DETERMINATION OF SPECIES DISTRIBUTION AND FORMATION CONSTANTS OF COMPLEXES BETWEEN ION Cu^{2+} AND AMINO ACIDS USING MULTIVARIATE REGRESSION ANALYSIS

Le Thi My Duyen[(1)] – Pham Van Tat [(2)]
(1) University of Dalat – (2) University of Thu Dau Mot

ABSTRACT

In present work, the formation constants, $log\beta_{110}$, $log\beta_{120}$ and the concentration of [M] and $[ML_i]$ in complex solutions of Cu^{2+} and the amino acids were determined by using the quantitative electron structure and properties relationships (QESPRs) and quantitative complex and complex relationships (QCCRs). The relative charge nets for complex structures were calculated by using molecular mechanics MM+ and semiempirical quantum chemistry calculations ZINDO/1. The QESPRs and QCCRs models were constructed by the atomic charge net on complex structures and the multivariate regression analysis. These were employed for approximate determination the formation constants $log\beta_{110}$, $log\beta_{120}$ and the distribution diagram of species [M], $[ML_i]$ in various solutions. These results were compared with those from literature [[3]]. They were also validated by the statistical method ANOVA. The dissimilarities between these models and experimental data are insignificant.

Keyworks: *formation constants, semiempirical quantum chemistry calculations ZINDO/1, multivariate regression analysis, quantitative complex and complex relationships*

*

1. INTRODUCTION

In recent years computer is becoming a helpful tool, an effective means of strong calculation in many different areas. It is used in the inorganic chemistry, analytical chemistry, organic chemistry, physical chemistry, material simulation and data mining [[1],[2]]. The multivariate analysis methods are becoming a convenient and an easy tool for building empirical and theoretical models. The linear correlation relationships can be assessed from different characteristics of the system.

Formation constants of complexes are one of the most important factors to explain reaction mechanisms, chemical properties of biological systems in nature. From the formation constants we can calculate the equilibrium concentration of components in a solution. These can forecast the changes of complex electronic structure in solution from the initial concentration of the central ion and ligand. In recent years the formation constants of the complexes can be determined by experimental ways using UV-Vis spectral data [[7]] and computational techniques. The theoretical methods used for predicting stability constants of complexes based on the relationship between structural and topological descriptors were introduced [[8]]. A few topological descriptors of complexes Cu^{2+} with amino acids were determined by molecular mechanics methods [[4],[5],[6]].

In this work, the linear relationship between topological parameters and formation constants of the complexes is not done. We focused only on constructing the quantitative electron structure and properties relationships (QESPRs) from the atomic charge nets and formation constants of complexes Cu^{2+} with amino acids. These linear models were carried out by using principal component analysis. The atomic charges are calculated using the semiempirical quantum chemical method ZINDO/1 SCF MO. We also reported the quantitative complex and complex relationships (QCCRs) using the atomic charges. The formation constants $\log\beta_{110}$ and $\log\beta_{120}$ of complexes Cu^{2+} and amino acids were predicted from these linear models. Those were also compared to predictive ability of artificial neural networks. The distribution diagram of ions in complex solution was built upon the predicted values of $\log\beta_{110}$ and $\log\beta_{120}$. All the results were also compared with experimental data from literature.

2. METHODS

2.1. Reaction equations

In aqueous solution, amino acid dissociates into anion L^{2-} then reacts with metal ion Cu^{2+}:

$$k\mathrm{Cu}^{2+} + l\mathrm{L}^{2-} + m\mathrm{H}^{+} = [\mathrm{Cu}_k\mathrm{L}_l\mathrm{H}_m] \quad (1)$$

Ions Cu^{2+} participate in reactions with L^{2-} ligands to form complexes $[Cu_kL_lH_m]$:

$$\beta_{klm} = \frac{[\mathrm{Cu}_k\mathrm{L}_l\mathrm{H}_m]}{[\mathrm{Cu}^{2+}]^k[\mathrm{L}^{2-}]^l[\mathrm{H}^+]^m} \quad (2)$$

2.2. Data and software

The values of $\log\beta_{110}$ and $\log\beta_{120}$ (with $k = 1$; $l = 1, 2$; $m = 0$) of complexes between Cu^{2+} ion and the corresponding amino acids were taken from the literature [[3]], given in Table 1.

