Prediction of Stability Constants of Lanthanide Complexes with Amino Acids by Model Based on Connectivity Index $^3\chi^v$

Ante Miličević and Nenad Raos

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

* Corresponding author: E-mail: antem@imi.hr
Tel.:+385 14682524; Fax: +385 14673303

Abstract

In order to model the stability of lanthanide complexes with amino acids, we used a set of 20 *mono*-complexes of La^{3+}, Ce^{3+}, Pr^{3+}, Nd^{3+} and Sm^{3+} with glycine, alanine, valine and leucine. The quadratic model based on the $^3\chi^v$ index gave $r = 0.978$, S.E. = 0.08. The predictive power of the model was tested by splitting the initial set of complexes on training ($N = 15$) and test set ($N = 5$). This enabled the logarithm of the stability constant, log K_1, of the leucine complex of neodymium(III) and all four complexes of samarium(III) to be predicted with S.E. = 0.11.

Keywords: Lanthanide coordination; Coordination compounds; Regression models; Topological indices; Ionic radii

1. Introduction

Models for the prediction of stability constants based on graph-theoretical indices, especially on the valence molecular connectivity index of the 3rd order ($^3\chi^v$),[1] are simple both conceptually and computationally but lack generality in contrast to more sophisticated QM,[2,3] and other QSPR[4] models. Usually, models are developed for a single metal (*e.g.* copper) and a narrow class of ligands (amino acids, aliphatic diamines, carboxylic acids etc.),[5–9] as well as for two (Cu^{2+}, Ni^{2+}),[10] four (Ni^{2+}, Co^{2+}, Fe^{2+}, Mn^{2+})[11] and five (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+})[12] bivalent metal complexes by introducing an indicator variable. Among models of a wider scope are also common models for *mono*-complexes of four lanthanides (La^{3+}, Ce^{3+}, Pr^{3+}, Nd^{3+}) with three monocarboxylic acids (ethanoic, propanoic and butanoic).[13] In modelling their stability constants, we used linear and quadratic functions, which were dependent not only on $^3\chi^v$, but also on the radius of the lanthanide cation. The linear model proved better for lower fractions of dioxane ($w = 0–20\%$) and the quadratic one for higher fractions ($w = 40–60\%$).

This paper will try to apply the already mentioned quadratic model on a more complex system, namely on lanthanide complexes with amino acids. Despite the significant number of papers dealing with chemistry, geochemistry,[14–16] medical chemistry[17] and the structure of lanthanide complexes, especially coordination of lanthanide cations,[18] there have not been many papers presenting the stability constants of their complexes with amino acids.[19–20] In lanthanide chemistry amino acids are monodendate ligands,[20] bound as zwiterions in the same way as monocarboxylic acids, but much stronger (Δlog $K_1 \approx 4$). This was tentatively attributed to hydrogen bonding between the charged amino group and the ligated water.[18,21] In crystal state, however, structures are much more complex, with 1–9 coordinated water and three kinds of C-O··Ln bridges.[20] An additional problem is that many stability constants were poorly determined, differing even to 1–2 log K units at principally the same experimental conditions.[20] As the best calibrations so far have been achieved by using a set of constants extracted from one particular scientific paper,[11] we chose to do the same with a representative number ($N = 20$) of determined log K_1 values and reasonably low declared standard errors of their determination (0.005–0.04).[19]

2. Methods

2. 1. Calculation of Topological Indices

We calculated topological indices using the E-DRAGON program system developed by R. Todeschini et al.,

which is capable of yielding 119 topological indices in a single run, along with many other molecular descriptors.[22,23] The connectivity matrices were constructed using the *Online SMILES Translator and Structure File Generator*.[24]

All of the models were developed using the $^3\chi^v$ index (the valence molecular connectivity index of the 3rd order), which was defined as:[25–28]

$$^3\chi^v = \sum_{\text{path}} [\delta(i)\ \delta(j)\ \delta(k)\ \delta(l)]^{-0.5} \qquad (1)$$

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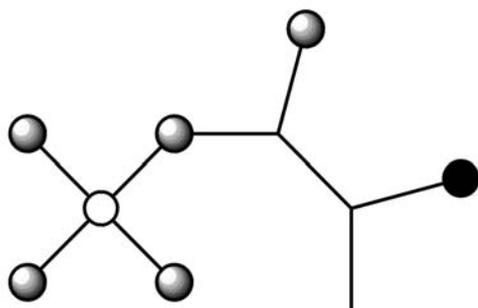

Figure 1. The graph representation for lanthanum(III) *mono*-complexes with alanine. Heteroatoms are marked with ○(La), ●(N), ●(O).

