Novel method for estimating isotope incorporation into peptides using the half-decimal place rule

Ingo Fetzer[1,*] **, Nico Jehmlich**[2]**, Frank Schmidt**[3]

1. Department of Environmental Microbiology, Helmholtz Center for Environmental Research - UMB, 04318 Leipzig, Germany
2. Department of Proteomics, Helmholtz Center for Environmental Research - UMB, 04318 Leipzig, Germany
3. Interfaculty Institute for Genetics and Functional Genomics, Ernst-Moritz-Arndt-University, 17487 Greifswald, Germany

* Contact author: ingo.fetzer@ufz.de

Keywords: biostatistics, robust linear modeling, k-means clustering, peptides, isotopes, incorporation rate estimation

The metabolic incorporation of stable isotopes such as ^{13}C or ^{15}N into proteins has become a powerful component to disentangle substrate fluxes in bacteria and understand microbial degradation processes. Incorporation of heavy isotopes into proteins can be used to analyze process parameters such as protein turnover rates. Here we present a new method for calculating the incorporation rate of ^{13}C into peptides by using the information given in the decimal places of peptides by making full use of the information acquired by high resolution mass spectrometry. Our method for estimating ^{13}C incorporations is based on the characteristics of the so called 'half decimal place rule' (Mann 1995, Schmidt 2003). The rule follows the observation that when molecular masses of (isotopic unlabelled) peptide sequences are plotted against their digital residuals (=numbers behind the decimal point) the resulting plot will follow a linear pattern with a characteristic slope. Gradual incorporation of heavy isotopes into peptides will produce the same linear structure but the regression slope will be also altered. Thus, the steepness of the slope contains information on the relative incorporation of heavy isotopes into the proteins. Our method and all calculations have been implemented in 'R' in form of three separate scripts: Primarily molecular weights as references for unlabelled and labeled ^{13}C peptides had to be calculated. For this an existing peptide database of *Mycobacterium tuberculosis* (initially containing >900.000 protein sequences with variable lengths of 2-40 amino acids) was used. In a first script, primarily the dataset was reduced to the relevant 90,637 amino acid sequences taking advantage of R's excellent data mining abilities. In a following step the molecular weights all sequences were determined by multiplying every amino acid from each sequence with its corresponding molecular weight. In a second step, after classification of the peptides by applying k-means clustering (kmeans() in stats package) we were able to plot molecular peptide weights against their digital residuals. With help of linear curve fitting (using the lm() application of the 'stats' package) the reference slopes for unlabelled and 13C-labeled amino acid sequences could be estimated. Including our reference slopes in a third script, a user friendly approach was developed using the tcl/tk package, providing windows to guide the user easily through the procedure calculating the incorporation rates of his own data series. Since user data often contain far less measurements than we used as reference and, moreover, often contain 'outliers' standard linear curve fitting was not applicable. Here the robust linear model procedure (rlm() of the MASS package) gave more reliable results with better isotope incorporation rate estimates. With help of the previously determined slopes of fully labeled/unlabeled data the relative incorporation rates of isotopes for user data could be computed. Additionally, the applicability of this approach is demonstrated by using *Pseudomonas putida* ML2 proteins labeled uniformly via the consumption of $^{13}C_6$-benzene. Based on several labeled peptides the incorporation of ^{13}C was calculated. As a result, the accuracy of the calculated incorporation rate depends on the number of used peptide masses, whereas only 100 peptide masses are required to get precision higher than 5 atom % of ^{13}C incorporation.

References

Mann, M. (1995). 43rd ASMS Conference on Mass Spectrometry and Applied Topics, Atlanta, GA.

Schmidt, F., Schmid, M., Mattow, J., Facius, A., Pleissner, K. P. and Jungblut, P. R. (2003) Iterative data analysis is the key for exhaustive analysis of peptide mass fingerprints from proteins separated by two-dimensional electrophoresis. *J Am Soc Mass Spectrom.* 14: 943-956.