The complexes Cu^{2+} with the amino acids were built and optimized by molecular mechanics MM+. The atomic charges of complexes were calculated by semiempirical quantum method ZINDO/1 SCF MO using Hyperchem 7.5[[12]]. The raw data were reduced by principal component analysis using Minitab 14.0[[11]]. The regression analysis and statistical evaluation were performed by the programs Regress 2006 [[10]] and MS-EXCEL [[1]]. The artificial neural network (ANN) was also constructed by INForm[[13]]. This was used to compare with those from the ordinary regression (OR) and principal component regression (PCR). Models were screened by using the values R^2-training and R^2-prediction.

Models were assessed by the formula:

$$R^2 = \left(1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}\right)100 \quad (3)$$

Where Y_i, $\hat{Y}_i$ and $\bar{Y}$ are the experimental, calculated and average values.

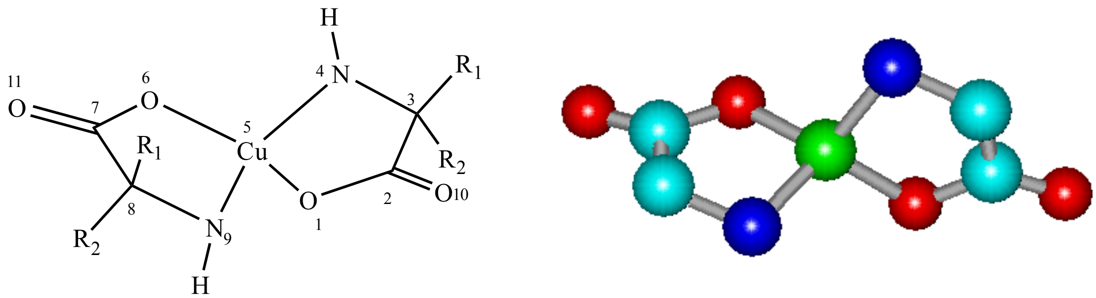

Figure 1: *The structure of complex between Cu^{2+} ion and amino acids.*

Table 1: The complexes between Cu2+ and amino acids, experimental formation constants [3]

Complex	Substitution		$\log\beta_{110}$	$\log\beta_{120}$	Complex	Substitution		$\log\beta_{110}$	$\log\beta_{120}$
	R_1	R_2				R_1	R_2		
Com-1	-H	-H	8.38	15.70	Com-5	$-C_2H_5$	$-C_2H_5$	6.88	12.86
Com-2	$-CH_3$	-H	7.94	14.59	Com-6	$-n-C_3H_7$	-H	7.25	13.31
Com-3	$-CH_3$	$-CH_3$	7.30	13.56	Com-7	$-n-C_4H_9$	-H	7.32	13.52
Com-4	$-C_2H_5$	-H	7.34	13.55	Com-8	$-izo-C_3H_7$	-H	6.70	12.45

3. RESULTS AND DISCUSSION

3.1. Constructing models QESRs

The atomic charge data of the complexes were divided into a training set and a test set. The atomic charge data were calculated by the semiempirical quantum method ZINDO/1, after optimizing by molecular mechanics MM+ with gradient 0.05, given in Table 2.

