Bioinformatics, 33(2), 2017, 300–302
doi: 10.1093/bioinformatics/btw588
Advance Access Publication Date: 14 September 2016
Applications Note

OXFORD

Data and text mining

IsotopicLabelling: an R package for the analysis of MS isotopic patterns of labelled analytes

Ruggero Ferrazza[1,2,3,*], Julian L. Griffin[3,4], Graziano Guella[1,5] and Pietro Franceschi[6]

[1]Bioorganic Chemistry Laboratory, Department of Physics and [2]Centre for Integrative Biology (CIBIO), University of Trento, 38123 Povo (TN), Italy, [3]Department of Biochemistry, University of Cambridge, Cambridge CB2 1GA, UK, [4]Medical Research Council (MRC) Human Nutrition Research (HNR), Cambridge CB1 9NL, UK, [5]Biophysical Institute, CNR, 38123 Povo (TN), Italy and [6]Edmund Mach Foundation, Research and Innovation Centre, Computational Biology Unit, I-38010 San Michele All'Adige (TN), Italy

*To whom correspondence should be addressed.
Associate Editor: Jonathan Wren

Received on May 6, 2016; revised on September 4, 2016; accepted on September 5, 2016

Abstract

Motivation: Labelling experiments in biology usually make use of isotopically enriched substrates, with the two most commonly employed isotopes for metabolism being ^{2}H and ^{13}C. At the end of the experiment some metabolites will have incorporated the labelling isotope, to a degree that depends on the metabolic turnover. In order to propose a meaningful biological interpretation, it is necessary to estimate the amount of labelling, and one possible route is to exploit the fact that MS isotopic patterns reflect the isotopic distributions.
Results: We developed the IsotopicLabelling R package, a tool able to extract and analyze isotopic patterns from liquid chromatography-mass spectrometry (LC-MS) and gas chromatography-MS (GC-MS) data relative to labelling experiments. This package estimates the isotopic abundance of the employed stable isotope (either ^{2}H or ^{13}C) within a specified list of analytes.
Availability and Implementation: The IsotopicLabelling R package is freely available at https://github.com/RuggeroFerrazza/IsotopicLabelling.
Contacts: r.ferrazza@unitn.it
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

In recent years there has been increasing interest in the application of metabolomics to medical and biological problems. However, many of the proposed metabolomics approaches turn out to be unsuitable to directly measure metabolic fluxes, which have to be indirectly inferred from pool sizes of metabolites. These limitations can be overcome by using stable isotopes to label metabolites in living cells, and stable isotope tracer-based approaches have been established for metabolite identification, quantification, and pathway analyses (Breitling, 2006; You, 2014). Mass Spectrometry (MS) is particularly suitable to detect stable isotopes (Klein and Heinzle, 2012), because the relative abundance of isotopes within a molecule has a direct impact on the isotopic pattern measured by the mass spectrometer.

Many software packages are available for predicting MS isotopic patterns (Ipsen, 2014; Rockwood, 1995; Valkenborg, 2012) and some exist for the analysis of experimental patterns resulting from labelling experiments (Capellades, 2016; Chokkathukalam, 2013; Huang, 2014; Millard, 2012). In addition, algorithms for the determination of isotope incorporation into substrates have been described (Lee, 1991). However, the already available software tools aim at either correcting for the contribution of naturally occurring isotopes, or finding the relative abundance between different isotopologues; a solution to directly estimate the abundance of the labelling isotope within target analytes is still lacking. To fill this gap, we developed the *IsotopicLabelling* R package, which can be used to assess the level of incorporation of the labelling isotope in a specific

metabolite (or list of metabolites) directly from the results of an MS-based targeted or untargeted metabolomics experiment. It is important to point out that a reliable estimate of the incorporation can be obtained only if the formula of the target metabolite is known in advance. In a completely untargeted scenario, indeed, the labelling makes the isotopic patterns useless for annotation, and mass accuracy alone cannot be used to infer a reliable elemental composition (Kind and Fiehn, 2006).

