Development of an electrospray-mass spectral database for annotating metabolomics datasets: application to the analysis of the adult human urinary metabolome

Aurélie Roux, Ying Xu, Jean-François Heillier, Marie-Françoise Olivier, Eric Ezan, Jean-Claude Tabet, Christophe Junot

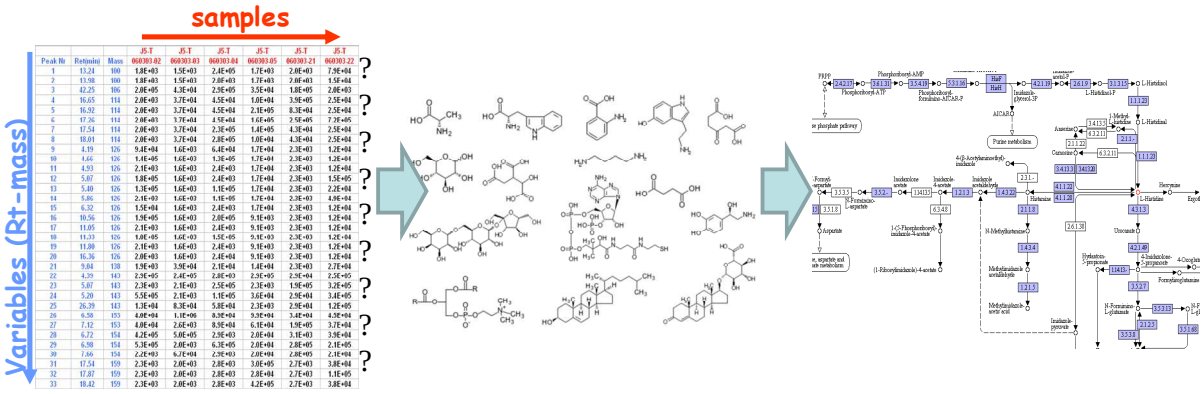
Development of an electrospray-mass spectral database for annotating metabolomics datasets: application to the analysis of the adult human urinary metabolome

Aurelie ROUX[1]; Ying XU[1]; Jean-François HEILIER[2]; Marie-Françoise OLIVIER[1]; Eric EZAN[1]; Jean-Claude TABET[3]; Christophe JUNOT[1]

[1]CEA - Centre d'Etude de Saclay, Gif-Sur-Yvette, France; [2]Institut National de Recherche et de Sécurité, Nancy, France; [3]LCSOB, IPCM, UMR-CNRS 7201, UPMC Paris Universitas, Paris, France

Introduction

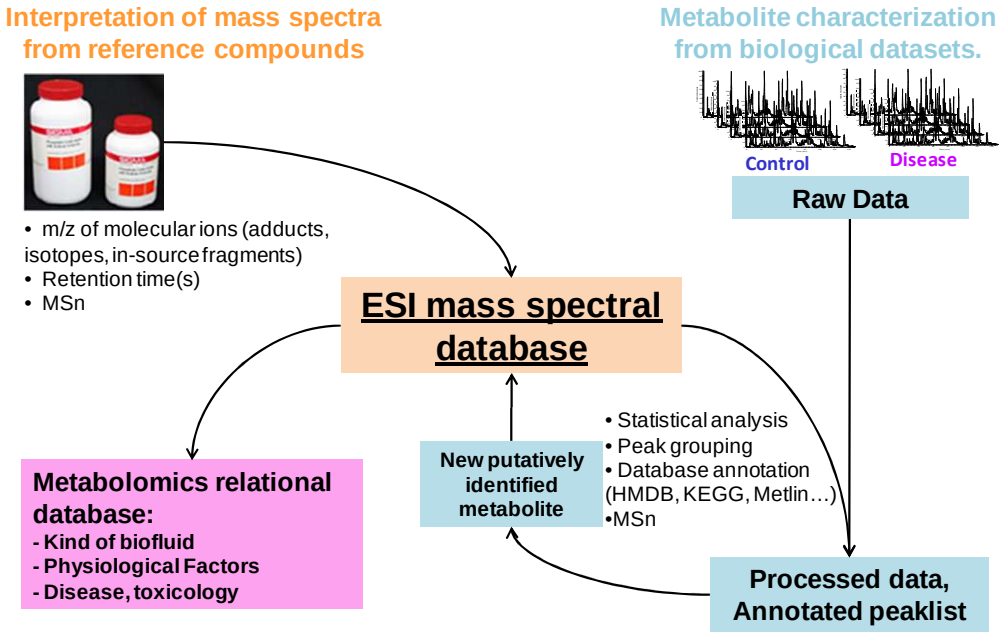
Metabolomics opens new perspectives for biomarker discovery in the field of nutrition and heath, and also in the development of system biology. The metabolic profiles of biofluids obtained by mass spectrometry or nuclear magnetic resonance contain a few hundreds to thousands of signals. However, a major part of this information remains unknown, or at least not characterized in the analytical systems, thus hampering the obtention of biologically meaningfull data. In this context, we here report on the development of an electrospray (ESI) spectral database for the annotation of high resolution mass spectrometry based metabolomics data sets and on its application to the analysis of urine samples from a cohort of healthy volunteers.



Methods

Urines samples were collected from 227 healthy volunteers matched according to gender (121 males, 106 females), age (18 to 81 years-old) and body mass index. They were randomly analyzed using ultra performance liquid chromatography coupled to a LTQ-Orbitrap Discovery mass spectrometer (ThermoFisher scientifics) operated in both negative and positive ion mode with the resolution set to 30000 (FWHM at m/z 400). Elution was performed in gradient conditions with mobile phases composed of water and acetonitrile containing 0.1% formic acid. Data treatment was achieved with XCMS R package, statistical analysis with SIMCA P11 and signals clustering with CAMERA R package and other home-made informatics tools, and signal annotation was performed using public metabolite databases and our in-house spectral database.

Preliminary Data



Two complementary approaches were designed to build the spectral database. The first one involved the analysis of mass spectra of reference compounds from our chemical library (1200 chemicals). Analytical information (compound related-ions recorded in both positive and negative mode and retention times in various chromatographic conditions) was collected from these spectra and then matched to that obtained from experimental metabolomics data sets. By these means, we formally annotated and identified about 163 metabolites (based on accurate mass, retention time and MSn spectral similarities to those of reference compounds) in urine samples from our cohort. The second approach was based on the characterization of new metabolites from experimental datasets. Automatic detection of ions was performed from raw data using XCMS software. Signals of interest were highlighted by multivariate analysis. The redundant information (i.e., isotope, in source product and adduct ions generated during the ionization process together with protonated or deprotonated ions) was then highlighted and further grouped by using mathematical and informatics tools. Finally, ions of interest were annotated thanks to queries of metabolomic public databases with the measured accurate masses and characterized by MSn experiments. By these means, 81 metabolites were putatively identified, based on accurate mass and interpretation of MSnspectra, Altogether 244 metabolites were identified or putatively identified including 78 never described in Human urines yet. 139 were only characterized by matching of the accurate measured mass to that of metabolites present in public databases, and the interpretation of their MSn spectra is under progress. In fact these 383 metabolites represent 1000 ions in both positive and negative ion mode. At last, we also address the impact of physiological factors such as gender, age and body mass index on the urinary concentrations of two hundred compounds for further building of a database of metabolic profiles.