Kopriva, I., Jerić, I., Hadžija, M. P., Hadžija, M., Lovrenčić, M. V., & Brkljačić, L. (2019). Library-assisted nonlinear blind separation and annotation of pure components from a single 1H nuclear magnetic resonance mixture spectra. Analytica Chimica Acta, 1080, 55–65. https://doi.org/10.1016/j.aca.2019.07.004

Library-Assisted Nonlinear Underdetermined Blind Separation and Annotation of Pure Components from ^{1}H Nuclear Magnetic Resonance Mixture Spectra

Ivica Kopriva[1], Ivanka Jerić[2], Marijana Popović Hadžija[3], Mirko Hadžija[3], Marijana Vučić Lovrenčić[4] and Lidija Brkljačić[2]*

[1]Division of Electronics

[2]Division of Organic Chemistry and Biochemistry

[3]Division of Molecular Medicine

Ruđer Bošković Institute, Bijenička cesta 54, HR-10000, Zagreb, Croatia

[4]Department of Medical Biochemistry and Laboratory Medicine

University Hospital Merkur, Zajčeva 19, HR-10000 Zagreb, Croatia

[*]ikopriva@irb.hr; Tel.: +385-1-4571-286. Fax: +385-1-4680-104

Abstract

Due to its capability for high-throughput screening ^{1}H nuclear magnetic resonance (NMR) spectroscopy is commonly used for metabolite research. However, the key problem in ^{1}H NMR spectroscopy of multicomponent mixtures is overlapping of component signals and that is increasing with the number of components, their complexity and structural similarity. This makes metabolic profiling, that is carried out through matching acquired spectra with metabolites from the library, a hard problem. Here, we propose a methodology for nonlinear blind separation of highly correlated components spectra from smaller number of, including one only, ^{1}H NMR mixture spectra. The method transforms related nonnegative underdetermined blind source separation problem into multiple high-dimensional reproducible kernel Hilbert Spaces (mRKHSs). Therein, highly correlated components are separated by sparseness constrained nonnegative matrix factorization (sNMF) in each induced RKHS. Afterwards, analytes are identified through comparison of separated components with the library comprised of 160 pure components, whereas significant number of them is expected to be related with diabetes. The method is exemplified on: (*i*) annotation of five components spectra separated from two and one ^{1}H NMR model mixture spectra; (*ii*) annotation of 55 metabolites separated from ^{1}H NMR mixture spectra of urine of subjects with and without diabetes type 2. Arguably, it is for the first time to propose method for blind separation of large number of components from single nonlinear mixture. Moreover, proposed method pinpoints urinary creatine, glutamic acid and 5-hydroxyindoleacetic acid as the most prominent metabolites in samples from diabetic subjects, when compared to healthy controls. We also provide metabolic interpretation of obtained results.

Keywords: nonlinear underdetermined blind source separation, single mixture, multiple reproducible kernel Hilbert spaces, nonnegative sparse matrix factorization, ^{1}H NMR spectroscopy, metabolic profiling.

1. Introduction

Metabolic profiling aims to identify and quantify small-molecule analytes (a.k.a. metabolites or pure components) present in complex multicomponent mixture samples acquired in drug development [1, 2], toxicology studies [3], disease diagnosis [4,5], food, nutrition and environmental sciences [6-8]. Metabolic profiling technologies are mainly based on nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry, because both techniques provides structural information on chemical classes in a single analysis. NMR spectroscopy is a quantitative, non-destructive, robust and reliable technique that provides detailed information of structurally diverse metabolites. NMR spectroscopy-based non-targeted metabolite profiling aims to identify as many metabolites as possible in a targeted sample [9]. Candidates for biomarkers are then obtained through matching acquired spectra with those from the library [9, 10], such as the BioMagResBank metabolomics database [11] or Wiley ^{1}H NMR database [12]. However, since many metabolites are structurally similar, their NMR spectra are highly correlated, with many overlapping peaks [13, 8]. That is especially true for ^{1}H NMR spectroscopy [9], which, due to its capability for high-throughput screening [14], is routinely used for metabolite biomarker research. The presence of a large number of metabolites in the studied samples makes metabolic profiling a notoriously difficult problem. ^{1}H-^{1}H J-couplings generates broad multiplets that keep the exact elucidation of the chemical structure ambiguous [15, 16, 8]. Many of metabolites are not species dependent, thus, allowing translation of some specific biomarkers from preclinical studies directly in clinical studies [17]. Quantitative metabolomic profiling of patients with

inflammatory bowel disease characterized 44 serum, 37 plasma, and 71 urine metabolites using ^{1}H NMR spectroscopy [18]. Therefore, for the presented study an in-house library comprised of 160 ^{1}H NMR spectra of pure components, whereas many of them are known to be present in the urine samples of diabetic patients, is built. The primary reason for building the in-house library was to solve problems associated with annotations of components recorded at NMR spectrometers with different strengths of the magnetic field [19]. That is, ^{1}H NMR spectra of the same compound recorded at different spectrometers will have peaks at slightly displaced chemical shifts. If similarity measures, such as correlation, are not invariant to these shifts that will affect accuracy of the annotation.

To extract metabolic information, and enable sample classification and biomarker discovery, computational methods for multivariate analysis of complex metabolomic datasets are of utmost importance [20, 21, 13, 8]. Algorithmic approaches to solve peak overlapping problem may be grouped in three main categories. The scoring methods assess the matches between the experimental and theoretical spectra. To reduce the false alarm rate, a various similarity scores are developed [23, 24]. It is clear that this approach fails if number of analytes in a mixture spectra increases. Machine learning approaches try to learn a classifier using reference components from the library and apply it to experimental spectra [24, 25]. Accuracy of this approach is affected by the size of the training set, but also by the overlapping of analytes spectra. The third category of methods is known as a source separation or deconvolution methods. It is properly pointed out in [26] that the term "deconvolution" is essentially wrong, since it actually denotes inversion of a convolution, a particular kind of integral transform that describes input-output relations of linear systems with memory [27]. As opposed to that, separation of analytes from mixtures of overlapped NMR spectra is related to solving of (non)linear equations that describe memoryless (instantaneous) system with multiple inputs

(analytes) and multiple outputs (mixtures spectra). The source separation methods, a.k.a. multivariate curve resolution (MCR) methods, extract concentration and spectra of individual components from multicomponent mixtures spectra [28]. In particular, blind source separation (BSS) [29, 13] refers to class of multivariate data analysis methods capable of blind (unsupervised) separation of analytes from mixtures spectra only. However, under stated conditions related inverse problem is highly ill-posed. To narrow-down infinite number of solutions to, ideally, unique one, constraints have to be imposed on analytes spectra. Typically, constraints include uncorrelatedness, statistical independence, sparseness and nonnegativity. This, respectively, leads to principal component analysis (PCA) [30], independent component analysis (ICA) [31, 32], sparse component analysis (SCA) [33, 34] and nonnegative matrix factorization (NMF) [35]. These methods have already been applied successfully to separation of components from various types of spectroscopic mixtures [36-41]. PCA, ICA and many NMF algorithms require that the *unknown* number of analytes is less than or equal to the number of mixtures spectra available. That makes them inapplicable for the analysis of complex multicomponent mixtures spectra. The same conclusion applies to many other "deconvolution" methods [42]. Sparseness-based approaches to BSS are presently highly active research area in signal processing. Unlike PCA and ICA, SCA enables solution of an underdetermined BSS problem, i.e. separation of more analytes than mixtures available [38-40]. Sparseness implies that at each chemical shift coordinate only very small number of analytes is present. Moreover, majority of SCA algorithms require that each analyte is present at certain chemical shift region alone [38-40, 43, 44]. In case of ^{1}H NMR spectroscopy, due to reasons elaborate previously, it is impossible to satisfy this assumption when complexity of multicomponent mixtures grows. Some development in blind separation of positive and partially overlapped sources requires that each analyte is dominant, instead of present alone, at a certain chemical shift region [45].

Nevertheless, for complex multicomponent mixture ^{1}H NMR spectra the same conclusion applies as above. Furthermore, blind separation of analytes from mixtures of ^{1}H NMR spectra by means of sparseness and nonnegativity constrained BSS methods is additionally limited by an assumption that ^{1}H NMR spectrum is linear mixture of spectra of analytes. That is true at chemical shifts where only one analyte is present. Otherwise, the mixture is becoming more nonlinear when complexity of the mixture grows, i.e. when number of overlapped peaks is increasing [46].

Herein, we propose a method for blind separation of nonnegative correlated sources from smaller number of nonlinear mixtures, (nonnegative nonlinear underdetermined BSS problem) including single nonlinear mixtures as a special, but clinically most relevant, case. Developed methods are applied to separation of correlated analytes spectra from model mixtures (laboratory made) as well as from experimental (urine of subjects with and without diabetes type II) ^{1}H NMR mixtures spectra. The method, through use of implicit (kernel-based) nonlinear transform, maps original underdetermined BSS problem into new one in reproducible kernel Hilbert space (RKHS) [47]. In so doing, the method increases significantly the number of mixtures, while number of new components generated by nonlinear transform is increased only modestly. That, in combination with sparse distribution of amplitudes of analytes ^{1}H NMR spectra, enables separation of highly correlated analytes spectra by means of sparseness constrained NMF (sNMF) in mapping induced RKHS. Analytes are identified through comparison of separated components with the pure components in the library. That distinguishes proposed methodology from RKHS-based BSS methods developed recently for separation of analytes from mixtures of mass spectra [48-50]. Proposed methodology is demonstrated on two experiments: (*i*) separation and annotation of five correlated components spectra from two as well as only one model ^{1}H NMR mixture spectra and (*ii*) separation and annotation of components present in ^{1}H NMR

mixtures spectra of diabetic and non-diabetic urine samples. To the best of our knowledge, we are the first to demonstrate method for blind separation of large number of components from single ^{1}H NMR nonlinear mixture spectra. Furthermore, proposed method emphasized urinary creatine, glutamic acid and 5-hydroxyindoleacetic acid as the most prominent metabolites in samples from diabetic subjects, when compared to healthy controls. The rest of the paper is organized as follows. Section 2 presents nonlinear mixture models of multicomponent ^{1}H NMR spectra, solvability conditions and analysis, nonlinear transformations of multiple mixtures and single mixture nonlinear BSS problem as well as criterions for evaluation of separation and annotation quality. Section 3 describes experiments and materials used for comparative performance analysis of methods for nonlinear blind separation of components from multiple and single ^{1}H NMR spectra of model and experimental mixtures. Results related to separation and annotation performances of developed algorithms are presented in section 4. Section 5 discusses metabolic interpretation of the most prominent metabolites in urine of diabetic patients. Conclusion is presented in section 6.

2. Theory and methods

2.1 Nonlinear mixture model of multicomponent ^{1}H NMR spectra

Linear mixture model (LMM) is commonly used in NMR spectroscopy [28, 36-40]. It is the model upon which linear instantaneous BSS methods are based, [29, 31-35]. In case of NMR signals, which are intrinsically time domain harmonic signals with amplitude decaying exponentially with some time constant, LMM applies to either time domain or Fourier transform

domain representations. The model in the Fourier (chemical shift) domain in the absence of additive noise reads out as:

$$\mathbf{X} = \mathbf{AS} \tag{1}$$

where $\mathbf{X} \in C^{N \times T} =: \left\{ \mathbf{X}_n = FT(\mathbf{x}_n) \in C^{1 \times T} \right\}_{n=1}^{N}$ represents mixture matrix such that each row of $\mathbf{X}$ contains one multicomponent complex ^{1}H NMR mixture signal, obtained as Fourier transform (*FT*) of related time domain equivalent $\mathbf{x}_n$, comprised of complex (real and imaginary parts) values at *T* chemical shift instants, and symbol " =:" means "by definition".

$\mathbf{A} \in R_{0+}^{N \times M} =: \left\{ \mathbf{a}_m \in R_{0+}^{N \times 1} \right\}_{m=1}^{M}$ represents mixture (a.k.a. concentration) matrix, whereas each column vector represents concentration profile of one of the *M* analytes across the *N* mixtures.

