J Fluoresc (2013) 23:605–610
DOI 10.1007/s10895-013-1183-0

ORIGINAL PAPER

Decomposition of Complex Fluorescence Spectra Containing Components with Close Emission Maxima Positions and Similar Quantum Yields. Application to Fluorescence Spectra of Proteins

Aleksandar Savić · Roland Kardos · Miklós Nyitrai · Ksenija Radotić

Received: 27 August 2012 / Accepted: 24 February 2013 / Published online: 8 March 2013

Abstract Despite of widely application of multivariate analysis in chemometrics, problem of resolving closely positioned components in the fluorescence spectra remained unsolved, thus limiting the usage of fluorescence spectroscopy in analytical purpose. In this paper we have described a novel procedure, adapted especially for the analysis of complex fluorescence spectra with multiple, closely positioned components' maxima. The method was first tested on the simulated spectra and then applied on the spectra of proteins whose fluorophores have similar properties of both the excitation and the emission spectra. In this paper, simple but efficient modification of the method was applied. Instead of analyzing full size emission matrix (12 spectra), 9 spectra wide windows were analyzed, and 4 factors (greatest possible number of factors with physical meaning both for actin and simulated spectra) were extracted in each pass. Obtained factor scores were grouped by using the K-means algorithm. Groups of factor scores obtained from K-means algorithm were passed through the one more factor analysis (FA) in order to find one factor that represents each group. Our approach provides resolution of extremely closed spectral components, which is a vital data for protein conformation analysis based on fluorescence spectroscopy.

Keywords Fluorescence spectra · Proteins · Fixed size window factor analysis · Clustering

A. Savić (✉) · K. Radotić
Institute for multidisciplinary research, University of Belgrade,
Kneza Višeslava 1,
11000 Belgrade, Serbia
e-mail: enzo1320@yahoo.com

A. Savić
e-mail: asavic@imsi.bg.ac.rs

K. Radotić
e-mail: xenia@imsi.bg.ac.rs

R. Kardos · M. Nyitrai
Department of Biophysics, University of Pecs, Medical School,
Pecs, Hungary

R. Kardos
e-mail: roland.kardos@aok.pte.hu

M. Nyitrai
e-mail: miklos.nyitrai@aok.pte.hu

M. Nyitrai
Szentágothai Research Center, PécsIfjúság str. 34,
7624, Hungary

Introduction

Fluorescence spectroscopy may be a powerful technique for resolving the structure of biomolecules - proteins and other polymers. In measurement of emission spectrum of a complex molecule, such as proteins, series of emission spectra are recorded by excitation at different wavelengths in a certain wavelength range [1–4]. By changing the excitation wavelength, the intensity of fluorophore emission spectra changes, which leads to a change of the integral spectrum. The aim is to find the number of spectral components (fluorophores) and positions of their maxima in an integral spectrum.

There have been reported various methods that can be successfully used for component analysis of the emission spectra of complex biological molecules, such as band deconvolution by nonlinear fitting [1, 2], as well as evolving factor analyais (EFA), multivariate curve resolution-alternating least squares (MCR-ALS) [5, 6] and self-modeling mixture analysis method (SIMPLISMA) [7–9].

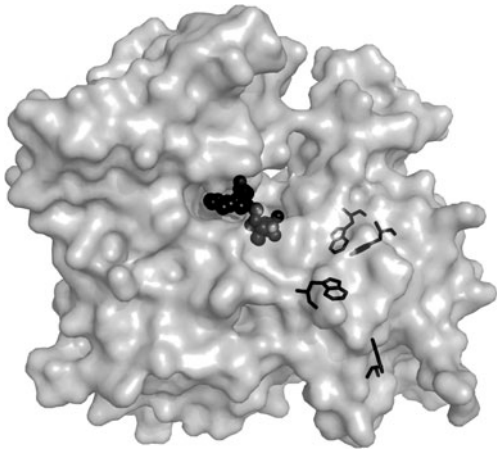

Fig. 1 3D model of actin molecule, showing positions of four tryptophans

Most presently used methods for decomposition of the emission spectra of proteins are not reliable, since the components are closely positioned, i.e. only a few nanometers apart. In this paper we have introduced a new approach for determination of both the number and the shape of spectral components with close maxima positions in a fluorescence spectrum. The core of the method is based on fixed size window factor analysis (FSW-FA), which was previously applied in fluorescence spectroscopy [10, 11]. Such an approach will enable decomposition of the emission spectra of complex molecules such as proteins. The genuine protein fluorescence originates from tryptophan (Trp) which can be situated in various microenvironments, being considered as different fluorophores. Their corresponding spectra in an integral spectrum may be closely positioned.

