FEATURE ARTICLE =



Croat. Chem. Acta 2016, 89(1), 1-6 Published online: June 14, 2016 DOI: 10.5562/cca2723



How Reliable are Models Based on Topological **Index ³χ^ν for the Prediction of Stability Constants?**

Nenad Raos,* Ante Miličević

Institute for Medical Research and Occupational Health, Ksaverska c. 2, P.O.B. 291, HR-10001 Zagreb, Croatia

* Corresponding author's e-mail address: raos@imi.hr

RECEIVED: August 26, 2015 * ACCEPTED: April 20, 2016

- Based on lecture delivered by N. Raos at $27^{ au h}$ math/chem/comp, Dubrovnik, Croatia, 2015. –

Abstract: The theoretical models based on valence connectivity index of the 3^{rd} order, ${}^{3}\chi^{v}$, have been discussed in terms of their ability to predict stability of coordination compounds. The key factors for the success are: (1) the choice of reliable experimental data for the calibration of the model, (2) writing an appropriate constitutional formula (i.e. graph) of the complex, and (3) development of proper form of regression function. If these requirements were met, it is possible to obtain theoretical results comensurable with the experimental ones, *i.e.* of the sufficient quality to evaluate experimental methods or to propose the best values for stability constants.

Keywords: coordination compounds, graph-theoretical indices, valence connectivity index, refinement of stability constants.

1. INTRODUCTION

HE FIRST systematic application of topological indices for the prediction of stability constants of coordination compounds occured in 1999, when appeared the paper of our research group on copper(II) chelates with N-alkylated glycines attempting to correlate measured stability constants with four topological indices - Wiener number (W) and three consecutive valence-connectivity indices $(1\chi^{\nu}, 2\chi^{\nu}, 3\chi^{\nu})$.^[1] The next paper in the series,^[2] dealing with the mixed amino acid complexes, confirmed that the best topological index for this purpose is valence-connectivity index of the 3rd order, ${}^{3}\chi^{\nu}$, and thus we developed our further models exclusively on it.

There are many factors that determine stability of coordination compounds. Analysis of these factors led first to general rules (Irving-Williams order, HSAB model, rules for chelate, trans- and ring-size effect, etc.), and latter to many theoretical methods of a various levels of sophistication (molecular-mechanics, DFT and various kinds of QSPR models).^[3] However, these methods use many molecular descriptors (e.g. QSPR)^[4,5] or deal with specific interactions (e.g. MM).^[6,7] From this point the use of a single molecular descriptor, valence-connectivity index of the 3rd order to represent all the variety of interactions determining the

stability of the complex looks a bit naive. Moreover, the true meaning of that index, as well as other graph-theoretical indices, is obscure, despite many attempts to interpret them as molecular surfaces, volumes and the like.^[8–12] The vague interpretation of graph-theoretical indices led to a sceptical attitude of many chemists towards them and especially to their application in QSPR and QSAR.^[13,14] In spite of that, we expoited the advantage of models with one descriptor and applied ${}^{3}\chi^{\nu}$ index, as a sole descriptor, to predict the stability constants of a variety of coordination compounds (Table 1), ranging from copper(II), nickel(II), and other heavy metal (including lanthanides) chelates with diamines, triamines, amino acids and their N-alkylated and fructose derivatives,^[15] to the complexes of monocarboylic acids,^[16] and smaller peptides.^[17,18] The models proved reasonably good, with a typical standard error (S.E.) about 0.3 log units.

The aim of this paper is not, however, to give a comprehensive review of all the applications and variants of our method - we will rather focus our attention on two vital points. The first is evaluation of experimental data, the second is - as the tittle suggests - to explore the suitability of our models to predict stability constants with appropriate accuracy.