$\mathbf{S} \in C^{M \times T} =: \left\{ \mathbf{S}_m = FT(\mathbf{s}_m) \in C^{1 \times T} \right\}_{m=1}^{M}$ is a matrix with the rows representing ^{1}H NMR complex signals of the analytes present in the mixture signals $\mathbf{X}$. However, as shown in [46], amplitudes of the NMR mixture spectra, $|\mathbf{X}| \in R_{0+}^{N \times T} =: \left\{ |\mathbf{X}_n| \in R_{0+}^{1 \times T} \right\}_{n=1}^{N}$, are nonlinear mixtures of the amplitudes of the components NMR spectra, $|\mathbf{S}| \in R_{0+}^{M \times T} =: \left\{ |\mathbf{S}_m| \in R_{0+}^{m \times T} \right\}_{m=1}^{M}$. Thus, instead of LMM (1) we assume nonlinear mixture model (NMM) for ^{1}H NMR amplitude spectra:

$$|\mathbf{X}| = \mathbf{f}\left(|\mathbf{S}|\right) \tag{2}$$

where $\mathbf{f}:\mathbb{R}_{0+}^{M}\to\mathbb{R}_{0+}^{N}$ stands for an unknown nonlinear mapping $\mathbf{f}\left(|\mathbf{S}|\right):=\left[\,f_1\left(|\mathbf{S}|\right)\ldots f_N\left(|\mathbf{S}|\right)\right]^T$ acting observation-wise. We also assume that $\left\{\left\|\left|\mathbf{S}_t\right|\in\mathbb{R}_{0+}^{M\times 1}\right\|_0\le K\right\}_{t=1}^{T}$, where $\left\|\left|\mathbf{S}_t\right|\right\|_0$ is indicator function that counts number of non-zero entries of $|\mathbf{S}_t|$ and K denotes maximal number of sources that can be present (active) at any observation coordinate t. Nonlinear BSS problem (2) implies that amplitude spectra of pure components $|\mathbf{S}|$ ought to be inferred from mixture amplitude ^{1}H NMR spectra $|\mathbf{X}|$ only. Since nonlinear BSS methods that will be developed herein are aimed to be used for metabolic profiling we assume:

A1) $N\ge 1$

A2) $M>N$

Thus, nonlinear BSS problem (2) is underdetermined. Since peaks in amplitude spectra are not statistically related, the pure components are treated as independent and identically distributed (i.i.d.) random variables. Hence, we propose herein methodology for blind separation of mutually dependent but individually i.i.d. nonnegative pure components from smaller number of, including only one, their nonlinear mixtures. To the best of our knowledge, existing methods cannot address the nonlinear BSS problem under assumed scenario. Compared with methods proposed herein existing methods either: (*i*) address determined case, where the number of sources equals the number of mixtures [51-60]; (*ii*) do not take into account nonnegativity constraint [51-63];

(*iii*) assume that sources [52-54, 57-59, 61-64] or their derivatives [60] are statistically independent or that sources are individually correlated [58, 61-63]. Herein, we use empirical kernel map (EKM), [47], for observation-wise mapping of mixture spectra in RKHS.

As it is seen from A1, nonlinear BSS problem (2) also includes clinically most relevant scenario of single mixture, i.e. N=1. Algorithms for single-mixture BSS first have to transform the single- to the pseudo multi-mixture BSS problem [65-75]. Subsequently, some existing multivariate algorithms are used to perform BSS. As in [50], we use an explicit feature map (EFM) for observation-wise nonlinear mapping of the recorded mixture ^{1}H NMR spectra into pseudo multiple mixtures spectra. Pseudo multiple mixture data are mapped observation-wise in high-dimensional RKHS using EKM.

The proposed single-mixture nonlinear BSS algorithm differs from the existing single-mixture BSS algorithms in the following aspects: (*i*) algorithms [65-75] address the linear BSS problem, while the proposed method addresses the nonlinear BSS problem, and (*ii*) the hard constraints imposed on the source signals by single-mixture BSS algorithms [65-75] do not apply to the pure component ^{1}H NMR amplitude spectra that are of interest in this study. This statement is supported through the following analysis. The method [65] assumed that the source signals have disjoint support. The method [65] partitions single-channel time series to yield a pseudo multichannel mixture, and an independent component analysis (ICA) algorithm was then applied to extract the sources. The disjoint support assumption does not hold for the overlapped pure component ^{1}H NMR amplitude spectra. The algorithm [66] used empirical mode decomposition to decompose the single-channel mixture into intrinsic mode functions (IMFs) that represent the pseudo multichannel mixture. For separation by ICA algorithms, sources of interest are required to be IMFs, which does not hold for the pure component ^{1}H NMR amplitude

spectra. In [67], the wavelet transform is used to generate a pseudo multichannel mixture from a single-channel version. In this way, mother wavelets have to be non-orthogonal and have to match the shapes to the sources of interest. Thus, this wavelet-ICA method is applicable to the separation of the specific source signals, such as vibration signals [68, 69]. Many of single-channel BSS algorithms are derived to separate acoustic signals by factorizing a nonnegative spectrogram (magnitude of the short time Fourier transform) [70-75].

The essential differences of multiple- and single-mixture nonlinear BSS methods proposed herein in comparison with multiple, [49], and single-mixture nonlinear BSS method [50] are: (*i*) while only one RKHS was induced in [49, 50] the methods proposed herein map mixture spectra onto multiples RKHSs induced by Gaussian and polynomial kernels or by Gaussian kernel with different values of the variance; (*ii*) after separation library of in-house recorded pure components ^{1}H NMR spectra is used to annotate separated components. Thus, proposed methods are based upon implicit assumption that spectral library is rich enough to contain pure components that correspond to metabolites expected to be present in mixture spectra. The nonlinear BSS method proposed in [76] is also based on mapping of nonlinear mixtures onto multiple RKHSs. However, BSS in mapped spaces is organized as joint sparseness constrained NMF such that LMMs in induced RKHSs have different basis (mixing) matrices but share the same representation (source) matrix. As opposed to [76] the methodology proposed herein performs BSS based on sparseness constrained NMF in each induced RKHS separately. Afterwards, library of pure components ^{1}H NMR mixture spectra is used to annotate components separated from all RKHSs. As it can be seen in Table 1, nonlinear BSS method proposed herein significantly outperforms the nonlinear BSS method [76].

2.2 Solvability of underdetermined nonlinear system and sparse probabilistic model of ^{1}H NMR components spectra

Without loss of generality, we further assume the following:

A3) $0 \le |\mathbf{S}_{mt}| < 1 \;\; \forall m = 1,...,M \;\; t = 1,...,T$,

A4) $|\mathbf{S}_{mt}|$ is i.i.d. random variable that obeys exponential distribution on (0, 1] interval and discrete distribution at zero, see Eq. (3),

A5) Components of the vector-valued function $\mathbf{f}(|\mathbf{S}|) := \left[f_1(|\mathbf{S}|) ... f_N(|\mathbf{S}|) \right]^T$ are differentiable up to second-order.

Assumptions A3 to A5 are shown in [49] to be relevant for separation of pure components from nonlinear mixtures of mass spectra. They hold for separation of pure components from amplitude ^{1}H NMR spectra as well, whereas A4 is confirmed below. To be useful solution of any BSS problem is expected to be essentially unique [29]. However, even for linear underdetermined BSS problem hard (sparseness) constraints ought to be imposed on pure components [77, 48-50] to obtain essentially unique solution. The quality of separation heavily depends on degree of sparseness, i.e. the value of *K*. To make nonlinear underdetermined BSS problem tractable we assume, as in [48], that amplitudes of the source signals comply with sparse probabilistic model [77]:

$$p(|\mathbf{S}_{mt}|) = \rho_m \delta(|\mathbf{S}_{mt}|) + (1-\rho_m)\delta^*(|\mathbf{S}_{mt}|)\, g(|\mathbf{S}_{mt}|) \;\; \forall m = 1,\ldots,M \text{ and } \forall t = 1,\ldots T \qquad (3)$$

Herein, we assume that the ^{1}H NMR components spectra comply with sparse probabilistic model represented by exponential distribution:

$$g(|\mathbf{S}_{mt}|) = (1/\mu_m)\exp(-|\mathbf{S}_{mt}|/\mu_m) \;\; \forall m=1,\ldots,M \qquad (4).$$

We performed least square fitting of exponential distribution (4) to histograms of the experimental analytes ^{1}H NMR amplitudes spectra and obtained $\mu_m \approx 0.0387$. For exponential prior (4) with given μ_m and given probability $p(0<|\mathbf{S}_{mt}|\leq s)$ the value of s is obtained as: $s \approx -\mu_m \ln(1-p)$. For $p=0.99$ and $\mu_m=0.0387$ it follows $s=0.1782$. Thus, in probability the ^{1}H NMR components spectra will have very small values. That will justify cancellation of the higher order terms in the nonlinear transform that follows. Under sparse probabilistic prior (4) the nonlinear mixture model (3) simplifies to [49]:

$$\left|\mathbf{X}\right| = \mathbf{J}\left|\mathbf{S}\right| + \tfrac{1}{2}\mathbf{H}_{(1)}\begin{bmatrix} \left|\mathbf{S}_1\right|^2 \\ \cdots \\ \left|\mathbf{S}_M\right|^2 \\ \cdots \\ \left\{\left|\mathbf{S}_i\right|\left|\mathbf{S}_j\right|\right\}_{i,j=1}^{M} \end{bmatrix} + HOT = \mathbf{B}\begin{bmatrix} \left|\mathbf{S}\right| \\ \left|\mathbf{S}_1\right|^2 \\ \cdots \\ \left|\mathbf{S}_M\right|^2 \\ \cdots \\ \left\{\left|\mathbf{S}_i\right|\left|\mathbf{S}_j\right|\right\}_{i,j=1}^{M} \end{bmatrix} + HOT \tag{5}$$

where $\mathbf{J}$ stands for Jacobian matrix, $\mathbf{H}_{(1)}$ stands for mode-1 unfolded third-order Hessian tensor, $\mathbf{B} = \left[\mathbf{J}\ \tfrac{1}{2}\mathbf{H}_{(1)}\right]$ stands for the overall mixing matrix and *HOT* stands for higher order terms. Since original nonlinear problem (2) is underdetermined the equivalent linear problem (5) is even more underdetermined because it is comprised of the same number of mixtures, N, but of the $P=2M + M(M-1)/2$ dependent sources. When degree of the overlap of the sources in (2) is K degree of the overlap of new sources in (5) is $Q \approx 2K + K(K-1)/2$. Uniqueness of the solution of (5) depends on the triplet (N,P,Q). For deterministic mixing matrix $\mathbf{B}$ the necessary condition for uniqueness is $N=O(Q^2)$ [78]. Thus, it becomes virtually impossible to obtain an essentially unique solution of the underdetermined nonlinear BSS problem (5) with overlapped sources. Separation quality can however, be increased through nonlinear mapping of mixture data:

$$\left\{\left|\mathbf{X}_t\right| \in \mathbb{R}_{0+}^{N\times 1} \to \phi\left(\left|\mathbf{X}_t\right|\right) \in \mathbb{R}_{0+}^{\bar{N}\times 1}\right\}_{t=1}^{T} \tag{6}$$

where explicit feature map (EFM) $\phi\left(\left|\mathbf{X}_t\right|\right)$ maps data into, in principle, infinite dimensional feature space. To make calculations in mapped space computationally tractable,

$\phi\left(|\mathbf{X}|\right) := \left\{\phi\left(|\mathbf{X}_t|\right)\right\}_{t=1}^{T}$ needs to be projected to a low-dimensional subspace of induced space spanned by $\phi\left(\mathbf{V}\right) := \left\{\phi\left(\mathbf{v}_d\right)\right\}_{d=1}^{D}$. Projection known as EKM, see definition 2.15 in [47], maps data from the input space onto RKHS:

$$\Psi\left(|\mathbf{X}|, \mathbf{V}\right) = \phi\left(\mathbf{V}\right)^T \phi\left(|\mathbf{X}|\right) = \mathrm{K}\left(|\mathbf{X}|, \mathbf{V}\right) \tag{7}$$

where $\mathrm{K}\left(|\mathbf{X}|, \mathbf{V}\right) \in \mathbb{R}_{0+}^{D \times T}$ denotes Gram or kernel matrix with the elements $\left\{\kappa\left(|\mathbf{X}_t|, \mathbf{v}_d\right) = \phi\left(\mathbf{v}_d\right)^T \phi\left(|\mathbf{X}_t|\right)\right\}_{d,t=1}^{D,T}$. It is shown in [48, 49] that under sparse probabilistic prior (3), Eq.(7) becomes:

$$\Psi\left(|\mathbf{X}|, \mathbf{V}\right) = \mathbf{G}\begin{bmatrix} \mathbf{0}_{1\times T} \\ |\mathbf{S}| \\ \left\{|\mathbf{S}_i||\mathbf{S}_j|\right\}_{i,j=1}^{M} \end{bmatrix} + \bar{\mathbf{E}} \tag{8}$$

where $\mathbf{G}$ denotes a nonnegative mixing matrix of appropriate dimensions, $\mathbf{0}_{1\times T}$ stands for row vector of zeros and $\bar{\mathbf{E}}$ stands for an approximation error. The uniqueness condition for system (8) becomes: $D=O(Q^2)$, [78]. For $D>>N$ uniqueness condition can be fulfilled with greater probability than uniqueness condition for system (5): $N=O(Q^2)$. Thus, the role of nonlinear EKM-based mapping is to "increase number of ^{1}H NMR mixture spectra".

