Anal Bioanal Chem (2014) 406:7967–7976
DOI 10.1007/s00216-014-8225-6

RESEARCH PAPER

Dolphin: a tool for automatic targeted metabolite profiling using 1D and 2D ^{1}H-NMR data

Josep Gómez · Jesús Brezmes · Roger Mallol · Miguel A. Rodríguez · Maria Vinaixa · Reza M. Salek · Xavier Correig · Nicolau Cañellas

Received: 30 July 2014 / Revised: 25 September 2014 / Accepted: 30 September 2014 / Published online: 5 November 2014

Abstract One of the main challenges in nuclear magnetic resonance (NMR) metabolomics is to obtain valuable metabolic information from large datasets of raw NMR spectra in a high throughput, automatic, and reproducible way. To date, established software packages used to match and quantify metabolites in NMR spectra remain mostly manually operated, leading to low resolution results and subject to inconsistencies not attributable to the NMR technique itself. Here, we introduce a new software package, called Dolphin, able to automatically quantify a set of target metabolites in multiple sample measurements using an approach based on 1D and 2D NMR techniques to overcome the inherent limitations of 1D ^{1}H-NMR spectra in metabolomics. Dolphin takes advantage of the 2D J-resolved NMR spectroscopy signal dispersion to avoid inconsistencies in signal position detection, enhancing the reliability and confidence in metabolite matching. Furthermore, in order to improve accuracy in quantification, Dolphin uses 2D NMR spectra to obtain additional information on all neighboring signals surrounding the target metabolite. We have compared the targeted profiling results of Dolphin, recorded from standard biological mixtures, with those of two well established approaches in NMR metabolomics. Overall, Dolphin produced more accurate results with the added advantage of being a fully automated and high throughput processing package.

Keywords Automatic metabolic profiling · Metabolite target analysis · ^{1}H NMR · 2D-JRES · Metabolomics tool

Introduction

The rapid growth of metabolomics, one of the "omic" science research fields [1], can be attributed to the fact that characterizing the metabolome allows the determination of what is actually happening inside a given biological system at specific time points.

In the context of metabolomics, nuclear magnetic resonance (NMR) spectroscopy provides an efficient tool for the analysis of biofluids and tissue extracts because of its rapid, highly accurate, nondestructive, and quantitative features [1–3]. Such advantages make it possible to get closer to the promise of high-throughput identification and quantification of large numbers of metabolites in large sample sets [4].

Currently, extracting metabolic information from NMR spectra is one of the most difficult challenges in NMR metabolomics. This is because in a ^{1}H NMR spectrum, each metabolite may contribute several individual resonances that often overlap with other signals from different metabolites. Moreover, the biofluid matrix in which we measure the metabolites (e.g., serum or urine) can affect the absolute and relative position of each peak (the so-called matrix effect). Therefore, the final shape of the resulting spectrum varies depending on the metabolite's proportions, which makes

Electronic supplementary material The online version of this article (doi:10.1007/s00216-014-8225-6) contains supplementary material, which is available to authorized users.

J. Gómez (✉) · J. Brezmes · R. Mallol · M. A. Rodríguez · M. Vinaixa · X. Correig · N. Cañellas
Metabolomics Platform, IISPV, Universitat Rovira i Virgili, Campus Sescelades, Carretera de Valls, s/n, 43007 Tarragona, Spain
e-mail: josep.gomez@urv.cat

J. Gómez · J. Brezmes · R. Mallol · M. A. Rodríguez · M. Vinaixa · X. Correig · N. Cañellas
CIBERDEM, Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders, Bonanova 69, 6th floor, 08017 Barcelona, Spain

R. M. Salek
European Bioinformatics Institute (EMBL-EBI) European Molecular Biology Laboratory, Wellcome Trust Genome Campus Hinxton, Cambridge CB10 1SD, UK

accurately deciphering every metabolite signal a daunting challenge.

The most widely employed NMR data analysis approach uses spectral binning [5–8] to overcome peak alignment and line width issues. Integrated bin regions containing important metabolites are manually annotated using metabolite database matching or by performing a literature search. However, a great degree of user expertise is needed to correlate bins to metabolites, and the reduction of the resolution associated with the method might hamper both metabolite identification and quantification precision.

With recent improvements in available software packages for NMR data analysis and processing, spectral alignment methods and curve fitting techniques are starting to replace the need for spectral binning. In bin-based methods, as in other untargeted approaches, this fitting process is performed without a reference library and the identification of signals occurs in subsequent steps. There are several methodologies available to obtain efficient deconvolutions in highly overlapped spectral regions. For example, one group of algorithms performs a non-negative matrix factorization using least squares [9] or Bayesian decompositions [10–13]. Promising results have also been reported with 'constrained-line-shape-fitting' methods using a linear combination of Lorentzian and Gaussian functions to model the previously detected peaks [14–17].

When using a targeted metabolite fitting approach, the NMR spectrum is decomposed as a linear combination of spectra from a reference library [18, 19], allowing the simultaneous identification and quantification of the metabolites by properly matching the reference peaks of the library to those found in each sample. However, this is a mainly a manual process , and most of the existing methods rely on user expertise to manually assign metabolites, making the results highly unreproducible [20], particularly in highly overlapped regions or when resonances from proteins, lipids, lipoproteins, or other metabolites are present [21].

Although new approaches try to identify and quantify in an automatic or semi-automatic way, they are still a long way from being truly automated. Most of the existing methods can identify and fit metabolites contained in their libraries, but they do not properly process the possible superposition of "unknown" signals not found in the libraries; these unmatched peaks represent a source of noise and often lead to inaccurate fitting results. Therefore, the challenges for high throughput metabolite identification and quantification remain an open topic for research until improvements in the way the available spectral information is used can help to solve these challenges.

