Predicting Carbon Spectrum in Heteronuclear Single Quantum Coherence Spectroscopy for Online Feedback During Surgery

E. Onur Karakaslar, Baris Coskun, Hassiba Outilaft, Izzie Jacques Namer, and A. Ercument Cicek

Abstract—^{1}H High-Resolution Magic Angle Spinning (HRMAS) Nuclear Magnetic Resonance (NMR) is a reliable technology used for detecting metabolites in solid tissues. Fast response time enables guiding surgeons in real time, for detecting tumor cells that are left over in the excision cavity. However, severe overlap of spectral resonances in 1D signal often render distinguishing metabolites impossible. In that case, Heteronuclear Single Quantum Coherence Spectroscopy (HSQC) NMR is applied which can distinguish metabolites by generating 2D spectra (^{1}H-^{13}C). Unfortunately, this analysis requires much longer time and prohibits real time analysis. Thus, obtaining 2D spectrum fast has major implications in medicine. In this study, we show that using multiple multivariate regression and statistical total correlation spectroscopy, we can learn the relation between the ^{1}H and ^{13}C dimensions. Learning is possible with small sample sizes and without the need for performing the HSQC analysis, we can predict the ^{13}C dimension by just performing ^{1}H HRMAS NMR experiment. We show on a rat model of central nervous system tissues (80 samples, 5 tissues) that our methods achieve 0.971 and 0.957 mean R^2 values, respectively. Our tests on 15 human brain tumor samples show that we can predict 104 groups of 39 metabolites with 97 percent accuracy. Finally, we show that we can predict the presence of a drug resistant tumor biomarker (creatine) despite obstructed signal in ^{1}H dimension. In practice, this information can provide valuable feedback to the surgeon to further resect the cavity to avoid potential recurrence.

Index Terms—Metabolomics, HRMAS NMR, HSQC NMR

1 Introduction

Metabolomics is a powerful omics platform, which reflect a snapshot of the state of the cell and provides the most direct cues about the phenotype, as it is the highest layer in the hierarchy of the omics. High Resolution Magic Angle Spinning (HRMAS) Nuclear Magnetic Resonance (NMR) spectroscopy is a technology that can efficiently detect and quantify metabolites in solid tissues [1]. HRMAS-NMR does not need any chemical extraction procedure, which is a must for MS technologies and liquid-state NMR. Thus, it is frequently used in biopsy analyses and provides high resolution [2], [3]. Sample preparation is fast and the results can be obtained in <20 minutes. Rapid response enables giving feedback to surgeons during an ongoing surgery. Recently, Battini et al. proposed using HRMAS-NMR for pancreatic adenocarcinoma surgeries [4]. During a surgery, even if it might seem like the tumor is completely removed, it is possible that residual tumor cells are left over in the excision cavity. Then there is the trade-off between removing healthy tissue, which risks the well being of the patient and leaving tumor cells in the body, which risks recurrence and further surgeries. In this system, the surgeon gets samples from the excision cavity for identifying possible left-over tumor cells. After HRMAS analysis, parts of the cavity that have tumor-like spectrum are reported for further resection. This pipeline is possible because the feedback is available within 20 minutes.

Even though ^{1}H is commonly used due to high sensitivity and natural abundance in samples, identification of biomarker metabolites can be impossible due to overlapping signal in ^{1}H-NMR spectrum. In that case, a second experiment called Heteronuclear Single Quantum Coherence Spectroscopy (HSQC)-NMR is performed. This analysis generates a 2D correlation plot for ^{1}H and ^{13}C spectra. However, this analysis requires around 15 hours to complete and is outside of the time frame of surgery.

There are algorithms in the literature to identify metabolites using a combination of 1D and 2D analyses [5], [6]. However these methods need both type of experiments to work on. One very widely used approach to identify metabolites is STOCSY - Statistical Total Correlation Spectroscopy [7], [8]. Using a set of independent samples, this method generates a pseudo 2D NMR spectrum for all analyzed samples that displays the correlation of the signals in two dimensions. The correlation plot is combined with O-PLS-DA to identify the compounds explaining the variation. Variants of this method have been developed for purposes like (i) assigning chemical structures, (ii) preprocessing datasets for downstream analysis, and (iii) identifying pathway relations between metabolites [9]. However, none of the above mentioned works aim at blindly predicting the outcome of a HSQC NMR experiment for a single sample, after learning the relations between two spectra from a mixed training cohort. Another approach to circumvent the time over head of 2D analysis is to accelerate the experiment via sampling [10], [11], [12], [13], [14], [15]. The main point of these algorithms is to reconstruct the signal from randomly selected acquisition points in the indirect dimensions [16]. Hoch et al. report that non-uniform sampling (NUS) based method can complete a 2D experiment 3 times faster [11] with comparable accuracy. However, this still is a long time for the duration of a surgery.

In this paper, we propose two methods to predict ^{13}C spectrum in the HSQC experiment, without performing the HSQC experiment at all. These methods are (i) performing multivariate multiple regression and (ii) repurposing STOCSY for a blind prediction of a single sample. Using a set of ^{1}H-^{13}C HSQC experiments, methods learn how each position in ^{1}H-dimension affects each position in ^{13}C-dimension. Applying these methods to a rat model of central nervous system, we show that average R^2 values of each model are 0.973 and 0.958 for regression and STOCSY, respectively. Then, using only ^{1}H HRMAS NMR for 14 human brain tissue samples and predicting their corresponding ^{13}C spectrum, we show that we can successfully identify presence and absence of 104 groups belonging 39 metabolites. Both methods achieve 97 percent accuracy in less than a second. We also show that regression model can be used to reconstruct the 2D HSQC experiment as accurately. We show that we are able to predict the presence of the creatine even though its position is overlapping with lysine in ^{1}H dimension. Creatine is an indicator of hypoxia and possibly drug resistant tumor tissue [17], [18]. Thus, our approach can make it possible to provide accurate feedback to the surgeon during the surgery even if ^{1}H HRMAS NMR results are inconclusive. Even though we experiment on ^{1}H - ^{13}C HSQC NMR dimensions in this paper, all methods can be used with any other 2D spectrum as well.