Table 2. *The atomic charge distribution Q_i in complex between Cu^{2+} and amino acids.*

Complex	O_1	C_2	C_3	N_4	Cu_5	O_6	C_7	C_8	N_9	O_{10}	O_{11}
Com-1	-0.0657	0.3415	-0.2127	0.3139	-0.5511	-0.0664	0.3408	-0.2135	0.3116	-0.3330	-0.3317
Com-2	-0.0673	0.3407	-0.1831	0.3092	-0.5440	-0.0671	0.3411	-0.1828	0.3101	-0.3309	-0.3314
Com-3	-0.0665	0.3418	-0.1502	0.3081	-0.5467	-0.0667	0.3415	-0.1504	0.3073	-0.3316	-0.3312
Com-4	-0.0854	0.3296	-0.1877	0.2472	-0.5368	-0.0625	0.3445	-0.1850	0.3477	-0.3205	-0.3375
Com-5	-0.0665	0.3345	-0.1426	0.3067	-0.5511	-0.0696	0.3484	-0.1484	0.3153	-0.3325	-0.3396
Com-6	-0.0685	0.3410	-0.1953	0.3114	-0.5527	-0.0661	0.3432	-0.1910	0.3086	-0.3312	-0.3348
Com-7	-0.0362	0.3431	-0.1706	0.3105	-0.5785	-0.0343	0.3405	-0.1687	0.3124	-0.3347	-0.3348
Com-8	-0.0636	0.3402	-0.1861	0.3073	-0.5516	-0.0707	0.3462	-0.1882	0.3219	-0.3321	-0.3351

QESPRs models were built from the training group with principal component analysis technique. The component scores were determined from covariance matrix. The components Z_i are founded by the equation (4). The formation constants of complexes were calculated by using the regression equation (5) for the components Z_i.

The regression model is represented in:

$$Z_{i,n,j} = \sum_{i,k,j} PC_{i,n,j} Q_{j,n} \quad \text{with } i = 1\text{-}5;\ j = 1\text{-}11;\ n = 1\text{-}8 \qquad (4)$$

Where PC_i is the principal component *i*th in coefficient matrix in which it includes 8 complexes and 11 atomic charge values Q_i.

$$\log \beta_{klm} = \sum_{i,k} b_{k,l,m} Z_{i,k} + b_{klm} \quad \text{with } k = 1;\ l = 1, 2;\ m = 0 \qquad (5)$$

***Table 3**. The principal component scores for the corresponding atomic charges.*

The atomic number	PC_1	PC_2	PC_3	PC_4	PC_5
O_1	-0.203	-0.316	0.429	-0.158	0.498
C_2	-0.031	-0.147	-0.004	0.203	0.309
C_3	-0.659	0.300	-0.082	-0.145	0.163
N_4	-0.273	-0.683	-0.364	-0.217	0.051
Cu_5	0.150	0.261	-0.461	0.309	0.395
O_6	-0.082	-0.094	0.632	0.391	-0.145
C_7	-0.014	0.050	-0.046	-0.399	-0.192
C_8	-0.616	0.295	-0.010	0.182	-0.125
N_9	0.189	0.370	0.246	-0.536	0.418
O_{10}	0.066	0.129	0.018	0.123	-0.069
O_{11}	0.014	-0.058	-0.040	0.351	0.467

From component equation (4), Z_i constituents were identified, and value Z_i was the combination of the principal components PC_i (*i* from 1 to 5). The coefficient matrix is given in Table 3, at each atomic position, respectively. The principal components Z_i (*i* from 1 to 5) were obtained from principal components PC_i, are depicted in Table 4. The importance of the principal components was validated using the eigenvalues, represented in Figure 2a.