Herein, we describe a new software package called Dolphin. Our main goal is to find and integrate additional information obtained from two-dimensional j-resolved NMR spectroscopy to improve the line-shape fitting algorithm, considering not only those signals present in its library, but also those unknown signals visualized in the 2D spectrum. In this way, Dolphin carries out the automatic quantification of a fixed target set of metabolites in a collection of experimental samples obtaining high throughput target metabolite profiling with minimum intervention from the operator.

In order to evaluate our approach, we used two different datasets to test Dolphin and compared them with the results from existing packages. The first dataset contained a pool mixture of standard compounds at various concentrations, whereas the second dataset was from a collection of 24 rat liver aqueous extracts. We then compared Dolphin's performance against other NMR approaches, namely an integration-based quantification method and the commercial package Chenomx NMR Suite 7.0, which is a line-fitting and library-based approach.

Materials and methods

Dolphin's approach and methodology

Dolphin utilizes a 1D line-shape fitting approach supported by a 2D complementary source of spectral data. There are a wide range of 2D NMR pulse sequences that address component separation in biological samples, such as COSY-TOCSY-HSQC [22], each one with its particular benefits when performing metabolite profiling. The common drawback of most of these 2D pulse sequences is that they require a lot of machine time to be acquired, a situation that can severely limit the throughput analysis of huge data sets. Two dimensional J-resolved (2D JRES) NMR spectroscopy [23] can acquire a second-dimension spectrum of a metabolite mixture with relatively little overlapping of signals. This approach separates the effects of chemical shift and J-coupling into independent dimensions [24, 25]. The drawback is that J-coupling can only be quantified for narrow signals, and so the use of Dolphin is limited to the automatic profiling of low molecular weight metabolites in samples free from broad resonances. This is less problematic while working with aqueous extracts of biofluids and tissues, since these samples are free of broad resonances arising from high molecular weight macromolecules. Otherwise, if it is not possible to work with extracts, the dominating broad resonances and background signals can be reduced using different NMR pulse sequences and editing techniques based on NMR relaxation times [26].

Dolphin's flowchart is represented in Fig. 1. The workflow starts by importing the NMR experiments and selecting the library to be used (see the next section). The line-shape-fitting algorithm works optimally when we have a precise referencing between 1D and 2D JRES spectra. By default, this referencing uses the alpha-glucose doublet [27], but it can be changed to the TSP signal or any other peak of interest (in this work we used the default referencing option). In addition,

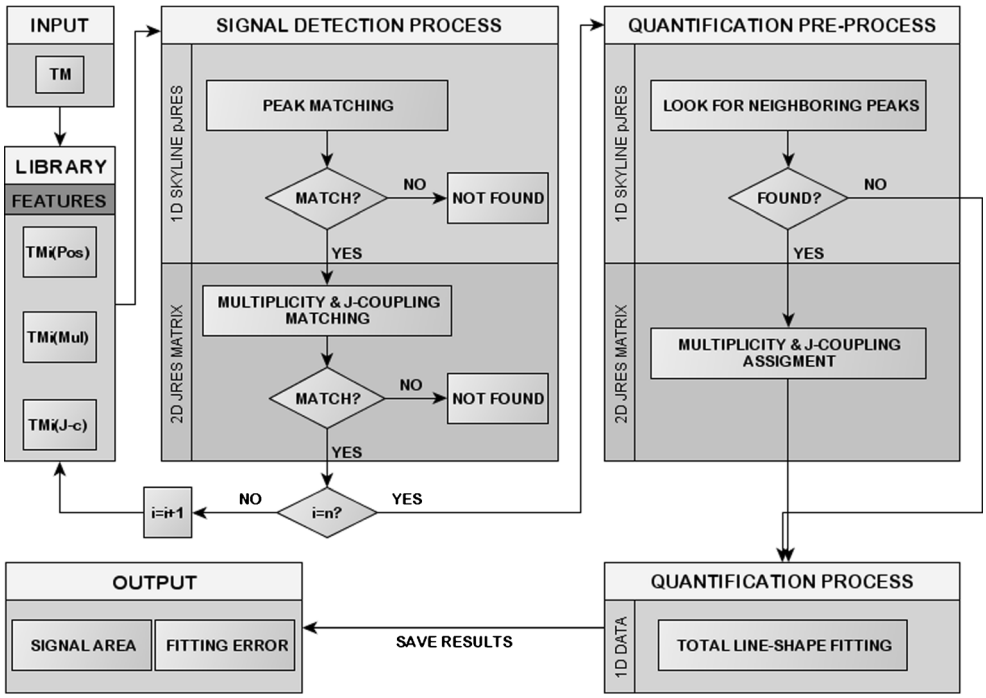

Fig. 1 Dolphin flowchart. Here we show the process of matching and quantification of a metabolite *TM* with *n* signals represented as *Mi*, where the *i* value goes from *1* to *n*. It is important to note that quantification is only performed after a validated match

during the data import process, the user can choose the data normalization method (selected peak, total area, or none). This step is especially important if quantification results have to be reported.

After these initial preprocessing steps, the user can run the automatic quantification of those metabolites selected in the user library. The process works in a two-stage approach detailed in the next two sections. First, in order to validate the presence of a target metabolite in the sample, the software checks for the metabolite pattern in the JRES spectrum. In a second phase, if the metabolite is detected, its abundance is quantified using one of its signals (peaks). The quantification is done by means of a line-shape-fitting approach in the 1D spectrum (see Figure S1 in the Supplementary Material for a more detailed explanation of the flowchart).

When all the steps are finished, the program displays the measured area, the chemical shift of the signal used for quantification, and the fitting error for the region containing the signal of interest in the selected sample. This fitting error will give us information on whether the modeling of known and unknown metabolites has been successful or not, giving an indicative value of the confidence in measurement. Finally, the results for the complete dataset can be exported as an Excel file for further analysis.