In Section 2, we describe the sample acquisition and prediction methods for metabolomics-guided surgery. In Section 3, we

- *E.O. Karakaslar is with the Computer Engineering Department, Bilkent University, Ankara 06800, Turkey. E-mail: onur.karakaslar@bilkent.edu.tr.*
- *B. Coskun is with TUSAS-TAI, Ankara 06800, Turkey. E-mail: bcoskun1993@gmail.com.*
- *H. Outilaft is with the University of Strasbourg, Strasbourg 67081, France. E-mail: hassiba2a@live.fr.*
- *I.J. Namer is with Strasbourg University, Strasbourg 67081, France. E-mail: IzzieJacques.NAMER@chru-strasbourg.fr.*
- *A.E. Cicek is with the Computer Engineering Department, Bilkent University, Ankara 06800, Turkey, and also with the Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213. E-mail: cicek@cs.bilkent.edu.tr.*

Manuscript received 16 Dec. 2018; revised 18 May 2019; accepted 28 May 2019. Date of publication 4 June 2019; date of current version 1 Apr. 2020.
(Corresponding author: A. Ercument Cicek.)
Digital Object Identifier no. 10.1109/TCBB.2019.2920646

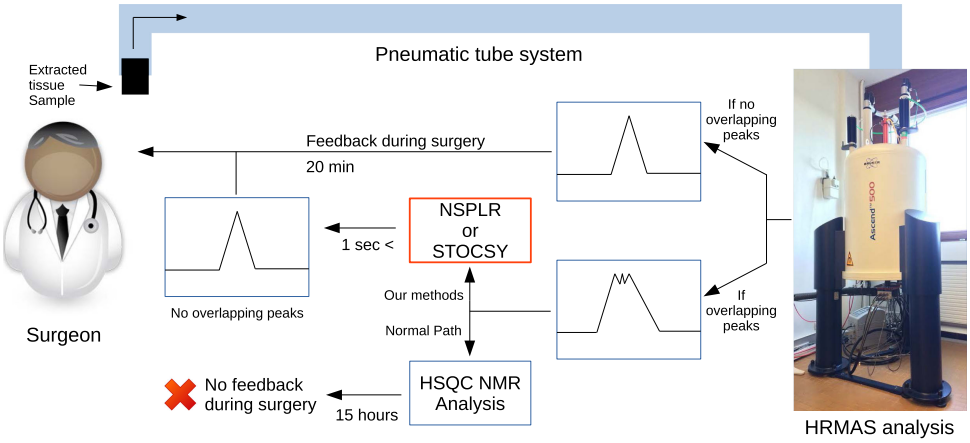

Fig. 1. This figure shows the workflow of the feedback mechanism that we suggest. Surgeon extracts a sample from excision cavity and sends it to the spectroscopy room where HRMAS NMR analysis is conducted. If there are no overlapping signals after the analysis, the results are then sent back to the surgeon during surgery. Otherwise, if there are overlapping signals, another procedure called HSQC NMR is conducted which approximately takes 15 hours or the feedback is provided with one of our methods which can be conducted less than a second.

present our results on the above-mentioned datasets and finally, in Section 4, we conclude.

2 Methods

2.1 Surgical Pipeline

After the removal of the tumor from the tissue, several samples are collected from the excision cavity by the surgeon. The samples are sent to the MRI room via pneumatic tube. HRMAS takes approximately 20 minutes. The learning stage is offline and the time requirement is irrelevant for the online analysis. Prediction stage takes time in the order of seconds, and thus, allows concluding presence/absence of bio-marker metabolites and giving feedback to the surgeon. Evaluation of both spectra takes less than 10 minutes. Fig. 1 shows the overall workflow of the procedure and this can be repeated as many times as the surgeon requests.

2.2 Tissue Sample Preparation for HRMAS NMR Spectroscopy

All tissue specimens were collected during surgery just after tumor removal and were snap-frozen in liquid nitrogen. Then, the sample preparation was performed at the temperature of -20°C. The amount of tissue used for the HRMAS NMR analysis ranged from 15 mg to 20 mg. Each tissue sample was placed in a 25 μl disposable insert. Next, 12 μl of deuterium oxide, were added in every biopsys insert in order to get a resonance frequency reference for the NMR spectrometer. Finally, inserts were kept at -20°C until the HRMAS NMR analysis was performed. The insert was placed in a 4 mm ZrO2 rotor just before the HRMAS NMR analysis.

2.3 HRMAS NMR Data Acquisition

All HRMAS NMR spectrum were obtained on a Bruker Avance III 500 spectrometer (installed at Hautepierre Hospital, Strasbourg) operating at a proton frequency of 500.13 MHz and equipped with a 4 mm quadruple resonance gradient HRMAS probe (^{1}H, ^{2}H, ^{13}C and ^{31}P).

The temperature was maintained at 4°C throughout the acquisition time in order to reduce the effects of tissue degradation during the spectrum acquisition. We realized: 1) A one-dimensional (1D) proton spectrum using a CarrPurcellMeiboomGill (CPMG) pulse sequence was acquired for each tissue sample. The inter-pulse delay between the 180 pulses of the CPMG module was set to 285 ms and the number of loops was set to 328, resulting in a total CPMG pulse train of 93 ms. 1D CPMG parameters are: Fid size: 32768; number of dummy scans: 4; number of scans: 4; spectral width (ppm): 14; acquisition time (s): 2.33; experiment time: 9 min 57 secs . The chemical shift was calibrated to the methyl proton of L-lactate at 1.33 ppm. 2) A two-dimensional (2D) heteronuclear single quantum coherence experiments (1H 13C) were also recorded immediately after ending the 1D spectrum acquisition in order to confirm resonance assignments in all the samples. HSQC parameters are: Fid size: F2:208 and F1: 256; number of dummy scans: 32; number of scans: 136; spectral width (ppm): F2:14.00 and F1: 165.65; acquisition time (s): F2:0.146 and F1:0.0066; experiment time: 16 hours 23 min 17 sec.