2.3 Nonlinear transformation of the original multiple mixtures nonlinear BSS problem

To increase probability of separation of highly correlated analytes from ^{1}H NMR spectra the number of mixture spectra *N* has to be increased to *D*>>*N*. For this purpose the nonlinear mapping, known as EKM, $\Psi : \mathbb{R}_{0+}^{N} \to \mathbb{R}_{0+}^{D}$ was proposed in (7)/(8). The mapping performed chemical shift-wise, i.e. $\left\{ \left|\mathbf{X}(\delta_t)\right| \in \mathbb{R}_{0+}^{N} \mapsto \Psi\left(\left|\mathbf{X}(\delta_t), \mathbf{V}\right|\right) \in \mathbb{R}_{0+}^{D} \right\}_{t=1}^{T}$ is kernel dependent:

$$\Psi_{\kappa}\left(\left|\mathbf{X}\right|, \mathbf{V}\right) = \begin{bmatrix} \kappa\left(\left|\mathbf{X}(\delta_1)\right|, \mathbf{v}_1\right) & \ldots & \kappa\left(\left|\mathbf{X}(\delta_T)\right|, \mathbf{v}_1\right) \\ \ldots & \ldots & \ldots \\ \kappa\left(\left|\mathbf{X}(\delta_1)\right|, \mathbf{v}_D\right) & \ldots & \kappa\left(\left|\mathbf{X}(\delta_T)\right|, \mathbf{v}_D\right) \end{bmatrix} \tag{9}$$

In machine learning problems the Gaussian kernel $\kappa\left(\left|\mathbf{X}(\delta)\right|, \mathbf{v}\right) = \exp\left(-\left\|\left|\mathbf{X}(\delta)\right| - \mathbf{v}\right\|_2 / \sigma^2\right)$ and the polynomial kernel $\kappa\left(\left|\mathbf{X}(\delta)\right|, \mathbf{v}\right) = \left(\mathbf{v}^T \left|\mathbf{X}(\delta)\right| + 1\right)^c$ are used most often. Use of Gaussian kernel, as well as other shift invariant kernels such as exponential and multi-quadratic, can be justified due to its universal approximation property [79]. It is, however, unclear how to select the optimal value of the kernel variance σ^2 for Gaussian kernel or degree *c* for polynomial kernel. It is known that value of σ^2 depends on signal-to-noise-ratio (SNR) [80]. If SNR is low, large value of σ^2 ought to be selected and vice versa. It is however hard to know SNR value in practice.

Hence, as opposed to [49] we proposed herein mapping of original ^{1}H NMR mixture spectra onto multiple RKHSs:

$$\left\{\left|\mathbf{X}\right| \in \mathbb{R}_{0+}^{N\times T} \mapsto \Psi_{\kappa_i}\left(\left|\mathbf{X}\right|, \mathbf{V}\right) \in \mathbb{R}_{0+}^{D\times T}\right\}_{i=1}^{I} \tag{10}$$

The role of basis $\left\{\mathbf{v}_d \in \mathbb{R}_{0+}^{N\times 1}\right\}_{d=1}^{D}$ is to approximately span the induced space:

$$span\left\{\phi\left(\mathbf{v}_d\right)\right\}_{d=1}^{D} \approx span\left\{\phi\left|\mathbf{X}\left(\delta_t\right)\right|\right\}_{t=1}^{T} \tag{11}$$

Eq. (11) holds under assumption that [48-50]:

$$span\left\{\mathbf{v}_d\right\}_{d=1}^{D} \approx span\left\{\left|\mathbf{X}\left(\delta_t\right)\right|\right\}_{t=1}^{T} \tag{12}$$

The basis $\mathbf{V} := \left\{\mathbf{v}_d \in \mathbb{R}_{0+}^{N\times 1}\right\}_{d=1}^{D}$ can be estimated from |**X**| by *k*-m*eans* clustering algorithm. Evidently, it can be used in all the mappings in (10). It is however demonstrated in Nyström approximation problem, [81], that more accurate approximation is obtained when *k*-means

centers are used in kernel domain. In relation to the basis selection problem considered herein, it makes sense to select basis that is kernel dependent, i.e. $\mathbf{V} := \mathbf{V}_{\kappa_i}$, such that *k*-means clustering is performed in induced RKHS. Hence, kernel *k*-means [82]. In the experimental section, we have used the *k-means* clustering algorithm, implemented with MATLAB function *kmeans*, to cluster $\left\{\left|\mathbf{X}\left(\delta_t\right)\right|\right\}_{t=1}^{T}$ into pre-specified number of *D* cluster centers which represent basis matrix **V**. To estimate $\mathbf{V}_{\kappa_i}$ we have used the *k-means* clustering algorithm, implemented with MATLAB function *kmeans*, to cluster kernel matrix $\Psi\left(|\mathbf{X}|,|\mathbf{X}|\right) = \mathrm{K}\left(|\mathbf{X}|,|\mathbf{X}|\right)$ into pre-specified number of *D* cluster centers. Indices of found cluster centers are used to identify $\mathbf{V}_{\kappa_i}$ in the input data space. Hence, mapping of original ^{1}H NMR mixture spectra onto multiple RKHSs has the form:

$$\left\{|\mathbf{X}| \in \mathbb{R}_{0+}^{N\times T} \mapsto \Psi_{\kappa_i}\left(|\mathbf{X}|, \mathbf{V}_{\kappa_i}\right) \in \mathbb{R}_{0+}^{D\times T}\right\}_{i=1}^{I} \tag{13}$$

The sNMF algorithms can now be applied to $\Psi_{\kappa_i}\left(|\mathbf{X}|, \mathbf{V}\right)$ in (10) or to $\Psi_{\kappa_i}\left(|\mathbf{X}|, \mathbf{V}_{\kappa_i}\right)$ in (13) in order to separate analytes amplitude spectra. By executing sNMF on data mapped into multiple RKHSs and by combining obtained results we can increase probability to separate correlated ^{1}H NMR component spectra from the small number, including one only, of mixtures spectra. Thus, we separate analytes ^{1}H NMR spectra through:

$$\left\{\left|\hat{\mathbf{S}}_{(m,i)}\right|\right\}_{m=1}^{D} = sNMF\left(\Psi_{\kappa_i}\left(|\mathbf{X}|,\mathbf{V}\right)\right) \quad \forall i = 1,...,I \tag{14}$$

$$\left\{\left|\hat{\mathbf{S}}_{(m,i)}\right|\right\}_{m=1}^{D} = sNMF\left(\Psi_{\kappa_i}\left(|\mathbf{X}|,\mathbf{V}_{\kappa_i}\right)\right) \quad \forall i = 1,...,I \tag{15}$$

Regarding SNMF algorithm, we have used, as in [48, 49], the nonnegative matrix underapproximation (NMU) algorithm [83] with a MATLAB code freely available [84]. A main reason for preferring the NMU algorithm over other sNMF algorithms is that there are no regularization constants related to sparseness constraint that require a tuning. It is important to notice that in (14) or (15) initial number of components to be extracted was set to *D* even though expected number of components is smaller. That comes as a benefit of using EKM-based mapping and alleviates difficult problem related to *a priori* setting of the number of components to be separated. That, in general, is a hard problem in computer science with, so far, no algorithm agreed to work well on data of diverse origins.

Components separated in (14) or (15) are compared with the pure components spectra stored in the library, $\left\{|\mathbf{S}_m|\right\}_{m=1}^{J}$, using normalized correlation coefficient as a similarity measure. That is, component m=1,...,D separated from either $\Psi_{\kappa_i}\left(|\mathbf{X}|,\mathbf{V}\right)$ or $\Psi_{\kappa_i}\left(|\mathbf{X}|,\mathbf{V}_{\kappa_i}\right)$, i=1,...,I is paired with pure component $j^* \in \{1, ..., J\}$ according to :

$$j^* = \arg\max_{j=1,...,J} \frac{\left\langle \left|\hat{\mathbf{S}}_{(m,i)}\right|, \left|\mathbf{S}_j\right| \right\rangle}{\left\|\hat{\mathbf{S}}_{(m,i)}\right\| \left\|\mathbf{S}_j\right\|} \quad \forall m = 1,...,D \quad \forall i = 1,...,I \tag{16}$$

Afterwards, normalized correlation coefficients:

$$c_{(m,i,j^*)} = \frac{\left\langle \left|\hat{\mathbf{S}}_{(m,i)}\right|, \left|\mathbf{S}_{j^*}\right| \right\rangle}{\left\|\hat{\mathbf{S}}_{(m,i)}\right\| \left\|\mathbf{S}_{j^*}\right\|} \quad \forall m = 1,...,D \quad \forall i = 1,...,I \tag{17}$$

are ranked in descending order. Finally, the list is refined by removing from it all the components $\left|\hat{\mathbf{S}}_{(m,i)}\right|$ paired with the same pure component $\left|\mathbf{S}_{j^*}\right|$ with the exception of one with the largest correlation coefficient. Thus, we obtain the final list of separated components annotated to only one pure component from the library to which it is most similar in terms of metric based on normalized correlation coefficient. Number of pure components J stored in the library can in general be large, for example $J \approx 100000$ for the Wiley ^{1}H NMR spectral library [12]. Herein, we used the in-house built library comprised of $J=160$ ^{1}H NMR spectra of pure components. The proposed algorithm based on (14) is named EKM-mRKHS-V_{Input}. The proposed algorithm based on (15) is named EKM-mRKHS-V_{RKHS}. When these algorithms are applied to data mapped in one RKHS only they are named, respectively, as EKM-sRKHS-V_{Input} and EKM-sRKHS-V_{RKHS}. The algorithms EKM-mRKHS-V_{Input} and EKM-mRKHS-V_{RKHS} are summarized in Algorithm 1.

Algorithm 1. Summary of the EKM-mRKHS-V_{Input} and EKM-mRKHS-V_{RKHS} algorithms.

Required:

$\mathbf{X} \in C_{0+}^{N \times T}$, D, $\left\{\sigma_1^2, ..., \sigma_I^2\right\}$ for Gaussian kernel and/or $\left\{c_1, ..., c_I\right\}$ for polynomial kernel.

1. Execute Fast Fourier transform on each row of $\mathbf{X}$: $\left\{\mathbf{X}_n \mapsto \mathbf{X}_n = FFT(\mathbf{X}_n)\right\}_{n=1}^{N}$. Scale $|\mathbf{X}|$ to satisfy A3.

2. Use kernel *k-means* algorithm to estimate bases $\left\{\mathbf{V}_{\kappa_i}\right\}_{i=1}^{I}$ that comply with (11). Alternatively, use *k-means* algorithm to estimate basis $\mathbf{V}$ that complies with (12).
3. Executed mappings according to (10) or (13).
4. Use NMU algorithm to separate ^{1}H NMR components spectra according to (14) or (15).
5. Annotate separated analytes spectra with the pure components spectra from the library according to procedure (16), (17) and the succeeding paragraph.

2.4 Nonlinear transformation of the original single mixture nonlinear BSS problem

As it is seen from A1, nonlinear BSS problem (2) also includes clinically most relevant scenario of single mixture, i.e. N=1. As mentioned previously, algorithms for single-mixture BSS first have to transform the single- to the pseudo multi-mixture BSS problem [65-75]. For single-mixture case, EFM (6) reduces to [50]:

$$\left\{|\mathbf{X}_t| \in \mathbb{R}_{0+} \rightarrow \phi\left(|\mathbf{X}_t|\right) \in \mathbb{R}_{0+}^{\bar{N} \times 1}\right\}_{t=1}^{T} \tag{18}$$

To obtain pseudo multi-mixture data, EFM $\phi\left(\left|\mathbf{X}_t\right|\right)$ has to satisfy two conditions: (*i*) it has to be of finite order and (*ii*) it has to have analytic form. Hence, we provide in (19) analytic expressions for EFM obtained by factorization of Gaussian kernel:

$$\phi\left(\left|\mathbf{X}\right|\right)=\mathrm{e}^{-\frac{\left\|\mathbf{X}\right\|_2^2}{\sigma^2}}\left\{\frac{1}{\sigma^r}\sqrt{\frac{2^r}{\boldsymbol{\alpha}!}}\left|\mathbf{X}\right|^{\boldsymbol{\alpha}}\right\}_{|\boldsymbol{\alpha}|=r,r=0}^{\infty} \tag{19}$$

where $\boldsymbol{\alpha}\in\mathbb{N}_0^r$, $\left|\boldsymbol{\alpha}\right|=\alpha_1+...+\alpha_r$, $\boldsymbol{\alpha}!=\alpha_1!\times...\times\alpha_r!$ and $\left|\mathbf{X}\right|^{\alpha}=\left|\mathbf{X}_1\right|^{\alpha_1}\times...\times\left|\mathbf{X}_r\right|^{\alpha_r}$. Approximate explicit feature map (aEFM) of order *d* is obtained for $0\leq r\leq d<\infty$. Hence, Gaussian kernel induces infinite dimensional RKHS, while aEFM associated with it induces RKHS of dimension *d* that determines order of the approximation. Hence, for mapping associated with RKHS induced with Gaussian kernel instead of (18) we use:

$$\left\{\left|\mathbf{X}_t\right|\in\mathbb{R}_{0+}\rightarrow\phi_d\left(\left|\mathbf{X}_t\right|\right)\in\mathbb{R}_{0+}^{(d+1)\times 1}\right\}_{t=1}^{T} \tag{20}$$

where:

$$\begin{aligned}\phi_i^d\left(\left|\mathbf{X}_t\right|\right)&=\mathrm{e}^{-\frac{\left\|\mathbf{X}_t\right\|_2^2}{\sigma_i^2}}\left\{\frac{1}{\sigma_i^r}\sqrt{\frac{2^r}{\boldsymbol{\alpha}!}}\left|\mathbf{X}_t\right|^{\boldsymbol{\alpha}}\right\}_{|\boldsymbol{\alpha}|=r,r=0}^{d}\\&=\mathrm{e}^{-\frac{\left\|\mathbf{X}_t\right\|_2^2}{\sigma_i^2}}\left[1\ \frac{\sqrt{2}}{\sigma_i}\left|\mathbf{X}_t\right|\ \frac{2}{\sigma_i^2}\left|\mathbf{X}_t\right|\ ...\ \frac{\sqrt{2^d}}{d!}\frac{1}{\sigma_i^d}\left|\mathbf{X}_t\right|^d\right]^T\quad\forall i=1,...,I\end{aligned} \tag{21}$$