| | Amines | Mono-carboxylic acids | Amino acids | N-alkylated amino acids | Other amino acid derivatives | Oligopeptides $(n = 2-5)$ |
|--------------------|----------|--------------------------|---------------------------------------------------------------|-----------------------------------------------|---------------------------------|---------------------------|
| Cu ²⁺ | 28,39,29 | 16 | 2 ^(a) ,40 ^(a) ,28,39, 20,29,37,46,38 | 1 ^(a) ,40 ^(a) ,28,39,20 | 32,30,24 ^(b) ,33 | 41,42,17,34,35,18 |
| Ni ²⁺ | | 16 | 44,46 | | 32,30,33 | 41 |
| Zn ²⁺ | | 16 | 48 | | 33,48 ^(b) | |
| Cd ²⁺ | | 31,16 | 43,44,31 | | | |
| Other heavy metals | | 16 | 44 | | | |
| Lanthanides | | 45,47 | 36 | | | |

Table 1. Complexes included in the models based on valence connectivity index ${}^{3}\chi^{\nu}$. Numbers correspond to references

^(a) Other topological indices were also used.

^(b) Aromatic ligands unrelated to amino acids.

2. EVALUATION OF EXPERIMENTAL METHODS

2.1. Case 1: Evaluation of Two Electroanalytical Methods

Here we present our research on the comparison and evaluation of the stability constants obtained by two electroanalytical methods, namely glass electrode potentiometry (GEP) and square wave voltammetry (SWV). The methods were used to measure stability constants of copper(II) *mono-* and *bis*-complexes with alanine and its five *N*-alkylated derivatives.^[19] All the constants were measured at the same temperature (T = 298 K), but in a slightly different background electrolyte: *I*(GEP) = 0.1 mol L⁻¹ solution of KNO₃, *I*(SWV) = 0.15 mol L⁻¹ solution of NaClO₄. However, they differed considerably.

The constants for *mono*-complexes, log K_1 , measured by the two methods differ 0.01–0.64 (mean = 0.26). The constants obtained by GEP were lower than those obtained by SWV method. They are also correlated to SWV values, but worse than values obtained by the best theoretical model (Figure 1).

Calculating ${}^{3}\chi^{\nu}$ index from two constitutional formulas (*i.e.* graphs) of the complex (see 3.1.) we obtained two linear regression models for each, GEP and SWV methods. Better agreement between theory and experiment was achieved for GEP than for SWV method; S.E. = 0.17, S.E._{cv} = 0.33 for GEP and S.E. = 0.32, S.E._{cv} = 0.50 for SWV (the reported values are averages of both models). However, similar analysis on the stability constants of *bis*-complexes (log β_2) showed just the opposite; the SWV constants (S.E. = 0.43, S.E._{cv} = 0.73) were better reproduced than GEP constants (S.E. = 0.53, S.E._{cv} = 0.97). (The reported values are averages of four models.)^[20]



Figure 1. Linear regression of experimental stability constants as obtained by GEP method against log K_1 values obtained by SWV (—, •), S.E. = 0.18, and by the best theoretical model, # 2, Ref. [20] (- - -, •), S.E. = 0.16.

Those findings were confirmed by an additional test. Instead of regression lines developed on copper(II) complexes with alanines, for prediction of their constants we used regression lines developed on a similar system, copper(II) chelates with glycine and its *N*-alkylated derivatives^[21] along with three aliphatic amino acids (Ala, Val, Leu). Despite these contants were measured by GEP, the same resulted. Namely, comparison of theoretical with experimental GEP and SWV values yielded *rms* 0.21 and 0.47, respecitvely, for log K_1 (the average of two models). The same comparison of log β_2 (the average of four models) gave *rms* 0.79 and 0.64 for GEP and SWV, respectively.

These findings were further elaborated by the analysis of both experimental methods.^[20] Data for log K_1 (SWV) were measured around the detection limits, and log K_1 (GEP) constants were determined in the pH range where the response of glass electrode is due only to the formation of ML species.





Figure 2. Regression models for aromatic ligands with a common thioflavin and clioquinol core (Ref. [24]). After refuting three tia-complexes, S.E. drops from 0.85 to 0.30, and S.E._{cv} from 0.97 to 0.39.

