Bioinformatics, 32(1), 2016, 154–156
doi: 10.1093/bioinformatics/btv514
Advance Access Publication Date: 17 September 2015
Applications Note

Data and text mining

ICT: isotope correction toolbox

Christian Jungreuthmayer[1,2,*], Stefan Neubauer[1,3], Teresa Mairinger[1,3], Jürgen Zanghellini[1,2,*] and Stephan Hann[1,3]

[1]Austrian Centre of Industrial Biotechnology (ACIB), Vienna, Austria, [2]Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria and [3]Department of Chemistry, University of Natural Resources and Life Sciences, Vienna, Austria

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

Received on April 27, 2015; revised on August 7, 2015; accepted on August 25, 2015

Abstract

Summary: Isotope tracer experiments are an invaluable technique to analyze and study the metabolism of biological systems. However, isotope labeling experiments are often affected by naturally abundant isotopes especially in cases where mass spectrometric methods make use of derivatization. The correction of these additive interferences—in particular for complex isotopic systems—is numerically challenging and still an emerging field of research. When positional information is generated via collision-induced dissociation, even more complex calculations for isotopic interference correction are necessary. So far, no freely available tools can handle tandem mass spectrometry data. We present *isotope correction toolbox*, a program that corrects tandem mass isotopomer data from tandem mass spectrometry experiments. Isotope correction toolbox is written in the multi-platform programming language Perl and, therefore, can be used on all commonly available computer platforms.

Availability and implementation: Source code and documentation can be freely obtained under the Artistic License or the GNU General Public License from: https://github.com/jungreuc/isotope_correction_toolbox/

Contact: {christian.jungreuthmayer,juergen.zanghellini}@boku.ac.at

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Isotope labeling experiments play a crucial role in exploring, studying and understanding metabolic pathways in biological systems. However, the obtained isotopolog and tandem mass isotopomer signals are biased by natural abundant isotopes of the native molecules and of atoms introduced via derivatization (Antoniewicz, 2013). Consequently, the natural abundance results in a distortion of the experimental data. The correction of mass spectrometry data is still an emerging field of research (Carreer et al., 2013; Choi et al., 2013; Moseley, 2010; Rantanen et al., 2002; van Winden et al., 2002; Wittman and Heinzle, 1999), as it is non-trivial and numerically challenging, in particular, if a large number of isotopes are considered during the correction procedure. There exist several software tools which are able to compute the contribution of intensities from the natural abundant isotopes on the measured intensities and hence are capable to correct the measured data. *Pynac* (Carreer *et al.*, 2013) (http://files.cesb.uky.edu/software/PYNAC/) is a correction software written in the programming language Python. *Pynac* benefits from the multi-platform capabilities of Python is highly flexible by using text files for configuration and setup and can account for any number of isotopes. However, it is not able to correct data from tandem mass spectrometry experiments. Another prominent available program is *IsoCor* (Millard *et al.*, 2012) (http://metasys.insa toulouse.fr/software/isocor/). *IsoCor* is also implemented in Python and is a fast and excellent tool to correct mass spectrometry data. It supports the correction of the natural abundance of the tracer at unlabeled positions and allows to take into account the isotopic purity of the labeling source. Furthermore, in *IsoCor*, any isotope and not only ^{13}C can be used as tracer. Though *IsoCor* is also not able to deal with tandem mass isotopomer data. In this application note, we

present a novel software tool named *isotope correction toolbox* (*ICT*), which is also able to deal with ion fragmentation and, hence, can correct mass spectra obtained by tandem mass spectrometry.

2 Methods

ICT is written in the programming language Perl. Perl runs on all commonly available operating system, such as Linux/Unix, BSD, Mac OS and Windows. Hence, *ICT* can be used on virtually all computer platforms. *ICT* is a command line tool. The input data (chemical composition of the fragments and the measured intensities) must be given in form of plain text files. After the execution of *ICT*, the corrected data are provided to the user as text files, as well. The main features of *ICT* are as follows:

- *ICT* can correct tandem mass spectrometry data.
- *ICT* supports a batch execution mode which allows to quickly correct many experimental datasets in a single run.
- Any labeling source can be used with *ICT*.
- Any element and number of isotopes can be considered with *ICT*.
- *ICT* can account for effects caused by the natural abundance of the tracer at unlabeled positions.
- *ICT* allows to consider the isotopic purity of the tracer.