As shown in Fig. 2, the user can graphically and visually verify the fitting performance. Dolphin's user interface plots four figures to show the relevant information for the selected sample. The upper graph contains the 1D spectrum plot of the fitting region; the middle segment contains the pJRES spectrum, where all signals are unified and displayed as a unique singlet, reducing congestions and overlapping; finally, the bottom figure contains the 2D JRES matrix contour plot, and the lower-right figure contains the sum projection of the vertical cut of the 2D JRES spectrum.

All the functions are programmed under the matrix calculation platform MATLAB (ver. 7.5.0; The Mathworks, Inc., Natick, MA, USA). Dolphin is available by request, but a compiled demo version of the package can be downloaded following the instructions provided in the Supplementary Material.

Automatic matching of target metabolites using 2D-JRES

In order to check for the presence of each target metabolite in a sample, Dolphin looks for a subset of its spectral pattern in the 2D-JRES NMR spectra. This characteristic subset of spectral resonances is stored in a proprietary library edited specifically for the kind of sample under study. The metabolite information about peak positions, signal multiplicity, and J-coupling values is included in the library, and is based on public domain databases such as the Human Metabolome Database (HMDB) [28], the Birmingham Metabolite Library [29], and the BioMagResBank (BMRB) [30], or commercial packages such as Chenomx NMR Suite (Chenomx Inc., Edmonton, Alberta, Canada) and Bioref Amix (Bruker, GmbH, Silberstreifen, Rheinstetten, Germany).

The spectral pattern of a metabolite can be very complex and, in some cases, we can observe more than 10 different types of signals for a metabolite (e.g., glucose); however, most metabolites have a fixed pattern of signals that can be used to identify a metabolite in a mixture using just one, two, or three

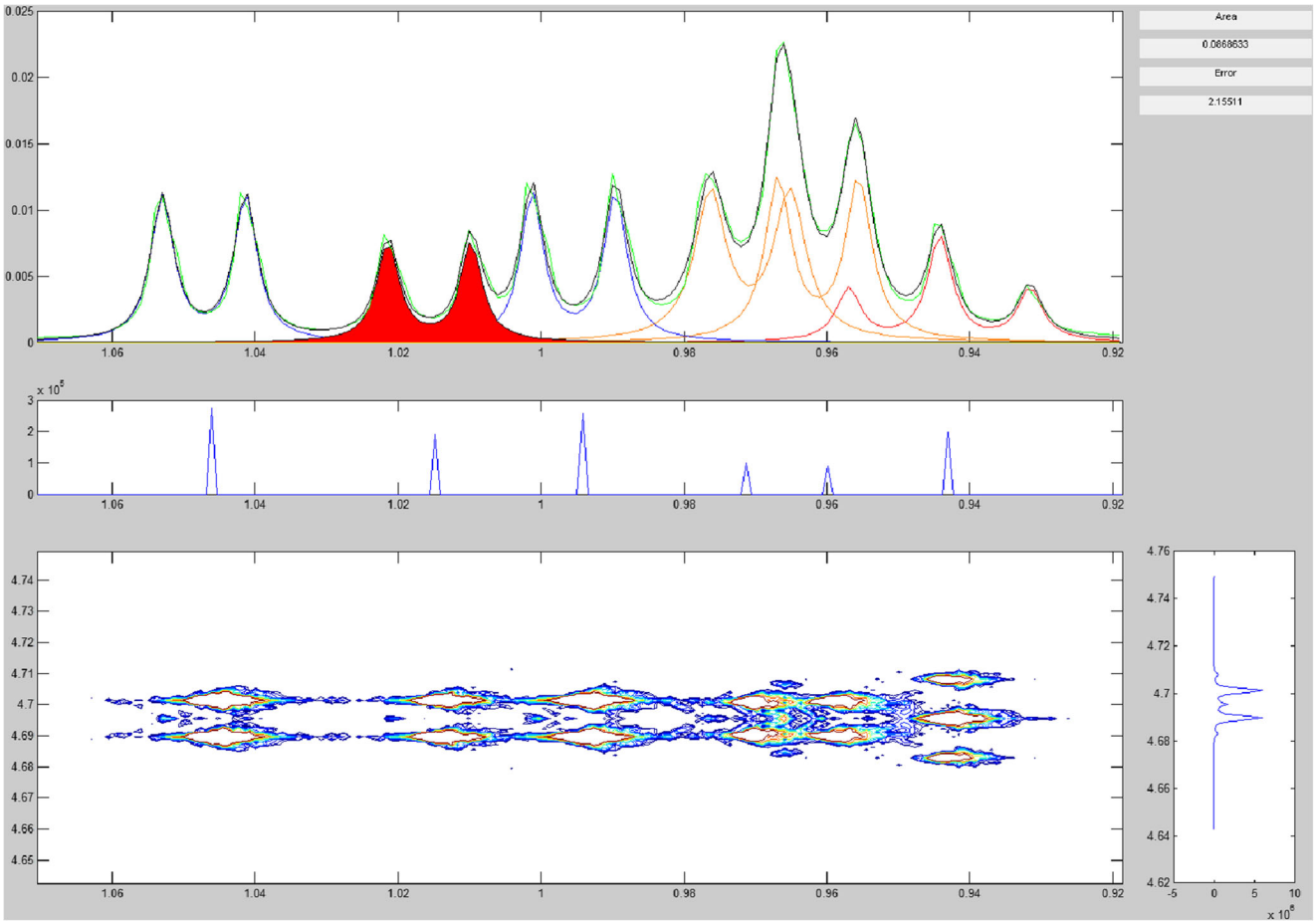

Fig. 2 Image of Dolphin's isoleucine quantification in STN1. We can see here how Dolphin uses its automatic signal annotation in 2D data to deconvolute our target signal (second doublet from the left of the image, near 1.02 ppm) in the 1D spectra successfully. Dolphin detects five more signals that may affect our final goal, without previous knowledge of what those signals are, just taking them as unknowns

of these features since they are unique for that metabolite and have a fixed location in the spectrum.