2.2. Case 2: Comparison with DFT Method

The second example is related to the study of copper(II) binding to aromatic ligands with a common core of the thioflavin T (ThT) and clioquinol (CQ) molecule, which were investigated as potential drugs against Alzheimer's disease (Figure 2).^[22,23] The authors applied various theoretical methods on those systems, calculating stability constants of *bis*-complexes (log β_2) by DFT, along with HOMA and aromatic indices (I_{NG}),^[22] but unfortunatelly they measured stability constants of only two of altogether ten complexes.^[23] Experimental (UV-Vis) and theoretical constants (DFT) agree only qualitatively; [Cu(No-O1)2] is more stable than $[Cu(N_S-O1)_2]$. But the experimental difference between $\log \beta_2$'s equals to 2.8, in contrast to the theoretical value of 4.1. Moreover, DFT calculations gave nearly ten orders of magnitude lower constants (experimental: 23.1 and 20.3, theoretical: 15.1 and 11.0 for $[Cu(N_0-O1)_2]$ and [Cu(N_S-O1)₂], respectively).^[23]

Linear correlation of DFT constants with ${}^{3}\chi^{v}$ index yielded S.E.= 0.85 and S.E._{cv} = 0.97 (r = 0.988) for all ten complexes.^[24] However, thia-complexes could be regarded as a separate group and if they were discharged (N = 7), S.E. drops to 0.30 with the absolute values of residuals in the range 0.0 to 0.6 (Figure 2).

Advantage of our method is its simplicity; disadvantage is its inability to predict stability constants without a proper set of experimental data. However, DFT method should be capable to predict experimental data without any experimental constants. As in this case it failed to do that, it cannot be judged as advantageous to our method.

3. RELIABILITY

As was said before, methods based on the valence connectivity index of 3rd order reproduce experimental constants with a typical standard error of 0.3 log units. But the success depends from the one hand on the quality of experimental data and from the other on the quality of regression model. It is recommended to use the values of stability constants from the same paper, or at least issued by the same research group. Despite the standardization of methods and experimental conditions (temperature, ionic stength, and background electrolyte), the best constants ("recommended", according to IUPAC criteria) are determined with S.D.≤ 0.05 log units, and the majority (denoted as "tentative") with 0.05 < S.D. < 0.2 log units.^[25] In a test case, constants determined in different laboratories differed up to 0.3 log units,^[26] that is close to log K_1 values for copper(II)/glycine system (T = 298 K, I = 0.1 mol L⁻¹, GEP), whose values were referred in the range 8.07-8.38 log units (all the values were denoted as "tentative").[27] An additional problem in choosing the appropriate set of stability constants is the usual practice that researchers focus their attention to ligands, not to metals, thus in the same paper usually stability constants of a few ligands with many metals are referred.

However, a wide range of values of the measured stability constants opens a possibility to test our method. From the histogram of 14 above mentioned "tentative" log K_1 values for copper(II) monoglycinate, the most probable value should be setled at log $K_1 = 8.21$ (Figure 3). (Note that the distribution is not Gaussian!) However, an analogeous histogram (Figure 4) for seven theoretical (estimated) values gave log $K_1 = 8.19$ – the difference is only 0.02 log units! ("The best" value for T = 298 K, I = 0.1-0.2 mol L⁻¹ was referred as log $K_1 = 8.20 \pm 0.10$.]^[27]

3.1. Problem of the Proper Constitutional Formula

The first problem in application of models based on topological indices on coordination compounds is the construction of appropriate constitution formula (or graph). In contrast to organic compounds, that are defined by their constitution, coordination compounds are actually defined by their composition; stability constant $K_1(ML)$ refers to equilibrium equation M + L \rightleftharpoons ML and to nothing else. The structure of the complex is usually unknown, and even if it was determined in the crystal state it is dubious if the complex persists in such a form in solution.

There is a simple way to find out, from the one hand, the proper structure of the complex in dissolved state, and from the other to prove the soundness and reliability of our





Figure 3. A histogram of 14 experimental ("tentative") values of log K_1 for copper(II) *mono*-complex with glycine, as measured by potentiometric titration at T = 298 K, I = 0.1 mol L⁻¹ (Ref. [27]). Mean value for the whole set is calculated as 8.19, but 8.21 is a better estimate, for it is calculated as a mean of the five values in the range 8.194 – 8.225.



