Software tool for automated processing of ^{13}C labeling data from mass spectrometric spectra

Priti Talwar[1], Christoph Wittmann[2], Thomas Lengauer[1], and Elmar Heinzle[2]
[1]Max-Planck-Institut für Informatik, Saarbrücken, Germany and [2]Institut für Technische Biochemie, Saarland University, Saarbrücken, Germany

BioTechniques 35:1214-1215 (December 2003)

In recent years, intensive research has been carried out concerning the quantitative investigation of metabolic networks as the basis for understanding metabolic functioning and regulation machinery of specific metabolic systems (1–3). A powerful approach to quantify metabolic fluxes is based on tracer studies with ^{13}C-labeled substrates combined with mass spectrometry (MS) measurement of ^{13}C labeling patterns of biomass constituents (4–6). In these tracer studies, the measured labeling pattern reflects the metabolic state of the cell and is used to calculate intracellular flux parameters. Metabolic flux analysis is especially useful when applied in comparative studies, so experimental and computational tools for efficient metabolic flux analysis on a broad level are highly desired. Efficient flux analysis on a broad level, however, requires a straightforward approach that can be parallelized and automated for all steps involved. A time-consuming and error prone step in the whole procedure of metabolic flux analysis is the extraction of labeling patterns from mass spectrometric spectra, which is currently done manually. Manual processing of these MS data sets is highly time-consuming and subject to error. In the present work, the software module CalSpec, for automatic processing of labeling data from MS spectra, is presented. It has been tested and applied to gas chromatography/mass spectrometry (GC/MS) analysis of t-butyldimethylsiloxy (TBDMS)-derivatized amino acids. Amino acids have been shown to provide valuable labeling information for flux calculations in ^{13}C tracer experiments (5,7). The module is flexible and can be easily adapted to other MS measurements. The module was developed and tested for GC/MS analysis of TBDMS-derivatized amino acids using a GC with HP-5MS capillary column (5% phenylmethylsiloxane diphenylpolysiloxane, 30 m × 250 µm, electron impact ionisation at 70 V and a quadrupole detector; Agilent Technologies, Waldbronn, Germany) as described previously (6). The instrument was equipped with MSD Productivity ChemStation software (G1701C, Rev. Code 00.01; Agilent Technologies) generating specific data files (*.ms files). These data files were first converted into comma separated value (*.csv) files using a macro that was supplied by Agilent Technologies and further modified by authors. Additional macros must be developed to use the program with non-Agilent systems.

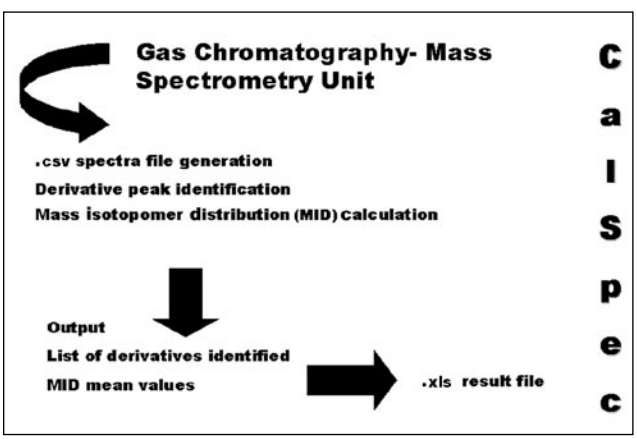

Figure 1. Data processing schema for CalSpec.

CalSpec automatically performs identification of specified analytes in the MS spectrum and the subsequent quantification of labeling patterns. An overview on the steps involved in the data processing by CalSpec is given in Figure 1.

The tool has an initial step for conversion of the *.ms file format originating from the GC/MS system into a *.csv file. In this step, a widely used platform independent file format is generated that can be further processed.