2.4 Predicting Carbon Spectrum in HSQC NMR

In this section, we describe two methods: STOCSY and multivariate multiple linear regression in order to predict 1D ^{13}C spectrum of a sample when ^{1}H spectrum is inconclusive and we also propose one regression based algorithm to reconstruct HSQC NMR.

2.4.1 1D-NMR Spectrum Prediction by Linear Regression (NSPLR)

Multivariate multiple regression is concerned with finding the linear relationship between multiple response variables (multivariate) and multiple predictor variables. In our setting, the response variables are ^{13}C signal values, and the predictor variables are ^{1}H signal for n samples. Let y_j be a c-dimensional vector, where c denotes the number of observed signal values in ^{13}C dimension for sample j such that $1 \leq j \leq n$. Similarly, let x_j be a h-dimensional vector corresponding to ^{1}H signal values for sample j where h denotes the number of observed signal values in ^{1}H dimension. Finally, let z_i be a $(h+1)$-dimensional vector which is same as x_i with an extra 1 padded to the beginning: $z_j = [x_{0j}, x_{1j}, .., x_{hj}]$, x_{ij} denotes the ith ^{1}H value for the jth sample and $x_{0j} = 1$ for all j. Then the regression model can be stated as follows:

$$y_j = z_j\beta + \epsilon_j, \tag{1}$$

where $\beta \in R^{(h+1)\times c}$ and represents the estimated coefficients and ϵ_j is the error vector. Then, the multivariate multiple regression model is defined as follows. Let Y be the response matrix such that $Y \in R^{n\times c}$. Similarly, let Z be the design matrix such that $Z \in R^{n\times(h+1)}$. Then,

$$Y = Z\beta + \epsilon, \tag{2}$$

where $\epsilon \in R^{n\times c}$. The β matrix is unknown and is estimated using ordinary least squares. Let $\beta = [b_1; ..; b_c]$, then each column vector b_j $(1 \leq j \leq c)$ is a vector of coefficients $[w_{0j}, w_{1j}, .., w_{hj}]^T$. w_{0j} is the

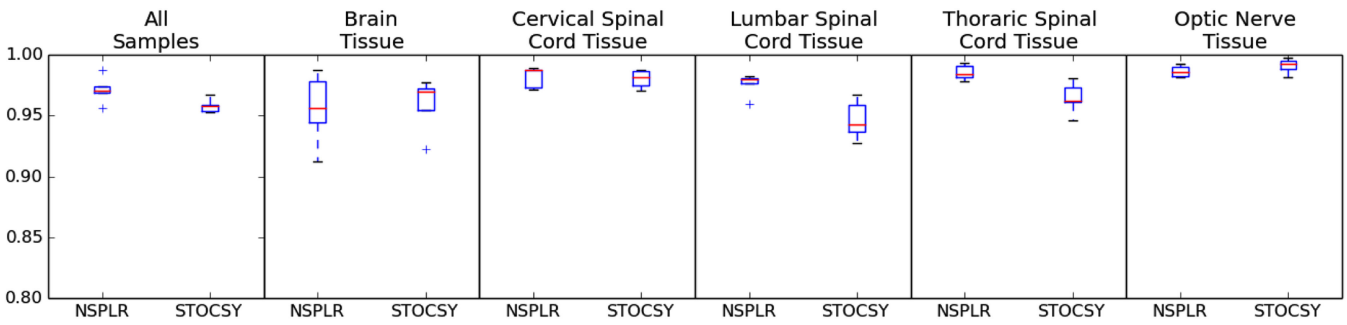

Fig. 2. The figure shows the box-plots of R^2 values of NSPLR and STOCSY for EAE rat cohort obtained via 5-fold cross validation. Left-most panel shows the results obtained on the full cohort of 80 samples. Following panels show the results obtained per tissue. The mean R^2 values for NSPLR and STOCSY, (i) are 0.971 and 0.957 for the full cohort; (ii) are 0.955 and 0.959 for brain tissue; (iii) are 0.981 and 0.980 for cervical tissue; (iv) are 0.975 and 0.946 for lumbar spinal tissue; and (v) are 0.985 and 0.964 for thoracic spinal tissue; and finally, (vi) are 0.986 and 0.990 for optic nerve tissue, respectively.

mean effect of all hydrogen values on the jth carbon value and w_{ij} $(1 \leq i \leq h)$ denotes the weight of the effect of the ith hydrogen value on the jth carbon value. The ^{13}C spectrum of a sample is then found by simply multiplying the ^{1}H spectrum of that sample (also $h+1$-dimensional vector with a 1 padded as the zeroth index) with the β matrix.

2.4.2 Statistical Total Correlation Spectroscopy - STOCSY

Using a set of independent samples, statistical total corre- lation spectroscopy (STOCSY) method generates a pseudo 2D NMR spectrum for all analyzed samples that displays the correlation of the signal intensities in two dimensions [7]. Here, we use C for a different purpose: To transform a ^{1}H spectrum into ^{13}C domain. In short, the method computes the correlation matrix C of the two dimensions (in our case ^{13}C spectrum and ^{1}H spectrum). A correlation matrix is a d_1 by d_2 matrix where d_1 and d_2 denote the number of variables (i.e., ppm) in each dimension. Each index (i, j) in this matrix denotes the correlation of the ith variable in dimension d_1 with the jth variable in dimension d_2 over all samples. Let $X_1 \in R^{n \times d_1}$ and $X_2 \in R^{n \times d_2}$; d_1 and d_2 are the number of variables in each spectra and n is the sample size. STOCSY calculates the correlation matrix as follows:

$$C = \frac{1}{n-1} X_1^T X_2. \tag{3}$$

In our setting, X_1 and X_2 represent ^{1}H and ^{13}C spectra of the samples, respectively. Only statistical assumptions are that the relationship between the ^{1}H and ^{13}C spectra is linear and the observations are independent.

