Influence of Chelate Ring Interactions on Copper(II) Chelate Stability Studied by Connectivity Index Functions

Ante Miličević and Nenad Raos*

Institute for Medical Research and Occupational Health, Ksaverska c. 2, P.O. Box 291, HR-10000 Zagreb, Croatia

Received: March 7, 2008; Revised Manuscript Received: May 30, 2008

Linear models for estimation of the first (K_1), second (K_2), and overall stability constant (β_2) based on the valence connectivity index of the third order ($^3\chi^v$) were developed and checked on four sets of copper(II) chelates (with diamines, *N*-alkylated glycines, and naturally occurring amino acids, including their mixed complexes). Univariate models were valid when log K_1 and log K_2 values were linearly correlated, i.e., when there was no interaction between chelate rings. The univariate models proved applicable for estimation of all three stability constants of complexes with diamines and *N*-alkylated glycines, but for complexes with amino acids additional terms were needed (bivariate models). Models reproduced stability constants with an error usually less than 0.3 log *K* units.

1. Introduction

Topological index is a half-a-century-old concept, but despite wide application of many kinds of topological indices in chemistry (for calculation of physicochemical parameters,[1–4] chromatography-related properties,[5] QSAR analysis,[6–9] etc.) and various attempts to explain their exact physical meaning,[10–16] some chemists are still sceptical about the suitability of these molecular descriptors for solving complex chemical problems. This scepticism probably stems from the simplicity of the indices; it seems naive to derive all the properties of a molecule from a number representing nothing but its constitutional formula. However, from the other point of view, the constitutional formula and index derived from it have to be representative of all molecular properties. In this simple assumption lays the possibility to successfully correlate topological indices to many chemical parameters.

In our effort to develop simple and efficient model(s) for estimation of stability constants of coordination compounds based on topological indices,[17–21] we detected a problem with overall stability constants, β_2, namely, that log β_2 values were usually reproduced with about twice as large error as log K_1. For instance, copper(II) complexes with diamines yielded $SE_{cv} = 0.38$ for log K_1 and $SE_{cv} = 0.93$ for log β_2.[19] It has to be pointed out that we could not estimate overall stability constants of mixed amino acid copper(II) complexes using the linear univariate model.[19] We also have yet to make a systematic study to develop a model for estimation of the second stability constants, log K_2.

The aim of this paper was to develop well-defined procedures for the estimation of K_1, K_2, and β_2 constants. For this purpose we used test sets of copper(II) chelates with diamines, *N*-alkylated glycines, and two sets with naturally occurring amino acids, which we had already considered.[19]

2. Methods

2.1. Calculation of Topological Indices. We calculated topological indices using a program system DRAGON 2.1, written by R. Todeschini and co-workers,[22] which is capable of yielding 262 topological indices in a single run along with many other molecular descriptors. Connectivity matrices were constructed with the aid of the Online SMILES Translator and Structure File Generator.[23]

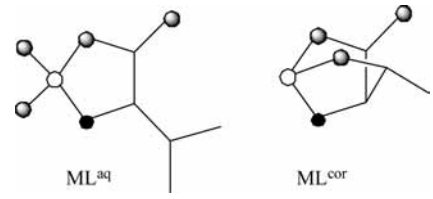

Figure 1. Two graph representations of copper(II) chelates with valine (ML^{aq}) and threonine (ML^{cor}). Heteroatoms are marked with (○) Cu, (●) N, and (gray circles) O.

All models were developed using the $^3\chi^v$ index (the valence molecular connectivity index of the third order), which was defined as[24–26]

$$^3\chi^v = \Sigma[\delta(i)\,\delta(j)\,\delta(k)\,\delta(l)]^{-0.5} \quad (1)$$

where $\delta(i)$, $\delta(j)$, $\delta(k)$, and $\delta(l)$ are weights (valence values) of vertices (atoms) *i*, *j*, *k*, and *l* making up the path of length 3 (three consecutive chemical bonds) in a vertex-weighted molecular graph. The valence value, $\delta(i)$, of a vertex *i* is defined by