Figure 4. A histogram of seven theoretical values for log K_1 for copper(II) complex with glycine, as estimated in Refs. [39] (Tables 3 and 4), [20] (Table 6) and [38] (from linear model for aliphatic and both linear and quadratic model for polar amino acids). Mean values: 8.25 (N = 7), 8.19 (8.15–8.24 range, N = 4).

method. In the first paper on systematic application of ${}^{3}\chi^{\nu}$ index to coordination compounds^[28] we checked four constitutional formulas, *i.e.* corresponding graphs. The first was the graph of free ligand (L), the second was the graph of metal-ligand complex (ML), and the third graph corresponds to metal-ligand complex with two ligated water molecules, ML^{aq}. The fourth graph was based on the presumtion that the side atoms of some ligands also bind to metal, or influence otherwise its coordination sphere (ML^{cor}). Namely, it had been assumed that the terminal

atom of the side chain was bound to metal, and from this assumption corresponding molecular graph was constructed. For copper(II) 1,2-diaminoethane complexes (N = 14) ${}^{3}\chi^{\nu}$ (ML^{cor}) proved best (S.E._{cv} = 0.38) and ${}^{3}\chi^{\nu}$ (ML^{aq}) gave acceptable results (S.E._{cv} =0.49). However, ${}^{3}\chi^{v}$ (ML) yielded S.E._{cv} = 0.54, and for ${}^{3}\chi^{\nu}$ (L) S.E._{cv} rise to 0.62 log units. The similar trend has been observed for $\log \beta_2$ constants, and also for both log K_1 and log β_2 constants of copper(II) amino acid complexes. It was even shown that some, presumablly existing, bonds between metal and ligand should be removed from the molecular graph, as in the case of copper(II) complexes with diethylenetriamines^[29] or copper(II) and nickel (II) complexes with N-phenylimidoacetic acid.^[30] For cadmium(II) mono-complexes with monocarboxylic acids (N = 9), the excellent results (r = 0.983, S.E. = 0.05, S.E._{cv} = 0.06) were obtained after supposition that only one ligand, 2-hydroxybutanoic acid is bidendate (other two 2-hydroxycarboxylic acids were taken as monodentate).^[31] In contrast to this, if all the ligands were taken as monodentate linear regression gave r = 0.778, S.E._{cv} = 0.26.

3.2. Problem of the Proper Regression Function

The most common regression functions that we used for the prediction of stability constants are linear and quadratic ones. They are also the simplest, but by using indicator variable(s) we succeded to develop even linear models for rather complex systems. Assuming the regression lines for copper(II) and nickel(II) complexes are parallel, it was possible to propose a common model for copper(II) and nickel(II) bis-complexes with amino acids, that yielded even (slightly) better S.E._{cv} value (= 0.24 log units) than the separate models for each metal.^[32] That approach was latter routinely applied in the common models for copper(II) and nickel(II) complexes of iminodiacetates and pyridyl derivatives of aspartic acid, [30] N-salycidene-aminoacidato complexes with Cu2+, Ni2+ and Zn2+, [33] and for the prediction of stability of copper(II)/peptide complexes.[17,18,34] By taking one of stability constants as a reference, it has been possible to build a common models for complexes of five metals (Co²⁺, Ni²⁺, Cu²⁺, Zn²⁺, Cd²⁺) with four monocarboxylic acids (methanoic, ethanoic, propanoic and butanoic)^[16] or for copper(II) complexes with tripepties containing glycine, histidine and glutamic acid residues.^[35]

However, we have also used models of higher complexity. The first is polynomial model with ${}^{3}\chi^{v}$, r and ${}^{3}\chi^{v}$ rvariables (r stands for atomic radius of the central atom) for lanthanide complexes.^[36] The second are models with variables calculated as differences between ${}^{3}\chi^{v}$ values of various molecules.^[37] Unfortunately, despite many models checked, the general form of the regression function has not yet been found. Majority of regression lines have negative



slope; actually the sole exception is the regression function for the cadmium(II) *mono*-complexes with monocarboxylic acids, having a positive slope.^[31] It seems that the form of regression function is determined by the nature of the ligand and consequently by its interactions with the central atom. In our recent paper^[38] we have shown that nonpolar amino acids fit best the line of negative slope, but polar amino acids fit parabola better. By using such a division, we were able to reproduce log K_1 and log β_2 values with S.E._{cv} 0.03 and 0.06 log units, respectively.