$$\hat{\beta} = corr(X_1, X_2) \frac{\sqrt{var(X_1)}}{\sqrt{var(X_2)}} = C \frac{\sqrt{var(X_1)}}{\sqrt{var(X_2)}} \propto C, \tag{4}$$

where var denotes the variance of a given signal, and $corr$ is the correlation matrix of two signals in which each index (i,j) denotes the correlation coefficient between two variables of X_1 and X_2. We predict the ^{13}C vector y_j that corresponds to ^{1}H vector x_j as follows: $y_j = z_j \hat{\beta}$. Thus, one can also use C instead of $\hat{\beta}$: $y_j = z_j C$.

Note that even if the $equal - variance$ assumption is violated, correlation matrix is a scaled version of the design matrix. Since, we are not interested in predicting the exact signal values, but presence and absence of the metabolite groups in ^{13}C spectrum of the signal, this scaling effect can be ignored.

2.4.3 HSQC NMR Reconstruction Based on NSPLR

Let matrix A be a HSQC NMR sample, $A \in R^{h \times c}$ where h and c are defined as in Section 2.4.1. Then, each kth column vector of a sample can be treated as the response variable, y_j^k, as y_j in Section 2.4.1 where $1 \leq k \leq h$ and $1 \leq j \leq n$. In this way, h regression matrices (β^k) are obtained for a given sample. So the regression model becomes:

$$y_j^k = z_i \beta^k + \epsilon_j^k, \tag{5}$$

where $\beta^k \in R^{(h+1) \times c}$ and represents the estimated coefficients and ϵ_j^k is the error vector. Then, the multivariate multiple regression model is defined as follows. Let Y be the response matrix such that $Y \in R^{n \times c}$. Similarly, let Z be the design matrix such that $Z \in R^{n \times (h+1)}$. Then,

$$Y = Z\beta^k + \epsilon, \tag{6}$$

where $\epsilon \in R^{n \times c}$. The β^k matrix is unknown and is estimated using ordinary least squares. Let $\beta^k = [b_1; ..; b_c]$, then each column vector b_j $(1 \leq j \leq c)$ is a vector of coefficients $[w_{0j}, w_{1j}, .., w_{hj}]^T$. w_{0j} is the mean effect of all hydrogen values on the jth carbon value and w_{ij} $(1 \leq i \leq h)$ denotes the weight of the effect of the ith hydrogen value on the jth carbon value. The kth carbon column vector of the HSQC NMR sample is then found by simply multiplying the ^{1}H spectrum of that sample (also $h+1$-dimensional vector with a 1 padded as the zeroth index) with the β^k matrix. Finally, HSQC NMR (matrix A) is reconstructed by concatenating these column vectors.

3 Results

We test our prediction scheme on two datasets. First, a rat cohort of experimental allergic encephalomyelitis (EAE), is used to create a baseline for further investigation. Next, we evaluated our scheme on 14 samples of epilepsy and cerebral tumor patients to predict presence and absence of metabolites as a simulation of a surgery. The ground truth is obtained by the manual inspection of domain scientists at Department of Nuclear Medicine, University Hospitals of Strasbourg, Hautepierre Hospital, Strasbourg, France.

3.1 Experimental Allergic Encephalomyelitis (EAE) Rat Cohort

This study included 20 female Lewis rats (Charles River, France), aged 6-8 weeks, (weight: 130-145 g). Ten rats were immunized with intradermal injection of a 0.1 mg of MBP in a complete Freund adjuvant containing 0.5 mg of attenuated Mycobacterium tuberculosis strain H37RA (EAE group). Ten other non-immunized rats constituted the control group. All rats were sacrificed the same day when clinical signs were maximal (appearance of typical paraplegia, on the 12th day) in the EAE group. The whole CNS and optic nerves were collected and snap-frozen in liquid nitrogen before storage. 84 samples (44 in the EAE group and 40 in the control group) were kept for NMR data processing: 19 brain tissue samples (respectively 10 and 9), 17 cervical spinal cord tissue samples (respectively 8 and 9), 20 thoracic spinal cord tissue samples (respectively 10 and 10), 20 lumbar spinal cord tissue samples (respectively 10 and 10) and 8 optic nerve tissue samples

