Chapter 2
Integrating Optical Spectroscopy and Chemometric Methods

Karolina Jagiełło, Anita Sosnowska, Jan Mazerski and Tomasz Puzyn

Abstract In this Chapter, we describe the usage of several chemometric and numerical techniques to analyse of UV-Vis sets of spectra. The fundamentals of each technique are briefly presented with examples of its applications. This approach allows obtaining deeper insight in studied system. These methods can be used not only to analyze aggregation process, as it was presented in the Chapter, but also to study the interaction between small ligands and macromolecules, such as DNA. Determination of the number of formed complexes and the binding constant of interaction ligand/macromolecules can be received with this methodology.

2.1 Introduction

In chemistry, we have been often dealing with a set of measured data, which are in fact a mixture of the information and the noise. In many cases, the magnitude of the noise is as much great as the information that significantly hinders the ability to find the interesting results in multidimensional data sets [1]. The solutions of such problems are proposed by chemometrics. By means of chemometric techniques we are able to: (i) delete as much noise as possible from the data sets; and, (ii) extract as much information as possible from the multidimensional data [1].

However, to obtain reliable results of chemometrics analysis some special rules should be applied.

The chemical/analytical problem that one would like to solve, has to be defined precisely before starting the experiment. Problems solved by means of chemometric methods can be grouped into four main families: making the visu-

T. Puzyn (✉) · K. Jagiełło · A. Sosnowska
Laboratory of Environmental Chemometrics, Faculty of Chemistry, University of Gdansk, Gdansk, Poland
e-mail: puzi@qsar.eu.org

J. Mazerski
Department of Pharmaceuticals Technology and Biochemistry, Chemical Faculty, Gdansk University of Technology, Gdansk, Poland

M. Baranska (ed.), *Optical Spectroscopy and Computational Methods in Biology and Medicine,* Challenges and Advances in Computational Chemistry and Physics 14,
DOI 10.1007/978-94-007-7832-0_2, © Springer Science+Business Media Dordrecht 2014

alisation of multivariate data set; searching the relationships between sets of data; recognition of internal data structure and making the classification [2]. Each of them needs adequate type of data sets and usage the appropriate chemometric approaches.

The next step is to design carefully the experiment and then perform measurements. The choice of the experimental design depends on the type of information required to verify the research hypothesis, which in fact is the solution of the defined problem. It is indicated to perform as little trials as possible, but selected samples should be representative [1, 2]. In such way, we do not only save time and cost of the whole project but also let make to increase significantly the quality of the results.

Collected data should undergo multi-aspect control procedures to: (i) identify the outliers; (ii) determine of the variable distribution that in some cases indicate the necessity to perform data transformation; and, (iii) determine of interdependences between variables. The main goal of this procedure is to eliminate errors and shorten the time necessary to get reliable results [2].

Only data prepared according to described above procedures can be analysed by means of chemometric techniques. The main approaches of chemometrics data analysis are presented in Fig. 2.1.

Summing up, **chemometrics is the field of science dealing with extracting valuable information from multidimensional data sets by means of mathematical and statistical methods**. However, appropriate usage of chemometrics requires thinking about its application in each step of experiment: from the problem defining to the analysis of obtained results.

2.2 Chemometrics and Numerical Approach to a Set of Spectra

One of a typical problem in absorption spectroscopy is to determine an influence of experimental conditions, such as temperature, ionic strength of the medium or pH, on the electronic spectrum of tested compounds. The experimental condition may affect intensity as well as shape of the spectrum. To detect such changes a set of spectra registered at different experimental condition should be analysed.

From chemometrics point of view each spectrum could be treated as a one-dimensional matrix. Absorptions at different wavelengths are elements of this matrix. A set of spectra registered at different condition forms two-dimensional matrix **X** [3, 4].

2.2.1 Preprocessing of Spectra

Registered sets of spectra, transformed to molar extinction scale according to the Lambert-Beer Eq. (2.1) have to be preprocessed.

