Automated Identification of Molecular Structures for NMR Based Metabolomics

Arianna Filntisi, George K. Matsopoulos
School of Electrical and Computer Engineering
National Technical University of Athens
Athens, Greece
Email: afilntisi@biomig.ntua.gr, gmatso@esd.ece.ntua.gr

Charalambos Fotakis, Panagiotis Zoumpoulakis
Institute of Biology, Medicinal Chemistry and Biotechnology
National Hellenic Research Foundation
Athens, Greece
Email: bfotakis@yahoo.com, pzoump@eie.gr

Dionisis Cavouras, Pantelis Asvestas
Department of Biomedical Engineering
Technological Educational Institute of Athens
Athens, Greece
Email: cavouras@teiath.gr, pasv@teiath.gr

***Abstract*—The Nuclear Magnetic Resonance (NMR) based metabolomics approach is implemented in medicine and pharmacology to assess, identify and quantify metabolites in different biological samples. Metabolite determination, which is a challenging task due to the complexity of the biological matrices, can benefit from bioinformatics tools. In this context, our research has focused on the development of a new computational metabolite identification method from ^{1}H-NMR spectra.**

Keywords- metabolomics; automated metabolite identification; ^{1}H NMR.

I. INTRODUCTION

Metabolomics is the research discipline that is concerned with the qualitative and quantitative assessment of the metabolic response of biological systems to pathophysiological stimuli or genetic modifications. Metabolomics provides information of *in vivo* multi-organ functional integrity in real time [1]. The NMR metabolomics approach has a number of applications, such as the identification of biomarkers of disease and pharmacodynamic response.

In alignment with the increasing applications of the NMR metabolomics, the chemoinformatics field has evolved focusing on data processing and the metabolite identification algorithms. Several methods have targeted the task of metabolite profiling with line fitting and Bayesian modelling, [2-4]. Another strategy has been based on the matching of the input spectrum with a set of reference compounds [5-6].

Herein, we introduce a new computational method for the automatic identification of metabolites from 1D ^{1}H-NMR spectra. In Section II, the main steps of the method are presented. In Section III, preliminary results from the analysis of two mixtures are described.

II. METHOD

The steps A, B presented below are preprocessing steps, while step C constitutes the core of our method.

A. Preprocessing

The input spectrum can be preprocessed by removing low significance regions. Denoising and thresholding can be performed if necessary. The mean, median and Gaussian denoising filters were tested and the last one has been chosen as the best. However, denoising was performed frugally as an auxiliary step, and therefore the filter selection did not affect significantly the results.

B. Data Reduction

The Adaptive Intelligent binning algorithm [7] is applied to the spectrum, resulting in a number of bins corresponding to the local minima across the frequency spectrum. Subsequently, peak picking is performed, selecting for every bin the frequency corresponding to the maximum intensity.

C. Metabolite Search

The input spectrum is screened for metabolites, as defined by a specific database (see Section II.D). For each multiplet peak of every candidate metabolite in the database, a number of peak combinations are being considered as a possible fit. A candidate peak combination $p = \{p_{c1}, p_{c2}, \ldots, p_{cn}\}$ for a multiplet is scored differently depending on the order of the spectral lines (first order or higher order multiplets, singlets and multiplets without rules). The scoring of a multiplet is based on its properties, such as the *j* coupling and the height ratios. The optimal peak combination is chosen for each metabolite.

D. Database

A database file containing 850 metabolites was synthesized from the available Human Metabolome Database (HMDB) [8]. For every metabolite, information such as the multiplet type, the expected frequency ranges, the number of hydrogen atoms, the j coupling values, the height ratios, as well as the number of peaks has been stored.

 ISBN: 978-1-61208-488-6

III. RESULTS

The method described in Section II has been tested on an amino acid mixture and a human amniotic fluid sample. The former comprised of L-Alanine, L-Valine, L-Methionine, L-Proline, L-Glutamic acid, L-Leucine, L-Isoleucine, L-Arginine, Trigonelline, which were all successfully identified. The latter was screened for the presence of 40 metabolites (experimentally identified), 36 of which were positively recognized.

The performance of our method upon those spectra has been compared to an existing metabolite recognition tool, MetaboHunter (Table 1) [5] and proved to be enhanced despite a higher execution time. Our method seems to be more robust seeking the optimal peaks for a candidate metabolite at each input spectrum with a four digit accuracy, as opposed to assigning predetermined peaks to a metabolite.

In Figure 1, we can see characteristic multiplet peaks assigned to metabolites in an area of the spectrum of the human amniotic fluid mix.

IV. DISCUSSION

This work presents briefly a new chemoinformatics method for metabolite recognition under development. Also, preliminary results from the application of the method on two spectra have been described.