The information on peak positions (or metabolite chemical shifts) is used in the first step of the metabolite match process. In this step, Dolphin matches the chemical shift of a signal annotated in the library with a peak in the processed 1D SEM-skyline pJRES spectrum (the 2D-JRES data was processed using a combined sine-bell-exponential (SEM) function, and a combination of SEM-skyline was applied for the 1D projection of the JRES data (pJRES) [31]). If a matching peak is present in the 2D JRES datasets, the algorithm tries to match the multiplicity and J-coupling of this signal annotated in the library with that of the vertical cut in the same peak position in the 2D JRES matrix. This is one of the main advantages of Dolphin, since all other approaches use identifications based only on 1D libraries that can lead to misidentifications.

Automatic quantification of peaks using line-shape fitting

Dolphin continues with the quantification process only if location, multiplicity, and J-coupling values match those annotated in the library for a given metabolite. The quantification step is carried out by measuring the area under one of the signals of the metabolite in question. The library defines which signal will be used in the peak quantification step, and a line-shape fitting algorithm based on the sum of Lorentzian functions modulated with different Gaussian proportions (Voigt profile with up to 10 % of Gaussian shape) is applied to discriminate the area of our target signal from other interferences sharing the same spectral region.

To achieve an optimal fitting, it is important to take into account the proportions between Lorentzians conforming the multiplet of each signal. Often, the presence of unknown signals surrounding or overlapping the signal of interest makes it difficult to accurately quantify a metabolite, especially in congested regions. In Dolphin, the fitting process is improved by using more detailed information from the position and the multiplicity of all the resonances surrounding the signal of interest as well. The processed 1D SEM-skyline pJRES spectrum gives the position of all the resonances surrounding our signal of interest, whereas the vertical cuts at these resonances indicate the multiplicity and J-coupling of

all the signals surrounding the signal. In this way, the software is aware not only of the resonances included in the metabolite library but also of the presence of potential unknown interfering signals that are overlapping with the signal in question; taking this information into account results in a more accurate fit.

The implemented line-shape fitting approach is constrained using well-known intensity relationships within the singlets, doublets, and triplets. Multiplets, however, are fitted according to a sum of singlets. When the position, multiplicity, and J-coupling values are calculated for all signals present in the region to fit, Dolphin is ready to proceed with the quantification step. Finally, both fitting errors and intensity values for all targeted metabolite quantifications are returned in an excel format file.

Samples preparation and NMR data acquisition

As a proof of concept, we have built a library of 15 hydrophilic metabolites that frequently appear in liver aqueous extracts (isoleucine, leucine, valine, β-hydroxybutyrate, alanine, acetate, succinate, creatine, glucose, fumarate, tyrosine, phenylalanine, uridine, formate, and adenosine). We tested Dolphin quantifying these target metabolites in two different real sample sets. The first test set comprised eight standard mixtures (labeled from STN1 to STN8) of pure metabolites, at different concentration ranges (see Supplementary Material, Table S1). The second dataset contained the NMR measurements of a set of 24 rat liver aqueous extracts (EXP1 to EXP24).

Three additional reference mixtures (REF1 to REF3) were also prepared for calibration purposes. Each one of the mixtures contained five of the library compounds, chosen to ensure a spectrum free of overlapped signals so that all three methods could obtain accurate quantification using their individual units of measure. In this manner, we were able to normalize each approach for comparison purposes (see Supplementary Material, TableS2).

Stock solutions for every single compound were prepared separately in 1 mL tubes with D_2O solution containing 0.6753 mM trimethylsilylpropionic acid (TSP), and the same D_2O solution was added to the final mixtures to obtain 700 μL for the NMR measurements. The liver extraction protocol was performed as described previously in [32].

The 1D nuclear Overhauser effect spectroscopy with a spoil gradient (NOESY) was used to record 1D ^{1}H NMR spectra using a 600.2 MHz frequency Avance III-600 Bruker spectrometer (Bruker, Germany) equipped with an inverse TCI 5 mm cryoprobe. A total of 256 transients were collected across a 12 kHz spectral width at 300 K into 64 k data points, and exponential line broadening of 0.3 Hz was applied before Fourier transformation. A recycling delay time of 8 s was applied between scans to ensure correct quantification. Two-dimensional J-RES spectra were acquired using the standard pulse sequence with pre-saturation and spoil gradients (jresgpprqf). Spectral widths of 7.5 kHz in the F2 dimension and 64 Hz in the F1 dimension were used; eight scans per increment and 32 increments were accumulated into 8 k data points. Zero-filled into 16 k points and linear prediction to 128 increments were applied prior to Fourier transformation (FT) followed by tilting and symmetrisation. The acquired NMR spectra were phased and baseline-corrected using TopSpin software (ver. 2.1; Bruker BioSpin GmbH, Silberstreifen, Rheinstetten, Germany).

Comparison of methods

We compared the automatic results obtained by Dolphin with two other approaches: one, a widely used integration-based quantification method (in our case we use that mode in Amix[1]) and second, using the commercial package Chenomx NMR Suite 7.0. Chenomx is a well-known commercial program that incorporates a large list of compounds in its library with the specific resonance signatures for each stored metabolite along the NMR spectra. Although there is a semi-automatic fitting mode available, the software is mainly manually operated and requires expertise to adjust the metabolite peaks correctly. Therefore, the process can be considered as "user supervised."

Area integration is a standard quantification method, easy to implement with homemade tools and often included in commercial packages. It allows the quantification of metabolites by means of measuring the area in the spectrum associated with known metabolites. However, quantification of crowded regions using this approach might be challenging because of the potential unknown contribution from overlapping peaks and unidentified signals in the same segment.