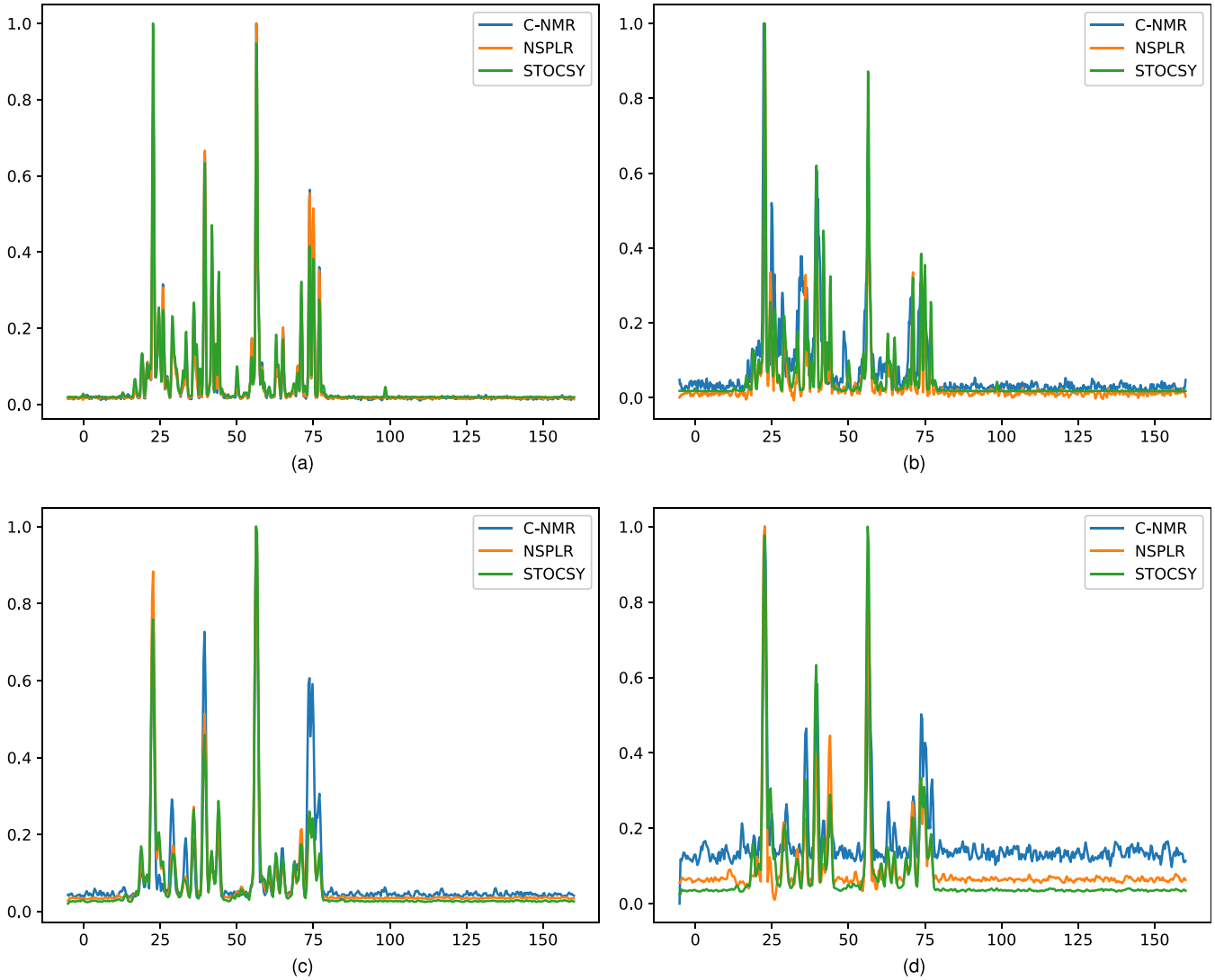

Fig. 3. This figure shows four predicted samples of ^{13}C-NMR spectrum (blue) and their corresponding predictions with NSPLR (orange) and STOCSY (green) methods. For all figures, *x*-axes show the ppm values and all *y*-axes values are normalized in order to be able to compare the locations of signal values. Panels (a) and (b) show the best and worst performing predictions of both methods for EAE rat cohort, respectively. Panels (c) and (d) show the best and worst performing predictions of both methods for epilepsy and cerebral tumor dataset, respectively.

(respectively 6 and 2). We excluded 4 samples due to high variance in the signal indicating systematic error.

3.1.1 Prediction Performances of NSPLR and STOCSY

Above mentioned, NMR Spectrum Prediction by Linear Regression (NSPLR) and STOCSY methods were used for blindly predicting the ^{13}C-NMR spectrum of 80 samples of the EAE rat cohort. We used 5-fold cross-validation. For each fold, a design matrix was trained using rest of the data. Then the left-out fold of ^{13}C-NMR spectrum was predicted via corresponding ^{1}H-NMR spectrum.

First subpanel of Fig. 2 displays the box plots of R^2 values of all and subject based separated versions of the EAE rat cohort for both methods. NSPLR's average R^2 for all rat samples was 0.971 and STOCSY's average R^2 was 0.957. We also repeated the same analysis within all 5 tissue types which are shown in the subsequent subpanels of Fig. 2. The mean R^2 values for NSPLR and STOCSY, (i) are 0.971 and 0.957 for the full cohort; (ii) are 0.955 and 0.959 for brain tissue; (iii) are 0.981 and 0.980 for cervical tissue; (iv) are 0.975 and 0.946 for lumbar spinal tissue; and (v) are 0.985 and 0.964 for thoracic spinal tissue; and finally, (vi) are 0.988 and 0.990 for optic nerve tissue, respectively. Also, we show the best and the worst performances of both methods on ^{13}C-NMR spectrum in Panels (a) and (b) of Fig. 3, respectively.

TABLE 1
This Table Demonstrates the Patient Characteristics

ID	Gender	Age (years)	Pathology
Sample 1	M	76	Glioblastoma
Sample 2	M	46	Glioblastoma
Sample 3	M	34	Epilepsy
Sample 4	M	34	Epilepsy
Sample 5	F	35	Epilepsy
Sample 6	M	66	Epilepsy
Sample 7	M	51	Epilepsy
Sample 8	M	44	Oligoastrocytoma grade II-III
Sample 9	M	37	Pineal tumor
Sample 10	F	22	Oligodendroglioma grade III
Sample 11	M	56	Glioblastoma
Sample 12	M	46	Oligodendroglioma grade III
Sample 13	M	42	Astrocytoma grade III
Sample 14	F	51	Oligodendroglioma grade III
Sample 15	M	47	Epilepsy

Names are concealed and each patient are given an ID to ensure their privacy.

