Journal of Magnetic Resonance 226 (2013) 93–99

Contents lists available at SciVerse ScienceDirect

Journal of Magnetic Resonance

journal homepage: www.elsevier.com/locate/jmr

MetaboID: A graphical user interface package for assignment of ^{1}H NMR spectra of bodyfluids and tissues

Neil MacKinnon [a,b,*], Bagganahalli S. Somashekar [a,b], Pratima Tripathi [a,b], Wencheng Ge [b], Thekkelnaycke M. Rajendiran [c], Arul M. Chinnaiyan [c,d,e,f], Ayyalusamy Ramamoorthy [a,b,*]

[a] Biophysics, University of Michigan, Ann Arbor, MI, USA
[b] Department of Chemistry, University of Michigan, Ann Arbor, MI, USA
[c] Michigan Center for Translational Pathology, Department of Pathology, University of Michigan Medical School, Ann Arbor, MI, USA
[d] Department of Urology, University of Michigan Medical School, Ann Arbor, MI, USA
[e] Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI, USA
[f] Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI, USA

ARTICLE INFO

Article history:
Received 21 August 2012
Revised 5 November 2012
Available online 21 November 2012

Keywords:
NMR
Metabolomics
Assignment
Mixture analysis

ABSTRACT

Nuclear magnetic resonance based measurements of small molecule mixtures continues to be confronted with the challenge of spectral assignment. While multi-dimensional experiments are capable of addressing this challenge, the imposed time constraint becomes prohibitive, particularly with the large sample sets commonly encountered in metabolomic studies. Thus, one-dimensional spectral assignment is routinely performed, guided by two-dimensional experiments on a selected sample subset; however, a publicly available graphical interface for aiding in this process is currently unavailable. We have collected spectral information for 360 unique compounds from publicly available databases including chemical shift lists and authentic full resolution spectra, supplemented with spectral information for 25 compounds collected in-house at a proton NMR frequency of 900 MHz. This library serves as the basis for MetaboID, a Matlab-based user interface designed to aid in the one-dimensional spectral assignment process. The tools of MetaboID were built to guide resonance assignment in order of increasing confidence, starting from cursory compound searches based on chemical shift positions to analysis of authentic spike experiments. Together, these tools streamline the often repetitive task of spectral assignment. The overarching goal of the integrated toolbox of MetaboID is to centralize the one dimensional spectral assignment process, from providing access to large chemical shift libraries to providing a straightforward, intuitive means of spectral comparison. Such a toolbox is expected to be attractive to both experienced and new metabolomic researchers as well as general complex mixture analysts.

1. Introduction

Metabolomics is increasingly utilized in biomedical and clinical studies to better understand various diseases. As an analytical tool, nuclear magnetic resonance (NMR) spectroscopy continues to make invaluable contributions to the field of metabolomics/metabonomics. With increasing availability of high magnetic-field spectrometers (e.g., 900 MHz) coupled with advanced data collection and analysis techniques (non-linear sampling [1], covariance [2–4], STOCSY [5]), and extension into selective one-dimensional (1D) and two-dimensional (2D) NMR techniques (e.g. sel-TOCSY [6,7], JRES [8]), NMR-based investigations now rival mass spectrometry metabolomic studies in terms of the large sample cohorts that can be analyzed [9,10]. A significant challenge for NMR-based metabolomic studies remains the assignment of all small molecules contributing to the 1D proton (^{1}H) spectrum.

One-dimensional ^{1}H NMR spectra are by far the most common data routinely collected on biological fluids, cell extracts, and tissues due to the high NMR sensitivity of the hydrogen nucleus. While several tools for processing and preparing raw NMR data for multivariate statistical analysis are publicly available (NMRLAB [11], ProMetab [12], matNMR [13], MetaboAnalyst [14],

Abbreviations: 1D/2D, one-dimensional/two-dimensional; BMRB, Biological Magnetic Resonance Bank; HMBC, heteronuclear multiple bond correlation; HMDB, Human Metabolome Database; HSQC, heteronuclear single quantum coherence; JRES, *J*-coupling resolved; NMR, nuclear magnetic resonance; SelTOCSY, selective total correlation spectroscopy; STOCSY, statistical total correlation spectroscopy; TOCSY, total correlation spectroscopy.

* Corresponding authors. Address: Department of Chemistry, University of Michigan, 930 North University Avenue, Ann Arbor, MI 48109-1055, USA. Fax: +1 734 764 3323.

E-mail addresses: nmackinn@umich.edu (N. MacKinnon), ramamoor@umich.edu (A. Ramamoorthy).

1090-7807/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmr.2012.11.008

MetaboLab [15], rNMR [16]), the challenge remains resonance assignment which is inhibited by the severe signal overlap (or spectral crowding). The traditional solution to the overlap problem is to extend the NMR experiment into higher frequency dimensions, either with homonuclear (e.g. JRES, COSY, and TOCSY) or heteronuclear (e.g. HSQC and HMBC) experiments. This strategy is limited due to the large time requirements of the 2D experiment, which may be as much as 10–40 times longer than that of the 1D ^{1}H experiment. The collection of a full suite of 2D experiments required for full spectral assignment of every sample is clearly a daunting task, particularly when considering the large sample cohorts typically utilized in metabolomics studies. However, it is not the case that every sample is completely unique within a cohort; often there is significant correspondence between the samples and thus selecting one or several representative samples to undergo the full battery of 2D experimental measurements is all that is necessary. Thus, the final analysis still requires resonance assignment of all the peaks in 1D ^{1}H spectra, with the knowledge of the 2D spectra of selected samples guiding the process.