The best results reported in the experimental section were obtained for order of aEFM *d*=2. Thus, single-mixture $|\mathbf{X}| \in \mathbb{R}_{0+}^{1\times T}$ is mapped into pseudo multi-mixture according to:

$$|\mathbf{X}| \in \mathbb{R}_{0+}^{1\times T} \rightarrow \phi_2\left(|\mathbf{X}|\right) \in \mathbb{R}_{0+}^{3\times T} \tag{22}$$

$\phi_2\left(|\mathbf{X}|\right)$ can now be mapped into RKHS in the manner equivalent to (10):

$$\left\{\phi_2\left(|\mathbf{X}|\right) \in \mathbb{R}_{0+}^{3\times T} \mapsto \Psi_{\kappa_i}\left(\phi_2\left(|\mathbf{X}|\right), \mathbf{V}\right) \in \mathbb{R}_{0+}^{D\times T}\right\}_{i=1}^{I} \tag{23}$$

where basis $\mathbf{V}$ is found by *k*-means clustering of $\phi_2\left(|\mathbf{X}|\right)$. Mapping of $\phi_2\left(|\mathbf{X}|\right)$ into RKHS equivalent to (13) was not possible because it was not possible to find basis $\mathbf{V}_{\kappa_i}$ by using kernel *k*-means clustering in RKHS induced by Gaussian kernel with $\sigma^2<1$. The reason is that there was not enough diversity in pseudo multi-mixture data $\phi_2\left(|\mathbf{X}|\right)$, i.e. in Gram matrix $\mathrm{K}\left(\phi_2\left(|\mathbf{X}|\right), \phi_2\left(|\mathbf{X}|\right)\right)$ when $\sigma^2<1$. Hence, we separate analytes ^{1}H NMR spectra through:

$$\left\{\left|\hat{\mathbf{S}}_{(m,i)}\right|\right\}_{m=1}^{D} = sNMF\left(\Psi_{\kappa_i}\left(\phi_2\left(|\mathbf{X}|\right), \mathbf{V}\right)\right) \quad \forall i = 1,...,I \tag{24}$$

Due to the same reasons as in section 2.3 we have used the NMU algorithm [83] for sparseness constrained NMF. Components separated in (24) are compared with the pure components spectra stored in the library, $\left\{\left|\mathbf{S}_m\right|\right\}_{m=1}^{J}$, using normalized correlation coefficient as a similarity measure. They are paired with pure components from the in-house library comprised of J=160 ^{1}H NMR spectra of pure components $j^* \in \{1, ..., J\}$ according to :

$$j^* = \arg\max_{j=1,...,J} \frac{\left\langle \left|\hat{\mathbf{S}}_{(m,i)}\right|, \left|\mathbf{S}_j\right| \right\rangle}{\left\|\hat{\mathbf{S}}_{(m,i)}\right\| \left\|\mathbf{S}_j\right\|} \quad \forall m = 1,...,D \quad \forall i = 1,...,I \tag{25}$$

Afterwards, normalized correlation coefficients:

$$c_{(m,i,j^*)} = \frac{\left\langle \left|\hat{\mathbf{S}}_{(m,i)}\right|, \left|\mathbf{S}_{j^*}\right| \right\rangle}{\left\|\hat{\mathbf{S}}_{(m,i)}\right\| \left\|\mathbf{S}_{j^*}\right\|} \quad \forall m = 1,...,D \quad \forall i = 1,...,I \tag{26}$$

are ranked in descending order. Finally, the list is refined by removing from it all the components $\left|\hat{\mathbf{S}}_{(m,i)}\right|$ paired with the same pure component $\left|\mathbf{S}_{j^*}\right|$ with the exception of one with the largest correlation coefficient. Thus, we obtain the final list of separated components annotated to only

one pure component from the library to which it is most similar in terms of metric based on normalized correlation coefficient.

The proposed algorithm for single-mixture nonlinear BSS is named aEFM-EKM-mRKHS-V_{Input}. When this algorithm is applied to data mapped in one RKHS only it is named aEFM- EKM-sRKHS-V_{Input}. Analogously, aEFM-EKM-sRKHS-V_{RKHS} stands for single-mixture nonlinear BSS algorithm for $\mathbf{V}_{\kappa_i}$ estimated by kernel *k*-means from $\mathrm{K}\left(\phi_2\left(\mathbf{X}\right),\phi_2\left(\left|\mathbf{X}\right|\right)\right)$ with σ^2=1. The algorithm aEFM-EKM-mRKHS-V_{Input} is summarized in Algorithm 2.

Algorithm 2. Summary of the single-mixture nonlinear BSS algorithm aEFM-EKM-mRKHS-V_{Input}.

Required:

$\mathbf{x} \in C_{0+}^{1\times T}$, *D*, $\left\{\sigma_1^2,...,\sigma_I^2\right\}$ for Gaussian kernel.

1. Execute Fast Fourier transform on **x**: $\mathbf{x} \mapsto \mathbf{X} = FFT(\mathbf{x})$. Scale $\left|\mathbf{X}\right|$ to satisfy A3.
2. Use *k-means* algorithm to estimate basis **V** from $\phi_2\left(\left|\mathbf{X}\right|\right)$.
3. Executed mappings according to (23).
4. Use NMU algorithm to separate ^{1}H NMR components spectra according to (24).
5. Annotate separated analytes spectra with the pure components spectra from the library according to (26) and the succeeding paragraph.

2.5 Criteria for evaluation of the qualities of separation and annotation

After the separated components are annotated and ranked, the most desirable outcome is that top *M* components on the ranking list correspond with the *M* pure components present in the mixture spectra. However, given the fact that possibly large number of correlated pure components ^{1}H NMR spectra ought to be separated from small number, including one only, of their nonlinear mixture spectra, it is certain that the quality of separation will be limited. Consequently, some number of separated components will be annotated incorrectly. Thus, we propose four criteria to compare the separation and annotation results achieved by nonlinear BSS methods proposed herein with the results of the state-of-the-art competitors:

Criterion 1 (C1) counts number of correctly annotated components out of *M* separated components ranked first on the list. If the separation is perfect, all first *M* separated components would be annotated correctly.

Criterion 2 (C2) is related to penalized mean normalized correlation between the first *M* separated components and pure components they are correctly annotated with:

$$penalized\ mean\ correlation = \left(\sum_{i \in I_c} c_i \left(\left| \hat{\mathbf{S}}_i \right|, \left| \mathbf{S}_{j^*} \right| \right) \right) \Big/ M \tag{27}$$

where I_c denotes index set of correctly annotated components among first M ranked separated components. Hence, it applies for cardinality of the set I_c: $\#I_c \leq M$. Thus, when all first M ranked separated components are annotated correctly we have $\#I_c = M$. Then, the penalized mean correlation equals the mean correlation.

Criterion 3 (C3) is related to penalized mean normalized correlation between all separated components and pure components they are correctly annotated with. Thus, the difference with respect to C2 is that in case of C3 the whole space of latent variables is considered. It applies C2≤C3 with the equality in case of perfect separation. For the EKM-mRKHS-V_{Input}, EKM-mRKHS-V_{RKHS} and aEFM-EKM-mRKHS-V_{Input} algorithms the overall dimensionality of induced RKHSs is $D \times I$. Hence, we want to benefit from mapping the original input mixture spectra onto multiple high-dimensional RKHSs.

Criterion 4 (C4) is related to the mean rank of correctly annotated separated components:

$$Mean\,rank = \left(\sum_{i=1}^{M} m_i \right) \Big/ R \;\; s.t. \;\; \forall \# I_c < i \leq M : m_i = R \tag{28}$$

where R equals dimensionality of the space of latent variables. As an example, for EKM-mRKHS-V_{Input}, EKM-mRKHS-V_{RKHS} and aEFM-EKM-mRKHS-V_{Input} algorithms it applies $R = D \times I$. C4 simultaneously takes into account two factors: (i) increase of dimensionality of

induced space increases probability that all separated components will be annotated with the pure components from the library; (*ii*) it penalizes annotated components with large indicies in the latent space as well as those components that are not annotated at all. Thus, if separation is perfect and all first *M* ranked components are annotated correctly the value of **C4** will be (very) small, i.e. $\lim_{D\to\infty} \mathbf{C4} = 0$. Since with the increase of dimension of induced space the probability of both correct and incorrect annotation is increased the **C4** is sensitive to (in)correct annotation related to dimension of induced spaces.

3.0 Experiment and materials

3.1 Algorithms to be compared

We have proposed family of algorithms for nonlinear BSS from multiple mixtures: EKM-mRKHS-V_{RKHS}, EKM-mRKHS-V_{Input}, EKM-sRKHS-V_{RKHS}, EKM-sRKHS-V_{Input}, as well as from single mixture: EFM-EKM-mRKHS-V_{Input}, aEFM- EKM-sRKHS-V_{Input} and aEFM- EKM-sRKHS-V_{RKHS}. These algorithms are compared with the following methods capable to address (non)linear underdetermined BSS problem: (*i*) the NMR-NMU algorithm [46], i.e. the NMU [83] with a MATLAB code provided at [84] is applied to the squares of amplitude spectra of mixtures. The NMR-NMU assumes LMM and can extract more sources than mixtures spectra. Thus, for benchmark problem comprised of *M* sources we have separated 2*M* components used for annotation with the pure components from the library. Thus, for NMR-NMU it applies *R*=2*M*.; (*ii*) Sparse component analysis (SCA) method [38] capable to extract multiple sources from two linear mixtures. The algorithm estimates mixing matrix in the wavelet domain and ^{1}H NMR components amplitude spectra in Fourier domain solving linear programs at each chemical

shift. The critical assumption upon which SCA method is built is that for each source component at least one point in wavelet domain exists where only this component is dominantly present. This assumption is hard to fulfill for metabolic components that are structurally similar and correlated. In the experiment reported below and related to two-mixtures problem dimensionality of the latent space was *R*=6. It corresponds with the number of source components inferred from data directly. (*iii*) multi-view NMF method (mvNMF) [76]. This method treats each mapped mixture data as one view of the original mixture data. It assumes LMM for each view with view dependent mixing matrix but the same source components matrix for all the views. Hence, the method is less general than the ones proposed herein because it assumes that RKHSs induced with different kernels or same kernel with different parameters are equally suitable for all source components. Dimensionality of the latent space for this method is *R=D.* Multiple RKHSs were induced with Gaussian kernel and variances $\sigma_i^2 \in \{1.0, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001\}$. When single RKHS was used it was generated with Gaussian kernel with variance $\sigma_1^2 = 1$. We also have combined RKHSs induced with Gaussian kernel with the RKHSs induced with polynomial kernel with degrees $c_i \in \{1, 2, 3, 4, 5, 6, 7\}$. We named such algorithms EKM-GP-mRKHS-V_{RKHS} and EKM-GP-mRKHS-V_{Input}.

3.2 Recording of ^{1}H NMR spectra of 160 pure components

We have recorded in-house library comprised of ^{1}H NMR spectra of 160 pure components expected to correspond with small organic molecules present in samples such as tissue, blood, urine etc. Among them there are six pairs with amplitude spectra correlated above 0.9, eight pairs with correlation above 0.8, twelve pairs with correlations above 0.7, twenty-two pairs with

correlations above 0.6, thirty-four with correlations above 0.5 and fifty-nine with correlations above 0.4. Thus, spectral library contains many structurally similar components because of what annotation will be incorrect when separation quality is modes or poor. Library content is presented in Table S-5 of the Supplementary material. All measurements were performed on a Bruker AVANCE 600 MHz spectrometer, operating at 298 K. Samples were dissolved in 700 μL phosphate buffer (100 mM, pH 7.2 prepared with D_2O) prior to NMR measurement. 3-(Trimethylsilyl)-1-propanesulfonic acid sodium salt was used as an internal standard. Water suppression using excitation sculpting with gradients was applied [85]. ^{1}H spectra at a spectral width of 6.700 Hz with 16K data points and a digital resolution of 0.41 Hz per point were measured with 64 scans (time delay 2 sec, acquisition time 1.22 sec, pulse with 90).

3.3 ^{1}H NMR spectroscopy measurements of two model mixtures

To validate methods proposed for nonlinear BSS problems two mixtures of five pure components were prepared in the laboratory. Compounds 4-aminoantipyrine ($\mathbf{S}_1$), 4-aminobutyric acid ($\mathbf{S}_2$), allantoin ($\mathbf{S}_3$), cholic acid ($\mathbf{S}_4$) and naphtoic acid ($\mathbf{S}_5$) 30 mg of each, were mixed together. From the resulting crude mixture, 2 samples of 10 mg were taken and their NMR spectra were recorded as described above for the pure components. Mixture $\mathbf{X}_2$ was used for validation of single-mixture nonlinear BSS methods aEFM-EKM-mRKHS-V_{Input}, aEFM- EKM-sRKHS-V_{Input} and aEFM- EKM-sRKHS-V_{RKHS}.

