BIOINFORMATICS APPLICATIONS NOTE
Vol. 28 no. 9 2012, pages 1294–1296
doi:10.1093/bioinformatics/bts127

Data and text mining

Advance Access publication March 13, 2012

IsoCor: correcting MS data in isotope labeling experiments

Pierre Millard[1,2,3], Fabien Letisse[1,2,3], Serguei Sokol[1,2,3] and Jean-Charles Portais[1,2,3,*]

[1]Université de Toulouse; INSA, UPS, INP; LISBP, 135 Avenue de Rangueil, F-31077 Toulouse, [2]INRA, UMR792, Ingénierie des Systèmes Biologiques et des Procédés, F-31400 Toulouse and [3]CNRS, UMR5504, F-31400 Toulouse, France

Associate Editor: John Quackenbush

ABSTRACT

Mass spectrometry (MS) is widely used for isotopic labeling studies of metabolism and other biological processes. Quantitative applications—e.g. metabolic flux analysis—require tools to correct the raw MS data for the contribution of all naturally abundant isotopes. IsoCor is a software that allows such correction to be applied to any chemical species. Hence it can be used to exploit any isotopic tracer, from well-known (^{13}C, ^{15}N, ^{18}O, etc) to unusual (^{57}Fe, ^{77}Se, etc) isotopes. It also provides new features—e.g. correction for the isotopic purity of the tracer—to improve the accuracy of quantitative isotopic studies, and implements an efficient algorithm to process large datasets. Its user-friendly interface makes isotope labeling experiments more accessible to a wider biological community.

Availability: IsoCor is distributed under OpenSource license at http://metasys.insa-toulouse.fr/software/isocor/

Contact: jean-charles.portais@insa-toulouse.fr

Received on November 23, 2011; revised on February 13, 2012; accepted on March 8, 2012

1 INTRODUCTION

Mass spectrometry (MS) is extensively used for stable isotopic studies of metabolism. During the last decade, quantitative approaches—such as ^{13}C metabolic flux analysis—have been increasingly used in the fields of systems biology and biotechnology to provide novel biological insights and improve industrial processes (Nicolas *et al.*, 2007; Sauer, 2006). In such experiments, the different labeled forms—or isotopologues—of metabolites are quantified by the fine exploitation of isotopic clusters in MS spectra (Kiefer *et al.*, 2007). These clusters also contain information on all other isotopes that occur naturally in the molecules. To extract meaningful labeling information—i.e. isotopologue distribution—the contribution of naturally occurring isotopes first has to be subtracted. This can be achieved using publicly available software such as MSCorr (Wahl *et al.*, 2004) developed under MATLAB or the algorithm developed in Perl by (Moseley, 2010). However, the former software is only applicable to ^{13}C-labeling experiments and the latter is restricted to the correction of ultra-high resolution MS data for natural abundance of ^{13}C or ^{15}N.

Here, we present a novel software to correct MS data, IsoCor, which includes the following features:

- Correction for any chemical element, thereby extending the range of isotopic tracers—from well-known (^{13}C, ^{15}N) to unusual (^{57}Fe, ^{77}Se, etc) isotopes—and chemical species that can be investigated. Tracer elements with more than two isotopes (e.g. ^{16}O, ^{17}O, ^{18}O) can be considered.
- When metabolite derivatization is required for analytical purposes [e.g. derivatization of amino acids for GC-MS analysis (Wittmann, 2007)], IsoCor performs appropriate correction by taking into account the contribution of naturally occurring isotopes brought by the derivatization reagent.
- IsoCor provides two options that allow flexible correction of isotopic clusters to account for (or not) the composition of the label input including: (i) correction for the isotopic purity of the label input and (ii) correction for the occurrence of unlabeled positions in the label input. These options are of particular value for specific applications (Rodriguez-Castrillon *et al.*, 2008; Wittmann and Heinzle, 1999).
- IsoCor calculates the mean isotopic enrichment of molecules, which refers to the molecular content in the isotope. This information is particularly useful for targeted metabolic investigations, such as the quantification of split ratios between two metabolic pathways.
- IsoCor can be applied to large datasets and can deal with the increasing number of data that are generated with modern methods in a single experiment.