The principal program flow of *ICT* is illustrated in Figure 1. After the input files are read, the correction matrix *CM* is generated which relates the measured intensities, I_{meas}, with the corrected intensity values, I_{corr}:

$$CM \times I_{corr} = I_{meas}. \quad (1)$$

The generation of *CM* is done by a multi-step procedure. First, for each considered isotope, all relevant *N*/*n*-pairs are computed, where *N* and *n* are the number of the isotopes in the precursor and the product ion, respectively. For each considered element, these isotope *N*/*n*-pairs are combined to create sets of *N*/*n*-pairs. Next, these sets of *N*/*n*-pairs are combined to obtain the complete list of relevant isotope combination which affects the measured spectrum. During these combining processes, all irrelevant entries are immediately removed (Fig. 1), to reduce the amount of consumed memory (RAM) and to speed up the execution time. For each set of isotope combination, the probability of occurrence is computed (see Supplementary Information for further details). The fragmentation of the precursor can be considered as a drawing process where *n* samples are drawn from a pool of *N* elements, which affects the probability of occurrence. The probability values are finally used to create the correction matrix. A lookup table is utilized to identify the index where a specific isotope combination contributes to the correction matrix. The lookup table is realized by a hash table (associative array). Hash tables are a native variable type in Perl and allow fast access to data that are stored as key-value-pairs. In a last step, Equation (1) is solved to obtain the corrected intensities I_{corr} and the results are written to a file. Negative corrected intensities which can arise from measurement uncertainties (Millard *et al.*, 2012) are set to zero (Moseley, 2010) during the solution process of Equation (1).

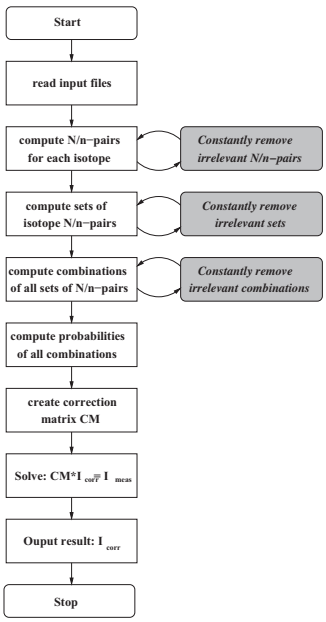

Fig. 1. Illustration of the principal workflow of *ICT*

3 Results

We have successfully used our program on Linux, Mac OS X and Windows computers. To verify the results of *ICT*, we performed numerous correction runs of non-fragmentation mass spectrometry data and compared them with results, which we obtained by using *IsoCor*. We found an excellent agreement between the corrected data computed by *ICT* and *IsoCor*. The verification of the correction procedure for tandem mass spectrometry data was performed by comparing the results of *ICT* with a large set of corrected data which we obtained by utilizing a (slow and inefficient) Microsoft Excel based method.

Typically, the correction for a smaller problem takes a fraction of a second. For instance, running a batch of 585 correction procedures taking into account the natural abundance of carbon and silicon for a variety of compounds, such as amino acids and low molecular weight organic acids, took approximately 20 s. Larger problems, such as the correction of interferences caused by the natural abundance of carbon, hydrogen, nitrogen, oxygen and silicon for ethoximated/trimethylsilylated sedoheptulose-7-phosphate (C29H73N1O10P1Si7), can usually be done within less than 1 s.

4 Discussion and conclusion

We present a novel isotope correction tool that is freely available and runs on virtually any computer platform. Our program supports numerous features, including but not limited to batch processing, considering the purity of the tracer, supporting any labeling source and accounting for any number of isotopes. However, the main advantages of our program are that it can deal with precursor ion fragmentation and hence, unlike other available programs, can correct tandem mass spectrometry data. *ICT* is an efficient and easy-to-use tool that is freely available as open source software. It might be useful to anyone who needs to correct the natural abundance in tracer experiments.

Funding

This work has been supported by the Federal Ministry of Science, Research and Economy (BMWFW), the Federal Ministry of Traffic, Innovation and

Technology (bmvit), the Styrian Business Promotion Agency SFG, the Standortagentur Tirol, the Government of Lower Austria and ZIT - Technology Agency of the City of Vienna through the COMET-Funding Program managed by the Austrian Research Promotion Agency FFG.

Conflict of Interest: none declared.

References

Antoniewicz,M.R. (2013) Tandem mass spectrometry for measuring stable-isotope labeling. *Curr. Opin. Biotechnol.*, **24**, 48–53.

Carreer,W.J. *et al.* (2013) A computational framework for high-throughput isotopic natural abundance correction of omics-level ultra-high resolution FT-MS datasets. *Metabolites*, **3**, 853–866.

Choi,J. *et al.* (2013) Measuring complete isotopomer distribution of aspartate using gas chromatography/tandem mass spectrometry. *Anal. Chem.*, **84**, 4628–4632.

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

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

Rantanen,A. *et al.* (2002) Computing positional isotopomer distributions from tandem mass spectrometric data. *Metab. Eng.*, **4**, 285–294.

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

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