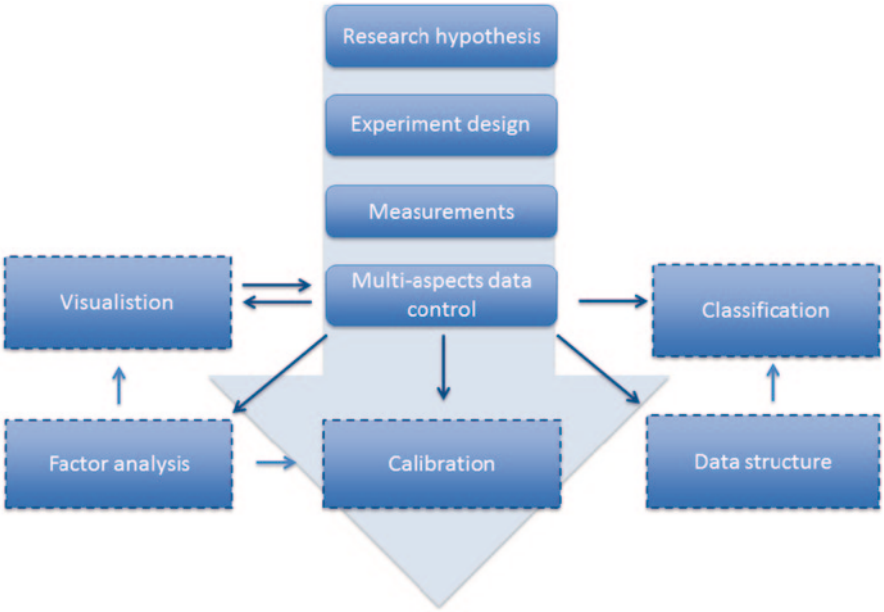

Fig. 2.1 The overview of chemometrics methods

$$A_\lambda = \varepsilon_\lambda cl \tag{2.1}$$

A_λ absorbance at a given wavelength, expressed in nanometers,
ε_λ the molar extinction coefficient at a given wavelength,
c molar concentration of compound,
l optical path length expressed in centimeters

The first steps of these operations include: baseline correction and dilution correction. Several techniques applied to perform baseline correction are summarized by Kohler et al. [5].

Obtain, corrected and transformed results are organize into matrix **X**, presented in Table 2.1 consists n rows (wavelengths, λ) and m columns (spectra obtained for the m increasing concentrations of the compound or spectra registered at different temperatures, pH, etc.). Each element of this matrix corresponds to extinction values calculated for each sample in following wavelengths.

For each column of matrix, the mean values and standard deviation values can be calculated for whole range of wavelengths or for selected range of wavelength, for example range characteristic for research chromophore. These two parameters can be used for measuring changes in intensity of the spectra [3, 4]. To minimalize influence of changes in intensity of spectra, and in this way study only changes into its shapes, sets of spectra should be transformed. The most popular transformations are centring and autoscaling.

Table 2.1 Matrix X obtain as a result of 1st step preprocessing operations of set of UV-Vis spectra

Wavelength	Sample 1	Sample 2	…	Sample m
λ_1	$\varepsilon_{1,1}$	$\varepsilon_{1,2}$	…	$\varepsilon_{1,m}$
λ_2	$\varepsilon_{2,1}$	$\varepsilon_{2,2}$	…	$\varepsilon_{2,m}$
.	.	.	…	.
λ_n	$\varepsilon_{n,1}$	$\varepsilon_{n,2}$	…	$\varepsilon_{n,m}$

Centring the data is the linear transformation performs to make all variables coincide with the beginning of the coordinate system. It is done by subtracting from each value x_{ij} of data organized into matrix **X** corresponding mean value of jth variable to obtain new matrix **X′**, according to formula [1]:

$$x'_{ij} = x_{ij} - \mu_j \tag{2.2}$$

The autoscaling is made by diving by centered data by the standard deviation of the jth variables according to formula (2.2). As a result, we obtained the normalized matrix **Z** [2]:

$$z_{ij} = \frac{x_{ij} - \mu_j}{\sigma_j} \tag{2.3}$$

where:

x_{ij} molar extinction coefficient for the i-th wavelength of the j-th spectrum
μ_j the average value of the molar extinction coefficients of the j-th spectrum
σ_j standard deviation of the molar extinction coefficients of the j-th spectrum.

Spectrum after standardization can be treated as a vector of unit length. A set of spectra creates a bunch of vectors which have common origin in the k-dimensional hyperspace, where k is the number of spectral forms present in the analyzed samples. A bunch of vectors for recorded spectra is limited by vectors representing the spectra of pure ingredients. An example for two component mixture (k = 2) is shown in Fig. 2.2.