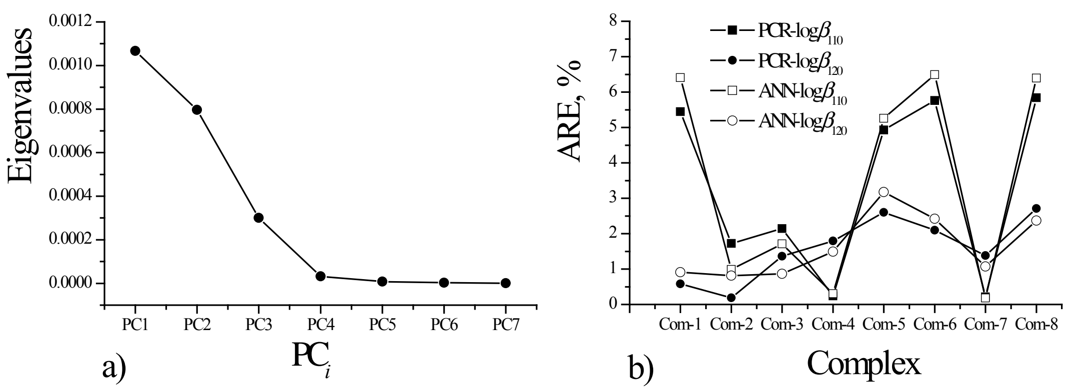

***Figure 2**. a) Eigenvalues change of principal components PC_i;*
b) Comparison of ARE% values of PCR models with artificial neural network (ANN).

The independent variables of component Z_i are illustrated in Table 4, which were used to build regression models with the dependent variable $\log\beta_{110}$ and $\log\beta_{120}$ ($k = 1$, $l = 1, 2$, $m = 0$), respectively. The general regression model (5) for values $\log\beta_{120}$ and $\log\beta_{110}$ were tested by using the leave-one-out cross-validation technique.

***Table 4**. The components Z_i obtained from relation (4) for complexes Cu^{2+} and amino acids.*

Complex	Z_1	Z_2	Z_3	Z_4	Z_5
Com-1	0.139	-0.400	0.156	-0.653	-0.194
Com-2	0.103	-0.376	0.150	-0.648	-0.192
Com-3	0.061	-0.358	0.149	-0.646	-0.193
Com-4	0.137	-0.311	0.175	-0.652	-0.195
Com-5	0.056	-0.351	0.151	-0.660	-0.197
Com-6	0.114	-0.386	0.154	-0.651	-0.200
Com-7	0.072	-0.391	0.199	-0.652	-0.194
Com-8	0.109	-0.375	0.157	-0.662	-0.191

The cross-validation results were carried out by the leave-one-out technique for the principal component regression model. These in turn were compared with the calculated results from the artificial neural net I(5)-HL(2)-O(2). The error back-propagation algorithm was used to train this neural net.

***Table 5**. Comparison of the predicted stability constants using the principal component regressions and neural network I(5)-HL(2)-O(2) in the leave-one-out case.*

Complex	Ref. [[3]]		Principal component regression				I(5)-HL(2)-O(2)			
	$\log\beta_{110}$	$\log\beta_{120}$	$\log\beta_{110}$	$\log\beta_{120}$	ARE,%		$\log\beta_{110}$	$\log\beta_{120}$	ARE,%	
					$\log\beta_{110}$	$\log\beta_{120}$			$\log\beta_{110}$	$\log\beta_{120}$
Com-1	8.380	15.700	7.924	14.694	5.445	6.408	8.332	15.557	0.579	0.909
Com-2	7.940	14.590	7.803	14.446	1.723	0.987	7.955	14.708	0.184	0.810
Com-3	7.300	13.560	7.457	13.792	2.144	1.709	7.400	13.677	1.364	0.862
Com-4	7.340	13.550	7.322	13.509	0.241	0.306	7.472	13.752	1.796	1.489
Com-5	6.880	12.860	6.541	12.184	4.927	5.259	6.701	12.452	2.603	3.173
Com-6	7.250	13.310	7.668	14.174	5.764	6.489	7.402	13.632	2.098	2.422
Com-7	7.320	13.520	7.304	13.496	0.215	0.178	7.421	13.665	1.378	1.070
Com-8	6.700	12.450	7.091	13.246	5.836	6.396	6.882	12.745	2.710	2.367

The topological structure of this neural network consists of three layers: an input layer I(5) with 5 nodes (components Z_1, Z_2, Z_3, Z_4, Z_5), an output layer with two nodes ($\log\beta_{110}$ and $\log\beta_{120}$), and a hidden layer HL(2), the optimum number of hidden layer nodes was found to be 2. The training parameters of this neural net were found by using a trial and error approach. The best parameters consist of the sigmoid transfer function on the hidden and output nodes, momentum 0.7, learning rate 0.7 and training epochs 1000. The MSE value of 0.000236 obtained from the training process for $\log\beta_{110}$ and $\log\beta_{120}$ together. The $\log\beta_{110}$ and $\log\beta_{120}$ values derived from the different models using the atomic charges were compared with those from the literature [[3]].