$$\delta(i) = [Z^v(i) - H(i)]/[Z(i) - Z^v(i) - 1] \quad (2)$$

where $Z^v(i)$ is the number of valence electrons belonging to the atom corresponding to vertex *i*, $Z(i)$ is its atomic number, and $H(i)$ is the number of hydrogen atoms attached to it. For instance, δ values for primary, secondary, tertiary, and quaternary carbon atoms are 1, 2, 3, and 4, respectively; for oxygen in the OH group it equals 5, and for the NH_2 group $\delta(N) = 3$. It has to be pointed out that $^3\chi^v$ is only a member of the family of valence connectivity indices $^n\chi^v$, which differ between each other by the path length, i.e., the number of δ′s in the summation term, eq 1.

Relying on our earlier experience,[19] in this study we used two kinds of graph representations of complexes ML, ML_2, and

* To whom correspondence should be addressed. E-mail: raos@imi.hr.

10.1021/jp802018m CCC: $40.75
Published on Web 07/30/2008

TABLE 1: Univariate and Bivariate Linear Regressions of Stability Constants on the Connectivity Index $^3\chi^v$

		regression coefficients (eqs 3, 5, 6)						
set	dependent variable	a_1(SE)	a_2(SE))	a_3(SE)	intercept(SE)	r	SE	SE_{cv}
1	log K_1	−0.74(14)			13.01(57)	0.837	0.44	0.49
	log K_2	−1.51(16)			13.97(64)	0.940	0.50	0.60
	log β_2	−2.25(28)			27.0(11)	0.918	0.88	1.03
2	log K_1	−0.68(14)			9.64(42)	0.858	0.28	0.37
	log K_2	−0.36(10)			7.41(31)	0.767	0.21	0.26
	log β_2	−1.04(22)			17.05(68)	0.845	0.45	0.59
3	log K_1	−0.129(37)			8.41(12)	0.799	0.09	0.11
	log K_2	0.136(57)	−0.076(25)		6.57(13)	0.782	0.07	0.11
	log β_2	0.100(74)	−0.124(32)		14.79(18)	0.905	0.09	0.14
4	log K_1	−0.220(44)			8.61(10)	0.803	0.08	0.08
	log K_2	0.244(35)	−0.141(23)		6.410(98)	0.898	0.06	0.08
	log β_2		−0.167(34)	0.197(56)	15.170(86)	0.845	0.12	0.15

TABLE 2: Experimental and Estimated Stability Constants of Copper(II) Complexes with 1,2-Diaminoethane and Its Derivatives (Set 1)

		experimental[28–30]	estimated (by cross validation)		
no.	ligands[a]	log β_2	log K_1	log K_2	log β_2
1	1,2-diaminoethane	20.13	11.31	10.52	21.83
2	*N*-methyl-1,2-diaminoethane	19.11	10.40	8.70	19.10
3	*N*-ethyl-1,2-diaminoethane	18.57	10.17	8.17	18.34
4	*N*-propyl-1,2-diaminoethane	18.14	10.15	8.11	18.26
5	*N*-butyl-1,2-diaminoethane	18.21	9.93	7.64	17.57
6	*N*-isopropyl-1,2-diaminoethane	16.52	10.04	7.79	17.83
7	*N*,*N*′-dimethyl-1,2-diaminoethane	18.10	9.65	7.24	16.89
8	*N*,*N*′-diethyl-1,2-diaminoethane	15.62	9.20	6.22	15.42
9	*N*,*N*′-dipropyl-1,2-diaminoethane	14.34	9.24	6.24	15.48
10	*N*,*N*′-dibutyl-1,2-diaminoethane	13.51	8.76	5.48	14.24
11	1,2-diaminopropane	20.06	10.95	9.80	20.74
12	(*R*,*S*)-2,3-diaminobutane	20.06	10.43	8.71	19.14
13	(*R*,*R*)-2,3-diaminobutane	21.21	10.36	8.66	19.02
14	2-methyl-1,2-diaminopropane	19.58	10.73	9.29	20.02
rms			0.49	0.60	1.03

[a] Representation ML^{aq} was applied for all ligands.