2.2.2 Internal Order of a Matrix and Number of Variety Sources

The Principal Component Analysis (PCA) [6] is a special example of projection pursuit techniques, in which variance is used as a projection index. PCA is mainly used for modeling, compressing and visualizing multidimensional data [7–10]. Application PCA in the relationship analysis involves two basic tasks: graphical

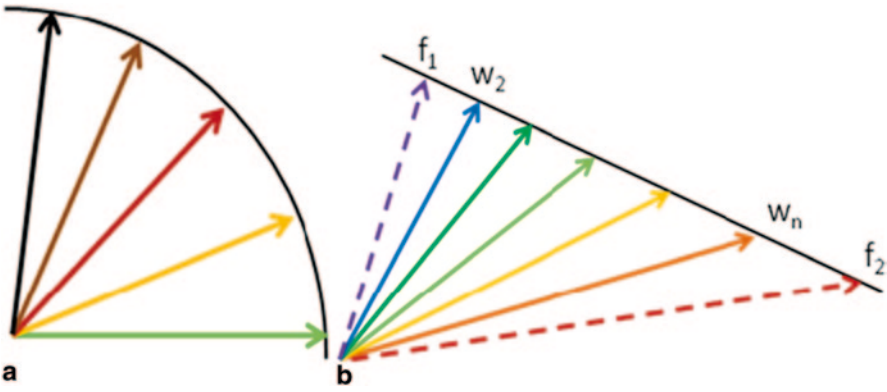

Fig. 2.2 Vectors of the two component mixture in the space of two main components: **a** standardized spectrum. **b** centered spectrum

presentation of the dimensional dependence and reducing the dimensionality of the problem. In this technique, set of correlation coefficient between variables in multivariate space might be transformed to equivalent set of orthogonal factors. The principal components are new orthogonal variables (which are expressed as a linear combination of original variables) and they maximized description of the data variance.

The PCA analysis allows to determine the internal dimension of the matrix of the standardized spectra. This dimension is equal to the number of spectral forms presented in the samples. If the spectra of individual species are not excessively correlated, it is equal to the number of significant principal components of this matrix. Unfortunately, when the spectral forms have very similar spectra, as is usual in the case of aggregation, the analysis of the residual spectra have to be used to determine the number of spectral species in the samples [3, 4].

The principal components and their loadings permit to complete reconstruction of the standardized spectra matrix according to the equation:

$$\mathbf{Z}_{nm} = \mathbf{P}_{nm}\mathbf{L}_{nm} \tag{2.4}$$

where:

$\mathbf{P}_{nm}$ matrix of m principal components,
$\mathbf{L}_{nm}$ matrix of loadings.

If in the spectra reconstruction we use only j first principal components we obtain following relationship:

$$\mathbf{Z}_{nm} = \mathbf{P}_{nj}\mathbf{L}_{jm} + \mathbf{E}^{(j)} \tag{2.5}$$

$\mathbf{E}^{(j)}$ matrix represents the residual spectra of j order. The residual spectra for $j<k$ have the character of the differences spectra, for $j \geq k$ the residual spectrum represents random noise. Thus, further analysis of the residual spectra allows to specify the correct value of k.

2.2.3 Numerical Spectrum Decomposition Technique and Physicochemical Model of a Process

Applying PCA analysis let us know not only the number of spectral forms presented in the samples, but also the first spectra approximation, however, does not allow to designate individual forms, because the principal components are obtained with the assumption of perfect orthogonality.

Therefore, next step in chemometrical analysis of sets of spectra is to obtain the spectra of indywidauals and estimation of its molar fractions. To reach this goal, **Z** matrix should be expressed as a multiplication of the two other matrices: the matrix of components **B**, in which the columns represent the spectrum of pure spectral forms, and matrix **X**, which columns represent different forms of molar fractions [3, 4].

$$\mathbf{Z} = \mathbf{BX} \tag{2.6}$$

Knowing that the ends of the centered vectors of two components mixtures are on a straight line (Fig. 2.2b), the position of the spectra of pure components on this line is unknown. Therefore, in order to solve the Eq. (2.6) there is need to use the iterative procedure called Numerical Spectrum Decomposition [1]. To determine spectra of the complexes, as well as relative amount of each species, an iterative self-consisting procedure—multivariate curve resolution-alternating least squares (MCR-ALS) has to be used [11]. The most outer vectors w1 i wm participate the most within the spectral forms according to the equation.

