Semana de Engenharia 2010
Guimarães, 11 a 15 de Outubro

X-METABOLOMICS: A SOFTWARE TOOL FOR DYNAMIC MASS SPECTROSCOPY AUTOMATED SIGNAL PROCESSING

Cristiana C. Castro[1], António C. Silva-Ferreira[2], José A. Teixeira[1], Rui C. Martins[3]
[1] Biological Engineering Centre – University of Minho, Portugal
[2] Biotechnology Research Centre - Interface A4 Catholic University, Portugal
[3] Molecular and Environmental Biology Centre – University of Minho, Portugal
Email: cristianacastro@deb.uminho.pt

KEYWORDS
Dynamic systems, GC-MS, Signal processing, Metabolomics, Multivariate analysis

ABSTRACT

One of the major challenges of today's biotechnology is to be able to obtain the maximum of metabolic information for the holistic interpretation of biological systems. Herein we present a new computational application for gas-chromatography mass spectroscopy automated signal processing, named 'X-Metabolomics', that shows to be a potential framework for dynamic systems as fermentation processes interpretation by compounds and pathways identification and quantification and moreover for new metabolites discovery.
Our GC-MS signal processing pipeline is implemented into an X-window interface using Tcl/Tk interface and based on R statistical programming environment for comprehensive statistical computing of results and access to 'Bioconductor' bioinformatics platform under Unix, Linux and MacOS.
This approach focuses on the robustness of peak extraction algorithms for further identification, quantification and biological interpretation by multivariate analysis, evolving the following steps: i) peaks extraction; ii) supervised filtering; iii) identification of candidate fragments and removal of possible contaminants; iv) compounds identification/quantification; v) compounds expression and co-expression in time-course; and vi) sample classification and biological interpretation by multivariate analysis.
'X-Metabolomics' can be an useful tool in different fields such as pharmacology, genetics, living cells systems, promising to be innovative and very helpful for new drug discovery and new advances in dynamic systems understanding.

INTRODUCTION

Fermentations are dynamic systems where metabolic changes occur while fermentation microrganisms adapt to the medium and reactor operational conditions.
Metabolomics is a technique, which aims at identifying metabolites and their reaction networks and dynamics in a biological cell, tissue, organ or organism, as a response to environmental changes. The metabolic profiling of a complex process as the fermentation process can give an instantaneous snapshot of the physiology of a cell in the system.
High-throughput metabolomic chromatography systems consisting of hardware or analytical equipment and bioinformatic tools for signal preprocessing, data storage and multivariate analysis are more widespread.
GC-MS is especially suited for a significant part of the yeast metabolome originating several hundreds of volatile metabolites. Classical mass spectroscopy analyses are highly laborious, and significant amount of time is needed for peak analysis, fingerprint recognition, identification and quantification by an human analyst (Christensen et al. 2005).
In this sense, methodologies for automatic processing of chromatograms are of great importance as they can aid metabolomics researchers by providing an usefull laboratory tool for a rapid GC-MS diagnosis.
In this research, we present 'X-Metabolomics', a software pipeline for GC-MS chromatograms processing.

SOFTWARE IMPLEMENTATION

'X-Metabolomics' is intended for GC-MS spectral validation by multivariate process analytical technology, compounds identification, quantification, time course and co-expression, providing metabolic pathways identification, as well as, new pathways discovery.
The processing pipeline works as follows: i) chromatograms import (directory or selected samples); ii) peak extraction and alignment; iii) supervised filtering; iv) fragment classification, identification and quantification; v) identification and composition tables building; vi) compounds expression in time-course and co-expression; and vii) multivariate statistics for data interpretation and classification. Figure 1 presents the 'X-Metabolomics' workflow for chromatographic signal processing and interpretation.