MLA, where L and A stand for ligands and M denotes the central atom (Cu^{2+}). The first, aq representation (ML^{aq}, etc.) is derived from the constitutional formula of a complex with additionally bound two water molecules (diaqua complexes). The second cor representation (ML^{cor}, etc.) is derived from the constitutional formula of a complex by drawing a new edge, i.e., bonds, to simulate additional interaction between ligand side chain and the central atom (Figure 1). For each stability constant we developed a regression model using any combination of these aq and cor representations depending on the nature of the complex in the set.

2.2. Regression Models. In this report we applied univariate and bivariate models. For estimation of the first stability constant we used the univariate linear model in all sets

$$\log K_1 = a_1[^3\chi^v(\mathrm{ML})] + b \quad (3)$$

where ML denotes the ML^{aq} or ML^{cor} representation of the molecular graph.[19]

For estimation of the second stability constant, i.e., equilibrium constant of the reaction

$$\mathrm{ML} + \mathrm{A} \rightleftarrows \mathrm{MLA} \quad (4)$$

we used a bivariate function in the form

$$\log K_2 = a_1[^3\chi^v(\mathrm{MA})] + a_2[^3\chi^v(\mathrm{MLA}) - {}^3\chi^v(\mathrm{ML}) - {}^3\chi^v(\mathrm{MA})] + b \quad (5)$$

which was reduced to a univariate model ($a_2 = 0$) if all ML and MA were represented as ML^{aq} (MA^{aq}). Obviously, if there is no additional apical interaction between the side chain and copper(II), coordination of the second ligand will not be influenced by the first ligand. Thus, the second term is added for mutual interactions of the ML and MA chelate rings.

For estimation of overall stability constants we used a general function

$$\log \beta_2 = a_1[^3\chi^v(\mathrm{MA})] + a_2[^3\chi^v(\mathrm{MLA}) - {}^3\chi^v(\mathrm{ML}) - {}^3\chi^v(\mathrm{MA})] + a_3|^3\chi^v(\mathrm{ML}) - {}^3\chi^v(\mathrm{MA})| + b \quad (6)$$

which is reduced to a bivariate or univariate function according to the proposed model. Namely, for estimation of log β_2 for ternary complexes ($L \neq A$, Set 4), $a_1 = 0$, and for binary complexes ($L \equiv A$, Set 3) the third term is omitted. If ML^{aq} representations are used, $a_2 = a_3 = 0$ (Sets 1 and 2), i.e., eq 6 is reduced to univariate function. As said above (section 2.1), the choice of ML^{aq} or ML^{cor} representation depends on the nature of the ligand (see Tables 2–5).

2.3. Regression Calculations. Regression calculations, including the leave-one-out procedure of cross validation, cv, were done using the CROMRsel program.[27] The standard error of cross-validation estimate is defined as

$$SE_{cv} = \sqrt{\sum_i \frac{\Delta X_i}{N}} \quad (7)$$

where ΔX and N denote cv residuals and the number of reference points, respectively.

TABLE 3: Experimental and Estimated Stability Constants of Copper(II) Complexes with Aliphatic Amino Acids and *N*-Alkylated Glycines (Set 2)

		experimental[31–33]	estimated (by cross validation)		
no.	ligands[a]	log β_2	log K_1	log K_2	log β_2
1	glycine	15.17	8.42	6.73	15.15
2	alanine	14.82	8.12	6.59	14.71
3	valine	14.79	7.78	6.42	14.20
4	leucine	14.34	7.72	6.44	14.15
5	*N*-methylglycine	14.59	7.79	6.41	14.20
6	*N*,*N*-dimethylglycine	13.65	7.32	6.16	13.48
7	*N*-Ethylglycine	13.55	7.56	6.31	13.87
8	*N*,*N*-diethylglycine	12.86	6.15	5.53	11.67
9	*N*-propylglycine	13.31	7.54	6.31	13.84
10	*N*-butylglycine	13.52	7.32	6.18	13.50
11	*N*-isopropylglycine	12.45	7.42	6.25	13.67
rms			0.37	0.26	0.59

[a] Representation ML^{aq} was applied for all ligands.