3.4 Urine samples collection, preparation and ^{1}H NMR spectroscopy measurements

Urine aliquots were obtained from residual routine samples from 33 unrelated patients with diabetes type 2 (age range: 30 – 84 years; 17 males). Urine samples were collected in the morning, during the regular outpatient checkup in the clinical laboratory affiliated to the tertiary-level diabetes clinic. Patients were categorized and treated according to the current World Health Organization (WHO) recommendations at the University Clinic Vuk Vrhovac, Zagreb, that is the WHO collaborating center for diabetes. The study protocol was approved by the institutional Ethics Committee and patients gave their written consent for using their residual samples. The group of control subjects included 30 healthy, unrelated consenting adult volunteers, matched for age and sex to diabetic subjects. To each of them glucose level was measured before taking of urine and they were all normoglycemic. All study subjects were Caucasians. Morning urine samples were stored at -20^{0}C until clean-up procedure that is performed by C18 SampliQ Solid Phase Extraction (SPE) (Agilent Technologies, USA). C18 polymer sorbents was first conditioned by passing MeOH (3x5 mL) and then equilibrated by passing QH_2O (3x5 mL). Urine sample (3x5 mL) was loaded into the column and fraction after cleaning was collected in separate tubes. All the steps were performed at a flow rate of 1 mL min^{-1}. Thereafter, samples were frozen by immersion in liquid nitrogen followed by evaporation in vacuum chamber of freeze dryer to dryness (under controlled temperature and reduced pressure). 10 mg of each dry sample was further used for spectroscopic analysis. NMR spectra of urine samples were recorded as described for the pure components. Due to the clinical character of this problem only single-mixture method aEFM-EKM-mRKHS-V_{Input} was applied to 33 ^{1}H NMR mixtures spectra of urine obtained from diabetic patients and 30 ^{1}H NMR mixtures spectra of urine collected from subjects without diabetes.

3.5 Software environment

All the experiments were executed on the PC running under 64-bits Windows 10 operating system with 256 GB of RAM using Intel Xeon CPU E5-2650 v4 2 processors and operating with a clock speed of 2.2 GHz. All codes are run using MATLAB 2017a environment.

4. Results

4.1 Blind separation and annotation of five correlated amplitude ^{1}H NMR component spectra from two ^{1}H NMR model spectra

^{1}H amplitude NMR spectra of two model mixtures are shown in Figure 1. ^{1}H NMR amplitude spectra of pure components $\mathbf{S}_1$ to $\mathbf{S}_5$ as well as spectra of annotated components separated by the EKM-mRKHS-V_{RKHS} algorithm are shown in Figure 2, whereas dimension of each induced RKHS is D=2000. Table 1 summarizes separation and annotation results for dimension of induced RKHSs D=2000 in terms of criteria C1 to C4 and additional information related to computation time and correlation coefficients and ranking of annotated components. Corresponding results for dimensions D=100 and D=1000 are presented in Tables S-1 and S-2 in Supplementary material. The following conclusions are drawn from results presented in Tables 1, S-1 and S-2: (*i*) there is insignificant increase in quality of separation and annotation between dimensions of induced RKHSs D=1000 and D=2000. However, there is roughly threefold increase in computational complexity between D=2000 and D=1000. (*ii*) Combination of RKHSs induced with Gaussian and polynomial kernels brings insignificant increase in performance in comparison with RKHSs induced with Gaussian kernels only (EKM-mRKHS-V_{RKHS} vs. EKM-GP-mRKHS-V_{RKHS}). (*iii*) There is notable increase in performance when basis

is estimated for each induced RKHS separately (EKM-mRKHS-V_{RKHS} vs. EKM-mRKHS-V_{Input}). (*iv*) There is very significant increase in performance when separation is performed in multiple RKHSs as opposed to single RKHS (EKM-mRKHS-V_{RKHS} vs. EKM-sRKHS-V_{RKHS}). Separated components shown in Figure 2 and annotated to pure components 1 to 5 were selected from RKHSs induced by Gaussian kernel with variances in respective order 0.05, 0.01, 0.1, 0.005 and 0.05. Hence, a recommendation for separation of pure components ^{1}H NMR amplitude spectra from small number of their mixtures is to use either EKM-mRKHS-V_{RKHS} or EKM-mRKHS-V_{Input} algorithms with Gaussian kernel and $\sigma_i^2 \in \{1.0, 0.5, 0.1, 0.05, 0.01, 0.005, 0.001\}$ with dimension of individual RKHS $1000 \leq D \leq 2000$.

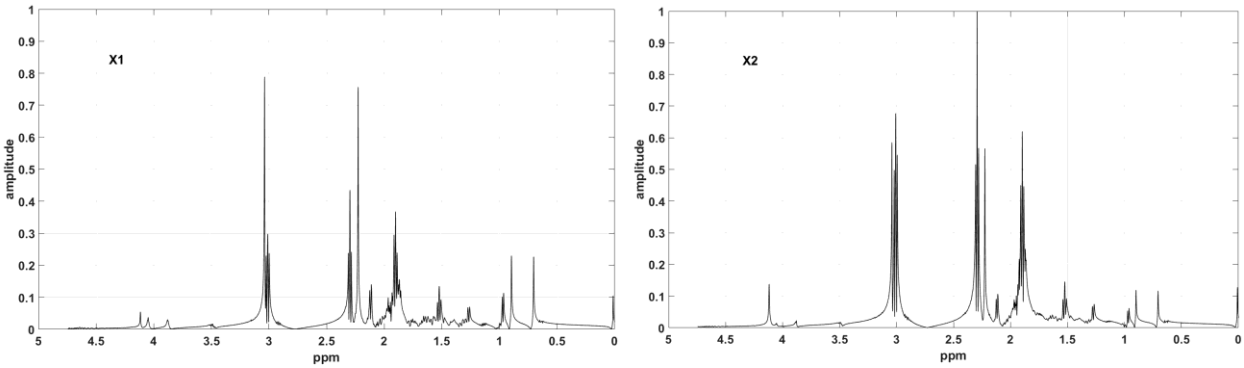

Figure 1. ^{1}H NMR amplitude spectra of two mixtures: $\mathbf{X}_1$ and $\mathbf{X}_2$.

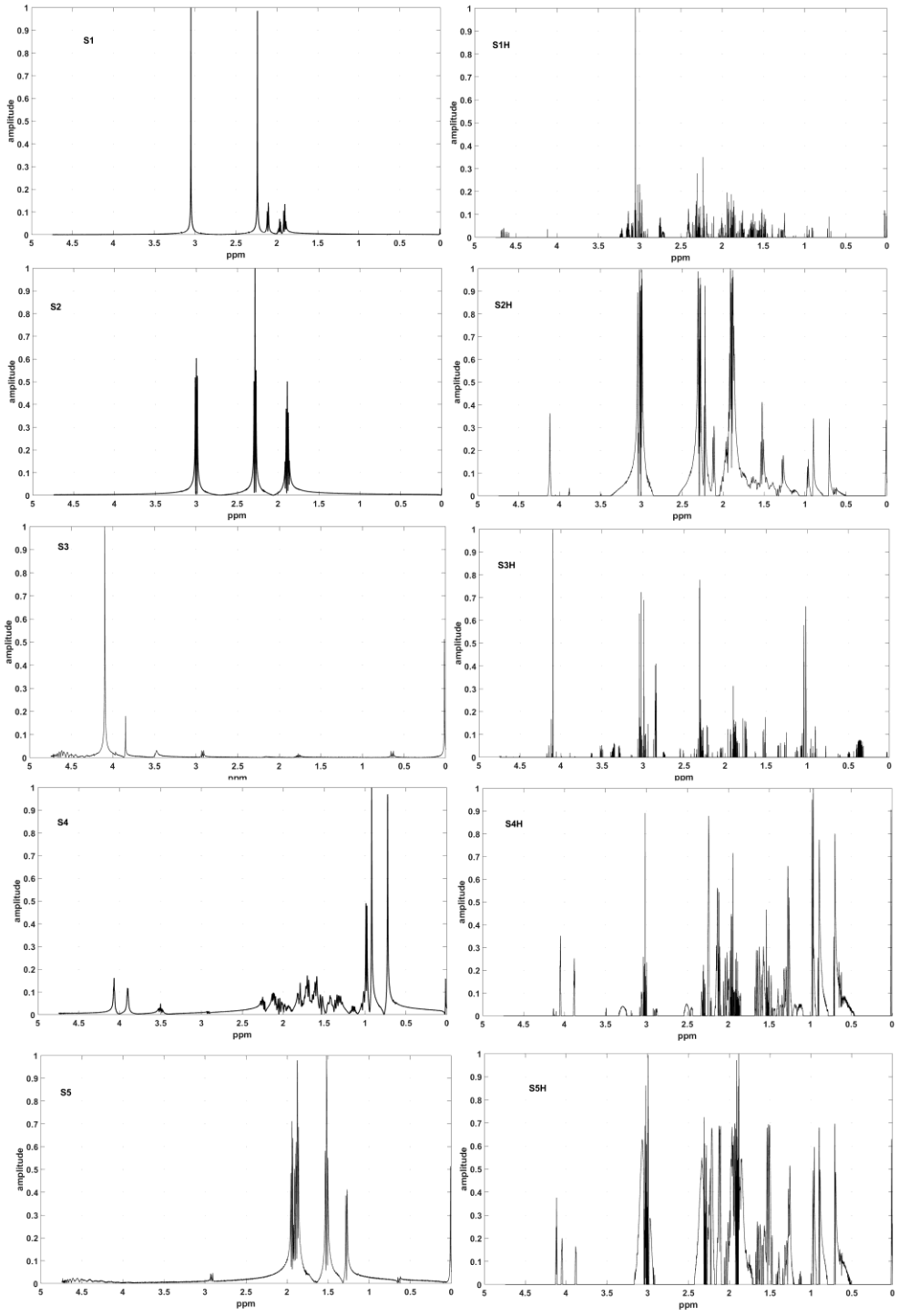

Figure 2. ^{1}H NMR amplitude spectra of pure components $\mathbf{S}_1$ to $\mathbf{S}_5$ (first column) and annotated components separated by the EKM-mRKHS-V_{RKHS} algorithm (second column). Dimension of each induced RKHS D=2000.

Table 1. Separation and annotation results from two ^{1}H NMR mixtures for dimension of induced RKHSs D=2000. Numerical codes assigned to acronyms of compared algorithms are as follows: 1: EKM-mRKHS-V_{RKHS} , 2: EKM-sRKHS-V_{RKHS}, 3: EKM-mRKHS-V_{Input}, 4: EKM-sRKHS-V_{Input} , 5: EKM-GP-mRKHS-V_{RKHS} , 6: EKM-GP-mRKHS-V_{Input} , 7: mvNMF-V_{RKHS} , 8: mvNMF -V_{Input}, 9: NMR-NMU and 10: SCA.

	1	**2**	**3**	**4**	**5**
C1	1	1	1	1	1
C2	0.1276	0.1354	0.1259	0.1354	0.1259
C3	0.4372	0.1354	0.3886	0.1354	0.4354
C4	0.0158	4.0005	0.0196	4.0005	0.0113
Ranking and correlations of correctly annotated components 1 to 5	23: 0.4461 1: 0.6295 64: 0.3491 114: 0.2341 19: 0.4641	NOT FOUND 1: 0.6772 NOT FOUND NOT FOUND NOT FOUND	16: 0.4457 1: 0.6297 125: 0.1946 119: 0.2125 14: 0.4769	NOT FOUND 1: 0.6772 NOT FOUND NOT FOUND NOT FOUND	27: 0.4753 1: 0.6295 67: 0.3684 123: 0.233 29: 0.4700
CPU time	167 745 s	24 366 s	150 159 s	21 667 s	333 977 s

Table 1. Continuation.

	6	**7**	**8**	**9**	**10**
C1	1	0	0	1	2
C2	0.1259	0	0	0.1151	0.1479
C3	0.4274	0.0293	0.0278	0.1151	0.1479
C4	0.0089	0.0223	0.0155	4.1	3.5
Ranking and correlations of correctly annotated pure components 1:5	15: 0.4894 2: 0.6297 82: 0.3193 131: 0.2285 19: 0.4701	78: 0.0282 113: 0.0267 18: 0.0318 91: 0.0284 12: 0.0327	77: 0.0256 17: 0.0293 18: 0.0292 78: 0.0255 27: 0.0286	NOT FOUND 1: 0.5754 NOT FOUND NOT FOUND NOT FOUND	1: 0.4570 2: 0.4305 NOT FOUND NOT FOUND NOT FOUND
CPU time	313 816 s	18 554 s	430 s	0.7155 s	291.9 s

4.2 Blind separation and annotation of five correlated amplitude ^{1}H NMR component spectra from one ^{1}H NMR mixture spectrum

As mentioned in section 3.3, mixture $\mathbf{X}_2$ was used for validation of single-mixture nonlinear BSS methods. Based on recommendations from section 4.3 only RKHSs induced with Gaussian kernels with variances $\sigma_i^2 \in \{1.0, 0.5, 0.1, 0.05\}$ were used to evaluate single-mixture BSS algorithms. There was not enough diversity in EKMs (23) generated with smaller variances. Due to the same reason it was infeasible to estimate basis $\mathbf{V}$ using kernel *k*-means algorithm for $\sigma^2<1$. Thus, only algorithms aEFM-EKM-mRKHS-V_{Input}, aEFM- EKM-sRKHS-V_{Input} and aEFM-

EKM-sRKHS-V_{RKHS} were compared herein. Evidently, clinically the most relevant single-mixture scenario has limitation in comparison with multiple mixtures scenario. Table 2 summarizes separation and annotation results for dimension of induced RKHSs D=2000 in terms of criteria C1 to C4 and additional information related to computation time and correlation coefficients and ranking of annotated components. Corresponding results for dimensions D=100 and D=1000 are presented in Tables S-3 and S-4 in Supplementary material. Only dimension of individual induced RKHSs equal to D=2000 enabled detection of three (out of five) pure components with all three BSS algorithms where the aEFM-EKM-mRKHS-V_{Input} has best performance. In agreement with the "no free lunch theorem" this algorithm has highest computational complexity. It can be seen from Tables 1 and 2 that separated components annotated with the true pure components are mostly not placed at the top of the ranking list. Thus, in the real world scenario related to separation and annotation of metabolites from single ^{1}H NMR mixture of biological sample such as urine, an interpretation of the list of ranked annotated components by domain expert will be necessary.