A limitation of our method is the fact that its performance has not been verified on different sample types. Our future goals include the refinement of the method and the default parameter values used. Further validation of the proposed method and comparison with other metabolite identification methods are also necessary.

ACKNOWLEDGMENT

Funding was received through an IKY Fellowship of Excellence for postgraduate studies in Greece – Siemens Program. The authors confirm that the funder had no influence over the study design, content of the paper, or selection of this conference.

REFERENCES

[1] J. K. Nicholson, J. C. Lindon, and E. Holmes, ""Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data". Xenobiotica, vol. 29(11), pp 1181-1189, 1999.

[2] C. Zheng, S. Zhang, S. Ragg, D. Raftery, and O. Vitek, "Identification and quantification of metabolites in (1)H NMR spectra by Bayesian model selection," Bioinformatics, vol. 27(12), pp. 1637–44, 2011.

[3] P. Mercier, M. J. Lewis, D. Chang, D. Baker, and D. S. Wishart, "Towards automatic metabolomic profiling of high-resolution one-dimensional proton NMR spectra," J. Biomol. NMR, vol. 49(3-4), pp. 307–23, 2011.

[4] W. Astle, M. De Iorio, S. Richardson, D. Stephens, and T. Ebbels, "A Bayesian Model of NMR Spectra for the Deconvolution and Quantification of Metabolites in Complex Biological Mixtures," J. Am. Stat. Assoc., vol. 107(500), pp. 1259–71, 2012.

[5] D. Tulpan, S. Léger, L. Belliveau, A. Culf, and M. Cuperlović-Culf, "MetaboHunter: an automatic approach for identification of metabolites from 1H-NMR spectra of complex mixtures," BMC Bioinformatics, vol. 12, pp. 400, 2011.

[6] D. Jacob, C. Deborde, and A. Moing, "An efficient spectra processing method for metabolite identification from 1H-NMR metabolomics data," Anal. Bioanal. Chem., vol. 405(15), pp. 5049-61, 2013.

[7] T. D. De Meyer, et al, "NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm," Anal. Chem., vol. 80(10), pp. 3783–90, 2008.

[8] D. S. Wishart, et al, "HMDB: the Human Metabolome Database," Nucleic Acids Res., vol. 35(Database issue), D521–6, 2007.

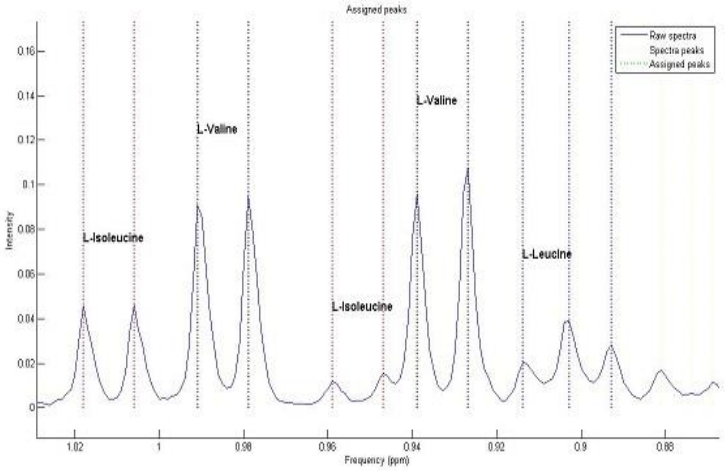

Figure 1. Peak assignment in an area of the human amniotic fluid spectrum.

TABLE I. COMPARATIVE PEAK ASSIGNMENT FOR THREE METABOLITES OF THE HUMAN AMNIOTIC FLUID SPECTRUM

Metabolites	Metabolite identification methods	
	Our method	***MetaboHunter***
L-Valine	0.9269, 0.9389, 0.9789, 0.9909, 3.5571, 3.5641, 2.1900, 2.2020, 2.2100, 2.2140, 2.2200, 2.2270, 2.2320	0.98, 1.02, 1.04, 2.23, 2.24, 2.25, 2.26, 2.27, 2.28, 2.29
L-Isoleucine	0.9469, 0.9589, 1.0059, 1.0179, 1.9430, 1.9570, 1.9690, 1.9790, 1.9910	0.91, 0.93, 0.94, 0.99, 1, 1.21, 1.22, 1.26, 1.28, 1.42, 1.43, 1.44, 1.45, 1.46, 1.47, 1.48, 1.49, 1.94, 1.96, 1.97, 1.98, 1.99, 2, 3.66
L-Leucine	0.8929, 0.9029, 0.9139, 1.6470, 1.6590, 1.6720, 1.6840, 1.6980, 3.7001, 3.7111, 3.7191	0.94, 0.95, 0.96, 1.65, 1.67, 1.69, 1.7, 1.71, 1.72, 1.73, 1.75, 3.71, 3.72, 3.73, 3.74

 ISBN: 978-1-61208-488-6