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

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, delta values for primary, secondary, tertiary, and quaternary carbon atoms are 1, 2, 3, and 4, respectively, while for the oxygen in the OH group, it equals 5 and for the NH_2 group 3. It should be pointed out that $^3\chi^v$ is only one of many members of the family of valence connectivity indices $^n\chi^v$, which differ amongst each other by path length, *i.e.* the number of δ´s in the summation term Eq. 1.

The $^3\chi^v$ index for all complexes was calculated from the graph representations of *aqua*-complexes, under the assumption that amino acids are monodendate ligands[20] and that all lanthanides bind an equal number of water molecules. Conventionally, we treated them as tetracoordinated (Figure 1).[13]

2. 2. Regression Calculations

Regression calculations, including the leave-one-out procedure (LOO) of cross validation, were done using the CROMRsel program.[29] The standard error of the cross validation estimate is defined as:

$$\text{S.E.}_{\text{cv}} = \sqrt{\sum_i \frac{\Delta X_i^2}{N}} \qquad (3)$$

Table 1. Logarithms of the stability constants of La^{3+}, Ce^{3+}, Pr^{3+}, Nd^{3+} and Sm^{3+} complexes with glycine, alanine, valine and lucine, their $^3\chi^v$ index, radii of metal ions (r), and the regression coefficients of parabolas (log $K_1 = A + B_1[^3\chi^v] + B_2[^3\chi^v]^2$) obtained for the complexes of each metal

Metal	Ligand	log K_1	r/Å*	$^3\chi^v$	A	B_1	B_2
La	Glycine	5.32	1.045	2.4106	−58.513	46.282	−8.218
	Alanine	5.82		2.5044			
	Valine	5.94		3.1230			
	Leucine	5.61		3.1601			
Ce	Glycine	5.38	1.010	2.4328	−70.449	54.552	−9.611
	Alanine	6.03		2.5255			
	Valine	6.05		3.1441			
	Leucine	5.84		3.1818			
Pr	Glycine	5.55	0.997	2.4535	−93.670	70.999	−12.455
	Alanine	6.36		2.5451			
	Valine	6.28		3.1637			
	Leucine	5.99		3.2014			
Nd	Glycine	5.68	0.983	2.4726	−108.502	81.177	−14.158
	Alanine	6.52		2.5633			
	Valine	6.52		3.1819			
	Leucine	6.03		3.2196			
Sm	Glycine	5.84	0.958	2.5146	−114.926	84.616	−14.555
	Alanine	6.68		2.6032			
	Valine	6.68		3.2218			
	Leucine	6.18		3.2595			

* Values for ionic radii have been taken from ref. 30.

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

3. Results and Discussion

We modelled the stability of the set of 20 amino acid *mono*-complexes with lanthanides (Table 1); the complexes of lanthanum(III), cerium(III), praseodymium(III), neodymium(III) and samarium (III) with four amino acids (glycine, alanine, valine and leucine).[19]

The dependence of log K_1 on $^3\chi^v$ shows parabolic dependence for complexes of every metal studied (Figure 2, Table 1). Thus, we used the previously applied function:[13]

$$\log K_1 = a([^3\chi^v] - [^3\chi^v]_0)^2 + b \tag{4}$$

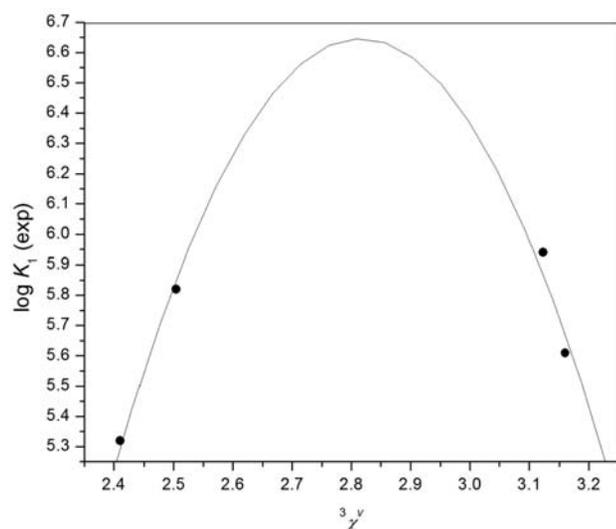to estimate the stability of all 20 complexes.