$$w_1 = x_{11}f_1 + x_{21}f_2 \quad x_{11} \approx 1, x_{21} \approx 0 \tag{2.7}$$

$$w_m = x_{1m}f_1 + x_{2m}f_2 \quad x_{1m} \approx 1, x_{2m} \approx 0 \tag{2.8}$$

In addition, in the two components mixtures, the molar ratios should not be negative, and the sum of the molar fractions of the two forms must be equal to 1:

$$x_{11} + x_{21} = 1 \tag{2.9}$$

$$x_{1m} + x_{2m} = 1 \tag{2.10}$$

Using the Eq. (2.9) and (2.10) molar fraction x_{11} i x_{2m} can be determined:

$$x_{11} = 1 - x_{21} \tag{2.11}$$

$$x_{2m} = 1 - x_{1m} \tag{2.12}$$

After substituting these molar fractions into formulas (2.11) and (2.12) set of two equations were obtained:

$$\begin{cases} w_1 = x_{11} f_1 + (1 - x_{11}) f_2 \\ w_m = (1 - x_{2m}) f_1 + x_{2m} f_2 \end{cases} \tag{2.13}$$

These equations can be solved due to the spectrum of pure form. Assuming further indications:

$$\alpha = \frac{1}{x_{11}} > 1 \tag{2.14}$$

$$\beta = \frac{1}{x_{2m}} > 1 \tag{2.15}$$

the following formulas are obtained:

$$f_1 = \frac{\alpha w_m - (\alpha - 1)\beta w_m}{1 - (\alpha - 1)(\beta - 1)} \tag{2.16}$$

$$f_2 = \frac{\beta w_m - \alpha(\beta - 1) w_1}{1 - (\alpha - 1)(\beta - 1)} \tag{2.17}$$

Coefficients α and β are matched numerically during fitting a physicochemical model of the process studied using the Nelder–Mead simplex method [12, 13]. Obtaining optimal values of α and β is equivalent to determine the spectra of pure species. Determination of the spectra of pure spectral forms allow to calculate the matrix Eq. (2.6) because of the mole fractions. Molar fraction of the individual components in various conditions obtained by this method, could be used to determine the specific parameters of the model used.

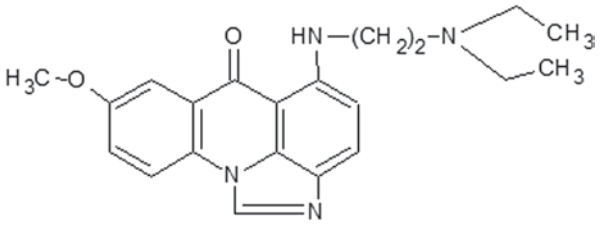

Fig. 2.3 Chemical structure of C-1330

2.3 Influence of Agregation on UV/Vis Spectrum—A Case Study

Acridine belong to a group of polycyclic heteroaromatic compounds and exhibit a broad spectrum of biological activity including antiprotozoal, antibacterial, antiviral and antitumor activity [7, 14]. Imidazoacridinones (IA) derivaties are a group of acridine antitumor drugs synthesized in Department of Pharmaceutical Technology and Biochemistry, Gdańsk University of Technology. The biological activity of the imidazoacridinones has now been extensively investigated. Published data suggested that imidazoacridinone drugs are capable of binding physicochemically to DNA [15, 16]. The nature of these investigation and their relevance to cytotoxic and antitumor properties of IA remains unknown.

Imidazoacrinononones tend to aggregate in diluted solutions, leading to dimer formation, and sometimes even higher order aggregates. Driving forces of this process are hydrophobic interactions. Kinetic information of aggregation process are very helpful to understand molecular interaction such as micelle formation of amphiphilic substances and the binding of small ligands to macromolecules [17–19]. In spite of many studies, the mechanism of compounds self-aggregation seems to be not fully understood because of the experimental conditions, which are different from study to study.

