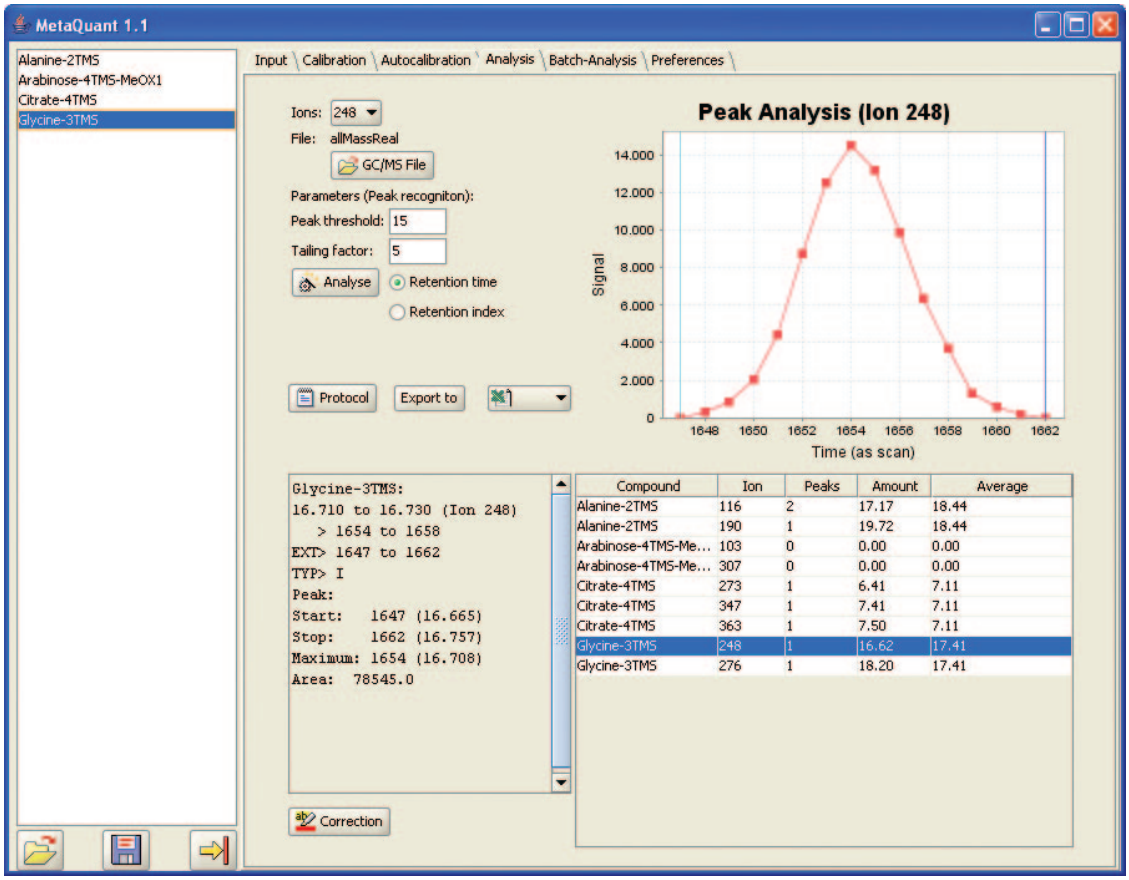

Fig. 2. Screenshot of the quantification dialog of MetaQuant: Peak detection and presentation of quantification results are shown here. The results can be exported to various file formats.

unusual patterns such as 'fused peaks', 'hidden peaks', 'shoulder peaks' or 'rider peaks'. For this purpose, the used algorithm analyzes parts of the chromatograms of several selected metabolite specific mass fragments eluted in a particular retention time window. These specific mass fragments in combination with the specific retention time windows need to be defined during metabolite definition. In contrast to the analysis of the complete TIC, the use of these fragments allows the differentiation between coeluting substances, since 'fused peaks', 'hidden peaks', 'shoulder peaks' and 'rider peaks' are absent in many of these ion chromatograms. Nevertheless, MetaQuant automatically determines the correct area of 'rider peaks' using tangential baselines. In order to minimize shifts of retention times, MetaQuant is able to use retention indices (see above) instead of retention times (Castello, 1999). Subsequently, in a second step, the program performs an integration of the detected substance specific peaks. The quantification is performed using the appropriate regression function defined in the calibration library of MetaQuant. If the metabolite definition contains more than one ion for the quantification, the program presents the results of the quantification for all ions of this metabolite individually (Fig. 2). Additionally, the average quantity of this metabolite is calculated.

2.4 Function 4 - Import and export of data

MetaQuant allows the import of GC/MS data either in a particular comma separated values (CSV) format or in the commonly used NetCDF format. Since the descriptions of the native export formats of most GC/MS recording softwares are not public, MetaQuant is not able to directly import these files. However, several popular GC/MS programs are able to export measured data either via the NetCDF or CSV format. For further analysis, the results of the quantification can be exported as SBML, CSV or Microsoft Excel (XLS) file as well as an MetaQuant-specific extensible markup language (XML) file.

2.5 Function 5 - Batch analysis

For a quantification analysis of several recorded chromatograms in parallel, we integrated a batch job function into MetaQuant. This feature offers the ability of selecting multiple chromatogram files and starting a fully automatic quantification process. Afterwards, the results can be combined and presented by MetaQuant or exported. For further automation we implemented a command-line version of the software for performing large scale analyses. This version of MetaQuant can be controlled by other software tools within a metabolomics data analysis pipeline.

3 CONCLUSION

MetaQuant is a new program for the quantitative analysis of high throughput GC/MS-based metabolomics data. The program is

intended to automatically determine the accurate intracellular amount of hundreds of metabolites.

ACKNOWLEDGEMENTS

We thank Dr Barbara Schulz for critical proofreading of the manuscript. This work was funded by the German Federal Ministry of Education and Research (BMBF) for the Bioinformatics Competence Center 'Intergenomics' (Grant No. 031U110A/031U210A).

Conflict of Interest: none declared.

REFERENCES

America,A.H.P. *et al*. (2006) Alignment and statistical difference analysis of complex peptide data sets generated by multidimensional lc-ms. *Proteomics*, **6**, 641–653.

Castello,G. (1999) Retention index systems: alternatives to the n-alkanes as calibration standards. *J. Chromatogr. A*, **842**, 51–64.

Fiehn,O. (2002) Metabolomics–the link between genotypes and phenotypes. *Plant Mol. Biol.*, **48**, 155–171.

Katz,J.E. *et al*. (2004) A new technique (COMSPARI) to facilitate the identification of minor compounds in complex mixtures by GC/MS and LC/MS: tools for the visualization of matched datasets. *J. Am. Soc. Mass Spectrom*, **15**, 580–584.

Nielsen,J. and Oliver,S. (2005) The next wave in metabolome analysis. *Trends Biotechnol.*, **23**, 544–546.

Oliver,S.G. *et al*. (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol.*, **16**, 373–378.

Stein,S. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *J. Am. Soc. Mass Spectrom.*, **10**, 770–781.

Strelkov,S. *et al*. (2004) Comprehensive analysis of metabolites in Corynebacterium glutamicum by gas chromatography/mass spectrometry. *Biol. Chem.*, **385**, 853–861.