Chenomx gives quantification results in terms of the absolute concentration of the metabolite, but Amix and Dolphin report their results in terms of non-normalized area units. Therefore, in order to compare the different values, we have to convert each method's own quantification units into absolute concentrations of the metabolite. Assuming zero error when quantifying clean and isolated signals in the reference mixtures, for each quantification signal of each metabolite we calculate a normalization factor as described in Eq. 1,

$$FMm = \frac{RCm}{QMm} \quad (1)$$

[1] Amix software has other quantification approaches, but in this work we only performed a comparison with its integration based quantification option.

where FM_m is the normalization factor of the method M for the metabolite m, QM_m are the quantification units of the method M quantifying the metabolite m, and RC_m is the real concentration of the metabolite m (all normalization factors are included in Table S3 in the Supplementary Material).

For the STN set, as we know the real concentration of each metabolite in the sample, we can evaluate the accuracy of each one of the methods (M) on every sample (i) for each one of the metabolites (m). Eq. 2 gives the relative error of quantification of a metabolite m in a sample i (EM_{im}):

$$EM_{im}(\%) = \frac{|(QM_{im} - C_{im})|}{C_{im}} \cdot 100 \quad (2)$$

where QM_{im} is the measured concentration of the metabolite m in the sample $_i$ of the method M, and C_{im} is the real absolute concentration of the metabolite m in the STN sample i (the final normalized values and quantification errors are included in the Supplementary Material as Tables S4 and S5) .

In the second experimental data sets used, the real concentration values in biological samples were not known. Therefore, we evaluated the differences of the three methods by comparing their results against each other in pairs, using the following relative difference estimation:

$$DM^{ab}_{im}(\%) = \frac{\left|\left(QM^{a}_{im} - QM^{b}_{im}\right)\right|}{QM^{b}_{im}} \cdot 100 \quad (3)$$

where DM^{ab}_{im} is the relative difference quantifying a metabolite m in the sample i between method a and method b, QM^{a}_{im} is the concentration of this metabolite m estimated by the method a in the sample i, and QM^{b}_{im} is the concentration of the metabolite m estimated by the method b in the sample i (see Tables S6–S8 in the Supplementary Material). Final concentrations of each metabolite in the biological dataset were obtained converting each method values using the normalization factor described in Eq. 1 (all normalized values are available in Table S9 in the Supplementary Material).

Results

Targeted profiling in standard pooled samples

All target metabolites were matched and quantified in the standard samples by all three approaches. In most cases, the quantifications were in good agreement with the known metabolite concentrations used to prepare the samples. When evaluating the results, it is important to distinguish those metabolites that can be quantified by an isolated signal from those that can be quantified by a signal that is present in overlapping region. For example, the region from 1.07 to 0.91 ppm, concentrates 11 peaks, corresponding to six signals (five doublets and 1 triplet) from three different metabolites (valine, isoleucine, and leucine). The fitting errors by methods used here in this region are very different and depend on the metabolite. The results from Dolphin were in very good agreement with actual concentrations in the standard pooled samples, showing very little quantification error for all above metabolites, even for the case of overlapped signals.

The results are summarized in Table 1, showing a maximum mean relative error of quantification less than 10 % for all those metabolites that are quantified by an isolated signal. In this case, none of the methods is better than the other two in precision terms, but Dolphin has the advantage of full automation. Focusing on the quantification of areas within the overlapped signals, both Dolphin and Chenomx performed similarly with less than 10 % of relative error. However, the integration approach performed worse in the overlapped signals region, with mean relative errors between 10 % and 20 %, with isoleucine as high as 46 %. See radar plot in Fig. 3 to graphically evaluate relative errors.

Targeted profiling in biological samples

The second dataset was composed by 24 rat liver aqueous extracts, (EXP1 to EXP24). In this case, we had no previous information about their metabolite content and, therefore, we did not know the correct identifications and quantification values. Absolute quantification errors could not be computed in this case and so we used Eq. 3 to perform comparisons by

Table 1 Quantification errors (in %) of the three methods in the STN mixtures. The table contains both the mean relative error (ME in %) and the standard deviation (SD in %) for every metabolite in the STN set

Method	Chenomx		Integration		Dolphin		All	
Statistic	ME	SD	ME	SD	ME	SD	ME	SD
Isoleucine	5	5	46	18	3	2	18	24
Leucine	9	5	12	7	8	4	10	2
Valine	5	2	5	7	4	2	5	1
DL-β-hydroxybutirate	5	3	4	3	4	3	4	0
Alanine	6	4	14	5	5	3	8	5
Acetate	3	2	6	4	4	4	4	1
Succinate	7	4	20	8	6	6	11	7
Creatine	5	4	13	5	7	1	8	4
Glucose	8	7	7	7	4	3	6	2
Fumarate	3	1	9	7	3	3	5	3
Tyrosine	6	3	4	5	2	2	4	2
Phenylalanine	6	5	5	5	5	4	5	0
Uridine	6	6	5	4	2	1	5	2
Adenosine	3	3	7	6	5	4	5	2
Formate	10	6	6	7	4	2	6	3
All	6	4	11	7	4	3	7	4