2 Methods

2.1 The input data

The *IsotopicLabelling* package requires MS pre-processed data which are fed to the algorithm as a data frame containing MS peak information (*m/z*, retention time (RT) and intensities or areas). In order to facilitate the use of *IsotopicLabelling* inside an *xcms* based pipeline (Smith, 2006), an *xcmsSet* object can be directly provided as input. Alternatively, the input list of detected features can be obtained in a number of other ways, including the use of proprietary software. For targeted analyses it is also possible to provide a list containing only the isotopic pattern of the compound of interest.

IsotopicLabelling also features the possibility of analysing isotopic patterns of multiply charged species.

In addition to the MS data, the user should provide: (*i*) the molecular formula of the charged species of interest, (*ii*) the charge state of the target adduct (if different from 1, the default), (*iii*) the isotope used for labelling (^{2}H or ^{13}C), (*iv*) the mass accuracy, (*v*) the expected RT and (*vi*) an estimated width of the chromatographic peaks.

It is also possible to process a list of target analytes (batch processing) by providing the previous parameters in tabular format (see the manual for details).

2.2 Isotopologue information

The package first lists all the possible isotopologues generated by the incorporation of the stable isotope in the target analyte, determining their *m/z*: the lightest one is the monoisotopic species, while the heaviest one is the species where all the atoms of the labelling element are present as their heaviest isotope (^{2}H or ^{13}C). To estimate a correct level of incorporation, the package also takes into account the natural isotopic distributions of all the elements present in the target analyte.

2.3 Extraction of the experimental patterns

Next, the package identifies the target experimental patterns by matching the *m/z* of the experimental features with the calculated isotopologue masses and applying an additional filter on the maximum tolerated retention time shift. The results are organized in a table of extracted MS patterns (one for each sample).

When extracting the signals, *IsotopicLabelling* properly takes into consideration the specified charge state of the adducts.

Importantly, the assumption being made during extraction and further analysis is that the mass resolution is not high enough to resolve the isotopic fine structure. Therefore, isobaric isotopologues are considered to give rise to a single mass peak.

2.4 Isotopic pattern analysis

The extracted patterns are then analyzed: for each sample, the abundance of the labelling isotope is estimated through a weighted non-linear least squares fitting whereby the theoretical pattern that best reproduces the experimental one is found; in order to correct for heteroskedasticity, the experimental signals are given weights proportional to the square root of their intensity. As a measure of the fitting accuracy, the standard error of the regression (SER) is also provided.

The patterns are computed using the *ecipex* package (Ipsen, 2014), since it relies on an algorithm with reduced computational demands.

During the analysis, it is possible to account for non-uniform incorporation by specifying how many atoms in the target analyte molecule are amenable to incorporation (see the package manual for details).

2.5 Inspecting the results

The fitting results can be assessed either graphically or exported as a *csv file. In the former case, a plot for each sample will show the extracted pattern superimposed to its best fitted theoretical pattern; in the latter case, the estimates are saved to a file containing information such as the estimated abundance of the label, the related SER and the deviation between experimental and fitted patterns.

In the package is also implemented a strategy to combine the estimates of the isotopic enrichment obtained on independent samples by taking into account both the SER associated to each single estimate and the distribution of the estimates across the different samples.

3 Results

3.1 The test dataset

To test the package, we used a dataset of 8 LC-MS analyses of the lipid extracts of a ^{13}C labelling experiment on YSBN1 yeast strain cultures; the first 4 samples have been extracted from unlabelled cell cultures (natural ^{13}C abundance), while in the last 4 the cells were grown using uniformly labelled ^{13}C glucose (99% ^{13}C labelling). Additional details can be found in the Supplementary Materials.

The test dataset is made available in the package as an *xcmsSet* object.