The challenge in assigning 1D ^{1}H spectra lies in streamlining the method and attempting to minimize the repetitive procedure of candidate resonance assignment. While public metabolite databases offer search functions based on NMR chemical shift lists (HMDB [17,18] and BMRB [19]), which is often the first step in the resonance assignment process, a convenient method to directly compare an experimental spectrum with an authentic compound spectrum is unavailable. Excellent commercial products such as Chenomx and Bruker BioSpin's AMIX are available; however, to the best of our knowledge only recently have publicly available tools designed for the assignment problem been proposed [20]. We present a graphical interface designed to aid in 1D ^{1}H spectral assignment, which is written within the technical programming environment of Matlab™ (R2010b, Mathworks, Natick, MA). The interface, named MetaboID, is built upon a library of 360 unique small molecule ^{1}H NMR spectra compiled from public metabolomic databases (HMDB [17,18] and BMRB [19]), and offers a convenient workspace for comparing experimental spectra with the spectra of authentic reference compounds to guide the resonance assignment process. Importantly, generating user-specific libraries (e.g. magnetic field specific, solvent specific, etc.) is straightforward and thus MetaboID is fully customizable to the particular user's requirements. Overall, the central goal of MetaboID is to provide a user friendly platform for the chemical assignment of metabolites by providing an interface to a large NMR spectral library of authentic compounds (360 compounds) with a seamless ability to plot individual authentic spectra while interacting with experimental NMR data of biofluids.

2. Implementation

MetaboID is a collection of three primary user interfaces designed to aid in the various aspects of 1D ^{1}H NMR spectral assignment. The design is such that assignment of each spectrum may be considered as an individual task with associated task-specific files (managed by a File Manager) including an authentic compound peak-list library, an experimental spectrum (an *.xls(x) file), compound sub-lists, and experimental spectra collected after authentic compound spiking. The overall envisioned workflow was divided into the following tasks (Scheme 1):

1. Preliminary authentic compound searching via chemical shift peak lists obtained, for example, directly from the 1D spectrum or from 2D homonuclear correlation spectra (e.g., COSY, TOCSY, NOESY).
2. Visualization of the authentic spectra of potential compounds (also *.xls(x) files), either as individual spectra or overlaid with an experimental spectrum.
3. Assignment confirmation by comparison of an experimental spectrum before and after spiking the sample with an authentic compound.
4. Generation of task-specific compound lists which may include compounds that are safely excluded, candidate assignments, or targeted profiling compounds. These lists can then be used to edit the main compound list and reduce redundant or repeated searches.

In order for such a tool to be useful, a comprehensive NMR spectral library of authentic compounds should be available. MetaboID has been built around a library of 360 unique compounds to date, including most common metabolites encountered in metabolomic studies (see Table 1).

A library of authentic compound data is critical in undertaking complex mixture analyses. The library provided with MetaboID is composed of publicly available compound data (HMDB and BMRB) including compound names and common alternative names, CAS number, KEGG compound identification [21], non-standard NMR sample conditions (standard conditions were considered to be aqueous at pH 7), chemical shift peak lists, authentic 1D ^{1}H NMR spectra, and compound structures. From this database, the user has the ability to search for particular chemical shift values, groups of chemical shift values, or compound names with optional display of chemical shift positions, authentic 1D ^{1}H spectra, and/or the chemical structure. Editing or creating a user-specific library to be used with MetaboID is straightforward, so long as the structure of the main compound chemical shift Excel file is maintained (see the User Manual, Supplemental file 2).

Assignment generally begins with library searches based on selected signal chemical shift values, followed by successive elimination of potential candidate compounds through authentic and experimental resonance comparisons (i.e. chemical shift, multiplicity and relative intensity). Using MetaboID, a coarse search may begin with a particular chemical shift value using the Peak Searching interface that is designed to efficiently identify candidate compounds at low computational cost. This is accomplished via searching and displaying chemical shift values instead of the complete high-resolution ^{1}H NMR spectra, which is the computationally expensive alternative. Refinement of the candidate compound list may be done by examination of the chemical shift positions or by loading the full resolution authentic spectra. Once a full resolution spectrum is imported into MetaboID, normalization to a total intensity of 1 is performed and the intensity within the region between 4.5 and 6.0 ppm is set to 0 (water and urea region) to eliminate the dominant solvent signal. The solvent region may be defined by the user through the File Management interface (see User Manual, Supplemental file 2).

Using the user-friendly Authentic Spectral Overlay interface, confidence is then gained by overlaying the authentic spectra with the experimental spectrum. At this stage of the assignment process, a short-list of candidate compounds for a given unknown resonance will be generated and a select series of authentic spike-in experiments can be performed to aid in the assignment verification. For convenience, an Authentic Spike Analysis interface has been included for simple comparison of experimental spectra pre- and post-authentic compound spiking. Alternatively, this interface could be utilized for a comparison of any two experimental spectra, for example an original and a selective-TOCSY spectrum. Together, the integrated toolbox of MetaboID is meant to centralize the 1D ^{1}H resonance assignment process by providing access to large chemical shift libraries to enable a straightforward, intuitive means of spectral comparison.