The calculated results were assessed by statistical method ANOVA, for the predicted value $\log\beta_{110}$ ($F = 0.034 < F_{0.05} = 3.467$), for $\log\beta_{120}$ ($F = 0.020 < F_{0.05} = 3.467$), for overall validation based on values ARE% for both $\log\beta_{110}$ and $\log\beta_{120}$ ($F = 2.058 < F_{0.05} = 2.947$). Consequently the formation constants resulting from PCR model and ANN I(5)-HL(2)-O(2) are not different.

3.2. Constructing models QCCRs

Besides the regression constructing technique and artificial neural network based on the atomic charge distribution of the complex, in this work we also built the regression models using the complex structure relationships, as illustrated in following equation (6):

$$\text{Com-}i = \sum_{j=1}^{m} b_j \text{Com-}j + b_0 \quad \text{with } m = 1 - 8 \quad (6)$$

where Com-i and Com-j are target complex i and predicted complexes j; b_j is the parameter for complex j; b_0 is the constant.

The QCCRs models are constructed by the ordinary regression techniques. Each complex in Table 2 was selected as a target complex, and independent variables were chosen from remaining compounds. The atomic net charge of complexes in Table 2 are used to establish the regression models using forward and elimination technique. The best models were found by this technique. The selected complex models QCCRs consist of the predicted complexes with the similar structural properties.

***Table 6**. The quantitative complex and complex relationships, and regression-statistical values.*

Statistical values, predictive complex	Target complex							
	Com-1	Com-2	Com-3	Com-4	Com-5	Com-6	Com-7	Com-8
R^2-training	99.999	100.000	99.994	99.537	99.978	99.996	99.883	99.996
R^2-adjusted	99.998	99.999	99.992	99.486	99.973	99.995	99.833	99.994
Standard error, SE	0.002	0.001	0.003	0.022	0.005	0.002	0.013	0.002
R^2-prediction	99.995	99.999	99.986	99.275	99.956	99.991	99.758	99.992
Constant	-0.001	0.001	0.001	-0.004	-0.002	-0.001	0.006	0.001
Com-1	-	0.536	-	-	-1.381	0.278	4.895	0.557
Com-2	1.859	-	0.952	-	2.399	0.727	-8.563	-
Com-3	-0.910	0.490	-	-	-	-	4.649	-
Com-4	-0.052	0.029	-	-	-	-	-	0.105
Com-5	-	-	0.757	-	-	-	-	0.344
Com-6	-	-	-	-	-	-	-	-
Com-7	0.108	-0.057	-	-	-	-	-	-
Com-8	-	-	-0.706	0.974	-	-	-	-

The complex model (7) for the Com-1 complex is shown in

$$\text{Com-1} = -0.001 + 1.859(\text{Com-2}) - 0.910(\text{Com-3}) - 0.052(\text{Com-4}) + 0.108(\text{Com-7}) \quad (7)$$

The 8 regression models between different complex structures with their statistical values depict the regression quality, shown in Table 6. All R^2-training and R^2-prediction values are larger than 99% from the standard statistical values. The complex structural models QCCRs were used to estimate the target complex properties using features of predicted complexes in the regression model. In this work we used the formation constants of the complexes Cu^{2+} with amino acids, as a important properties for calculating the stability constant of target complex in the respective models. The predicted results for $\log\beta_{120}$ and $\log\beta_{110}$ were validated by the values ARE% for the models, are given in Table 7.