3.2 The outputs

We performed the isotopic pattern analysis on several lipid species, taking the phosphatidylcholine proton-adduct $[PC\ 32{:}2 + H]^+$ as an example. For unlabelled samples *IsotopicLabelling* returned an average ^{13}C abundance of $(1.078 \pm 0.003)\%$, in fair agreement with the natural occurrence of 1.070%, whereas for the labelled cases the ^{13}C abundance was estimated at $(98.943 \pm 0.005)\%$. A scheme of the data analysis workflow can be found in the Supplementary Materials.

3.3 Evaluation of limitations

In order to evaluate the limitations of *IsotopicLabelling*, we simulated a set of synthetic isotopic patterns, and looked for the minimum amount of label enrichment for which the patterns can be clearly told apart from their corresponding natural patterns. Such lower limit of enrichment strongly depends on both signal-to-noise ratio (SNR) and on the metabolite size (see Supplementary Materials for more information): the higher the SNR and *m/z*, the higher the sensitivity of the approach.

4 Conclusions

The *IsotopicLabelling* R package is a tool aimed at analysing MS data from isotopic labelling experiments; it has been designed to be easy to use and it provides a number of functions to quickly look at and save the results.

IsotopicLabelling is based on the assumption of having uniform labelling, and therefore the value it returns is the average ^{13}C occurrence within each target analyte. As a consequence, position-specific labelling information cannot be achieved. At the same time however, the number of atoms that can incorporate the label can be specified. Worthy of note, *IsotopicLabelling* can even be used to analyze data from MS/MS experiments, which have the advantage of providing more structural information, an important feature in isotopologue analysis.

To promote the use of *IsotopicLabelling* in the metabolomics community, it is integrated with *xcms*; an extensive step-by-step guide is provided at the project website, where all the details about how to practically use it can be found.

Acknowledgements

We would like to thank Dr. Nyasha Clarence Munjoma and Dr. Nianshu Zhang (Department of Biochemistry, University of Cambridge, UK) for kindly providing the LC-MS data.

Funding

J.L.G. is funded by the Medical Research Council, UK (Programme Grant: Lipid Profiling and Signalling).

Conflict of Interest: none declared.

References

Breitling,R. *et al*. (2006) Precision mapping of the metabolome. *Trends Biotechnol*., **24**, 543–548.

Capellades,J. *et al*. (2016) geoRge: a computational tool to detect the presence of stable isotope labeling in LC/MS-based untargeted metabolomics. *Anal. Chem*., **88**, 621–628.

Chokkathukalam,A. *et al*. (2013) mzMatch-ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. *Bioinformatics*, **29**, 281–283.

Huang,X. *et al*. (2014) X13CMS: global tracking of isotopic labels in untargeted metabolomics. *Anal. Chem*., **86**, 1632–1639.

Ipsen,A. (2014) Efficient calculation of exact fine structure isotope patterns via the multidimensional Fourier transform. *Anal. Chem*., **86**, 5316–5322.

Kind,T. and Fiehn,O. (2006) Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics*, **7**, 234.

Klein,S. and Heinzle,E. (2012) Isotope labeling experiments in metabolomics and fluxomics. *WIREs Syst. Biol. Med*., **4**, 261–272.

Lee,W.N.P. *et al*. (1991) Mass isotopomer analysis: theoretical and practical considerations. *Biol. Mass Spectrom*., **20**, 451–458.

Millard,P. *et al*. (2012) IsoCor: correcting MS data in isotope labeling experiments. *Bioinformatics*, **28**, 1294–1296.

Rockwood,A. *et al*. (1995) Rapid calculation of isotope distributions. *Anal. Chem*., **67**, 2699–2704.

Smith,C.A. *et al*. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem*., **78**, 779–787.

Valkenborg,D. *et al*. (2012) The isotopic distribution conundrum. *Mass Spectrom. Rev*., **31**, 96–109.

You,L. *et al*. (2014) Application of stable isotope-assisted metabolomics for cell metabolism studies. *Metabolites*, **4**, 142–165.