Table 2. Separation and annotation results from one ^{1}H NMR mixture for dimension of induced RKHSs D=2000. Numerical codes assigned to acronyms of compared algorithms are as follows: 1: aEFM-EKM-mRKHS-V_{Input}, 2: aEFM- EKM-sRKHS-V_{Input} and 3: aEFM- EKM-sRKHS-V_{RKHS}.

	1	**2**	**3**
C1	1	1	1
C2	0.1185	0.1186	0.1188
C3	0.2290	0.2234	0.2420
C4	2.005	2.454	2.1625
Ranking and correlations of correctly annotated components 1 to 5	28: 0.2327 1: 0.5925 NOT FOUND NOT FOUND 13: 0.3198	634: 0.2072 1: 0.5931 1985: 0.0355 NOT FOUND 288: 0.2810	276: 0.2594 1: 0.5934 NOT FOUND NOT FOUND 48: 0.3572
CPU time	42 915 s	24 261 s	26 583 s

4.3 Blind separation and annotation of correlated amplitude ^{1}H NMR component spectra from one ^{1}H NMR mixture spectrum of urine of diabetic and non-diabetic subjects

Following discussion in section 4.2 we applied the aEFM-EKM-mRKHS-V_{Input} algorithm to separate and annotate components present in the single ^{1}H NMR spectra of 33 urine samples of patients with diabetes type II and 30 urine samples of non-diabetic subjects. Also based on discussion in section 4.2, dimension of individual induced RKHs was selected to be D=2000. We

provide in Tables S-6 to S-68 in Supplementary materials results of separation and annotation of 55 metabolites, expected to be related to diabetes, obtained by means of the aEFM-EKM-mRKHS-V_{Input} algorithm from each individual ^{1}H NMR spectra. Summarized results for all 55 metabolites are presented in Table S-69 in Supplementary materials. The most prominent metabolites in samples from diabetic subjects, when compared to healthy controls, were urinary creatine, glutamic acid and 5-hydroxyindoleacetic acid. Table 3 presents aggregate separation and annotation related performance measures: number of times detected, mean and median ranks in the latent space composed of 160 pure components (size of the library) as well as mean and median correlation between separated and annotated pure components. It is seen from the correlation values that related nonlinear single mixture BSS problem is very hard. Nevertheless, metabolites such as urinary creatine, glutamic acid and 5-hydroxyindoleacetic acid are detected in practically all the spectra and are more distinguished in spectra of urine of diabetic patients.

Table 3. Separation and annotation performance of metabolites urinary creatine, glutamic acid and 5-hydroxyindoleacetic acid obtained by means of aEFM-EKM-mRKHS-V_{Input} algorithm from ^{1}H NMR spectra of 33 urine samples of patients with diabetes type II and 30 urine samples of non-diabetic subjects.

Metabolite	**Number of times detected**		**Mean rank / Median rank**		**Mean correlation / Median correlation**	
	33 diabetic patients	**30 control subjects**	**Diabetic patients**	**Control subjects**	**Diabetic patients**	**Control subjects**
creatine	32	30	10.78 / 7	21.37 / 13	0.319 / 0.318	0.287 / 0.286
glutamic acid	32	29	20.93 / 8	31.45 /21	0.274 / 0.299	0.224 / 0.211
5-hydroxyindoleacetic acid	32	30	31.5 / 15.5	39.17 / 28	0.226 / 0.260	0.194 / 0.193

5. Discussion

Metabolomic studies of diabetes and metabolic syndrome, using both targeted and non-targeted approach by either mass spectrometry or ^{1}H NMR spectroscopy so far demonstrated the significant association of plasma branched chain amino acids: isoleucine, leucine and valine, as well as two aromatic amino acids: tyrosine and phenylalanine with the development of type 2 diabetes [86]. Furthermore, lipidomic-oriented studies identified plasma glycine, lysophosphatidylcholine 18:2 and acetylcarnitine as predictors of prediabetes and type 2 diabetes, [87]. Several studies reported on the associations between various phospholipids, hexoses and metabolites generated from oxidative damage, such as 2-aminoadipic acid, with incident diabetes

[88, 89]. These plasma metabolites were linked to the organ-specific processes and pathways involved in the pathogenesis of type 2 diabetes [86]. Urinary metabolic profiling in diabetes is less prominent. That is partly because of the complexity of matrix, containing approximately 3100 so far identified metabolites [90] and partly because of the limitations of current methodology, both analytical and computational in separation of the signals generated by structurally similar molecules. The nonlinear single-mixture BSS method proposed herein was able to distinguish 3 metabolites which are involved in diverse pathways relevant for diabetes pathogenesis: urinary creatine, glutamic acid and 5-hydroxyindoleacetic acid. Glutamic acid, in the form of its monosodium salt is a well-established neurotransmitter responsible for the synaptic plasticity. It has been hypothesized that abnormal glutamate homeostasis might contribute to diabetes pathogenesis by direct and indirect mechanisms mediating a progressive loss of insulin-producing pancreatic β-cells [91]. Recent study provided evidence on an increased plasma glutamate level in diabetic patients and mice, as well as β-cell lines following short-term exposure to high glucose *in vitro.* Enzymatic degradation of glutamate was able to normalize insulin secretion [92]. A toxic effect of an excess of glutamate in retinal cells was proposed as one of the mechanisms involved in the pathogenesis of diabetic retinopathy [93]. Thus, it seems that elevated level of glutamate plays a significant role in diabetes pathology. Urinary 5-hydroxyindoleacetate (5-HIAA) is an established indicator of serotonin levels and is routinely used as a laboratory test for carcinoid tumor diagnosis. Serotonin, synthesized by tryptophane hydroxylation in the brainstem serves as a neurotransmitter involved in regulation of multiple physiological functions of the brain, such as behavior and learning, as well as appetite and glucose homeostasis. However, peripherally produced serotonin serves as a hormone, which is involved in the regulation of function of the organs involved in the metabolic homeostasis at both glucose and lipid level [94]. A process called serotonylation was identified as an important

modulating mechanism of the insulin production and secretion within the β-cells [95]. It was reported that high level of plasma 5-HIAA in the stage of metabolic syndrome indicate a deranged serotonin metabolism with a presumed significant role for the development of cardiovascular complications via serotonin-mediated enhanced platelet aggregation and vasoconstriction [96]. Furthermore, regarding diabetes, it was recently proposed that an increased plasma 5-HIAA level in diabetic patients may play a role in the pathogenesis of microvascular complications [97]. The accumulated body of evidence pinpoints serotonin as a potential therapeutic target for type 2 diabetes and obesity [98]. Creatine (N-methyl-N-guanyxlglycine) is an essential guanidine compound widely distributed throughout human cells, which is equally provided by dietary sources and endogenous synthesis from arginine and glycine [99]. Phosphorylated creatine serves as the major endogenous phosphagenic substrate necessary for ATP synthesis within pathway catalyzed by creatine kinase. Creatine depletion, either acquired or inherited, seems to affect a variety of organs, with muscle and brain being the most interesting targets [100, 101]. Despite a pronounced popularity, presumed improvement of muscle mass and athletic performance by the oral supplementation of creatine remained ambiguous, but the widespread use of creatine for fitness purposes demonstrated its safety in healthy adults [102]. It was recently proposed that creatine deficiency due to the aging-related reduction of muscular mass may be responsible for age-related neurodegenerative diseases, and creatine supplementation emerged as an interesting treatment approach for a variety of geriatric disorders [103]. Pleiotropic effects of creatine seem to go beyond the creatine-kinase system of energy metabolism and involve various metabolic pathways, including glucose homeostasis [104]. Studies carried out in newly-diagnosed patients with type 2 diabetes demonstrated that short-term oral ingestion of creatine elicited a reduction of plasma glucose in which was equal to the effects obtained by two common oral antihyperglycaemic agents: sulfonylurea [105] and metformin

[106]. As evidenced in the recent meta-analysis [107], longer-term supplementation of creatine yielded indeterminate results regarding glycemic control, but creatine supplementation could be regarded as an adjuvant nutritional therapy with hypoglycemic effects, particularly when used in combination with exercise. *In vitro* studies revealed that creatine was able to improve glucose-stimulate insulin release [108] and to facilitate translocation of muscular glucose transporter GLUT4 [109]. More recent research showed that AMPK signaling may be implicated in the GLUT4 effects of creatine supplementation on glucose uptake in type 2 diabetes [110]. However, the mechanism(s) involved in the glucoregulatory action of creatine is far from being elucidated. Results of the present study indicate that urinary creatine secretion was significantly more pronounced in diabetic patients than healthy controls, which is a novel finding. Considering so far collected evidence on the role of creatine on the glucose homeostasis, it could be speculated that type 2 diabetes may be associated with a disturbed utilization of creatine associated with an increased renal loss, possibly due to glomerular hyperfiltration, which is commonly associated with diabetes [111].

6. Conclusions

Blind separation of structurally similar (overlapping) components from small number of their nonlinear mixtures is a hard inverse problem. It becomes notoriously difficult when only single mixture is at a disposal. Yet, separation of structurally similar components from a single nonlinear mixture is of potentially high clinical relevance and it is known as metabolic profiling. Driven by this motivation, the paper presented methodology for the blind separation and annotation of components present in the ^{1}H NMR amplitude mixture spectra. In addition to

model (laboratory prepared) mixtures the methodology was tested on separation and annotation of metabolites present in urinary samples collected from diabetic patients and healthy controls. The ability of our method to identify metabolite-related differences between the groups, albeit in the very early pilot-stage, revealed an interesting and novel pattern of metabolic components within various pathways, which are known to be influenced by diabetes. In particular, the method pinpointed urinary creatine, glutamic acid and 5-hydroxyindoleacetic acid as the most prominent metabolites in samples from diabetic subjects, when compared to healthy controls. Since presented study is at a pilot stage, our results do not allow any metabolic interpretation. However, our method was able to differentiate diabetic from non-diabetic subjects by identifying potentially relevant metabolites depicting pathways relevant for diabetes pathology. Further studies are needed to validate this method in terms of obtaining reproducible and clinically relevant results.

Conflict of interest

Authors declare no conflict of interest.

Acknowledgments

The work performed has been supported through grant IP-2016-06-5235 "Structured decompositions of empirical data for computationally-assisted diagnosis of disease" funded by the Croatian Science Foundation.

Appendix A. Supplementary material

Supplementary material associated with this article can be found, in the online version, at http://

References

[1] J. K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Metabonomics: a platform for studying drug toxicity and gene function, Nat. Rev. Drug Discovery 1 (2002) 153-161. https://doi.org/10.1038/nrd728.

[2] J. Keiser, U. Duthale, J. Utzinger, Update on the diagnosis and treatment of food-bone trematode infections, Curr. Opin. Infect. Dis. 23 (2010) 513-520. https://doi.org/10.1097/QCO.0b013e32833de06a.

[3] D. G. Robertson, Metabonomics in Toxicology: A Review, Toxicol. Sci. 82 (2005) 809-822. https://doi.org/10.1093/toxsci/kfi102.

[4] T. Hyotylainen, Novel methods in metabolic profiling with a focus on molecular diagnostic applications, Expert Rev. Mol. Diagn. 12 (2012) 527-538. https://doi.org/10.1586/erm.12.33.

[5] N.R. Patel, M.J.W. McPhail, M.I.F. Shariff, H.C. Keun, S.D. Taylor-Robinson, Biofluid metabpnomics using ^{1}H NMR spectroscopy: the road to biomarker discovery in gastroenterology and hepatology, Expert Rev. Gastroenterol. Hepatol. 6 (2012) 239-251. https://doi.org/10.1586/egh.12.1.

[6] D. S. Wishart, Metabonomics: applications to food science and nutrition research, Trends Food Sci. Technol. 19 (2008) 482-493. https://doi.org/10.1016/j.tifs.2008.03.003.

[7] S. Durand, M. Sancelme, P. Besse-Hoggan, B. Combourieu, Biodegradation pathaway of mesotrione: Complementaries of NMR, LC-NMR and LC-MS for qualitative and quantitaive

metabolic profiling, Chemosphere 81 (2010) 372-380. https://doi.org/10.1016/j.chemosphere.2010.07.017.

[8] P.-M.Nguyen, C. Lyathaud, O. Vitrac, A Two-Scale Pursuit Method for the Tailored Identification and Quantification of Unknown Polymer Additives and Contaminants by ^{1}H NMR, Indust. & Eng. Chem. Res. 54 (2015) 2667-2681. https://doi.org/10.1021/ie503592z.

[9] T. Gebregiworgis, R. Powers, Application of NMR metabolomics to search for human disease biomarkers, Combinatorial Chemistry & High Throughput Screening 15 (2012) 595-610. https://doi.org/10.2174/138620712802650522.

