COMPUTATIONALLY EFFICINET SEPARATION OF LARGE NUMBER OF ANALYTES FROM SMALL NUMBER OF MIXTURE MASS SPECTRA

IVANKA JERIĆ, LIDIJA BRKLJAČIĆ, IVICA KOPRIVA

Ruđer Bošković Institute, Bijenička cesta 54, Zagreb, CROATIA

Background

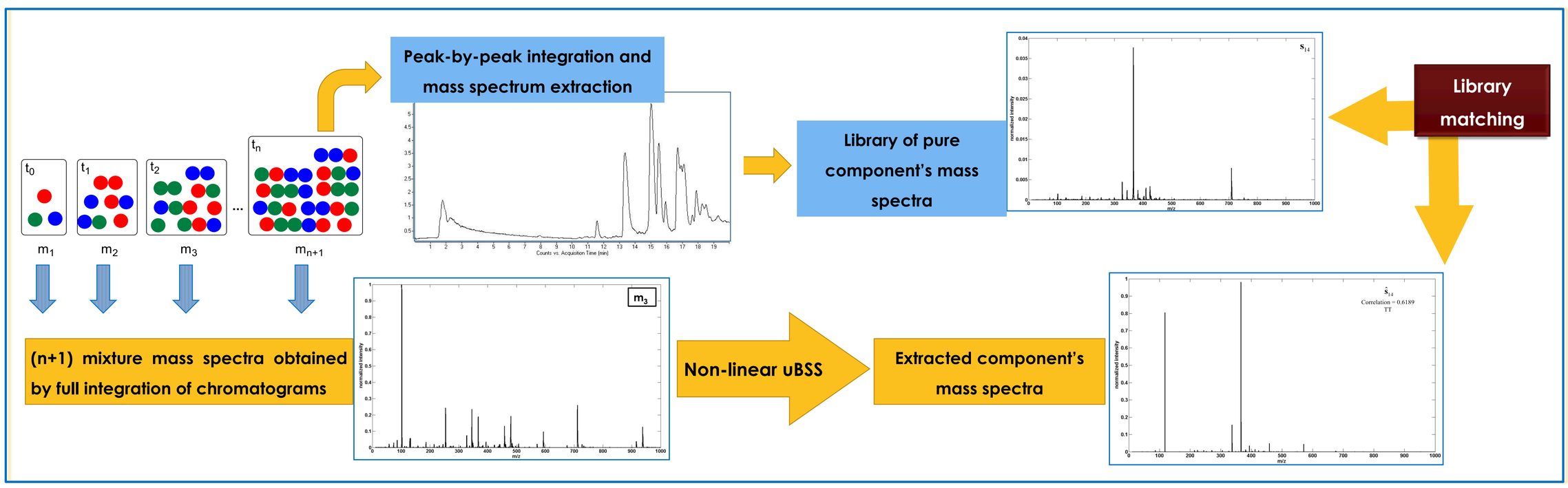
Identification of pure components present in mixtures is a traditional problem in spectroscopy (nuclear magnetic resonance, infrared and Raman) and mass spectrometry. Identification proceeds often by matching separated component spectra with a library of reference compounds, whereas the degree of correlation depends on how well pure components are separated from each other. Thereby, of interest are **blind source separation (BSS) methods** that use only the matrix with recorded mixture spectra as input information. In majority of scenarios, separation of pure components is performed by assuming that mixture spectra are linear combinations of pure components. While a linear mixture model is adequate for many scenarios, a **nonlinear model** offers a more accurate description of processes and interactions occurring in biological systems. Living organisms are the best examples of complex nonlinear systems that function far from equilibrium. Internal and external stimuli (disease, drug treatment and environmental changes) cause perturbations in the system as a result of highly synchronized molecular interactions. As opposed to many BSS methods developed for linear problems, the number of methods that address nonlinear BSS problem is considerably smaller. This number is reduced further when a related nonlinear BSS problem is underdetermined, that is, when the number of components is greater than the number of mixtures. That is why metabolic profiling, which aims to identify and quantify small-molecule analytes (a.k.a. pure components or sources) present in biological samples (typically urine, serum or biological tissue extract), is seen as one of the most challenging tasks in systems biology.

Aim

The aim of this study is to present a method for blind separation of pure components from a smaller number of multicomponent nonlinear mixtures. The model is validated on experimental mass spectrometry data recorded in nonlinear chemical reactions of peptide synthesis.