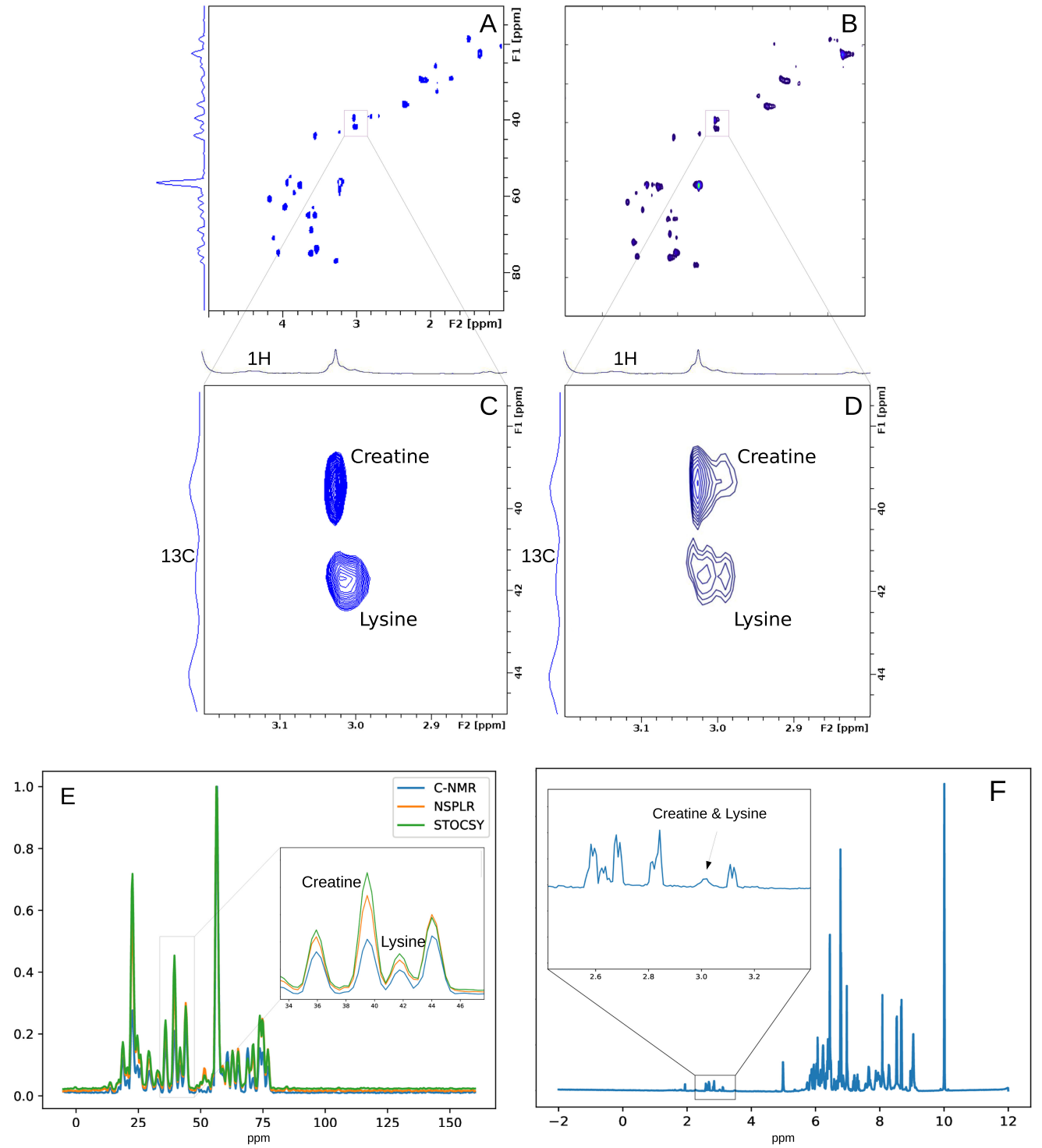

Fig. 4. This figure shows the ^{1}H-^{13}C HSQC NMR of Sample 3 and its reconstructed version. (A) Original spectrum captured with Bruker TopSpin 3.5. (B) Reconstructed version of the spectrum in (A) predicted using only ^{1}H-NMR sample. (C) Zoomed version of sample in (A), this figure shows metabolite groups of Creatine and Lysine overlapping on ^{1}H dimension of HSQC NMR, yet they are distinguishable on ^{13}C dimension. (D) Zoomed version of (B). (E) Reconstructed ^{13}C NMR spectrum of the same sample using NSPLR and STOCSY. Both metabolite groups are shown. (F) ^{1}H-HRMAS NMR spectrum of Sample 3, the overlapping metabolite group signals of Creatine and Lysine are shown near 3 ppm.

3.2 Epilepsy and Cerebral Tumor Dataset

This study included 15 samples obtained from 14 patients retrospectively selected after they had undergone epilepsy and cerebral tumors surgery, from February 2015 to February 2017, in the Department of Neurosurgery (University Hospitals of Strasbourg, Hautepierre Hospital, Strasbourg, France). Patients characteristics are detailed in Table 1. Among the 15 samples obtained from 14 patients:

- 6 samples from patients who had undergone epilepsy surgery (normal tissue)
- 9 samples from patients who had undergone cerebral tumors surgery (tumor tissue)

All sample tissues were collected just after resection by a pneumatic system connecting the neurosurgery operative room to the spectrometer room and were then snap-frozen in liquid nitrogen before storage. A written informed consent was given

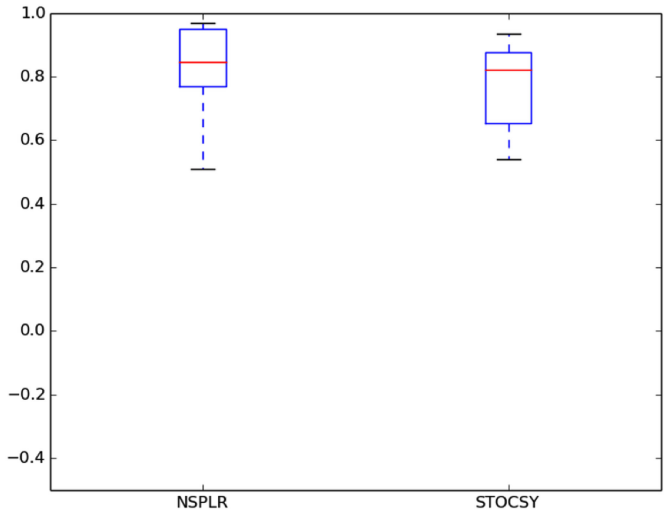

Fig. 5. This figure shows the boxplot of R^2 values of 14 human cancer patients each obtained via leave-one-out cross validation method. The mean of NSPLR method is 0.812 and mean of STOCSY is 0.774 which are indicated by the red lines.

by all the included patients. We excluded one sample (Sample 15 in Table 1) due to high variance in the signal indicating a systematic error.