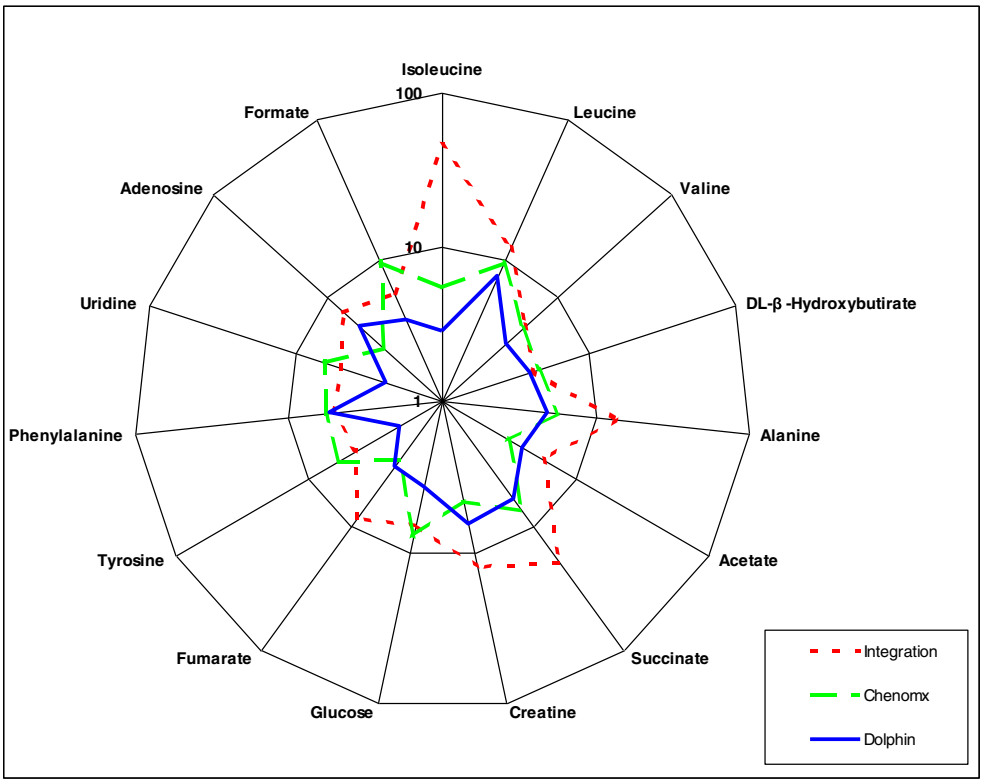

Fig. 3 Radar plot of the quantification errors (in %) in the STN set. Here we can check graphically the performance of the three approaches quantifying metabolites in the STN samples. The graph shows the mean quantification error values (in %) of Chenomx, Dolphin, and the standard integration approach on a logarithmic scale

pairing methods. The main goal of this second study was to evaluate Dolphin's performance in biological samples, where overlapping and base lining problems are more pronounced. All 15 target metabolites (those included in this Dolphin library) were matched and quantified by all three approaches in all the 24 biological samples. As expected, the quantification differences between methods (summarized in Table 2) were higher than in the case of the standard mixtures. There was a big difference comparing integration vs. the two fitting approaches, but the differences were not so important when comparing between the two fitting approaches. Although the mean difference in quantification of the whole set of metabolites between Chenomx and Dolphin is around 7 %, the same difference rises to 18 % and 23 % when Integration is compared with Dolphin and Chenomx, respectively. While only adenosine presents the same difference value between methods (4 %), 12 metabolites present lower values when the two fitting approaches are compared. In fact, only in the case of glucose, the difference value between the Integration approach and Dolphin (10 %) is lower than the difference between Dolphin and Chenomx (23 %). When taking Chenomx as reference against Integration, only eight of the whole set present quantification differences under 20 %, highlighting a difference of almost 80 % in the case of creatine. The comparisons between Dolphin and the Integration method followed a similar pattern, but with lower difference values, with the exception of phenylalanine.

Discussion

The differences in quantification accuracies for the three options are directly related to the approaches used and in the particular signal characteristics in the regions of interest. Area integration is a good approach when quantifying isolated peaks with intensities clearly higher than background noise; on the other hand, this approach is less accurate when used in congested areas or when weak signals need to be measured. These considerations explain the method accuracy in the case of metabolites that can be quantified using isolated signals. In fact, valine and phenylalanine present the lowest errors in quantification for the STN samples using the metabolite integration approach (see Table 1). On the other hand, and due to the shape constraints of its spectral library, Chenomx is the tool that usually presents higher quantifying errors for these metabolites. Chenomx uses reference deconvolutions to adjust the shim for the spectra in its library for all metabolites, but it is only effective in removing line-shape distortions that affect all signals in the spectrum equally [33]; in addition, there are other factors affecting the shape of he signals [34, 35], especially if they are unusually intense and, consequently, the Chenomx library signals rarely fit accurately such peaks present in the real spectrum. This lack of shape flexibility makes it difficult to achieve a precise quantification and increases error.

Table 2 Quantification differences (in %) of each method vs. the others by pairs in the EXP dataset. The table presents the mean relative difference (MD in %) and the standard deviation (SD in %) for every metabolite in the EXP set

Method	Chenomx vs. Dolphin		Integration vs. Dolphin		Chenomx vs. Integration	
Statistic	MD	SD	MD	SD	MD	SD
Isoleucine	5	5	18	13	25	21
Leucine	8	8	29	15	35	27
Valine	5	3	9	5	14	9
DL-β-hydroxybutirate	3	4	12	9	13	13
Alanine	3	5	16	5	16	3
Acetate	5	4	19	10	22	12
Succinate	6	5	6	6	11	8
Creatine	4	3	45	12	79	32
Glucose	23	5	10	5	36	7
Fumarate	7	6	13	11	14	7
Tyrosine	8	5	20	13	21	8
Phenylalanine	10	5	39	15	35	7
Uridine	4	3	7	6	8	8
Adenosine	4	2	4	3	4	3
Formate	8	5	19	9	19	4
All	7	5	18	9	23	11

Dolphin, on the other hand, quantifies using a line-shape fitting algorithm based on the addition of Lorentzian signals modulated with a different Gaussian proportion (to correct such problems associated with magnetic field inhomogeneity while acquiring data). The results show that this option works better in metabolite quantification; in fact, Dolphin presents even lower error values compared with the two software tools tested for most of the cases applied (see Table 1).