[10] A. Roux, Y. Xu, J.-F. Heilier, M.-F. Olivier, E. Ezan, J.-C. Tabet, C. Junot, Annotation of the human adult urinary metabolome and metabolite identification using ultra high performance liquid chromatography coupled to a linear quadrupole ion trap-orbitrap mass spectrometer, Anal. Chem. 84 (2012) 6429−6437. https://doi.org/10.1021/ac300829f.

[11] B. R. Seavey, E. A. Farr, W. M. Westler, J. L. Markley, A relational database for sequence-specific protein NMR data, J. Biomol. NMR. 1 (1991) 217-236. https://doi.org/10.1007/BF01875516.

[12] SpecInfo on the Internet NMR, https://application.wiley-vch.de/stmdata/specinfo.php (accessed December 28 2018).

[13] I. Toumi, S. Caldarelli, B. Torrésani, A review of blind source separation in NMR spectroscopy, Progress in Nuc. Mag. Res. Spec. 81 (2014) 37-64. https://doi.org/10.1016/j.pnmrs.2014.06.002.

[14] N. M. Jukarainen, S. -P. Korhonen, M. P. Laakso, M. A. Korolainen, M. Niemitz, P. P. Soininen, K. Tuppurainen, J. Vepsäläinen, T. Pirttilä, R. Laatikainen, Quantification of ^{1}H NMR spectra of human cerebrospinal fluid: a protocol based on constrained total-line-shape analysis, Metabolomics 1 (2008) 150-160. https://doi.org/10.1007/s11306-008-0106-6.

[15] E. E. Kwan, S. G. Huang, Structural Elucidation with NMR Spectroscopy: Practical Strategies for Organic Chemists. Eur. J. org. Chem. 39 (2008) 2671-2688. https://doi.org/10.1002/ejoc.200700966.

[16] G. F. Pauli, B. U. Jaki, D. C. Lankin, Quantitative 1H NMR: development and potential of a method for natural products analysis, J. Nat. Prod. 68 (2004) 133-149. https://doi.org/10.1021/np0497301.

[17] A. Smolinksa, L. Blanchet, L. M. C. Buydens, S. S. Wijmenga, NMR and pattern recognition methods in metabolomics: From data acquistion to biomarker discovery: A review, Anal. Chim. Acta. 750 (2012) 82-97. https://doi.org/10.1016/j.aca.2012.05.049.

[18] R. Schicho, R. Shaykhutdinov, J. Ngo, A. Nazyrova, C. Schneider, R. Panaccione, G. G. Kaplan, H. J. Vogel, M. Storr, Quantitaive Metabolomic Profiling of Serum, Plasma and Urine by ^{1}H NMR Spectroscopy Discriminates between Patients with Inflammatory Bowel Disease and Healthy Individuals, J. Proteome Res. 11 (2012) 3344-3357. https://doi.org/10.1021/pr300139q.

[19] N. MacKinnon, B. S. Somashekar, P. Tripathi, W. G. Thekkelnaycke, M. R. Arul, M. Chinnaiyan, A. Ramamoorthy, MetabolID: A graphical user interface package for assignment of ^{1}H NMR spectra of bodyfluids and tissues, J. Magn Reson. 226 (2013) 93-99. https://doi.org/10.1016/j.jmr.2012.11.008.

[20] G. A. Gowda Nagana, S. Zhang, H. Gu, V. Asiago, S. Narasimhamurty, D. Raftery, Metabolomics-based methods for early disease diagnostics, Expert Rev. Mol. Diag. 8 (2008) 617-633. https://doi.org/10.1586/14737159.8.5.617.

[21] M. E. Elyashberg, A. J. Williams, G. E. Martin, Computer-assisted structure verification and elucidation tools in NMR-based structure elucidation, Prog. Nucl. Magn. Reson. Spectrosc. 53 (2008) 1-104. https://doi.org/10.1016/j.pnmrs.2007.04.003.

[22] S. Kim, I Koo, J. Jeong, S. Wu, X. Shi, X. Zhang, Compound Identification Using Partial and Semipartial Correlations for Gas Chromatography-Mass Spectrometry Data, Anal. Chem. 84 (2012) 6477-6487. https://doi.org/10.1021/ac301350n.

[23] C. Shao, W. Sun, F. Li, R. Yang, L. Hnag, Y. Gao, Oscore: a combined score to reduce false negative rates for peptide identification in tandem mass spectrometry analysis, J. Mass Spectrom. 44 (2009) 25-31. https://doi.org/10.1002/jms.1466.

[24] J. Razumovskaya, V. Olman, D. Xu, E. C. Uberbacher, N. C. VerBerkmoes, R. L. Hettich, Y. Xu, A computational method for assessing peptide-identification reliability in tandem mass spectrometry analysis with SEQUEST, Proteomics, 4 (2004) 961-969. https://doi.org/10.1002/pmic.200300656.

[25] T. Baczek, A. Bucinski, A. R. Ivanov, R. Kaliszan, Artificial Neural Network Analysis for Evaluation of Peptide MS/MS Spectra in Proteomics, Anal. Chem. 76 (2004) 1726-1732. https://doi.org/10.1021/ac030297u.

[26] V. A. Likić, Extraction of pure components from overlapped signals in gas chromatography-mass spectrometry (GC-MS), BioData Min. 2 (2009) 6, https://doi.org/10.1186/1756-0381-2-6.

[27] B. Bracewell, Fourier transform and its applications, McGraw-Hill, New York, US, 1999.

[28] V. A. Shashilov, I. K. Lednev, Advanced Statistical and Numerical Methods for Spectroscopic Characterization of Protein Structural Evaluation, Chem. Rev. 110 (2010) 5692-5712. https://doi.org/10.1021/cr900152h.

[29] P. Comon, C. Jutted (Eds), Handbook of Blind Source Separation, Academic Press, Oxford, UK, 2010.

[30] I. T. Joliffe, Principal Component Analysis, Springer Series in Statistics, second ed., Springer, New York, 2002.

[31] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.

[32] P. Comon, Independent component analysis, A new concept?, Sig. Proc. 36 (1994) 287-314. https://doi.org/10.1016/0165-1684(94)90029-9.

[33] M. Zibulevsky, B. A. Pearlmutter, Blind Source Separation by Sparse Decomposition, Neural Comput. 13 (2001) 863-882. https://doi.org/10.1162/089976601300014385.

[34] P. Georgiev, F. Theis, A. Cichocki, Sparse Component Analysis and Blind Source Separation of Underdetermined Mixtures, IEEE Trans. Neural Net. 16 (2005) 992-996. https://doi.org/10.1109/TNN.2005.849840.

[35] A. Cichocki, R. Zdunek, A. H. Phan, S. I. Amari, Nonnegative Matrix and Tensor Factorizations, John Wiley, Chichester, 2009.

[36] D. Nuzillard, S. Bourg, J. M. Nuzilard, Model-Free Analysis of Mixtures by NMR Using Blind Source Separation, J. Magn. Res. 133 (1998) 358-363. https://doi.org/10.1006/jmre.1998.1481.

[37] E. Visser, T. W. Lee, An information-theoretic methodology for the resolution of pure component spectra without prior information using spectroscopic measurements, Chemom. Int. Lab. Syst. 70 (2004) 147-155. https://doi.org/10.1016/j.chemolab.2003.11.003.

[38] I. Kopriva, I. Jerić, Multi-component Analysis: Blind Extraction of Pure Components Mass Spectra using Sparse Component Analysis, J. Mass Spectrom. 44 (2009) 1378-1388. https://doi.org/10.1002/jms.1627.

[39] I. Kopriva, I. Jerić, Blind Separation of Analytes in Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry: Sparseness-Based Robust Multicomponent Analysis, Anal. Chem. 82 (2010) 1911-1920. https://doi.org/10.1021/ac902640y.

[40] I. Kopriva, I. Jerić, V. Smrečki, Extraction of multiple pure component ^{1}H and ^{13}C NMR spectra from two mixtures: Novel solution obtained by sparse component analysis-based blind decomposition, Anal. Cim. Acta 653 (2009) 143-153. https://doi.org/10.1016/j.aca.2009.09.019.

[41] Snyder, D. A., Zhang, F., Robinette, S. L., Brüschweiler-Li, L., Brüschweiler, R., 2008. Non-negative matrix factorization of two-dimensional NMR spectra: Application to complex mixture analysis. J. Chem. Phys. 128, 052313. https://doi.org/10.1063/1.2816782.

[42] L. Guo, A. Wiesmath, P. Sprenger, M. Garland, Development of 2D Band-Target Entropy Minimization and Application to Deconvolution of Multicomponent 2D Nuclear Magnetic Resonance Spectra, Anal. Chem. 77 (2005) 1655-1662. https://doi.org/10.1021/ac0491814.

[43] W. Naanaa, J. M. Nuzillard, Blind source separation of positive and partially correlated data, Sig. Proc. 85 (2005) 1711-1722. https://doi.org/10.1016/j.sigpro.2005.03.006.

[44] M. S. Karoui, Y. Deville, S. Hosseini, S. Ouamri, Blind spatial unmixing of multispectral images: New methods combining sparse component analysis, clustering and nonnegativity constraints, Patt. Recog. 45 (2012) 4263-4278. https://doi.org/10.1016/j.patcog.2012.05.008.

[45] Y. Sun, C. Ridge, F. del Rio, A. J. Shaka, J. Xin, Postprocessing and sparse blind source separation of positive and partially overlapped data, Sig. Proc. 91 (2011) 1838-1851. https://doi.org/10.1016/j.sigpro.2011.02.007.

[46] I. Kopriva, I. Jerić, Blind Separation of Analytes in Nuclear Magnetic Resonance Spectroscopy: Improved Model for Nonnegative Matrix Factorization, Chem. Int. Lab. Syst. 137 (2014) 47-56. https://doi.org/10.1016/j.chemolab.2014.06.004.

[47] B. Schölkopf, A. Smola, Learning with kernels. MIT Press, Cambridge, MA, US, 2002.

[48] I. Kopriva, I. Jerić, L. Brkljačić, Nonlinear mixture-wise expansion approach to underdetermined blind separation of nonnegative dependent sources, J. Chemometrics 27 (2013) 189-197. https://doi.org/10.1002/cem.2512.

[49] I. Kopriva, I. Jerić, M. Filipović, L. Brkljačić, Empirical Kernel Map Approach to Nonlinear Underdetermined Blind Separation of Sparse Nonnegative Dependent Sources: Pure Components Extraction from Nonlinear Mixtures Mass Spectra, J. Chemometrics 28 (2014) 704-715. https://doi.org/10.1002/cem.2635.

[50] I. Kopriva, I. Jerić, L. Brkljačić, Explicit-Implicit Mapping Approach to Nonlinear Blind Separation of Sparse Nonnegative Dependent Sources from a Single-Mixture: Pure Components

Extraction from Nonlinear Mixture Mass Spectr, J. Chemometrics 29 (2015) 615-626. https://doi.org/10.1002/cem.2745.

[51] K. Zhang, L. Chan, Minimal Nonlinear Distortion Principle for Nonlinear Independent Component Analysis, J. Mach. Learn. Res. 9 (2008) 2455-248.

[52] Levin, D. N., 2008. Using state space differential geometry for nonlinear blind source separation, J. Appl. Phys. 103, 044906. https://doi.org/10.1063/1.2826943.

[53] D. N. Levin, Performing Nonlinear Blind Source Separation with Signal Invariants, IEEE Trans. Sig. Proc. 58 (2010) 2132-2140. https://doi.org/10.1109/TSP.2009.2034916.

[54] A. Taleb, C, Jutten, Source Separation in Post-Nonlinear Mixtures, IEEE Trans. Sig. Proc. 47 (1999) 2807-2820. https://doi.org/10.1109/78.790661.

[55] L. T. Duarte, R. Suyama, B. Rivet, R. Attux, J. M. T. Romano, C. Jutten, Blind compensation of nonlinear distortions: applications to source separation of post-nonlinear mixtures, IEEE Trans. Sig. Proc. 60 (2012) 5832-5844. https://doi.org/10.1109/TSP.2012.2208953.

[56] E. F. S. Filho, J. M. de Seixas, L. P. Calôba, Modified post-nonlinear ICA model for online neural discrimination, Neurocomputing 73 (2010) 2820-2828. https://doi.org/10.1016/j.neucom.2010.03.025.

[57] V. T. Nguyen, J. C. Patra, A. Das, A post nonlinear geometric algorithm for independent component analysis, Digital Sig. Proc. 15 (2005) 276-294. https://doi.org/10.1016/j.dsp.2004.12.006.

[58] A. Ziehe, M. Kawanabe, S. Harmeling, K. R. Müller, Blind Separation of Post-Nonlinear Mixtures Using Gaussianizing Transformations And Temporal Decorrelation, J. Mach. Learn. Res. 4 (2003) 1319-1338.

[59] K. Zhang, L. W. Chan, Extended Gaussianization Method for Blind Separation of Post-Nonlinear Mixtures, Neural Comput. 17 (2005) 425-452. https://doi.org/10.1162/0899766053011500.

[60] B. Ehsandoust, M. Babaie-Zadeh, B. Rivet, C. Jutten, Blind Source Separation in Nonlinear Mixtures: Separability and a Basic Algorithm, IEEE Trans. Sig. Proc. 65 (2017) 4339-4352. https://doi.org/10.1109/TSP.2017.2708025.