TABLE 4: Experimental and Estimated Stability Constants of Binary Copper(II) Complexes with Naturally Occurring Amino Acids (Set 3)

		experimental[34–37]	estimated (by cross validation)			
no.	ligands[a]	log β_2	log K_1	log K_2	log K_1K_2	log β_2
1	glycine (aq)	15.11	8.15	6.78	14.93	14.91
2	alanine (aq)	14.99	8.10	6.89	14.99	15.02
3	serine (cor)	14.57	8.13	6.72	14.85	14.81
4	valine (aq)	14.91	8.06	6.96	15.02	15.12
5	threonine (cor)	14.77	8.04	6.75	14.79	14.73
6	leucine (aq)	15.13	8.03	6.94	14.97	15.03
7	phenylalanine (cor)	14.77	7.89	6.77	14.66	14.65
8	tyrosine (cor)	14.78	7.96	6.84	14.80	14.84
9	methionine (cor)	14.53	7.72	6.88	14.60	14.54
rms			0.11	0.11	0.14	0.14

[a] For the ligands marked with aq representation ML^{aq} was used, and those marked with cor were represented with graph ML^{cor}.

TABLE 5: Experimental and Estimated Stability Constants of Ternary Copper(II) Complexes with Naturally Occurring Amino Acids (Set 4)

		experimental[34,37]	estimated (by cross validation)				
no.	ligands[a]	log β_2	log K_1	log K_2	log K_1K_2	log β_2	log β_2[b]
1	glycine, alanine	15.10	8.21	6.99	15.20	15.33	15.02
2	glycine, serine	15.09	8.21	6.79	15.00	15.06	14.94
3	glycine, threonine	15.13	8.21	6.92	15.13	15.16	15.00
4	glycine, tyrosine	15.35	8.21	7.11	15.31	15.30	15.11
5	glycine, phenylalanine	15.36	8.21	7.16	15.37	15.31	15.14
6	alanine, serine	15.07	8.11	6.79	14.90	14.97	14.89
7	alanine, threonine	15.08	8.11	6.93	15.04	15.09	14.96
8	alanine, tyrosine	15.30	8.11	7.11	15.23	15.23	15.06
9	alanine, phenylalanine	15.26	8.11	7.19	15.30	15.25	15.09
10	valine, tyrosine	15.25	8.02	7.10	15.12	15.14	15.02
11	serine, threonine	14.55	8.09	6.75	14.84	14.85	14.87
12	serine, tyrosine	14.78	8.09	6.90	15.00	14.97	14.97
13	serine, phenylalanine	14.77	8.09	6.94	15.03	14.98	14.99
14	threonine, tyrosine	14.98	7.98	6.91	14.89	14.87	14.89
15	threonine, phenylalanine	14.96	7.98	6.93	14.91	14.83	14.90
16	tyrosine, phenylalanine	14.92	7.80	6.91	14.70	14.64	14.83
rms			0.08	0.08	0.14	0.15	0.18

[a] The representations ML^{aq} and ML^{cor} were applied as for the ligands in Table 4. [b] Calculated from the regressions for binary complexes (Set 3), as log K_1K_2.

3. Results and Discussion

3.1. Estimation of log K_1 and log K_2. In development of the regression models for estimation of log K_1 and log K_2 we used four sets of copper(II) chelates. The first set consisted of chelates with 1,2-diaminoethane and its *N*-alkyl, *N*,*N*′-dialkyl, and *C*-methylated derivatives (Set 1, $N = 14$). Along with the complexes of seven *N*-alkylated and *N*,*N*-dialkylated glycines, in the second set we also calculated the stability constants of glycine and three aliphatic amino acids Ala, Val, and Leu (Set 2, $N = 11$). The next set consisted of copper(II) chelates with naturally occurring amino acids (Set 3, $N = 9$), and the last one contained mixed complexes of aliphatic, aromatic, and hydroxyamino acids (Set 4, $N = 16$).