2 METHOD AND IMPLEMENTATION

The correction is performed with the matrix-based method introduced by (van Winden *et al.*, 2002), which requires solving the following equation:

$$IC_{cor} = CM^{-1} \cdot IC_{meas} \quad (1)$$

where IC_{cor} denotes the corrected isotopic cluster, CM is a correction matrix and IC_{meas} is the measured isotopic cluster.

This formalism has been implemented in Python programming language (http://python.org) which enables seamless usage of IsoCor on Windows, MacOS, Linux and other platforms supporting Python.

2.1 Construction of the correction matrix

The size of the correction matrix is $m \times n$, with $m \geq n$, where m is the length of the measured isotopic cluster and n is the number of isotopologues, i.e. $a + 1$ where a is the number of atoms of the tracer element in the molecule. The i-th column of the correction matrix ($1 \leq i \leq n$) is the isotopic cluster of the molecule containing i - 1 atoms of the tracer. It is commonly calculated using

*To whom correspondence should be addressed.

combinatorial probabilities (Hellerstein and Neese, 1999). For molecules containing a high number of atoms, the number of possible combinations to be calculated is high which can cause a bottleneck in the correction process. In probability theory, the probability distribution of the sum of independent variables is the convolution of their individual distributions. Therefore, the calculation can be performed faster by iteratively convolving the isotopic vectors $v_A = [p(A(1)),\ldots,p(A(x))]$ of all the atoms—except those of the label tracer—where $p(A(x))$ denotes the natural abundance of the x-th isotope of the atom A. Finally, the contribution of the tracer is added by convolving the previous isotopic cluster i - 1 times by the purity vector of the tracer (e.g. [0.01, 0.99] for a purity of 99%), and n - i times by its natural abundance (e.g. [0.9893, 0.0107] in a ^{13}C-labeling experiment) when it has to be corrected.

2.2 Optimization algorithm

Because the measurement of isotopic clusters contains noise, Equation (1) has to be solved using a least squares method. This process can result in biases, such as the calculation of negative isotopic fractions sometimes observed with MSCorr. Because negative fractions cannot exist, they are just flattened to 0 in (Moseley, 2010), which can lead to errors in the estimated parameters and in their interpretation. To avoid such biases, the optimization method used in IsoCor includes constraints on the isotopic fraction values. We applied the L–BFGS–B algorithm described in (Byrd *et al*., 1995) and implemented in the *optimize.l_bfgs_g()* function of the SciPy module. The cost function of the minimization process is defined as the sum of the squared weighted errors. The lower boundary of corrected isotopic fractions is constrained to 0, and the stopping criterion is fixed to 10^{-10} of relative reduction in cost value. The fitted isotopologue distribution is normalized to 1. The residuum vector may be used for an independent quality control (Moseley, 2010).

The calculation speed and correction accuracy of IsoCor were evaluated by processing several sets of simulated isotopologues with varying numbers of tracer atoms (up to 100 to stress the algorithm) and levels of labeling. The correction was performed after addition of the theoretical contributions of other naturally abundant elements to these isotopologues, with or without addition of measurement noise. The implementation was quite robust and highly accurate in both cases (see the IsoCor tutorial), showing that IsoCor provides high calculation speed together with high precision. Hence IsoCor is capable to efficiently process the large datasets that can now be acquired with modern, high-throughput MS technologies. This is critical for high-throughput metabolomics and even more for high-throughput fluxomics, which is a current major challenge for many biological applications.

3 USAGE

IsoCor is provided with plain text files that contain the information required to perform the correction. The file Isotopes.dat contains the values of the natural abundances of isotopes commonly found in biological compounds (Rosman and Taylor, 1998). The files Metabolites.dat and Derivatives.dat contain the elemental formula of most common metabolites and derivative residues, respectively. These files can be easily edited and implemented according to the user's needs.