We have used the simulated spectra to test our method for resolving the number of components and their properties (spectral shape, maximum position). Basic idea was to use the drawback of FA, reduction of number of components as the advantage. Instead of analysis of full size spectral matrix (12 spectra total), 9 spectra were analyzed in one step and 4 factors were extracted. If n factors should be extracted, at least $2n+1$ integral spectra should be used. As the window is moving through the spectral matrix (1st-9th, 2nd-10th… 4th-12th…12th-8th), and factors being extracted, it was found that factors have shown grouping in very distinctive clusters. Proper window size was determined in such a manner that the estimated components have physical meaning. Although none of the extracted factors does match individual components perfectly, groups of similar extracted factors obtained by application of K-means algorithm to factor scores do match individual components after the factors recognized as group were input data for one more FA with 1 factor extraction.

Described procedure provides highly reliable estimation of number of present components, its shape and consequently its contribution in integral spectra. Method is reliable for both extreme cases, if the components are too closely related originating from the same molecule in different micro environments (proteins), and too much separated as in case of extracts which contain chemically different components with wide range of fluorescence emission maxima positions.

We further applied our method in analysis of the emission spectra of a protein. We have chosen actin as the example of such complex system. Actin contains 4 tryptophans, positioned in different microenvironments (Fig. 1), thus only small changes in position of emission spectra maximum (up to 5 nm) occurs between them. The sensitivity of the emission maxima of tryptophans for different microenvironments can be used for investigating the interaction between actin and actin-binding proteins or drugs without labeling the actin with external reporter molecule. More generally, the Trp emission may be used for studies of protein-protein interactions.

Materials and Methods

Theoretical Background

Since the shape of fluorescence spectra is asymmetric, log-normal distribution was applied for simulation of fluorescence spectra in a range of 1–300 nm with increment of 1 nm. The wavelength values were arbitrary chosen because the position on X axis does not affects the results interpretation. Function of density of log-normal distribution used for the simulation of spectra can be described as:

$$f(x) = f(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln x-\mu)^2}{2\sigma^2}} \; x,\sigma \in R^+ \; \mu \in R \qquad (1)$$

Parameters μ and σ are chosen to create components of the integral spectra which correspond to experimentally obtained spectra (Table 1).

Table 1 Parameters of the log-normal distribution used for the simulations

Parameters of spectral components used in the simulations					
Four component model	μ	4.90	4.93	4.96	4.99
	σ	0.2	0.2	0.2	0.2
	Position of maxima	130	133	138	142