The univariate linear model for calculating log K_2 implies linear correlation between log K_2 and log K_1 for the binary complexes (ML_2). (If (1) log $K_2 = a(\log K_1) + b$ and (2) log

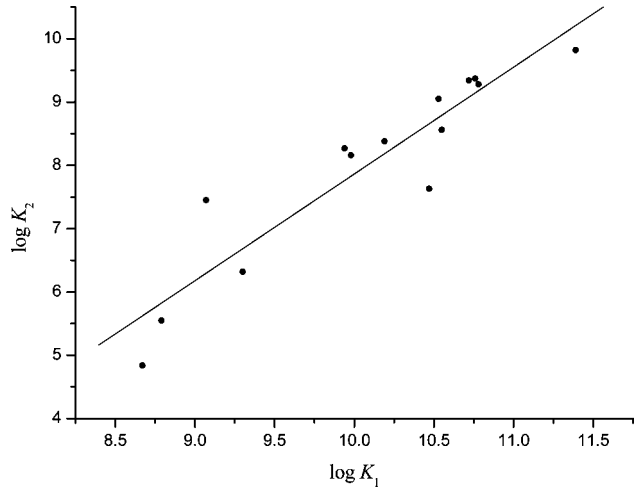

Figure 2. Linear regression of experimental log K_2 on log K_1 values for copper(II) chelates with diamines (Set 1). Slope = 1.69(19), intercept = −9.0(20), r = 0.929.

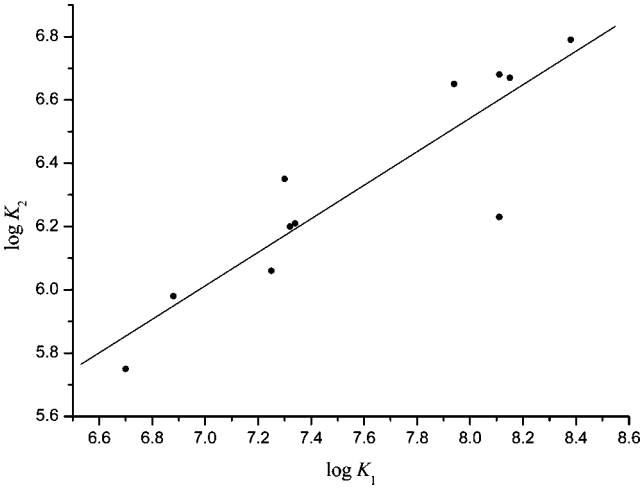

Figure 3. Linear regression of experimental log K_2 on log K_1 values for copper(II) chelates with *N*-alkylated glycines (Set 2). Slope = 0.530(87), intercept = 2.31(66), r = 0.897.

$K_1 = c[^3\chi^v(\mathrm{ML})] + d$, then (3) log $K_2 = e[^3\chi^v(\mathrm{ML})] + f$.) To check this assumption, we correlated log K_2 to log K_1 for the sets of copper(II) chelates with diamines (Figure 2), *N*-alkylated glycines (Figure 3), and binary complexes with naturally occurring amino acids (Figure 4). A very high correlation coefficient was obtained for chelates with diamines and *N*-alkylated glycines ($r \approx 0.9$) but very poor for complexes with amino acids (r = 0.294).

The results presented in Figures 2 – 4 speak in favor of the hypothesis that the univariate linear model ($a_2 = 0$, eq 5) works well with complexes with diamines and *N*-alkylated glycines but not with binary complexes with amino acids. Furthermore, as log K_1 and log K_2 cannot be correlated for ternary complexes, the univariate model is of no use for estimating log K_2 for the MLA complexes.

Our hypothesis proved right. Namely, the univariate model yielded regressions for diamine (Set 1) and *N*-alkylglycine complexes (Set 2) and the bivariate model for amino acidates (Table 1). Application of the bivariate model for the first two sets of complexes and the univariate model for the second two sets yielded bad regressions. It is also noteworthy that only the ML^{aq} representation is applicable for the univariate model, and ML^{cor} (along with ML^{aq} representation) is suitable for the