3.3. Problem of the Range of Stability Constants

As we tried our models on many systems, we were doing regressions with many sets of stability constants and finished our research with a various success. In short, S.E. of models varied from 0.03 to even 1.39 log units, but the difference between the highest and lowest constant employed in regression varied also considerably, from 0.32 to 30.62 log units. However, plot of stability constant ranges *vs.* S.E./log *K* range ratio (Figure 5) reveals that the relative standard error (S.E./ Δ log *K*) is in the range 0.02–0.05 for most cases. That means that our models are generally capable to predict stability constants with S.E. of about five percent of the range of constants employed in the regression.

Regression of the function S.E./ $\Delta \log K = a/\Delta \log K$ (Figure 5) reveals the grouping of points around the curve $a \equiv <$ S.E. > = 0.06 up to $\Delta \log K = 10$. The separate group of points $\Delta \log K > 15$ corresponds to the dissocitation constants of peptide complexes (*i.e.* constants for ML, MLH₋₁, MLH₋₂ etc. complexes calculated for separate ligands, L).^[17,18,34,35]

4. CONCLUSION

The method for the prediction of stability constants of coordination compounds from topological indices is strictly empirical. That means it cannot be developed for unrelated systems, i.e. for the compounds differing much in the structure of ligand or in the nature of central atom. But from the other hand, it is a very valuable tool in studying the stability of molecules composed of similar metals (e.g. Cu(II) and Ni(II), and similar ligands (e.g. α -amino acids and their Nalkylated derivatives). In this case it is possible, as was shown in this contribution, to obtain results of the same quality as those worked out by DFT method, to compare reliability of methods for determination of stability constants, or to find out the best estimate of their values. Therefore we hope that this simple method, in both conceptual and computational sense, will find its way to the people dealing with stability constants, especially with their measurement and rafinement.



Figure 5. Plot of stability constants range, $\Delta \log K$ vs. S.E./ $\Delta \log K$ for 40 models. Curve corresponds to the plot of regression function S.E./ $\Delta \log K = a / \Delta \log K$; a = 0.064(5), $R^2 = 0.579$.

REFERENCES

- B. Grgas, S. Nikolić, N. Paulić, N. Raos, Croat. Chem. Acta 1999, 72, 885.
- [2] S. Nikolić, N. Raos, Croat. Chem. Acta 2001, 74, 621.
- N. Raos, A. Miličević, Prediction of stability constants in Handbook of Inorganic Chemistry Research (Ed.: D. A. Morrison), Nova Science Publishers, New York, 2010, pp. 267–294.
- [4] H.-S. Kim, Bull. Korean Chem. Soc. 2008, 29, 431.
- [5] J. Ghasemi, S. Saaidpour, J. Incl. Phenom. Macrocycl. Chem. 2008, 60, 339.
- [6] J. Sabolović, N. Raos, *Polyhedron* **1990**, *9*, 2419.
- [7] D. E. Reichert, R. D. Hancock, M. J. Welch, *Inorg. Chem.* **1996**, *35*, 7013.
- [8] L. H. Hall, L. B. Kier, Eur. J. Med. Chem. Chim. Therapeut. 1981, 16, 399.
- [9] H. Bögel, J. Dettmann, M. Randić, Croat. Chem. Acta 1997, 70, 827.
- [10] I. Gutman, T. Körtwélyesi, Z. Naturforsch. 1995, 50a, 669.
- [11] E. Estrada, J. Phys. Chem. A 2002, 106, 9085.
- [12] N. Raos, Croat. Chem. Acta 2003, 76, 81.
- [13] M. Randić, MATCH 2008, 59, 5.
- P. L. Ayers, R. J. Boyd, P. Bultinck, M. Caffarel, R. Carbó-Dorca, M. Causá, J. Cioslowski, J. Contreras-Garcia, D. L. Cooper, P. Coppens, C. Gatti, S. Grabowsky, P. Lazzeretti, P. Macchi, Á. M. Pendás, P. L. A. Popelier, K. Ruedenberg, H. Rzepa, A. Savin, A. Sax, W. H. E. Schwarz, A. Shahbazin, B. Silvi, M. Solà, V. Tsirelson, *Comput. Theor. Chem.* **2015**, *1053*, 2.
- [15] N. Raos, A. Miličević, Arh. Hig. Rada Toksikol. 2009, 60, 123.