Figure 2. Dependences of experimental log K_1 stability constant on $^3\chi^v$ index for La^{3+} *mono*-complexes with glycine, alanine, valine and lucine (Table 1).

Because the maxima of dependence functions of log K_1 on $^3\chi^v$ fell within the range 2.816–2.907, to obtain the set of parabolas one above the other (Figure 3), we used the $^3\chi^v$ index insensitive to metal, *i.e.* we assumed the same δ value (Eq. 2) for all the metals (arbitrary, we took δ

value of La^{3+}). In this way, all of the parabolas had their maxima in the range $[^3\chi^v]_0 = 2.807–2.817$, which could be considered constant. Therefore, Eq. (4) could be written (assuming $a = a'r$, $b = a''r + b'$, and $[^3\chi^v]_0 = $ const.) as:

$$\log K_1 = a_1[^3\chi^v]^2r + a_2[^3\chi^v]r + a_3r + b \tag{5}$$

The regression gave $r = 0.972$, S.E.$_{cv} = 0.12$ (Model 1, Table 2), which is the same as for the set of lanthanide complexes with carboxylic acids.[13]

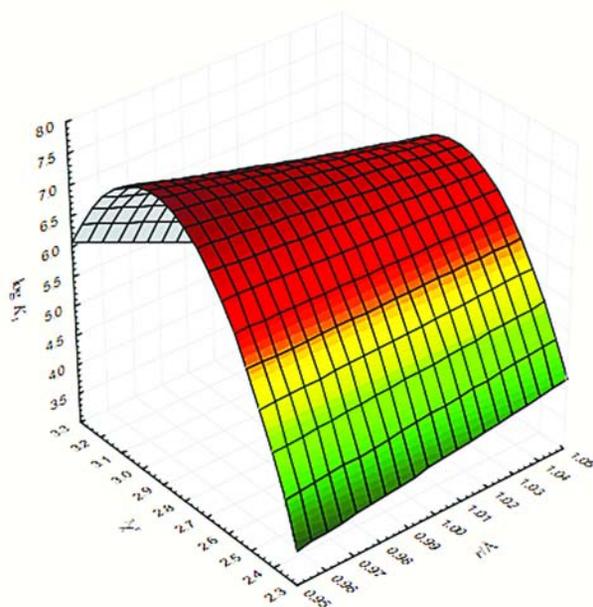**Figure 3.** Dependence of experimental log K_1 stability constant of La^{3+}, Ce^{3+}, Pr^{3+}, Nd^{3+} and Sm^{3+} complexes with glycine, alanine, valine and lucine on their $^3\chi^v$ index (δ(M) = δ(La)) for all the metals, M) and on metal radii (r).

In addition, we modified Eq. (5) by using $1/r$ instead of r:

$$\log K_1 = a_1[^3\chi^v]^2 1/r + a_2[^3\chi^v]1/r + a_3 1/r + b \tag{6}$$

The rationale for this modification is that $1/r$ is proportional to the Coulomb potential, which obviously influences the stability of coordination compounds.[31] Besides better physical meaning, the new model also gave better results ($r = 0.978$, S.E.$_{cv} = 0.10$; Model 2, Table 2). We

Table 2. Regression models for the estimation of the log K_1 of La^{3+}, Ce^{3+}, Pr^{3+}, Nd^{3+} and Sm^{3+} complexes with glycine, alanine, valine and leucine

| Model | Eq | N | Regression coefficients | | | | r | S.E. | S.E.$_{cv}$ |
			a_1(S.E.)	a_2(S.E.)	a_3(S.E.)	b(S.E.)			
1	5	20	–11.44(95)	64.3(53)	–97.2(74)	14.19(80)	0.972	0.09	0.12
2	6	20	–11.52(84)	64.7(47)	–81.4(65)	–2.22(71)	0.978	0.08	0.10
3	6	15	–10.9(10)	61.2(56)	–75.9(78)	–2.9(11)	0.975	0.08	0.11

also tried Eq. (6) on the set of carboxylic acids,[13] but the statistics remained unchanged ($r = 0.979$, S.E.$_{cv}$ = 0.12).