The most useful method to check if some compounds self-aggregate in aqua solution is spectroscopic analysis with increasing drug concentration. The following procedure of results analysis may be used to study the aggregation of various derivatives. An example of such derivative is imidazoacridinon C-1330 (Fig. 2.3), synthesized in Department of Pharmaceutical Technology and Biochemistry, Gdańsk University of Technology [20]. The samples of imidazoacridinon C-1330 were prepare in various concentrations (20 μM to 1 mM) by dilutions. The absorption spectra at different concentration were recorded. Obtained spectra were transformed into the molar scale using Lambert-Beer Eq. (2.1), and then standardized according to Eq. (2.3).

Figure 2.4 represents the set of standardized spectra obtained for C-1330 in different drug concentration. The compound has a strong absorption band in the visible field of the spectrum. Two maximum are observed in the spectra. For the lowest concentration of the compound, they are located at $\lambda = 371$ and $\lambda = 423$ nm. In real spectra, there are two isosbestic points at a wavelength $\lambda = 383$ and $\lambda = 431$ nm.

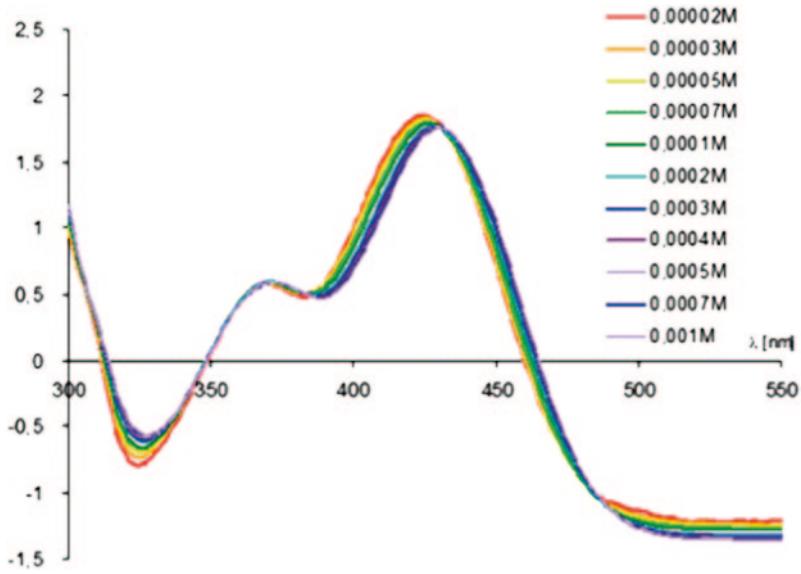

Fig. 2.4 Set of standardized spectra representative for C-1330 in buffer with 5 mM NaCl

Table 2.2 Eigen-values obtained for set of the C-1330 spectra

	No of PC				
	1	2	3	4	5
Eigen-value	9.9827	0.0170	0.0002	0.0001	0.0000
% of variance	99.83	0.16	0.01	0.00	0.00
Cumulative % of variance	99.83	99.99	100.00	100.00	100.00

Numerical analysis of the spectra for C-1330 was started by determining the number of spectral forms present in the solution. Table 2.2 presents the data obtained by using the PCA method. The following values suggest that the first two principal components (explaining 99.99 % of the total variability) are significant.

Similar results were obtained after careful analysis of the residual spectra (Fig. 2.5). The 1st and 2nd order residual spectra are relatively intensive and demonstrate the presence of absorption bands. Experimental noise prevails only in 3rd or higher order residual spectra. This indicates that there are two spectral forms in the solution: monomer and aggregate.

Using the appropriate transformations as it was presented in Chap. 2.3. there were possible to reproduce spectra of pure spectral forms. Figure 2.6 shows such spectra for C-1330. The maximum absorption of the monomer occurs at 423 nm. In the spectra of the aggregate exhibit characteristic batochromic effect with the maximum absorption at the 431 nm. The results of our studies indicate that the presence as well as proportions of particular spectral forms is dependent on the overall concentration of studied compound.