- [16] A. Miličević, N. Raos, Acta Chim. Slov. 2013, 60, 120.
- [17] A. Miličević, N. Raos, J. Coord. Chem. 2014, 67, 623.
 [18] A. Miličević, N. Raos, Arh. Hig. Rada Toksikol. 2015,
- 66, 165.
 [19] G. Branica, N. Paulić, B. Grgas, D. Omanović, *Chem.* Spec. Bioavabil. **1999**, *11*, 125.
- [20] N. Raos, G. Branica, A. Miličević, Croat. Chem. Acta 2008, 81, 511.
- [21] F. Basolo, Y. T. Chen, J. Am. Chem. Soc. 1954, 76, 953.
- [22] C. Rodríguez-Rodríguez, N. S. De Groot, A. Rimola, Á. Álvarez-Larena, V. Lloveras, J. Vidal-Gancedo, S. Ventura, J. Vendrell, M. Sodupe, P. González-Duarte, *J. Am. Chem. Soc.* 2009, 131, 1436.
- [23] A. Rimola, J. Ali-Torres, C. Rodríguez-Rodríguez, J. Poater, E. Matito, M. Solá, M. Sodupe, *J. Phys. Chem.* 2011, *115*, 12659.
- [24] A. Miličević, N. Raos, Arh. Hig. Rada Toksikol. 2013, 64, 539.
- [25] I. Sóvágó, T. Kiss, A. Gergely, Pure Appl. Chem. 1993, 65(5), 1029.
- [26] K. Popov, I. Pletnev, H. Wanner, A. Vendilo, *The first International Proficiency Testing Conference, Sinia, Romania, 11th-13th October 2007*, pp. 324–334.
- [27] T. Kiss, I. Sóvágó, A. Gergely, Pure Appl. Chem. 1991, 63(4), 597.
- [28] A. Miličević, N. Raos, Polyhedron 2006, 25, 2800.
- [29] A. Miličević, N. Raos, Acta Chim. Slov. 2009, 56, 373.

- [30] A. Miličević, N. Raos, Int. J. Chem. Model. 2008, 1(3/4), 395.
- [31] A. Miličević, N. Raos, Acta Chim. Slov. 2012, 59, 194.
- [32] A. Miličević, N. Raos, Croat. Chem. Acta 2007, 80, 557.
- [33] A. Miličević, N. Raos, Cent. Eur. J. Chem. 2014, 12(1), 74.
- [34] A. Miličević, N. Raos, Int. J. Chem. Model. 2014, 6(2–3), 301.
- [35] A. Miličević, N. Raos, Bull. Chem. Soc. Jpn. 2015, 88, 490.
- [36] A. Miličević, N. Raos, Acta Chim. Slov. 2014, 61, 904.
- [37] A. Miličević, N. Raos, Chin. J. Chem. 2011, 29, 2421.
- [38] A. Miličević, N. Raos, *Topics in Chemical Graph Theory* (Ed.: I. Gutman), Mathematical Chemistry Monographs 16a, University of Kragujevac, Kragujevac (Serbia), **2014**, pp. 193–204.
- [39] A. Miličević, N. Raos, J. Phys. Chem. A 2008, 112, 7745.
- [40] N. Raos, Croat. Chem. Acta 2002, 75, 117.
- [41] A. Miličević, N. Raos, Polyhedron 2008, 27, 887.
- [42] A. Miličević, N. Raos, Croat. Chem. Acta 2009, 82, 633.
- [43] A. Miličević, N. Raos, Acta Chim. Slov. 2010, 57, 866.
- [44] A. Miličević, G. Branica, N. Raos, *Molecules* 2011, 16, 1103.
- [45] A. Miličević, N. Raos, Chem. Phys. Lett. 2012, 528, 63.
- [46] A. Miličević, N. Raos, J. Mol. Lig. **2012**, 165, 139.
- [47] A. Miličević, N. Raos, J. Mol. Liq. 2013, 177, 60.
- [48] A. Miličević and N. Raos, Acta Chim. Slov., 2015, 62, 784.