The problem is completely different in the case of areas of the spectra with moderate to high signal congestion. It is in these regions where both Dolphin and Chenomx software present better results than the integration-based approach. Chenomx, through its manually adjusted fit, allows a trained user to solve the fitting puzzle really well, even though its library signals rarely fit in an accurate manner for those peaks present in the real spectrum. However, this manual approach is time-consuming and prone to subjective quantification. For example, in the branched chain amino acids area (from 1.07 to 0.91 ppm) the program presents a list of 21 possible compounds that may be adjusted for each sample. On the other hand, Dolphin's line-shape fitting approach does not require manual supervision since all the necessary information to deconvolute the signal of interest is obtained from the 2D JRES analysis. Figure 1 presents the Dolphin analysis of the branched chain amino acids region of one of the STN mixtures (Figures S2 and S3 in the Supplementary Material show Dolphin's performance in the EXP samples). In this example, we fit the doublet near 1.00 ppm coming from isoleucine to quantify this metabolite concentration, and the area that corresponds to the signal is adjusted taking into account all neighboring signals as unknowns (five doublets and one triplet, which were adjusted automatically with our line-shape fitting algorithm). Therefore, Dolphin was more accurate in these situations, even without human intervention. This enables the user to calculate the concentrations of numerous metabolites in an automated, high-throughput, and accurate manner.

The quantifications carried out by the peak integration method produced several inaccuracies. The main problem of this approach was associated with the integration technique, which measures the area under the curve assigned to a metabolite, ignoring the interferences generated by signals from adjacent overlapped metabolites. The quantification of isoleucine in the STN set (see Table 1) is a representative example of such a problem; the measured value is clearly affected by the amount of valine, which has a signal overlapped with our target signal (see Supplementary Material, Figure S4)

It is important to note that Dolphin performed well while quantifying metabolites in the STN dataset compared with the other two methods. This fact highlights the combined accuracy of the Chenomx line-fitting approach for congested signals regions and the efficient area integration for isolated peaks that are combined in the Dolphin software package.

Dolphin's quantification accuracy is directly related to the quality of the spectra. In complex biological matrices, especially in the presence of inhomogeneities in the samples, some fitting errors arise, mostly produced by spectral shapes that are quite far from being Lorentzian (the addition of 10 % of Gaussian shape does not resolve this issue totally). In the Supplementary Material we present the comparison of the three methods in a spike-in experiment in human urine samples.

Conclusions

This work presents a new NMR data processing tool (Dolphin) that assists with identification and quantification of target metabolites in an automatic manner once datasets are imported with the import parameters manually entered by the user. The tool has been compared against two popular NMR metabolite quantification approaches. Dolphin quantifications were more accurate compared with those obtained with bin integration approaches, and in good agreement with the manual line-shape fitting solutions. Our software is based on both the 1D and 2D spectra for the same sample, matching data to a library for metabolite identification, and automatically fitting peaks for quantification purposes. One advantage of such workflow is the increase in identification confidence. Moreover, this approach takes into account all neighboring signals, even if they are unknown, which leads to a more accurate quantification. This is especially relevant and useful in very congested regions, where current automatic tools and methods mostly fail. The maximum number of automatically identifiable and quantifiable metabolites is strongly related to the 2D-JRES spectra resolution and the type of biofluid matrix under study. Nevertheless, it is important to remark that Dolphin has been tested quantifying up to 60 different metabolites in different biological samples (urine, plasma, serum, tissue extracts, and cellular cultures) and the limitation to 16 metabolites in the library is self-imposed in this case study. Future versions of Dolphin will incorporate better filters and spectral processing routines in order to enlarge the target metabolite list without increasing the processing time requirements.

Acknowledgments The authors acknowledge CIBER de Diabetes y Enfermedades Metabólicas, an initiative of ISCIII (Ministerio de Ciencia e Innovación), and MINECO grant TEC2012-31074 for partially funding this work.

