Asian Journal of Chemistry; Vol. 25, No. 8 (2013), 4643-4645

ASIAN JOURNAL OF CHEMISTRY

http://dx.doi.org/10.14233/ajchem.2013.14401

QSPR Study on Enhanced Resonance Light Scattering of Dye-Human Serum Albumin

Y.K. LI

College of Environment Science and Engineering, North China Electric Power University, Baoding, P.R. China

Corresponding author: Tel: +86 312 7522243; E-mail: lyk800@tom.com

(*Received*: 7 August 2012; *Accepted*: 18 February 2013) AJC-13032

Quantitative structure-property relationships models of correlating structure information of dye molecule, with enhanced resonance light scattering intensity and the maximum resonance light scattering wavelength of dye-human serum albumin aggregation were established respectively by partial least squares regression method. The results show that the quantitative relationship models with good fitness and strong predicting power are obtained on small number of sample sets. Also the primary structural parameters of dye molecule affecting the enhanced resonance light scattering intensity and resonance light scattering wavelength were respectively selected and verified.

Key Words: Quantitative structure-property relationships, Resonance light scattering, Dye structure, Human serum albumin.

INTRODUCTION

Quantitative structure-activity relationship/quantitative structure property relationship (QSAR/QSPR) is a chemometrics analysis approach based on the use of regression methods related to structural parameters and material properties, to establish the quantitative structure- activity/property relationships equation[1,2]. In present, a large number of quantitative structure-activity/property relationships models were studies and applied in different areas such as computer-aided drug designs, new product development, environment chemistry and food chemistry *etc*.[3,4]. The quantitative structure-activity/property relationships methods usually includes 2D-QSAR, *e.g.* multiple linear regression, principle component analysis, genetic algorithm, partial least squares and 3D-QSAR method, *e.g.* molecular shape analysis, distance geometry and comparative molecular field analysis *etc*.[5,6].

However, quantitative structure property relationships model applied in resonance light scattering technique are less reported[7,8]. Resonance light scattering is an effective assay method for the aggregation of dyes and macromolecule including human serum albumin. In this work, the relationship between the structure of small dyes and the enhanced resonance light scattering response of dye-human serum albumin combination was researched and quantitative structure property relationships models by partial least squares regression with good fitness and strong predictive ability were obtained on small sample sets of 22 dyes. Furthermore, structure parameters of dyes which have better correlations with the enhanced resonance light scattering responses of dye-human serum albumin were selected and verified.

Theory and algorithm

Selection of quantitative structure property relationships modeling method: In the conventional 2D-QSAR methods, multiple linear regression is the most common statistical method. But how to choose proper parameters has been a problem during the course of multiple linear regression modeling and it may lead to model overfitting. At the same time, when the research system has strong noise or serious interference, it likely lead to the distortion of the model. The PCR method assumes mutual independence of each feature, once the hypothesis does not hold, the character may not be the most predictive and therefore can not construct meaningful quantitative structure property relationships model[9,10].

Partial least squares compared with multiple linear regression and PCR methods, although the differences of residual square sum of three models are not big, but prediction residual error sum of squares (PRESS) of partial least squares model is smaller and therefore have a higher prediction stability. On the other hand, due to the ratio between number of samples and number of variables (parameters) does not qualify, partial least squares model is more suitable for processing the system of more variables and fewer samples, which is often met in some experimental studies. And quantitative structure property relationships model of partial least squares is able to deal with serious multiple correlation data and tolerate some defect data[11,12]. So in this work, quantitative structure property relationships model of partial least squares is selected to construct the function between the structure parameters of 22 dyes and the enhanced resonance light scattering response of dye-human serum albumin.

Quantitative structure property relationships study includes the main steps of the calculation and selection of molecular structure descriptors, then as well as model construction and validation.

The detailed procedures in this essay can be described as follows: (1) The calculation process of the molecular descriptors can refer to literature[5]; (2) The data are randomly divided into a calibration set, an assessing set and a prediction set; (3) Quantitative structure property relationships model-1 of dye structure parameters and the maximum resonance light scattering wavelengths of dye-human serum albumin was constructed. First of all, all of calculated molecular descriptors of dyes were adopted to modeling and then several molecular descriptors mainly responsible for the resonance light scattering wavelength were selected to modeling. And at optimum number of PCs decided by assessing set in partial least squares model, the samples of prediction set were predicted; (4) Quantitative structure property relationships model-2 of dye structure parameters and the enhanced resonance light scattering intensities of dye-human serum albumin was constructed. In the building of the model, the above process (3) was repeated and the same number of PCs in partial least squares model was chosen.