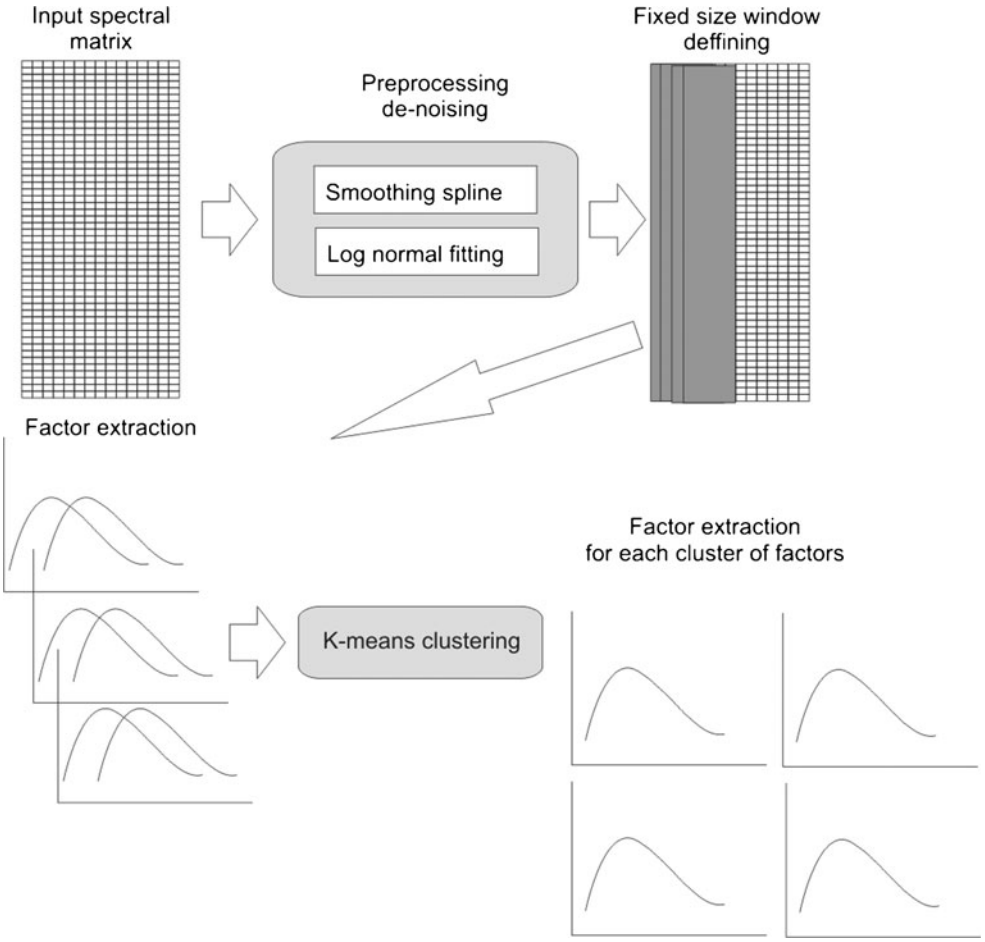

Fig. 2 Scheme of the analytical procedure: preprocessing, fixed size window factor analysis, clustering of extracted factors, factor analysis of estimated components, members of the clusters

Simulated spectral matrix was built according to:

$$\mathbf{D} = \mathbf{S}\mathbf{C}^T + \mathbf{E} \quad (2)$$

$\mathbf{D}_{n\times p}$ denotes matrix, whose columns represent the integral simulated spectra, $\mathbf{S}_{n\times m}$ is a matrix of source spectra obtained from Eq. 1, $\mathbf{C}_{p\times m}$ is matrix of coefficients of linear combinations, chosen in physically correct meaning, $\mathbf{E}_{n\times p}$ represents the matrix of errors (noise). Different levels of noise were applied, in order to check the performances of proposed method in comparison with classical FA, T is a symbol of the transpon matrix.

Elements for **C** were chosen to properly simulate the excitation spectra of fluorofores in order to make the simulation realistic (Fig. 1).

Task of the factor analysis is reverse from Eq. 2, to obtain source components and coefficients of linear combination from the matrix of integral spectra. As the components are highly correlated, for the best analytical results, application of oblique rotation of factor scores was required. It was

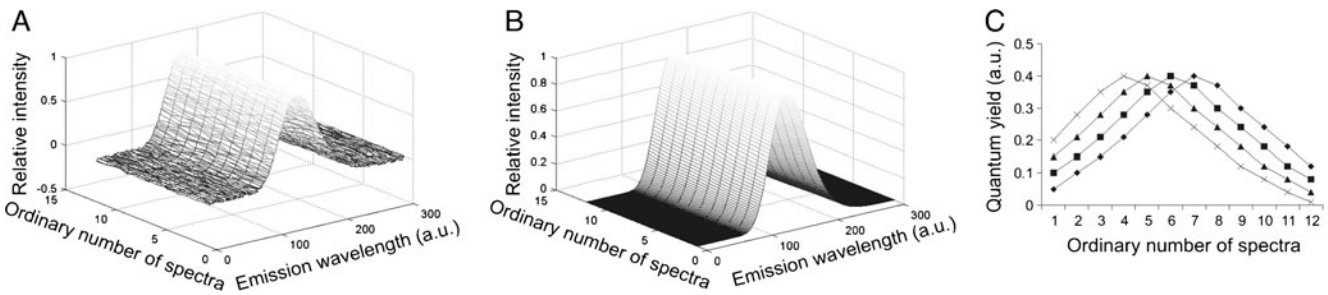

Fig. 3 Simulated spectral matrix, normalized (**a**) before denoising and (**b**) after denoising. Coefficients of linear combinations are presented in figure **c**