3.2.1 Prediction Performances of 1D-NSPLR and STOCSY

We tested NSPLR and STOCSY on NMR spectrum of human brain samples. Again, using leave-one-out cross-validation, each ^{13}C-NMR spectrum was predicted with both methods. Panels (c) and (d) in Fig. 3 display prediction performance of both methods on two ^{13}C-NMR spectrum (best performance on the left, worst performance on the right). We also provide boxplot of R^2 values of each human sample for both methods in Fig. 5. For R^2 values of human samples, NSPLR's average was 0.81 and STOCSY's average was 0.77. NSPLR and STOCSY yielded similar results, they both have 97.1 percent accuracy, and 94.1-94.0 percent recall rates, respectively.

Specifically, we predicted the presence and absence of 104 metabolite groups belonging to 39 metabolites in these 14 patients ($>$2100 predictions). Supplementary Table 1, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2019.2920646 shows all detected/undetected metabolite groups in each ^{13}C-NMR sample with respect to our database (ground truth). "In Database Predictions" tab shows whether those groups from the database are predicted by our methods given the ground truth. "Out of database" tab lists problematic signals which are not in the database but observed in the ^{13}C spectrum (called Unknown) and signals which are neither in the database nor observed but predicted to exist by one of our methods (called False positive prediction).

3.2.2 Prediction Performance of HSQC NMR Reconstruction

Using leave one out cross validation, we predicted the 2D spectrum for all 14 samples in the epilepsy and cerebral tumor dataset. To plot HSQC NMR predictions, we used NMRglue toolkit [19] with default parameters: 20 contours for each reconstruction starting from 30,000 ppm in z-axis with a scaling factor of 1.2. After normalization, we calculated the mean squared error (MSE) for all 14 samples which is $\sim$0.04% on average. Observing that our predictions fit well to the 2D signal, we checked if we correctly predicted the presence/absence of the 104 metabolite groups of 39 metabolites as also done for 1D reconstructions above. We report 97.26 percent accuracy for $>$2100 predictions (see the details in 2D Reconstruction - In database tab in Supplementary Table 1, available online) which shows that our NSPLR approach is also performing well in reconstructing two dimensional spectrum. Additionally, when we only focus on the metabolites that have overlapping signals in the ^{1}H dimension and check if we correctly predicted the signals of these metabolite groups in 2D reconstruction, we observed that our method correctly differentiated 106 metabolites out of 109 in ^{13}C dimension (see 2D Reconstruction - 1H overlaps tab in Supplementary Table 1, available online).

3.2.3 Predicting the Presence of Creatine as a Hypoxia Biomarker

We reconstructed the HSQC NMR of Sample 3 using the method described in Section 2.4.3. Rest of the dataset is used for training. Panel A in Fig. 4 shows the actual HSQC experiment and Panel C shows the close up to 2 signals which correspond to creatine and lysine's overlapping metabolite groups. Panel C clearly shows that the 1 dimensional ^{1}H signal cannot distinguish these two metabolites. This is because the CH3 group of the creatine overlaps with the CH2 group of lysine, the two metabolites having an identical hydrogen chemical displacement of 3,03 ppm. If HSQC is performed we can distinguish these two metabolites thanks to their chemical carbon displacement: 39.61 ppm for creatine and 41.9 ppm for lysine, respectively. Panels B and D show our reconstruction for the same experiment. Figure suggests that without the need to perform HSQC, we can distinguish overlapping metabolite groups accurately. Panel E shows our one-dimensional NSPLR prediction for the same sample (Section 2.4.1) and Panel F shows the original ^{1}H-HRMAS NMR spectrum and overlapping metabolite groups of creatine and lysine. This approach also clearly predicts the existence of two distinct metabolites. This distinction is important because creatine is a biomarker for tumor cells that are hypoxic since the tumor cells use phosphocreatine as a source of high-energy phosphate that can be transferred to ADP to generate ATP and creatine [17]. As hypoxic cells are resistant to chemotherapy and photodynamic therapy [18], leaving those cells in the excision cavity is a major risk for the patient which suggests recurrence with drug resistance. Thus, distinguishing creatine and lysine in this example has implications for this patient.

3.3 Time Performance

Training time to obtain all β_k matrices, defined in Section 2.4.3, for a given sample of HSQC NMR takes approximately 70 seconds, yet this is irrelevant for the time frame of surgery. Analysis of the ^{1}H NMR spectrum can be conducted in matter of seconds for all methods described in Section 2.

4 Discussion and Conclusion

Metabolomics-guided surgery is a promising technique to guide the surgeons on distinguishing tumor and normal tissue. HRMAS NMR spectroscopy can quantify biomarker metabolites in solid tissues and its rapid response time fits very well into this surgical pipeline. However, overlapping signals in one dimensional spectrum might prohibit observing presence/absence of metabolites using this technique. We proposed two techniques to overcome this bottleneck and resolve those ambiguous cases. We showed on a rat model of central nervous system as well as on a human brain dataset that our proposed methods work with high accuracy. Our work addresses an important challenge in the realization of metabolomics guided surgery.

In the current state of the pipeline, making a binary prediction (i.e., whether a metabolite is present) is sufficient for the tumors we considered. However, in more complicated biomarkers where

concentration of a metabolite matters, then precision of the height of the predicted signal is also going to be an important aspect in assessing the performance of the method. We show that on the rat model we achieve high R^2 values in regression, even though that was not the primary evaluation metric in our pipeline. Still, this aspect of the method needs further research depending on the precision requirement of the application at hand.