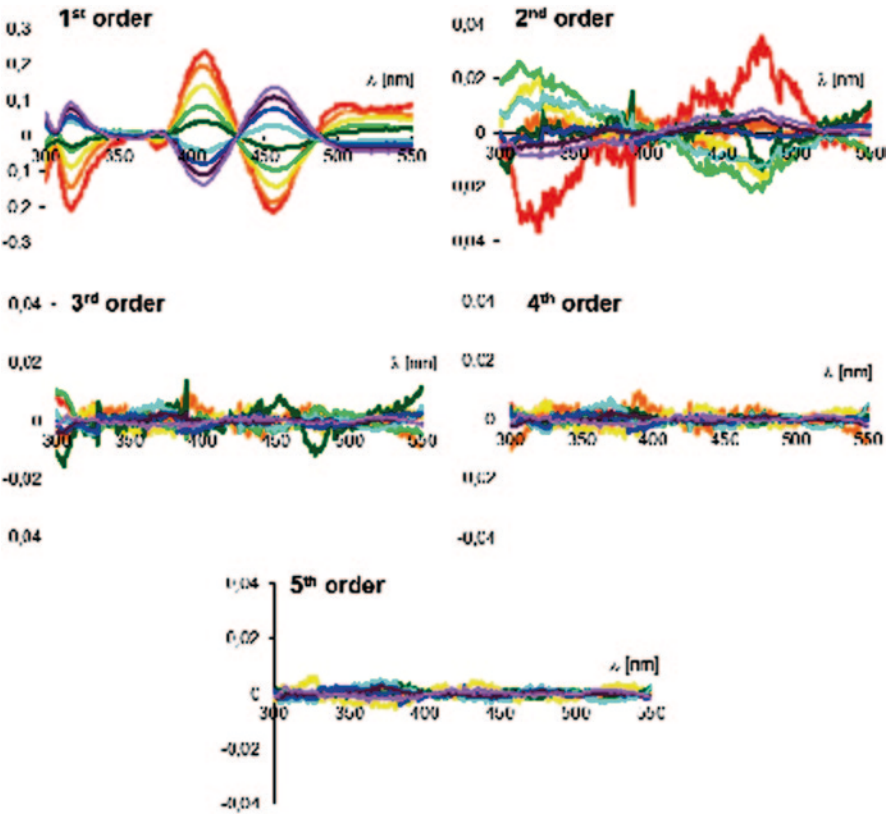

Fig. 2.5 Further residual spectrum obtained for the C-1330 in 5 mM NaCl solution in the PCA method

Obtained molar fraction of the individual components in various concentrations was used to determine the aggregation constant K_A according to the formula:

$$K_{Ai} = \frac{1 - \sqrt{x_{mi}}}{C_{ti} x_{mi}} \tag{2.18}$$

Where

C_{ti} the total concentration of the compound in the i-th solution,
x_{mi} the molar fraction of the monomer calculated from the Eq. (2.6).

The estimation of the m aggregation constant—one for each compound concentrations (for each spectra) were obtained. Than the average value and the standard deviation were calculated: $K_A = (6,327 \pm 0,218) * 10^3 M^{-1}$.

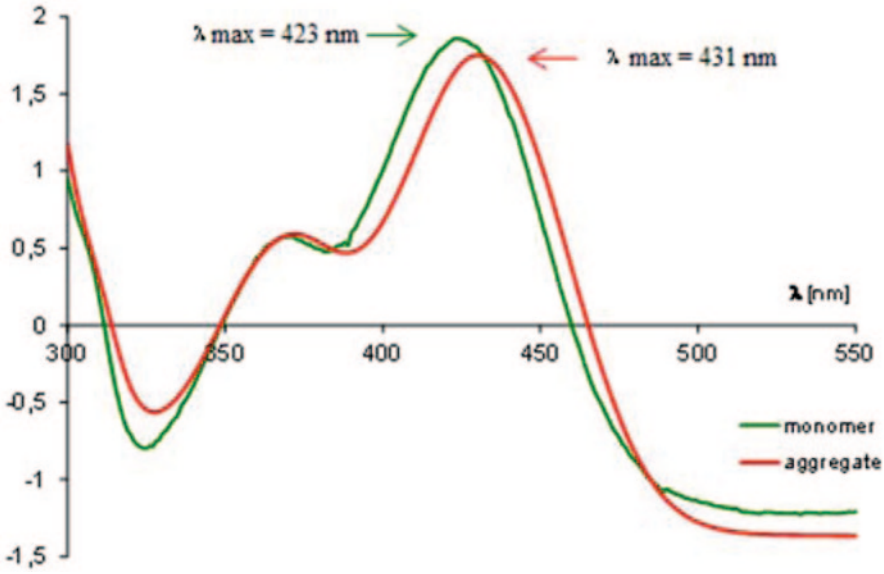

Fig. 2.6 Standardized spectra for monomer (*red line*) and aggregate (*green line*) for C-1330 in 5 mM NaCl