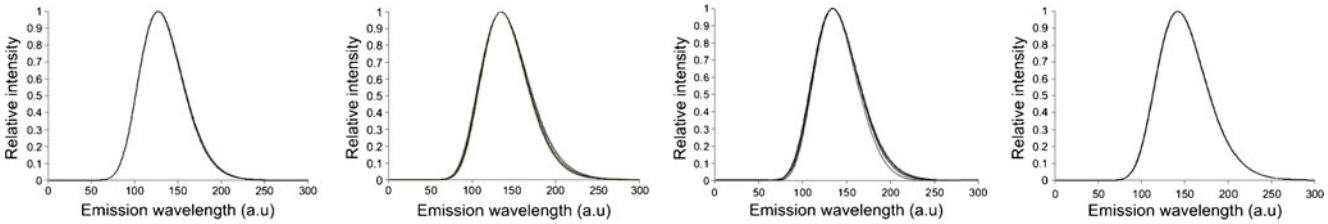

Fig. 4 K-means clustering applied to extracted factor scores and specified groups of factors

empirically found that the promax rotation with kappa value $k=4$ provides the best results. Setting of higher kappa values leads to positioning of components closer than expected, and lower kappa values can give the solutions with no physical meaning. Spectral simulation, fixed size window definition and factor analysis were performed in self developped script written for Matlab 2007b.

K-means algorithm was applied for clusterization of obtained factor scores. The task of k-means algorithm is to classify N objects into K classes, with the condition $K \leq N$, according to minimization of cost function defined as:

$$J = \sum_{j=1}^{N} \sum_{i=1}^{c} \left\| x_j - v_i \right\|^2 \quad (3)$$

x_j- specified intensity, v_i is the centroid of ith cluster, N – total number of objects (spectral length), c – number of clusters (fluorophores), and $\|\cdot\|$ denotes the norm.

Steady-State Fluorescence Spectroscopy

Actin was extracted from rabit muscle and purified as indicated previously [12, 13].

Fluorescence spectra were collected by using Fluorolog-3 spectrofluorimeter (Jobin Yvon Horiba, Paris, France) equipped with a 450 W xenon lamp and a photomultiplier tube. The tryptophan (Sigma, Germany) and actin were dissolved in deionized water (both in concentration 10^{-6} M), in a 1-cm optical path length quartz cuvette. The slits on the excitation and emission beams were fixed at 4 and 2 nm, respectively. The spectra were corrected for dark counts. In each measurement, five scans with 0.1 s integration time, were averaged. All measurements were performed at controlled temperature of 25 °C by means of a Peltier element.

For each sample series of emission spectra were measured by varying excitation wavelengths with 5 nm steps. In order to include all geometric characteristics of an emission spectrum into modeling, initial excitation wavelength was set in such a way that recording of emission spectra started 30 – 40 nm before the sharp rise in emission.

A total of 12 emission spectra were collected for actin, respectively.

Results and Discussion

Simulated Spectra

Analytical procedure can be described in scheme presented in Fig. 2.

The first analytical step, spectral denoising can be successfully performed by curve fitting with smoothing spline method or by fitting with asymmetrical distributions, such as log-normal or 4 parameter gamma distribution.

Simulated spectral matrix calculated by applying Eqs. 1 and 2 and parameters from Table 1 before and after denoising procedure was presented in Fig. 3.

Parameters used in simulations were chosen according to physical properties of fluorescence. Source components correspond to Trp emission data available in literature, while the coefficients of linear combinations were chosen to represent emission spectra of Trp. Factor analysis is generally sensitive to noise, because the noise component leads to interpretation of principal components with noisy details, and also the number of extracted components can be reduced.

After the second step in analytical procedure, the fixed size window factor analysis (window width 9 spectra, 4 factors extracted) was applied and it was obvious that some estimated

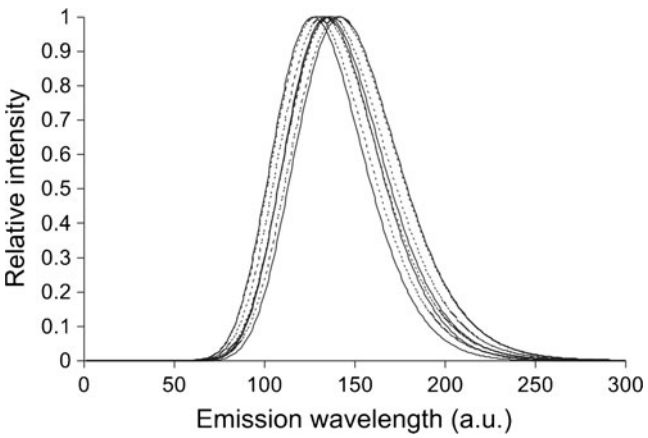

Fig. 5 Comparison between the starting components and the estimated components. *Dashed lines*, original components, *solid lines*, estimated components