References

1. Nicholson JK, Lindon JC, Holmes E (1999) Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29(11):1181–1189
2. Chen HW, Pan ZZ, Talaty N, Raftery D, Cooks RG (2006) Combining desorption electrospray ionization mass spectrometry and nuclear magnetic resonance for differential metabolomics without sample preparation. Rapid Commun Mass Spectrom 20(10): 1577–1584
3. Beckonert O, Keun HC, Ebbels TMD, Bundy JG, Holmes E, Lindon JC, Nicholson JK (2007) Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum, and tissue extracts. Nat Protoc 2(11):2692–2703
4. Fiehn O (2002) Metabolomics—the link between genotypes and phenotypes. Plant Mol Biol 48(1/2):155–171
5. De Meyer T, Sinnaeve D, Van Gasse B, Tsiporkova E, Rietzschel ER, De Buyzere ML, Gillebert TC, Bekaert S, Martins JC, Van Criekinge W (2008) NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. Anal Chem 80(10):3783–3790
6. Alves AC, Rantalainen M, Holmes E, Nicholson JK, Ebbels TMD (2009) Analytic properties of statistical total correlation spectroscopy based information recovery in H-1 NMR metabolic data sets. Anal Chem 81(6):2075–2084
7. Anderson PE, Reo NV, DelRaso NJ, Doom TE, Raymer ML (2008) Gaussian binning: a new kernel-based method for processing NMR spectroscopic data for metabolomics. Metabolomics 4(3):261–272
8. Jacob D, Deborde C, Moing A (2013) An efficient spectra processing method for metabolite identification from H-1-NMR metabolomics data. Anal Bioanal Chem 405(15):5049–5061
9. Eads CD, Furnish CM, Noda I, Juhlin KD, Cooper DA, Morrall SW (2004) Molecular factor analysis applied to collections of NMR spectra. Anal Chem 76(7):1982–1990
10. Ochs MF, Stoyanova RS, Arias-Mendoza F, Brown TR (1999) A new method for spectral decomposition using a bilinear Bayesian approach. J Magn Reson 137(1):161–176
11. Stoyanova R, Nicholson JK, Lindon JC, Brown TR (2004) Sample classification based on Bayesian spectral decomposition of metabonomic NMR data sets. Anal Chem 76(13):3666–3674
12. Hao J, Astle W, De Iorio M, Ebbels TMD (2012) BATMAN—an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. Bioinformatics 28(15):2088–2090
13. Hao J, Liebeke M, Astle W, De Iorio M, Bundy JG, Ebbels TMD (2014) Bayesian deconvolution and quantification of metabolites in complex 1D NMR spectra using BATMAN. Nat Protoc 9(6):1416–1427
14. Laatikainen R, Niemitz M, Malaisse WJ, Biesemans M, Willem R (1996) A computational strategy for the deconvolution of NMR spectra with multiplet structures and constraints: Analysis of overlapping C-13-H-2 multiplets of C-13 enriched metabolites from cell suspensions incubated in deuterated media. Magn Reson Med 36(3): 359–365
15. Soininen P, Haarala J, Vepsalainen J, Niemitz M, Laatikainen R (2005) Strategies for organic impurity quantification by H-1 NMR spectroscopy: constrained total-line-shape fitting. Anal Chim Acta 542(2):178–185
16. Jukarainen NM, Korhonen SP, Laakso MP, Korolainen MA, Niemitz M, Soininen PP, Tuppurainen K, Vepsalainen J, Pirttila T, Laatikainen R (2008) Quantification of H-1 NMR spectra of human cerebrospinal fluid: a protocol based on constrained total-line-shape analysis. Metabolomics 4(2):150–160
17. Mihaleva VV, Korhonen S-P, van Duynhoven J, Niemitz M, Vervoort J, Jacobs DM (2014) Automated quantum mechanical total line shape fitting model for quantitative NMR-based profiling of human serum metabolites. Anal Bioanal Chem 406(13):3091–3102
18. Wishart DS, Querengesser LMM, Lefebvre BA, Epstein NA, Greiner R, Newton JB (2001) Magnetic resonance diagnostics: a new technology for high-throughput clinical diagnostics. Clin Chem 47(10): 1918–1921
19. Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM (2006) Targeted profiling: quantitative analysis of H-1 NMR metabolomics data. Anal Chem 78(13):4430–4442
20. Tredwell GD, Behrends V, Geier FM, Liebeke M, Bundy JG (2011) Between-person comparison of metabolite fitting for NMR-based quantitative metabolomics. Anal Chem 83(22):8683–8687
21. Nicholson JK, Foxall PJD, Spraul M, Farrant RD, Lindon JC (1995) 750-MHz H-1 and H-1-C-13 NMR-Spectroscopy of human blood-plasma. Anal Chem 67(5):793–811
22. Bingol K, Bruschweiler R (2014) Multidimensional approaches to NMR-based metabolomics. Anal Chem 86(1):47–57

23. Aue WP, Karhan J, Ernst RR (1976) Homonuclear broad-band decoupling and 2-dimensional J-resolved NMR-spectroscopy. J Chem Phys 64(10):4226–4227
24. Ludwig C, Viant MR (2010) Two-dimensional J-resolved NMR spectroscopy: review of a key methodology in the metabolomics toolbox. Phytochem Anal 21(1):22–32
25. Huang Y, Cal S, Zhang Z, Chen Z (2014) High-resolution two-dimensional J-resolved NMR spectroscopy for biological systems. Biophys J 106(9):2061–2070
26. Tang HR, Wang YL, Nicholson JK, Lindon JC (2004) Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. Anal Bioch 325(2):260–272
27. Pearce JTM, Athersuch TJ, Ebbels TMD, Lindon JC, Nicholson JK, Keun HC (2008) Robust algorithms for automated chemical shift calibration of 1D H-1 NMR spectra of blood serum. Anal Chem 80(18):7158–7162
28. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A (2013) HMDB 3.0-the human metabolome database in 2013. Nucleic Acids Res 41(D1): D801–D807
29. Ludwig C, Easton JM, Lodi A, Tiziani S, Manzoor SE, Southam AD, Byrne JJ, Bishop LM, He S, Arvanitis TN, Guenther UL, Viant MR (2012) Birmingham Metabolite Library: a publicly accessible database of 1-D H-1 and 2-D H-1 J-resolved NMR spectra of authentic metabolite standards (BML-NMR). Metabolomics 8(1):8–18
30. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL (2008) BioMagResBank. Nucleic Acids Res 36:D402–D408
31. Tiziani S, Lodi A, Ludwig C, Parsons HM, Viant MR (2008) Effects of the application of different window functions and projection methods on processing of (1)H J-resolved nuclear magnetic resonance spectra for metabolomics. Analy Chim Acta 610(1):80–88
32. Vinaixa M, Rodriguez MA, Rull A, Beltran R, Blade C, Brezmes J, Canellas N, Joven J, Correig X (2010) Metabolomic assessment of the effect of dietary cholesterol in the progressive development of fatty liver disease. J Proteome Res 9(5):2527–2538
33. Vitols C, Mercier P (2006) Correcting line shapes in NMR Spectra. CHENOMX. Available at: http://www.chenomx.com/news/img/page23/ChenomxAppNote.an004.pdf. Accessed 1 Sept 2014
34. Alam TM, Alam MK, McIntyre SK, Volk DE, Neerathilingam M, Luxon BA (2009) Investigation of chemometric instrumental transfer methods for high-resolution NMR. Anal Chem 81(11):4433–4443
35. Saude EJ, Slupsky CM, Sykes BD (2006) Optimization of NMR analysis of biological fluids for quantitative accuracy. Metabolomics 2(3):113–123