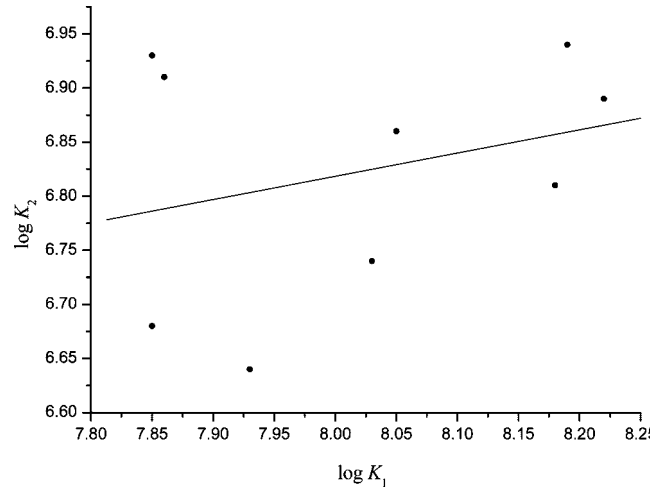

Figure 4. Linear regression of experimental log K_2 on log K_1 values for binary copper(II) chelates with naturally occurring amino acids (Set 3). Slope = 0.21(26), intercept = 5.1(21), r = 0.294.

bivariate model. This is another indication that interactions with the side chains are essential for the bivariate model.

Results in Table 1 show that log K_1 and log K_2 were equally well reproduced in all the sets. Moreover, for binary amino acidates (Set 3), log β_2 (SE_{cv} = 0.14) was reproduced with virtually the same error as for the other two constants (SE_{cv} = 0.11).

3.2. Estimation of log β_2. Tables 2–5 show the final results of our calculations. log β_2 was calculated directly from eq 6 and indirectly by adding up log K_1 and log K_2 values. As expected, both methods gave the same estimates for β_2's for the first two sets (Tables 2 and 3), while for the other two sets (Tables 4 and 5) the estimates slightly differed. However, in one case (Table 4) the direct approach gave lower maximal error (0.24 vs 0.28), and another (Table 5) indirect model proved slightly better in this respect (max. error = 0.29 vs 0.30). Averaging the estimates had a negligible influence on the results, indicating that the two approaches are essentially the same. (Note that eq 3 + eq 5 = eq 6, for binary complexes!)

The function for explicit calculation of log K_2, eq 5, makes it possible to calculate log β_2 values of ternary complexes from the calibration function developed on binary chelates. One such estimate (Table 5) yielded a higher rms (0.18) and a slightly higher maximal error (0.32, for Ser,Thr) than the other two procedures.

Most log β_2 values for binary and ternary chelates with amino acids (Tables 4 and 5) were reproduced with an error of less than 0.2, i.e., within the limits of experimental error. In the set of *N*-alkylated glycinates (Table 3) there are only two extreme estimate errors: 1.19 and 1.22 for *N*,*N*-diethyl- and *N*-isopropylglycinate, respectively. After discarding these two values, the rms error dropped from 0.59 to 0.32. In the set of diamines, there are 4 of 14 errors greater than 1.0 (Table 2). This makes estimation of log β_2 values in this set problematic.

4. Conclusion

Our results suggest that the univariate model is applicable for chelates if interactions of side groups between each other and with the central atom are negligible. If this is not the case, it should be replaced by the bivariate model for estimation of both log K_2 and log β_2 constants.

The bivariate model better reproduced log β_2 values. It is noteworthy that in comparison with the best results published

so far,[19] our SE_{cv} (log β_2) dropped from 0.16 to 0.14 for the binary (Set 3) and from 0.18 to 0.15 for the ternary complexes of amino acids (Set 4). Direct calculation makes it also possible to estimate log K_2 values when log β_2 values are not known as well as to estimate log β_2 for the ternary complexes from the log K_1 and log K_2 models for the binary chelates.

Acknowledgment. This work was supported by the Croatian Ministry of Science, Technology, Education and Sport (project 022-1770495-2901).