ACKNOWLEDGMENTS

The authors would like to thank Alper Eroglu and Gizem Caylak for their valuable comments. This research has been supported by the French Embassy in Turkey via a visiting researcher fellowship to AEC.

REFERENCES

[1] L. L. Cheng, C. L. Lean, A. Bogdanova, S. C. Wright Jr, J. L. Ackerman, T. J. Brady, and L. Garrido, "Enhanced resolution of proton nmr spectra of malignant lymph nodes using magic-angle spinning," *Magn. Resonance Med.*, vol. 36, no. 5, pp. 653–658, 1996.

[2] G. Erb, K. Elbayed, M. Piotto, J. Raya, A. Neuville, M. Mohr, D. Maitrot, P. Kehrli, and I. Namer, "Toward improved grading of malignancy in oligodendrogliomas using metabolomics," *Magn. Resonance Med.: An Official J. Int. Soc. Magn. Resonance Med.*, vol. 59, no. 5, pp. 959–965, 2008.

[3] S. Battini, A. Imperiale, D. Taïeb, K. Elbayed, A. E. Cicek, F. Sebag, L. Brunaud, and I.-J. Namer, "High-resolution magic angle spinning 1h nuclear magnetic resonance spectroscopy metabolomics of hyperfunctioning parathyroid glands," *Surg.*, vol. 160, no. 2, pp. 384–394, 2016.

[4] S. Battini, F. Faitot, A. Imperiale, A. Cicek, C. Heimburger, G. Averous, P. Bachellier, and I. Namer, "Metabolomics approaches in pancreatic adenocarcinoma: tumor metabolism profiling predicts clinical outcome of patients," *BMC Med.*, vol. 15, no. 1, 2017, Art. no. 56.

[5] Y. Xi, J. S. de Ropp, M. R. Viant, D. L. Woodruff, and P. Yu, "Automated screening for metabolites in complex mixtures using 2d cosy nmr spectroscopy," *Metabolomics*, vol. 2, no. 4, pp. 221–233, 2006.

[6] Y. Xi, J. S. de Ropp, M. R. Viant, D. L. Woodruff, and P. Yu, "Improved identification of metabolites in complex mixtures using hsqc nmr spectroscopy," *Analytica Chimica Acta*, vol. 614, no. 2, pp. 127–133, 2008.

[7] O. Cloarec, M.-E. Dumas, A. Craig, R. H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J. C. Lindon, E. Holmes, et al., "Statistical total correlation spectroscopy: An exploratory approach for latent biomarker identification from metabolic 1h nmr data sets," *Analytical Chemistry*, vol. 77, no. 5, pp. 1282–1289, 2005.

[8] E. Holmes, O. Cloarec, and J. Nicholson, "Probing latent biomarker signatures and in vivo pathway activity in experimental disease states via statistical total correlation spectroscopy (stocsy) of biofluids: Application to hgcl2 toxicity," *J. Proteome Res.*, vol. 5, no. 6, pp. 1313–1320, 2006.

[9] S. L. Robinette, J. C. Lindon, and J. K. Nicholson, "Statistical spectroscopic tools for biomarker discovery and systems medicine," *Analytical Chemistry*, vol. 85, no. 11, pp. 5297–5303, 2013.

[10] B. E. Coggins and P. Zhou, "High resolution 4-d spectroscopy with sparse concentric shell sampling and fft-clean," *J. Biomolecular NMR*, vol. 42, no. 4, pp. 225–239, 2008.

[11] J. C. Hoch, M. W. Maciejewski, M. Mobli, A. D. Schuyler, and A. S. Stern, "Nonuniform sampling and maximum entropy reconstruction in multidimensional nmr," *Accounts Chemical Res.*, vol. 47, no. 2, pp. 708–717, 2014.

[12] X. Qu, M. Mayzel, J.-F. Cai, Z. Chen, and V. Orekhov, "Accelerated nmr spectroscopy with low-rank reconstruction," *Angewandte Chemie*, vol. 127, no. 3, pp. 866–868, 2015.

[13] Y. Shrot and L. Frydman, "Compressed sensing and the reconstruction of ultrafast 2d nmr data: Principles and biomolecular applications," *J. Magn. Resonance*, vol. 209, no. 2, pp. 352–358, 2011.

[14] J. Ying, F. Delaglio, D. A. Torchia, and A. Bax, "Sparse multidimensional iterative lineshape-enhanced (smile) reconstruction of both non-uniformly sampled and conventional nmr data," *J. Biomolecular NMR*, vol. 68, no. 2, pp. 101–118, 2017.

[15] X. Qu, T. Qiu, D. Guo, H. Lu, J. Ying, M. Shen, B. Hu, V. Orekhov, and Z. Chen, "High-fidelity spectroscopy reconstruction in accelerated nmr," *Chemical Commun.*, vol. 54, no. 78, pp. 10 958–10 961, 2018.

[16] M. Billeter, "Non-uniform sampling in biomolecular nmr," *J. Biomolecular NMR*, vol. 62, pp. 65–66, 2017.

[17] M. Wyss and R. Kaddurah-Daouk, "Creatine and creatinine metabolism," *Physiological Reviews*, vol. 80, no. 3, pp. 1107–1213, 2000.

[18] Z. Luo, H. Tian, L. Liu, Z. Chen, R. Liang, Z. Chen, Z. Wu, A. Ma, M. Zheng, and L. Cai, "Tumor-targeted hybrid protein oxygen carrier to simultaneously enhance hypoxia-dampened chemotherapy and photodynamic therapy at a single dose," *Theranostics*, vol. 8, no. 13, 2018, Art. no. 3584.

[19] J. J. Helmus and C. P. Jaroniec, "Nmrglue: An open source python package for the analysis of multidimensional nmr data," *J. Biomolecular NMR*, vol. 55, no. 4, pp. 355–367, 2013.