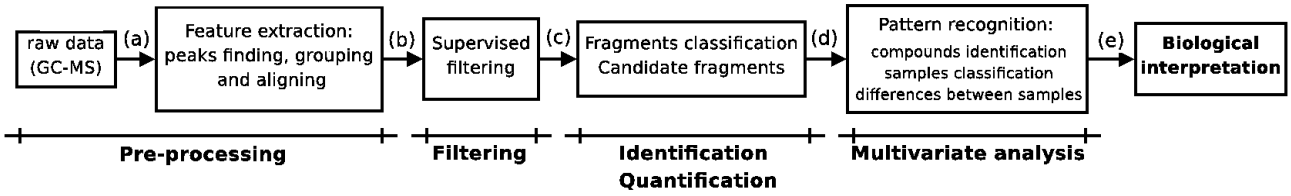

Figure 1 'X-Metabolomics' software workflow.

Pre-processing

Pre-processing task is performed as the methodology published by Smith et al. (2006), the XCMS methodology, and consists essentially in chromatograms feature extraction by peaks detection, grouping and non-linear retention time aligning.

Supervised filtering

Scans resulting from pre-processing task can correspond to real and interesting compounds or to contaminant compounds. The last mentioned can correspond to, p.e., SPME fiber material and, usually, have a typical pattern of occurrence throughout samples. In this sense, filtering is a crucial task for its recognition and identification. 'X-Metabolomics' allows the elimination of the contaminant scans prior to identification and multivariate analysis by: 1) threshold filtering; and 2) hierarchical cluster analysis (HCA) filtering (Tikunov et al. 2005).

Metabolites identification and quantification

'X-Metabolomics' produces two tables: i) identification table; and ii) composition table. The identification consists only in identifiable and coherent peaks among all samples. Data in this table is presented for each candidate compound with corresponding normalized fragments for annotation. The composition table comprises compounds concentrations that are the direct linear relationship to the internal standard in the linear regime of the mass detector for all samples.

Compounds expression in time course and co-expression

'X-Metabolomics' also allows observing compounds and samples co-expression, by scans-scans and samples-samples correlations mapping. This application seems to be essential for pathways and compounds identification and discovery. Metabolites kinetics during fermentations can be also obtained.

Multivariate analysis

Relevant principal component analysis (RPCA) (Martins et al. 2008), and multivariate control charts (MCC) (Hotelling 1947), with corresponding cluster analysis of samples (scores) and variables (loadings) are implemented for a quick non-supervised result interpretation, diagnosis and quality control of the process.

REFERENCES

Christensen, J.H..; G. Tomasi; and A.B. Hansen. 2005. "Chemical fingerprint of petroleum biomarkers using time warping and pca". *Environmental Science and Technology*, 39(1), 255-260.

Smith, C.A.; E.J. Want; G. O'Maille; R. Abagyan; and G. Siuzdak. 2006. "XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peaks alignment, matching, and identification". *Analytical Chemistry*, 78, 779-787.

Tikunov, Y; A. Lommen; C.H. Ric de Vos; H.A. Verhoeven; R.J. Bino; R.D. Hall; and A.G. Bovy. 2005. "A novel approach for nontargeted data analysis for metabolomics. Large-scale profiling of tomato fruit volatiles". *Breakthrough Technologies*. 139, 1125-1137.

Martins, R.C.; V.V. Lopes; A.A.Vicente; and J.A. Teixeira. 2008. "Computational shelf-life dating: complex systems approaches to food quality and safety". *Food Bioprocess Technology*, 1(3), 207-222.

AUTHOR BIOGRAPHIES

CRISTIANA C. CASTRO was born in Miranda do Douro, Portugal and went to University of Minho, where she studied Chemical and Biological Engineering and obtained her MSc degree in 2007. She worked for two years in the OPENMICROBIO project (PTDC/BIO/69310/2006) before start her PhD in Centre of Biological Engineering in 2009. Her email address is: cristianacastro@deb.uminho.pt and her web-page can be found in http://www.ceb.uminho.pt/pessoas/pid.aspx?id=328.