2.4 Conclusions

Application of chemometric and numerical methods can be used to analyse spectroscopic spectra. We presented these applications for the aggregations process that give interesting results even when the spectra of spectral forms presented in the sample differ slightly. The presented approach is universal if the model of the process is known. This methods can be also used to study the interaction between small ligands and macromolecules, such as DNA, for determining the number of formed complexes and the binding constant of interaction ligand/macromolecules. The chemometrics approach to an experimental data allows obtaining deeper insight in studied system. Application of various statistical methods is particular useful when classical simplified assumption is not valid.

Acknowledgements This work was supported by the Polish Ministry of Science and Higher Education, Grant No. DS/8180-4-0171-13

References

1. Kramer R (1998) Chemometric techniques for quantitative analysis. Dekker, New York, United States of America
2. Mazerski J (2006) Introduction to chemometrics. In: Zuba D, Parczewski A (eds) Chemometrics—methods and applications. Institute for Forensic Research Publisher, Krakow, Poland, pp 11–16
3. Slisz M, Mazerski J (2006) Chemometrics analysis of UV-Vis spectra of associated forms of amphotericin B and its derivatives. In: Zuba D, Parczewski A (eds) Chemometrics—methods and applications. Institute for Forensic Research Publisher, Krakow, Poland, pp 307–312
4. Jagiello K, Mazerski J (2006) Spectrophotometric titration of dsDNA ny nitroacridines—chemometrics approach. In: Zuba, D, Parczewski A (eds) Chemometrics—methods and applications. Institute for Forensic Research Publisher, Krakow, Poland, pp 301–306
5. Kohler A, Zimonja M, Segtnan V, Martens H (2009) Standard normal variate, mutliplicative signal correction and extended multiplicative signal correction preprocessing in biospectroscopy.In: Brown SD, Tauler R, Walczak B (eds) Comprehensive chemometrics: chemical and biochemical data analysis, vol 2. Elsevier, The Netherlands, pp 139–163
6. Sharaf MA, Ullman DL, Kowalski BR (1998) Chemometrics. Wiley, New York
7. Wolfe AD (1975) Quinacrine and other acridines. In: Corocoran JW (ed) Antibiotics Vol. III. mechanism of action of antumicrobial and antitumor agents. Springer Berlin, Heidelberg
8. Gemperline P (2006) Practical guide to chemometrics. Taylor, New York
9. Massat DL, Vander Heyden Y (2004) LC-GC Europe, 17, 586
10. Massat DL, Vander Heyden Y (2004) LC-GC Europe, 18, 84
11. de Juan A, Casassas E, Tauler R (2000) Encyclopedia of analytical chemistry: intrumantation and applications. Wiley, New York
12. Findeisen W, Szymanowski J, Wierzbicki A (1980) Teoria i metody obliczeniowe optymalizacji. Państwowe Wydawnictwo Naukowe, Warszawa
13. Brandt S (1998) Analiza danych. Państwowe Wydawnictwo Naukowe, Warszawa
14. Acheson RM (1956) Acridines. Wiles, New York
15. Dzięgielewski J, Składanowski A, Konopa J (1996) Ann Oncol 7(Suppl., 82)
16. Berger B, Marquardt H, Westecdorf J (1996) Cancer Res 56:2094
17. Moriyama Y, Takeda K, Murakami K (2000) Langmuir 16:7629
18. Murakami K (1998) Bull Chem Soc Jpn 71:2293
19. Murakami K, Tsurufuji K (1999) Bull Chem Soc Jpn 72:653
20. Burger AM, Double JA, Konopa J, Bibby MC (1996) Br J Cancer 74:1369

http://www.springer.com/978-94-007-7831-3

Optical Spectroscopy and Computational Methods in Biology and Medicine

Baranska, M. (Ed.)
2014, XII, 540 p. 204 illus., 134 illus. in color.,
Hardcover
ISBN: 978-94-007-7831-3