***Table** 7. The predicted formation constants by complex models QCCRs with values ARE,%.*

Complex	Ref.[[3]]		Models QCCRs		ARE,%	
	$\log\beta_{110}$	$\log\beta_{120}$	$\log\beta_{110}$	$\log\beta_{120}$	$\log\beta_{110}$	$\log\beta_{120}$
Com-1	8.380	15.700	8.524	15.535	1.717	1.050
Com-2	7.940	14.590	7.861	14.675	0.997	0.583
Com-3	7.300	13.560	7.822	14.482	7.150	6.796
Com-4	7.340	13.550	6.523	12.125	11.131	10.520
Com-5	6.880	12.860	7.475	13.321	8.650	3.582
Com-6	7.250	13.310	8.104	14.976	11.777	12.515
Com-7	7.320	13.520	6.978	14.973	4.669	10.746
Com-8	6.700	12.450	7.804	14.589	16.483	17.182

The absolute values of relative errors ARE% are calculated by

$$\text{ARE,\%} = \frac{\left|\log\beta_{k,l,m-\text{exp}} - \log\beta_{k,l,m-\text{cal}}\right|}{\left|\log\beta_{k,l,m-\text{exp}}\right|}.100 \quad (8)$$

Where $\log\beta_{k,l,m\text{-exp}}$ and $\log\beta_{k,l,m\text{-cal}}$ are the experimental and calculated formation constants.

From the obtained results for $\log\beta_{120}$ $\log\beta_{110}$ in Table 7, distribution diagram of ions is illustrated for the complex $Cu(Gly)_2$ and Cu $(GlyMe)_2$, as is shown in Figure 3.

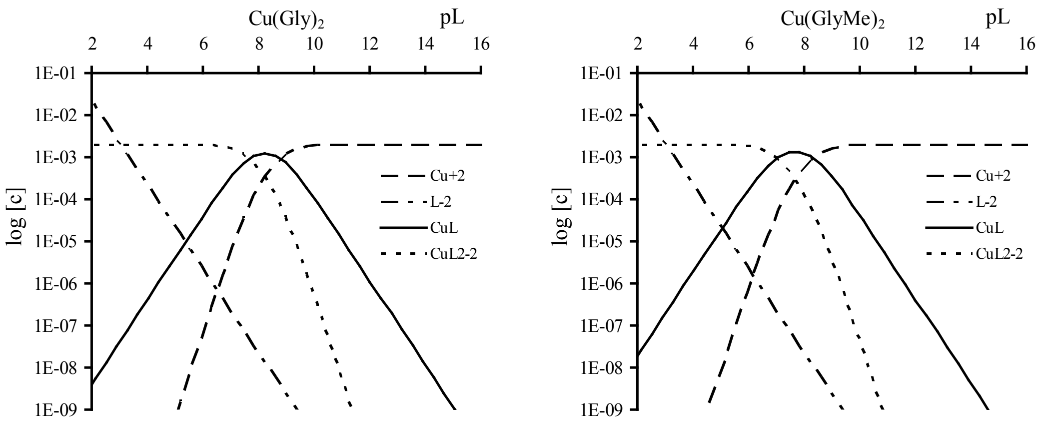

***Figure 3**. Species distribution of the complex solution $Cu(Gly)_2$ and $Cu(GlyMe)_2$*

The $\log\beta_{120}$ $\log\beta_{110}$ values in Table 7 obtained from the ordinary regression techniques are in very good agreement with the reference values [[3]]. The one-way ANOVA is used to evaluate $\log\beta_{110}$ values ($F = 0.705 < F_{0.05} = 4.600$) and $\log\beta_{120}$ ($F = 1.473 < F_{0.05} = 4.600$), and values ARE% ($F = 0.0003 < F_{0.05} = 4.6001$). Thus, PCR model for $\log\beta_{110}$, $\log\beta_{120}$ and QCCRs model fitted well with those from neural network I(5)-HL(2)-O(2) and literature [[3]].

4. CONCLUSION

This work has successfully built the quantitative electron structure and properties (QESPRs) and the quantitative complex and complex relationships (QCCRs) from complexes Cu^{2+} and amino acids using the atomic charge net. The formation constant values and values ARE% were assessed by ANOVA.