In the next step, identification of the amino acid TBDMS-derivatized fragments present in the sample is carried out using the presence of typical mass-to-charge (m/z) signals observed in the spectra. In this way, the sample can be checked for the presence of specified analytes, and thus, the preceding experimental protocol can be evaluated. For identification, the user should modify (*i*) the elution time (T_e) of an analyte in the GC run, and (*ii*) the m/z values of corresponding specific ion clusters to be observed, in the param.txt parameter file according to the format. These parameters are used to check for the presence of an analyte. Identification of peaks is currently performed in a window of (T_e - 0.25) to (T_e + 0.25) min for all analytes. In each time window, the spectrum is scanned at the specified m/z values, whereby the user can define a threshold for each signal that has to be exceeded to indicate presence of the corresponding analyte. In this way, signals with low abundance can be excluded that are subjected to interference with background noise and therefore should not be considered for flux estimation (7). To ensure for the presence of an analyte, the specific m/z values, correlating to typical fragments such as [M-57], [M-85], or [M-159] for TBDMS-derivatized amino acids in the time window should exceed the threshold set. First, abundance levels of different mass fractions of the analytes are calculated. Currently, CalSpec calculates mass isotopomer distributions (MID) according to Equation 1 in vector form, whereby each mass isotopomer fraction x_{m+y} represents the relative abundance of all isotopomers with y ^{13}C atoms corresponding to a mass shift $+y$. A compound with i carbon atoms can occur in $i + 1$ mass isotopomer fractions.

$$\mathrm{MID} = \begin{bmatrix} x_m \\ x_{m+1} \\ x_{m+2} \\ . \\ . \\ x_{m+\mathrm{i}} \end{bmatrix} \qquad \text{[Eq. 1]}$$

It is known that the high resolution of GC separation can lead to isotope fractionation, which results in gradients for the relative abundance of different mass isotopomers over a peak (7). To correctly extract labeling information from a peak, all mass scans performed by the MS detector during the elution of the peak have to be taken into account. CalSpec therefore integrates the different m/z signals, by calculating mean abundances for all mass isotopomer fractions over the entire peak. The automated specification of the time window ensures that the same signals are considered in every measurement. By contrast, manual integration is error prone and tedious. The output file is generated, which contains a list of the specified analytes, information about their presence, and the abundance of mass isotopomer fractions. This file has *.xls format and therefore can be easily imported into any text editing application. The software module is especially useful to routinely analyze samples derived from, for example, protein hydrolysates or cultivation supernatants. It should be noticed that care has to be taken regarding isobaric interference of the target analytes with other compounds, which might, for example, occur in highly complex mixtures such as cell extracts.

The developed software tool, CalSpec, is useful for efficient processing of ^{13}C labeling data from MS measurements in ^{13}C flux analysis. These MS data sets are generated in huge numbers due to (*i*) replicate measurements of one sample to assess the confidence in the measured values and estimation of error; (*ii*) replicate measurements of one experiment to check for isotopic steady-state; or (*iii*) different measurements of one sample with different protocols to obtain additional labeling information via alternative fragments. Data processing by CalSpec takes only a few seconds per spectrum, whereas the same task requires up to 30 min or more if done manually. The present development is a step within a larger effort to speed up the cultivation, analysis, and evaluation phases for application to huge experimental setup. The module is freely available at bioinf.mpi-sb.mpg.de/projects/CalSpec/.

ACKNOWLEDGMENTS

The authors wish to thank Andreas Kaemper and Jochen Maydt for their help in different stages of development of the system. We also acknowledge Agilent Technologies (Waldbronn, Germany) for supplying the macro for conversion of the .ms files to .csv, which was further modified.

REFERENCES

1. **Bailey, J.E.** 1991. Toward a science of metabolic engineering. Science *252*:1668-1675.
2. **Bailey, J.E.** 1998. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol. Prog. *14*:8-20.
3. **Cameron, D.C. and F.W.R. Chaplen.** 1997. Developments in metabolic engineering. Curr. Opin. Biotechnol. *8*:75-180.
4. **Christensen, B. and J. Nielsen.** 1999. Isotopomer analysis using GC-MS. Metab. Eng. *1*:282-290.
5. **Wittmann, C. and E. Heinzle.** 1999. Mass spectrometry for metabolic flux analysis. Biotechnol. Bioeng. *62*:739-750.
6. **Wittmann, C., M. Hans, and E. Heinzle.** 2002. In vivo analysis of intracellular amino acid labeling by GC/MS. Anal. Biochem. *307*:379-382.
7. **Dauner, M. and U. Sauer.** 2000. GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. Biotechnol. Prog. *16*:642-649.

Received 31 March 2003; accepted 14 August 2003.

Address correspondence to Priti Talwar, Stuhlsatzenhausweg 85, Saarbrücken 66123, Germany. e-mail: ptalwar@mpi-sb.mpg.de