Table 2 Comparison between classical factor analysis and proposed method based on fixed-size window factor analysis when applied to data sets with additional white noise

	White noise (%)	Number of clear components (4 required)	Number of noisy components	Number of components with no physical meaning
Classical analysis	0	4	0	0
	1	2	1 (2 overlapped)	0
	3	2 (low noise)	0	2
	5	0	2 (difficult to determine position of maxima)	2
	10	0	2 (difficult to determine position of maxima)	2
	15	0	2 (partially overlapped)	2
Proposed procedure	0	4	0	0
	1	2	2	0
	3	2	0	2
	5	2 (low noise)	1 (high noise)	1
	10	1 (low noise)	1 (difficult to determine position of maxima)	1
	15	0	4 (all components with physical meaning, difficult to determine position of maxima, but distinguishable)	0

components show tendency to group in clusters. Results of factor scores grouping in clusters are presented in Fig. 4.

K-means algorithm was efficient for determination of grouping for estimated factor scores.

In the final step of the procedure, it was necessary to extract single representative factor for each cluster, and compare them with the starting components (Fig. 5).

According to results obtained on simulated spectra, it was found that the method was efficient and accurate, thus applicable to real experimental data. Factor analysis of full size matrix, as the result gave the solution with 3 components whose maxima positions were 127, 133 and 141. Procedure described in this paper gave the solution with 4 components, whose maxima positions were 128, 134, 137, 142. Obtained result was very close to source components maxima positions (129, 132, 136, 141).

In order to check if the method is reliable in presence of white noise, several simulated noise levels were analyzed (Table 2).

Presence of white noise significantly affects reliability of factor analysis. Even with white noise with amplitude of 1 % in comparison to signal maximum peak, reduction of number of components emerges. Proposed variation of FSW-FA sustains to estimate right number of components. If the noise level was increased, estimated factor scores also suffer from presence of noise. In such cases, determination of emission maxima position becomes impossible. Our method has shown better results in comparison with classical factor analysis. Although still suffering from noise, it provides greater robustness than FA.

Measured Spectra

As the complex biological system with 4 possible components, protein molecule, actin, was chosen. Actin contains four Trps, the first buried inside the protein, the second inside the protein but close to the protein surface, and the two Trps are on the protein surface in similar microenvironments (Fig. 1). These four Trps are considered as four fluorophores. It was possible to measure 12 spectra and use them as input spectra for the analysis.

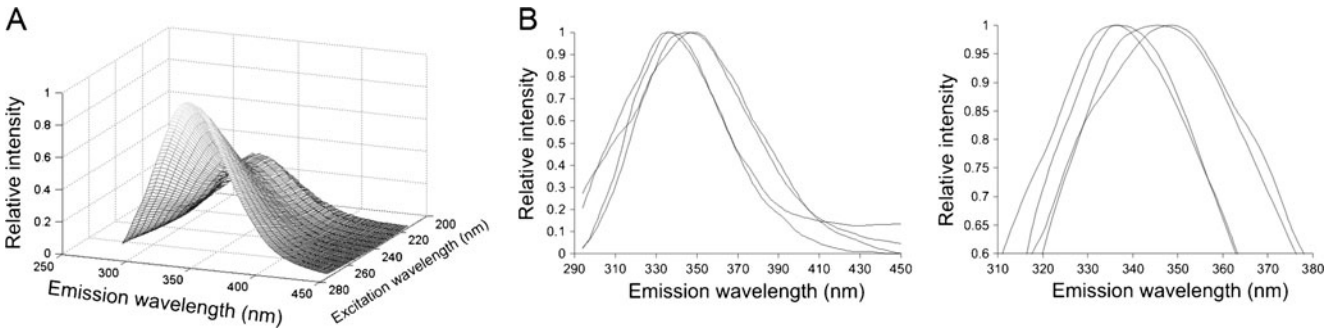

Fig. 6 **a** Emission spectra of actin. **b** Extracted components after described procedure was applied

Starting spectral matrix is shown in Fig. 6.

If the factor analysis with promax rotation was applied on full size spectral matrix, only two components with physical meaning could be extracted (maxima at 338 nm and 353 nm). Application of proposed method allows identification of all possible components (four) corresponding to the number of Trps. According to the data available in literature [14], 335 nm and 337 nm components correspond to the Trps buried inside the protein and 344 nm and 348 nm components are related to Trps located on the surface of the protein molecule, but in slightly different microenvironments.