[61] S. Harmeling, A. Ziehe, M. Kawanabe, Kernel-Based Nonlinear Blind Source Separation, Neural Comput. 15 (2003) 1089-1124. https://doi.org/10.1162/089976603765202677.

[62] D. Martinez, A. Bray, Nonlinear Blind Source Separation Using Kernels, IEEE Tr. Neural Net. 14 (2003) 228-235. https://doi.org/10.1109/TNN.2002.806624.

[63] H. G. Yu, G. M. Huang, J. Gao, Nonlinear Blind Source Separation Using Kernel Multi-set Cannonical Correlation Analysis, Int. J. Comp. Net. Inf. Sec. 1 (2010) 1-8.

[64] L. Almeida, MISEP-Linear and nonlinear ICA based on mutual information, J. Mach. Learn. Res. 4 (2003) 1297-1318.

[65] M. E. Davies, C. I. James, Source separation using single channel ICA, Sig. Proc. 87 (2007) 1819-1832. https://doi.org/10.1016/j.sigpro.2007.01.011.

[66] Y. C. Ouyang H. M. Chen, J. W. Chai, C. C. C. Chen, S. K. Poon C. W. Yang, S. K. Lee, C. I. Chang, Band Expansion-Based Over-Complete Independent Component Analysis for

Multispectral Processing of Magnetic Resonance Image, IEEE Trans. Biomed. Eng. 55 (2008) 1666-1677. https://doi.org/10.1109/TBME.2008.919107.

[67] B. Mijović, M. De Vos, I. Gligorijević J. Taelman , S. Van Huffel, Source Separation from Single-Channel Recordings by Combining Empirical Mode Decomposition and Independent Component Analysis, IEEE Trans. on Biomed. Eng. 57 (2010) 2188-2196. https://doi.org/10.1109/TBME.2010.2051440.

[68] J. Lin, A. Zhang, Fault feature separation using wavelet-ICA filter, NDT&E International 38 (2005) 421-427. https://doi.org/10.1016/j.ndteint.2004.11.005.

[69] Q. He, S. Su, R. Du, Separating mixed multi-component signal with an application in mechanical watch movement, Dig. Sig. Proc. 18 (2008) 1013-1028. https://doi.org/10.1016/j.dsp.2008.04.009.

[70] D. Gunawan, D. Sen, Iterative Phase Estimation for the Synthesis of Separated Sources from Single-Channel Mixtures, IEEE Sig. Proc. Let. 17 (2010) 421-424. https://doi.org/10.1109/LSP.2010.2042530.

[71] R. M. Parry, I. Essa I, Phase-Aware Non-negative Spectrogram Factorization, Lect. Notes Comp. Sci. 4666 (2007) 536-543. https://doi.org/10.1007/978-3-540-74494-8_67.

[72] B. Gao, W. L. Woo, B. W. K. Ling, Machine Learning Source Separation Using Maximum A Posteriori Nonnegative Matrix Factorization, IEEE Trans. on Cybernetics 44 (2014) 1169-1179. https://doi.org/10.1109/TCYB.2013.2281332.

[73] G. J. Jang, T. W. Lee, A Maximum Likelihood Approach to Single-channel Source Separation, J. Machine Learn. Res. 4 (2003) 1365-1392.

[74] S. T. Roweis, One microphone source separation, Advance in Neural Information Processing Systems 13 (2000) 793-799.

[75] E. M. Grai, H. Erdogan, Source separation using regularized NMF with MMSE estimates under GMM priors with online learning from uncertainties, Dig. Sig. Proc. 29 (2014) 20-34. https://doi.org/10.1016/j.dsp.2014.02.018.

[76] I. Kopriva, Joint Nonnegative Matrix Factorization for Underdetermined Blind Source Separation in Nonlinear Mixtures, LNCS 10891 (2018) 107–115. https://doi.org/10.1007/978-3-319-93764-9_11.

[77] C. Caifa, A. Cichocki, Estimation of Sparse Nonnegative Sources from Noisy Overcomplete Mixtures Using MAP, Neural Comput. 21 (2009) 3487-3518. https://doi.org/10.1162/neco.2009.08-08-846.

[78] R. A. DeVore, Deterministic constructions of compressed sensing matrices, J. Complexity 23 (2007) 918-925. https://doi.org/10.1016/j.jco.2007.04.002.

[79] C. A. Micchelli, Y. Xu, H. Zhang, Universal Kernels, J. Mach. Learn. Res. 7 (2006) 2651-2667.

[80] Kopriva, I., Popović Hadžija, M., Hadžija, M., Aralica, G., 2015. Unsupervised segmentation of low-contrast multichannel images: discrimination of tissue components in microscopic image of unstained specimen. Scientific Reports 5, 11576. https://doi.org/10.1038/srep11576.

[81] L. He, H. Zhang, Kernel K-Means Sampling for Nyström Approximation, IEEE Trams. Image Proc. 27 (2018) 2108-2120. https://doi.org/10.1109/TIP.2018.2796860.

[82] R. Chitta, R. Jin, T. C. Havens, A. K. Jain, Approximate Kernel *k*-means: Solution to Large Scale Kernel Clustering, Proceedings of the 17th ACM SIGKDD conference on Knowledge Discovery and Data mining 2011, pp. 895–903, https://doi.org/10.1145/2020408.2020558.

[83] N. Gillis, F. Glineur, Using underapproximations for sparse nonnegative matrix factorization, Pattern Recog. 43 (2010) 1676-1687. https://doi.org/10.1016/j.patcog.2009.11.013.

[84] The Nicolas Gillis web site: https://sites.google.com/site/nicolasgillis/code (accessed December 24 2018).

[85] T. L. Hwang, A. J. Shaka, Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients, J. Magn. Res. Series A 112 (1995) 275-279. https://doi.org/10.1006/jmra.1995.1047.

[86] K. Suhre, Metabolic profiling in diabetes, J Endocrinol 221 (2014) R75–85. https://doi.org/10.1530/JOE-14-0024.

[87] R. Wang-Sattler, Z. Yu, C. Herder, A. C. Messias, A. Floegel, Y. He, et al., Novel biomarkers for pre-diabetes identified by metabolomics, Mol Syst Biol 8 (2012) 615, https://doi.org/10.1038/msb.2012.43.

[88] A. Floegel, A. von Ruesten, D. Drogan, M. B Schulze, C. Prehn, J. Adamski, et al., Variation of serum metabolites related to habitual diet: a targeted metabolomic approach in EPIC-Potsdam, Eur J Clin Nutr 67 (2013) 1100–1108. https://doi.org/10.1038/ejcn.2013.147.

[89] T. J. Wang, D. Ngo, N. Psychogios, A. Dejam, M. G. Larson, R. S. Vasan, et al., 2-Aminoadipic acid is a biomarker for diabetes risk, J. Clin. Invest. 123 (2013) 4309–4317. https://doi.org/0.1172/JCI64801.

[90] Bouatra, S., Aziat, A., Mandal, R., Guo, A. C., Wilson, M. R., Knox, C., et al., 2013. The Human Urine Metabolome 8, e73076. https://doi.org/10.1371/journal.pone.0073076.

[91] A. M. Davalli, C. Perego, F. B. Folli, The potential role of glutamate in the current diabetes epidemic, Acta Diabetol 49 (2012) 167–183. https://doi.org/10.1007/s00592-011-0364-z.

[92] Huang, X. T., Li, C., Peng, X. P., Guo, J., Yue, S. J., Liu, W., et al., 2017. An excessive increase in glutamate contributes to glucose-toxicity in β-cells via activation of pancreatic NMDA receptors in rodent diabetes. Sci Rep 7. https://doi.org/10.1038/srep44120.

[93] K. Gowda, W. J. Zinnantif, K. F. LaNoue, The influence of diabetes on glutamate metabolism in retinas, J Neurochem 117 (2011) 309-320. https://doi.org/10.1111/j.1471-4159.2011.07206.x.

[94] R. El-Merahbi, M. Löffler, A. Mayer, G. Sumara, The roles of peripheral serotonin in metabolic homeostasis, FEBS Lett 15 (2015) 1728–1734. https://doi.org/10.1016/j.febslet.2015.05.054.

[95] Paulmann, N., Grohmann, M., Voigt, J. P.,Bert, B., Vowinckel, J., Bader, M., Skelin, M., Jevsek, M., Fink, H., Rupnik, M., Walther, D. J., 2009. Intracellular Serotonin Modulates Insulin Secretion from Pancreatic β-Cells by Protein Serotonylation. PLoS Biol 7, e1000229. https://doi.org/ 10.1371/journal.pbio.1000229.

[96] M. Fukui, M. Tanaka, H. Toda, M. Asano, M. Yamazaki, G. Hasegawa, S. Imai, N. Nakamura, High plasma 5-hydroxyindole-3-acetic acid concentrations in subjects with metabolic syndrome, Diabetes Care 5 (2012) 163-167. https://doi.org/10.2337/dc11-1619.

[97] J. Saito, E. Suzuki, Y. Tajima, K. Takami, Y. Horikawa, J. Takeda, Increased plasma serotonin metabolite 5-hydroxyindole acetic acid concentrations are associated with impaired systolic and late diastolic forward flows during cardiac cycle and elevated resistive index at popliteal artery and renal insufficiency in type 2 diabetic patients with microalbuminuria, Endocr J 63 (2016) 69-76. https://doi.org/10.1507/endocrj.EJ15-0343.

[98] C. M. Oh, S, Park, H. Kim, Serotonin as a New Therapeutic Target for Diabetes Mellitus and Obesity, 40 (2016) 89-98. https://doi.org/10.4093/dmj.2016.40.2.89.

[99] M. Joncquel-Chevalier Curt, P. M. Voicu, M. Fontaine, A. F. Dessein, N. Porchet, K. Mention-Mulliez, D. Dobbelaere, G. Soto-Ares, D. Cheillan, J. Vamecq, Creatine biosynthesis and transport in health and disease, Biochimie 119 (2018) 146-165. https://doi.org/10.1016/j.biochi.2015.10.022.

[100] C. I. Nabuurs, C. U. Choe, A. Veltien, H. E. Kan, L. J. C. van Loon, R. J. T. Rodenburg, J. Matscjke, B. Wieringa, G. J. Kemp, D. Isbrandt, A. Heerschap, Disturbed energy metabolism and muscular dystrophy caused by pure creatine deficiency are reversible by creatine intake, J Physiol 591 (2013) 571–592. https://doi.org/ 10.1113/jphysiol.2012.241760.

[101] L. Hanna-El-Daher, O. Braissant, Creatine synthesis and exchanges between brain cells: What can be learned from human creatine deficiencies and various experimental models ?, Amino Acids 48 (2016) 1877–1895. https://doi.org/ 10.1007/s00726-016-2189-0.

[102] J. Butts, B. Jacobs, M. Silvis, Creatine Use in Sports. Sports Health 10 (2018) 31-34. https://doi.org/10.1177/1941738117737248.

[103] R. N. Smith, A. S. Agharkar, E. B. Gonzales, A review of creatine supplementation in age-related diseases: more than a supplement for athletes, F1000Research 3 (2014) 222. https://doi.org/10.12688/f1000research.5218.1.

[104] T. Wallimann, M. Tokarska-Schlattner, U. Schlattner, The creatine kinase system and pleiotropic effects of creatine, Amino Acids 40 (2011) 1271-1296. https://doi.org/10.1007/s00726-011-0877-3.

[105] B. Ročić, A. Znaor, P. Ročić, D. Weber, M. Vučić Lovrenčić, Comparison of antihyperglycemic effects of creatine and glibenclamide in type II diabetic patients, Wien Med Wochenschr 161 (2011) 519-523. https://doi.ord/ 10.1007/s10354-011-0905-7.

[106] Rocic, B., Bajuk, N. B., Rocic, P., Weber, D. S., Boras, J., Lovrencic, M. V., 2009. Comparison of antihyperglycemic effects of creatine and metformin in type II diabetic patients. Clin Invest Med 32,E322. https://doi.org/10.25011/cim.v32i6.10669.

[107] C. L. Pinto, P. B. Botelho, G. D. Pimentel, P. L. Campos-Ferraz, J. F. Mota, Creatine supplementation and glycemic control: a systematic review, Amino Acids 48 (2016) 2103-2129. https://doi.org/10.1007/s00726-016-2277-1.

[108] B. Ročić, M. Lovrenčić, M. Poje, S. Ashcroft, Effect of Creatine on the Pancreatic ß-Cell, Exp Clin Endocrinol Diabetes 115 (2007) 29-32. https://doi.org/10.1055/s-2007-949591.

[109] B. Op ’t Eijnde, B. Ursø, E. A. Richter, P. L. Greenhaff, P. Hespel, Effect of oral creatine supplementation on human muscle GLUT4 protein content after immobilization, Diabetes 50 (2001) 18-23. https://doi.org/10.2337/diabetes.50.1.18.

[110] C. R. R. Alves, J. C. Ferreira, M. A. de Siqueira-Filho, C. R. Carvalho, A. H. Lancha, B. Gualano, Creatine-induced glucose uptake in type 2 diabetes: a role for AMPK-α?, Amino Acids 43 (2012) 1803-1807. https://doi.org/10.1007/s00726-012-1246-6.

[111] G. Jerums, E. Premaratne, S. Panagiotopoulos, R. J. MacIsaac, The clinical significance of hyperfiltration in diabetes, Diabetologia 53 (2010) 2093-2104. https://doi.org/10.1007/s00125010-1794-9.