EXPERIMENTAL

The detail steps of the experiment can refer to literature[5]. The resonance light scattering spectra of all the samples were measured at room temperature by a Hitachi M-850 fluorescence spectrophotometer. The enhanced intensities of 22 dye-HSA aggregates and the corresponding maximum resonance light scattering wavelengths were recorded. Quantum structural parameters (Wiener index and Balaban index *etc*.) of 22 dyes were obtained from literature[5] and which have been standardized.

Twenty two samples were used for calibration data set, 6 samples were arbitrarily chosen for validation data set and predict each of the samples. In optimization of procedure parameters, the root mean squared error of prediction (RMSEP) of assessing set is used as evaluation criterion.

$$RMSEP = \left[\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2\right]^{1/2} \quad (1)$$

where $\hat{y}_i$ is the prediction value of the ith sample, y_i is the true value of the *i*th sample, n is the number of prediction samples.

Matlab 6.5 was used as model calculation software.

RESULTS AND DISCUSSION

Determination of number of principal components: The number of principal factor (n_f) is an important parameter in partial least squares modeling. Too little of principal factor number will cause the model underfitting and on the contrary, too much of factor number will lead to the model overfitting, thus reduce the model prediction precision[13]. Principal factor numbers 1-19 are investigated in this work. Fig. 1 shows the variation of root mean squared error of prediction of assessing data set with n_f.

Fig. 1 showed the accuracy of model between structure parameters and resonance light scattering wavelengths changes with different principal factor number of partial least squares. At first, values of root mean squared error of prediction are large and root mean squared error of prediction descends sharply with the increase of principal factor number. After n_f >14, root mean squared error of predictions have a slowly descending trend with the increase of principal factor number and even almost be constant. Therefore, n_f =15 was used further calculations in all of quantitative structure property relationships models in this study.

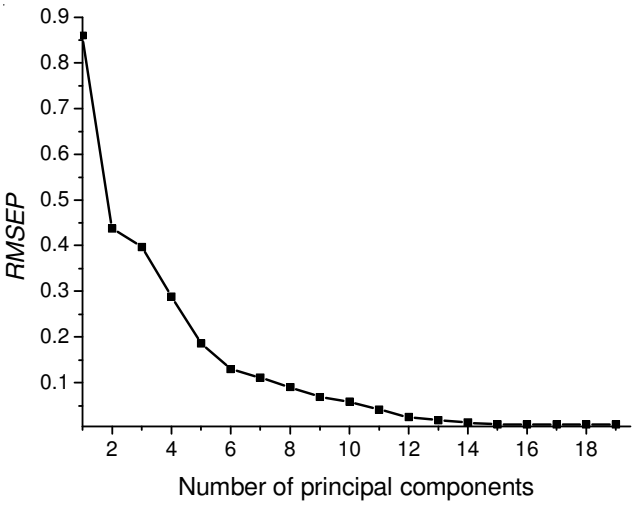

Fig. 1. Variation of root mean squared error of prediction *versus* the number of the principal components

Effect of structural parameters on resonance light scattering wavelength: First of all, all of 16 molecular descriptors of dyes including binding energy, Wiener index and HOMO *etc*. were adopted as independent variable to partial least squares modeling (named as QSPR model-1-1). Root mean squared error of prediction and recovery range of the predicted maximum resonance light scattering wavelengths of dye-human serum albumin were list in Table-1. And then according to literature 5, only three molecular descriptors containing Wiener index, Balaban index and binding energy mainly responsible for the resonance light scattering wavelength were selected to modeling(named as QSPR model-1-2). Root mean squared error of prediction and recovery range of the predicted maximum resonance light scattering wavelengths of dye-human serum albumin were also given in Table-1.

TABLE-1
PREDICTION RESULTS OF MAXIMUM RESONANCE LIGHT SCATTERING WAVELENGTHS OF DYE-HUMAN SERUM ALBUMIN

Model	RMSEP	Recovery (%)
QSPR model-1-1	0.010	100.11-99.09
QSPR model-1-2	0.998	100. 65-98.01

From Table-1, it can be seen that the proposed quantitative structure property relationships model could establish accurate intrinsic correlation between maximum resonance light scattering wavelength and dye structure. Furthermore, how to choose proper parameters has been a key problem of quantitative structure-activity relationship/quantitative structure property relationship. Only the independent variables which

have better correlations with the dependent variables, can be eventually put into the equation, so generate equation of good fitting performance and strong predictive power, especially when number of datasets is small[14]. Therefore, from the results of quantitative structure property relationships model-1-2, it can be concluded that the primary structural parameters of dye molecule affecting resonance light scattering wavelength were verified.