In comparison with classical FA, it is proved that FA applied to full size spectral matrix tends to reduce the number of components. It is obvious that the 335 nm- and 337 nm- maxima obtained in our analysis correspond to the 338 nm maximum obtained by our tests with classical FA. The component maximum at 353 nm obtained by FA has similar shape and position to the 348 nm component in our analysis. The shift in classical analysis emerges from the mis- estimation of the first two components.

Conclusion

Although based on simple modification of the existing analytical procedures, described method presents new, efficient and reliable procedure applicable for the analysis of the most challenging task in fluorescence spectroscopy – analysis of closely related components.

The efficiency has proven on simulated spectra and finally achieved the resolution of complex protein molecule fluorescence spectra, previously unsolvable by most commonly applicable algorithms.

Proposed method is much more reliable if the noisy spectra were analyzed. According to this, the proposed method could find implementation in other chemometrics analyses where good performances are required for noisy data sets.

Acknowledgments Grant 173017 from the Ministry of Education and Science of the Republic of Serbia supported this study. We thank to Mr Dragosav Mutavdžić for critical reading of this paper. This study was supported by the Hungarian Science Foundation (OTKA grant K77840 to MN) and also by the 'Science, Please! Research Team on Innovation' (SROP-4.2.2/08/1/2008-0011).

References

1. Radotić K, Kalauzi A, Djikanović D, Jeremić M, Leblanc R, Cerović Z (2006) Component analysis of the fluorescence spectra of the lignin model compound-dehydrogenative polimer. J Photochem Photobiol B Biol 83:1–10
2. Kalauzi A, Mutavdžić D, Djikanović D, Radotić K, Jeremić M (2007) Application of asymmetric model in analysis of fluorescence spectra of biologically important molecules. J Fluoresc 17:319–329
3. Djikanović D, Kalauzi A, Jeremić M, Mićić M, Radotić K (2007) Deconvolution of fluorescence spectra: Contribution to the structural analysis of complex molecules. Colloids Surf B: Biointerfaces 54:188–192
4. Šašić S, Antić-Jovanović A, Kuzmanović M, Jeremić M (1999) Quantitative analysis of the Raman spectra of mixtures of weakly interacting components by factor analysis methods. Analyst 124:1481–1487
5. Tauler R, Kowalski B (1993) Multivariate curve resolution applied to spectral data from multiple runs of an industrial process. Anal Chem 65:2040–2047
6. Garrido M, Rius FX, Larrechi MS (2008) Multivariate curve resolution-alternating least squares (MCR-ALS) applied to spectroscopic data from monitoring chemical reactions processes. Anal Bioanal Chem 390:2059–2066
7. Esteves Da Silva JCG, Machado AASC (1996) Interaction of fulvic acids with Al(III) studied by self-modeling curve resolution of second-derivative synchronous fluorescence spectra. Appl Spectrosc 50:436–443
8. Esteves Da Silva JCG, Machado AASC (1995) Self-modelling curve resolution analysis of synchronous fluorescence spectroscopy data for characterization of acid mixtures and study of acid–base equilibria. Analyst 120:2553–2560
9. Esteves Da Silva JCG, Machado AASC, Oliveira CJS (1996) Study of the interaction of a soil fulvic acid with UO22+ by self-modelling mixture analysis of synchronous molecular fluorescence spectra. Analyst 121:1373–1379
10. Keller HR, Massart DL (1991) Peak purity control in liquid chromatography with photodiode-array detection by a fixed size moving window evolving factor analysis. Anal Chim Acta 246(2):379–390
11. De Juan A, Maeder M, Hancewicz T, Tauler R (2005) Local rank analysis for exploratory spectroscopic image analysis. Fixed Size Image Window-Evolving Factor Analysis. 77(1):64–74
12. Feuer G, Molnár F, Pettkó E, Straub FB (1948) Studies on the composition and polymerization of actin. Hung Acta Physiol 1:150–163
13. Spudich JA, Watt S (1971) The regulation of rabbit skeletal muscle contraction. J Biol Chem 246:4866–4871
14. Ladokhin AS (2000) Fluorescence spectroscopy in peptides and protein analysis. In: Meyers RA (ed) Encyclopedia of analytical chemistry. John Wiley&Sons Ltd, Chichester, pp 5762–5779