Determination of formation constants of complexes Cu^{2+} and amino acids is one important direction to understand and to explain many biological properties. This research can be applied in different ways as a potential method to quickly determine the formation constants of complexes between metal and amino acids combining theory and experimental way. The ion H^+ affects for complex formation, this will be carried out by next work.

*

XÁC ĐỊNH PHÂN BỐ CÁC CẤU TỬ VÀ HẰNG SỐ TẠO THÀNH CỦA CÁC PHỨC GIỮA ION Cu^{2+} VÀ CÁC AXIT AMINO SỬ DỤNG PHƯƠNG PHÁP PHÂN TÍCH HỒI QUY ĐA BIẾN

Lê Thị Mỹ Duyên[(1)] – Phạm Văn Tất[(2)]

(1) Trường Đại học Đà Lạt - (2) Trường Đại học Thủ Dầu Một

TÓM TẮT

Trong công trình này, các hằng số tạo thành $\log\beta_{110}$, $\log\beta_{120}$ *và nồng độ [M] và* $[ML_i]$ *trong các dung dịch phức của* Cu^{2+} *với các acid amino được xác định bằng mối quan hệ định lượng cấu trúc điện tử và tính chất (QESPRs) và quan hệ định lượng phức chất và phức chất (QCCRs). Mạng lưới điện tích tương đối của các cấu trúc phức được tính toán bằng cơ học phân tử MM+ và hóa lượng tử bán kinh nghiệm ZINDO/1. Các mô hình QESPRs và QCCRs được xây dựng bằng mạng điện tích nguyên tử của phức chất và phân tích hồi quy đa biến số. Những mô hình này được dùng để xác định gần đúng hằng số tạo thành* $\log\beta_{110}$, $\log\beta_{120}$ *và giản đồ phân bố các cấu tử [M] và* $[ML_i]$ *trong các dung dịch. Các kết quả này được so sánh với những giá trị thực nghiệm tham khảo[[3]] và cũng được đánh giá bằng phương pháp thống kê ANOVA. Sự khác nhau giữa các phương pháp lý thuyết và dữ liệu thực nghiệm tham khảo là không có ý nghĩa.*

Từ khóa: *hằng số tạo thành, tính toán lượng tử bán thực nghiệm ZINDO/1, phân tích hồi quy, quan hệ phức chất và phức chất*

REFERENCES

[1] E. J. Billo., *Excel For Scientists And Engineers-Numerical Methods.,* Wiley, 2007.

[2] D. Harvey, *Modern analytical Chemistry,* Mc.Graw Hill, Boston, Toronto, 2000.

[3] B. Grgas, S. Nikolic, N. Paulic, and N. Raos., *Croatica Chemica Acta,* 72, 885-895, 1999.

[4] A. Milicevic and N. Raos., *Acta. Chim.* Slov, 56, 373-378, 2009.

[5] M. Ante and N. Raos., *Croatica Chemica Acta,* 79, 281-290, 2006.

[6] S. Nikolic and N. Raos., *Croatica Chemica Acta,* 74, 621-631, 2001.

[7] N. Raos., *Croatica Chemica Acta,* 75, 117-120, 2002.

[8] Pham Van Tat., *Development of QSAR and QSPR,* Publisher of Natural sciences and Technique, HaNoi, 2009.

[9] Ha Tan Loc, Pham Van Tat., *J. Analytical Sciences,* Vol. 15, No 4, 2010.

[10] D. D. J.Werner, P. R.Yeater, *Essential Regression and Experimental Design for Chemists and Engineer,* 2000.

[11] *MINITAB v 14 for Windows,* Minitab. Inc, Ltd, 2010.

[12] *HyperChem Release 7.5 for Windows,* Hypercube Inc. Getting Started., USA, 2002.

[13] *INForm v2.0,* Intelligensys Ltd., UK, 2002.