References and Notes

(1) Konstantinova, E. V.; Diudea, M. V. *Croat. Chem. Acta* **2000**, *73*, 383.
(2) Lučić, B.; Nikolić, S.; Trinajstić, N. *Croat. Chem. Acta* **1995**, *68*, 417.
(3) Balaban, A. T.; Mills, D.; Ivanciuc, O.; Basak, S. C. *Croat. Chem. Acta* **2000**, *73*, 923.
(4) Lučić, B.; Miličević, A.; Nikolić, S.; Trinajstić, N. *Croat. Chem. Acta* **2002**, *75*, 847.
(5) Bošnjak, N.; Adler, N.; Perić, M.; Trinajstić, N. On the structural origin of chromatographic retention data: alkanes and cycloalkanes. In *Modelling of Structure and Properties of Molecules*; Maksić, Z. B., Ed.; Wiley: New York, 1987; p 103.
(6) Kier, L. B.; Hall, L. H.; *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
(7) Hall, L. H.; Kier, L. B. *Eur. J. Med. Chem.* **1981**, *16*, 399.
(8) Medić-Šarić, M.; Malešé., Ž; Šarić, S.; Brantner, A. *Croat. Chem. Acta* **1996**, *69*, 1603.
(9) Bajaj, S.; Sambi, S. S.; Madan, A. K. *Croat. Chem. Acta* **2005**, *78*, 165.
(10) Stankevich, I. V.; Skvortsova, M. I.; Zefirov, N. S. *J. Mol. Struct.(Theochem)* **1995**, *342*, 173.
(11) Galvez, J. *J. Mol. Struct. (Theochem)* **1998**, *429*, 255.
(12) Galvez, J. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1231.
(13) Estrada, E. *J. Phys. Chem. A* **2002**, *106*, 9085.
(14) Estrada, E. *J. Phys. Chem. A* **2003**, *107*, 7482.
(15) Estrada, E. *J. Phys. Chem. A* **2004**, *108*, 5468.
(16) Raos, N. *Croat. Chem. Acta* **2003**, *76*, 81.
(17) Grgas, B.; Nikolić, S.; Paulić, N.; Raos, N. *Croat. Chem. Acta* **1999**, *72*, 885.
(18) Nikolić, S.; Raos, N. *Croat. Chem. Acta* **2001**, *74*, 621.
(19) Miličević, A.; Raos, N. *Polyhedron* **2006**, *25*, 2800.
(20) Miličević, A.; Raos, N. *Polyhedron* **2008**, *27*, 887.
(21) Miličević, A.; Raos, N. *Croat. Chem. Acta* **2007**, *80*, 557.
(22) 22. http://www.disat.unimib.it/chm/.
(23) http://cactus.nci.nih.gov/services/translate/.
(24) Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1976**, *65*, 1806.
(25) Kier, L. B.; Hall, L. H.; *Molecular Connectivity in Structure-Activity Analysis*, Wiley, New York, 1986.
(26) Randić, M. *MATCH Commun. Math. Comput. Chem.* **2008**, *59*, 5.
(27) Lučić, B.; Trinajstić, N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121.
(28) Basolo, F.; Murmann, R. K. *J. Am. Chem. Soc.* **1952**, *74*, 5243.
(29) Basolo, F.; Murmann, R. K. *J. Am. Chem. Soc.* **1954**, *76*, 211.
(30) Basolo, F.; Chem, Y. T.; Murmann, R. K. *J. Am. Chem. Soc.* **1954**, *76*, 956.
(31) Basolo, F.; Chen, Y. T.; Murmann, R. K. *J. Am. Chem. Soc.* **1954**, *76*, 953.
(32) Nancolas, I, T. P. G. H. *Inorg. Chem.* **1972**, *11*, 2414.
(33) Bretton, P. *J. Chim. Phys.* **1957**, *54*, 827.
(34) Gergely, A.; Sovago, I.; Nagypal, I.; Kiraly, R. *Inorg. Chim. Acta* **1972**, *6*, 435.
(35) Brookes, G.; Pettit, L. D. *J. Chem. Soc., Dalton Trans.* **1977**, 1918.
(36) Sigel, H.; Tribolet, R.; Scheller, K. H. *Inorg. Chim. Acta* **1985**, *100*, 151.
(37) Yamauchi, O.; Odani, A. *Inorg. Chim. Acta* **1985**, *100*, 165.

JP802018M