Effect of structural parameters on resonance light scattering intensity: With the same means as study of resonance light scattering wavelengths, relationships between resonance light scattering intensity and dyes structure were also investigated. Firstly, all of 16 molecular descriptors of dyes were adopted to partial least squares modeling(named as QSPR model-2-1). Then three molecular descriptors containing polarizability (polar), refractivity index (RI) and mass mainly responsible for the resonance light scattering wavelength were selected to modeling(named as QSPR model-2-2). Root mean squared error of predictions and recovery ranges of the predicted results were list in Table-2.

From Table-2, it can be seen that, although results of quantitative structure property relationships model-2-2 are inferior to those of quantitative structure property relationships model-2-1, it still proves that the molecular descriptors selected are the primary structural parameters affecting resonance light scattering intensity. Compared with the results of Table-1, performance of quantitative structure property relationships model-2 is a little worse than that of quantitative structure property relationships model-1, which may be interpreted that the enhanced resonance light scattering intensity of dye-human serum albumin was not only attributed to the state of dye molecule, but also attributed to the spatial configuration of human serum albumin. So it still can be concluded that the proposed quantitative structure property relationships model could establish good intrinsic correlation between enhanced resonance light scattering intensity and dye structure.

TABLE-2
PREDICTION RESULTS OF ENHANCED RESONANCE LIGHT SCATTERING INTENSITY OF DYE-HUMAN SERUM ALBUMIN

Model	RMSEP	Recovery (%)
QSPR model-2-1	0.023	101.50-97.94
QSPR model-2-2	1.608	105.51-93.63

Conclusion

Quantitative structure property relationships models of partial least squares were established between the resonance light scattering responses of dye-human serum albumin and dye structure parameters, with good fitness and strong predicting power on small sample datasets. Furthermore, concise and accurate quantitative structure property relationships models were obtained by choosing few parameters from the numerous parameters. Therefore, the proposed two models may be an effective tool for finding new resonance light scattering probes for human serum albumin determination and action mechanism of aggregations in resonance light scattering technique.

ACKNOWLEDGEMENTS

This study is supported by the Natural Science Foundation of Hebei Province, China (Grant No. B2011502061) and the Fundamental Research Funds for the Central Universities (No. 12MS105).

REFERENCES

1. J.J. Sutherland, L.A. O'Brien and D.F. Weaver, *J. Med. Chem.*, **47**, 5541 (2004).
2. R.P. Sheridan, B.P. Feuston, V.N. Meiorov and S.K. Kearsley, *J. Chem. Inf. Comput. Sci.*, **44**, 1912 (2004).
3. E.L. Piparo, F. Fratev, F. Lemke, P. Mazzatorta, M. Smiesko, J.I. Fritz and E. Benfenati, *J. Agric. Food. Chem.*, **54**, 1111 (2006).
4. S. Ajmani, S.C. Rogers, M.H. Barley and D.J. Livingstone, *J. Chem. Inf. Model.*, **46**, 2043 (2006).
5. H.M. Sun, Q. Xie, J.J. Zhou and Z.H. Xu, *Prog. Chem. (Chinese)*, **8**, 79 (1996).
6. R. Cramer, D. Patterson and J.D. Bunce, *J. Am. Chem. Soc.*, **110**, 5959 (1988).
7. R.P. Jia, H.L. Zhai, Y. Shen, X.G. Chen and Z.D. Hu, *Talanta*, **64**, 355 (2004).
8. L.J. Dong, X.G. Chen and Z.D. Hu, *J. Luminescence*, **124**, 85 (2007).
9. Y.K. Li and J. Jing, *Asian J. Chem.*, **23**, 3214 (2011).
10. H. Abdi and L.J. Williams, *Computation. Stat.*, **433**, 2 (2010).
11. D.V. Haaland and D.V. Thomas, *Anal. Chem.*, **1193**, 60 (1988).
12. R.W. Gerlach, B.R. Kowalski and H.O.A. Wold, *Anal. Chim. Acta*, **112**, 417 (1979).
13. Y.K. Li, *Anal. Methods*, **4**, 254 (2012).
14. R.P. Sheridan, *J. Chem. Inf. Comput. Sci.*, **40**, 1456 (2000).