Complete usage of IsoCor is detailed in the tutorial provided with the software. Figure 1 shows its graphic user interface. Briefly, the user can load datasets containing one or more isotopic clusters to be corrected (Panel C). These data files should be generated according to the format described in the tutorial. The options—i.e. selection of the isotopic tracer, correction for the isotopic purity of the label input, correction for unlabeled positions in the label input and calculation of mean enrichment—of the correction process have to be selected (Panel B) before running the calculation. When correction is applied to single isotopic clusters, the calculated

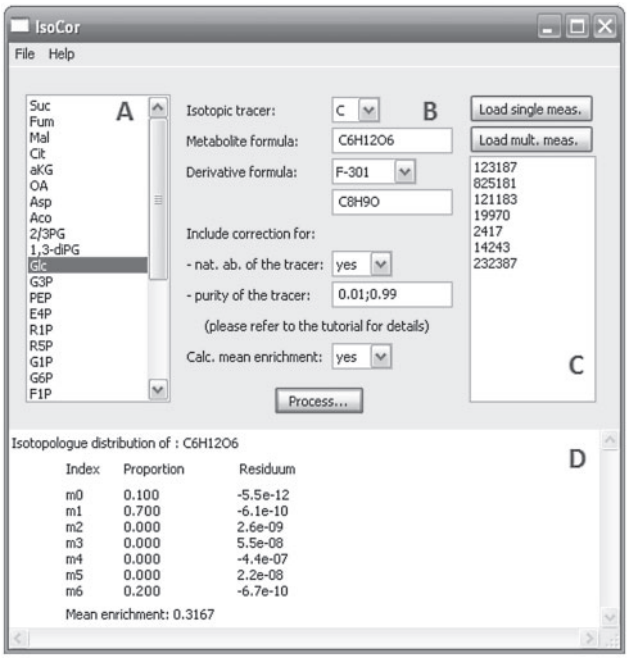

Fig. 1. Screenshot of IsoCor after correction of a single isotopic cluster. (**A**) List of metabolites; (**B**) correction parameters panel; (**C**) selection of experimental data files; and (**D**) calculation results.

data—corrected isotopic fractions and minimization residuum—are displayed in Panel D. When correction is applied to a dataset containing several measurements, the corrected data are written in a text file and other information—e.g. errors that occur during correction—is written in a log file and displayed in Panel D.

ACKNOWLEDGEMENTS

The authors are thankful to the members of the MetaSys team (LISBP, Toulouse, France) for fruitful discussions and to Daphne Goodfellow for careful reading of the manuscript.

Funding: INRA (Program CJS, PhD fellowship of P.M.) and Agence Nationale de la Recherche (ANR-08-BIOE-002-02).

Conflict of Interest: none declared.

REFERENCES

Byrd,R.H. *et al.* (1995) A limited memory algorithm for bound constrained optimization. *Siam. J. Sci. Comput.*, **16**, 1190–1208.

Hellerstein,M.K. and Neese,R.A. (1999) Mass isotopomer distribution analysis at eight years: theoretical, analytic, and experimental considerations. *Am. J. Physiol.*, **276**, E1146–E1170.

Kiefer,P. *et al.* (2007) Determination of carbon labeling distribution of intracellular metabolites from single fragment ions by ion chromatography tandem mass spectrometry. *Anal. Biochem.*, **360**, 182–188.

Moseley,H.N. (2010) Correcting for the effects of natural abundance in stable isotope resolved metabolomics experiments involving ultra-high resolution mass spectrometry. *BMC Bioinformatics*, **11**, 139.

Nicolas,C. *et al.* (2007) Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase. *FEBS Lett.*, **581**, 3771–3776.

Rodriguez-Castrillon,J.A. *et al.* (2008) Isotope pattern deconvolution as a tool to study iron metabolism in plants. *Anal. Bioanal. Chem.*, **390**, 579–590.

Rosman,K.J.R. and Taylor,P.D.P. (1998) Isotopic compositions of the elements 1997. *Pure Appl. Chem.*, **70**, 217–235.

Sauer,U. (2006) Metabolic networks in motion: 13C-based flux analysis. *Mol. Syst. Biol.*, **2**, 62.

van Winden,W.A. *et al.* (2002) Correcting mass isotopomer distributions for naturally occurring isotopes. *Biotechnol. Bioeng.*, **80**, 477–479.

Wahl,S.A., Dauner,M. and Wiechert,W. (2004) New tools for mass isotopomer data evaluation in (13)C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. *Biotechnol. Bioeng.*, **85**, 259–268.

Wittmann,C. (2007) Fluxome analysis using GC-MS. *Microb. Cell Fact.*, **6**, 6.

Wittmann,C. and Heinzle,E. (1999) Mass spectrometry for metabolic flux analysis. *Biotechnol. Bioeng.*, **62**, 739–750.