Setting-up experiments

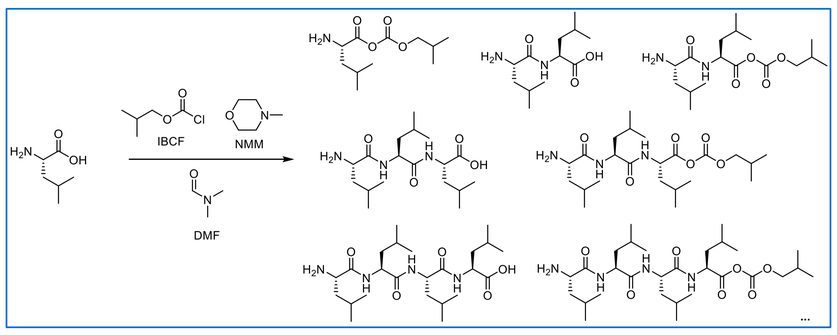
Amino acids were allowed to react under basic conditions giving various products: dipeptides, tripeptides, tetrapeptides, etc. Aliquots of the reaction mixture were withdrawn at different time points (t_0-t_n) and analyzed by the HPLC-(ESI)MS. A library of compounds required for the validation of the algorithm was built by integration of each peak in the chromatogram and subsequent extraction of mass spectrum. The mass spectra of mixtures ((m_1-m_{n+1}) were obtained by full integration of chromatograms.



Results

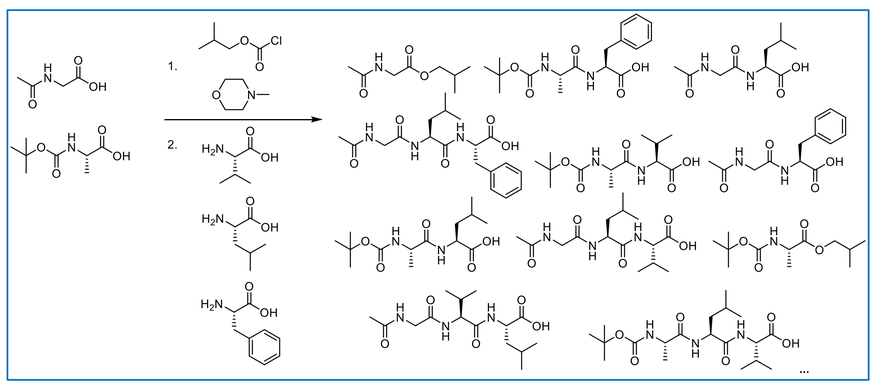
9 mixtures / 25 pure components

Normalized correlation coefficients for 30 combinations of pure components > 0.1



12 mixtures / 19 pure components

Normalized correlation coefficients for 42 combinations of pure components > 0.1



	NMU	NMF_L0	EKM-NMU	PTs_EKM-NMU	PTs-EKM-NMU	PTs-EKM-NMU	PTs-EKM-NMU	PTs-EKM-NMU
			D=*T*=9901	*D*=*T*=9901	*D*=4000	*D*=2000	*D*=1000	*D*=500
correlation G.E. 0.6	8	14	16	**18**	**18**	17	**18**	14
mean correlation	0.342	0.518	0.673	**0.7125**	**0.7095**	0.6983	0.6739	0.6569
minimal correlation	0.038	0.039	0.267	**0.4187**	0.2834	0.3378	0.1207	0.108
inccorect assignments	15	7	**0**	**0**	1	**0**	**0**	1
CPU time	1.3s	40 s	78.78h	4×78h	4×13.7h	4×3.5h	4×0.9h	4×0.23h

	NMU	NMF_L0	EKM-NMU	PTs_EKM-NMU	PTs-EKM-NMU	PTs-EKM-NMU	PTs-EKM-NMU	PTs-EKM-NMU
		K=4	*D*=*T*=9901	*D*=*T*=9901	*D*=4000	*D*=2000	*D*=1000	*D*=500
correlation G.E. 0.6	9	14	16	**18**	17	17	17	16
mean correlation	0.4602	0.575	0.6731	0.7018	0.7505	**0.7568**	0.7429	0.7403
minimal correlation	0.0379	0.0386	0.2673	0.4187	**0.5140**	0.4321	0.5540	0.4116
inccorect assignments	15	7	**0**	**0**	**0**	**0**	1	**0**
CPU time	1.3s	40 s	78.78h	4×78h	4×13.7h	4×3.5h	4×0.9h	4×0.23h

NMU: nonnegative matrix underapproximation ; NMF: nonnegative matrix factorization; EKM-NMU: empirical kernel map based nonnegative matrix underapproximation; PTs-EKM-NMU: preprocessing transforms and EKM-NMU. PTs includes robust principal component analysis (RPCA), and hard (HT), soft (ST) and trimmed tresholding (TT).

Conclusions

Extraction of pure component mass spectra from nonlinear mixtures of mass spectra was performed by an approch that combines four preprocessing methods and EKM-NMU. In comparison with originally developed method, formulation of the problem in a subspace of mapping induced space exhibits significant decrease in computational complexity while retaining quality of pure components extraction.

Journal of Chemometrics, 2014, **28** (2014) 704–715