To test the predictive power of Eq. (6), we split the initial set to training ($N = 15$) and test set ($N = 5$; all four complexes with Sm^{3+} and leucine complex with Nd^{3+}). The regression developed on the training set yielded $r = 0.975$, S.E.$_{cv}$ = 0.11 (Model 3, Table 2, Figure 4), and the standard error for the test set was 0.11 as well.
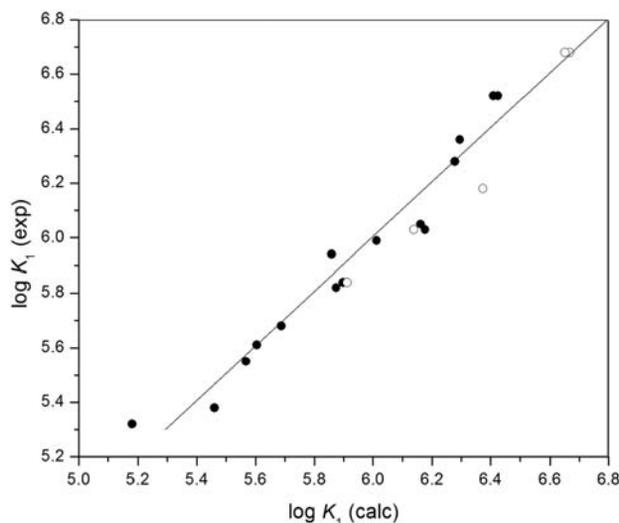

Figure 4. Experimental *vs.* calculated values of log K_1 for complexes of La^{3+}, Ce^{3+}, Pr^{3+} and Nd^{3+} (training set, $N = 15$; marked with ●) and 5 predicted log K_1 values for Nd^{3+} and Sm^{3+} complexes (test set marked with ○), obtained by Eq. 6 (Model 3, Table 2); S.E.$_{test}$ = 0.11.

4. Conclusion

The models presented in this paper are of the same quality as previously published results on lanthanide complexes with monocarboxylic acids.[13] The results obtained by the function with variable $1/r$ (Model 2, Table 2) yielded better results than the function with r (Model 1, Table 2); S.E.$_{cv}$ = 0.10 *vs.* 0.12 and the maximal cv error for a complex was 0.21 *vs.* 0.27. Our model did not depend much on the values of ionic radii. To be more precise, the standard error was not substantially changed by the introduction of other sets of ionic radii[32,33] in Eq. 6 (S.E.$_{cv}$ = 0.10–0.15). This did not occur even after replacing r with δ values for respective lanthanide (model yielded S.E.$_{cv}$ = 0.11), or by treating r as an indicator variable ($In = 1, 2, 3, 4$ and 5 for La^{3+}, Ce^{3+}, Pr^{3+}, Nd^{3+} and Sm^{3+}, respectively; S.E.$_{cv}$ = 0.14). However, in spite of the seemingly small differences, the results obtained by Model 2 speak strongly in favour of its soundness.

As the maximal cv error for Model 2 was 0.21, the general standards for stability constants (errors ≤ 0.05 for „recommended" and ≤ 0.2 for „tentative" constants)[34] were thus fulfilled. However, the predicted constants are about five times worse than the measured ones (judging from the declared standard error of the measurments); S.E. = 0.10 *vs.* 0.021.[19]

5. Acknowledgment

6. References

1. N. Raos, A. Miličević, *Arh. Hig. Rada Toksikol.* **2009**, *60*, 123–128.
2. O. Gutten, I. Beššeová, L Rulíšek, *J. Phys. Chem. A* **2011**, *115*, 11394–11402.
3. O. Gutten, L Rulíšek, *Inorg. Chem.* **2013**, *52*, 10347–10355.
4. V. P. Solov'ev, A. Yu. Tsivadze, A. A. Varnek, *Macroheterocycles,* **2012**, *5*, 404–410.
5. A. Miličević, N. Raos, *Polyhedron* **2006**, *25*, 2800–2808.
6. A. Miličević, N. Raos, *J. Phys. Chem. A* **2008**, *112*, 7745–7749.
7. N. Raos, G. Branica, A. Miličević, *Croat. Chem. Acta* **2008**, *81*, 511–517.
8. A. Miličević, N. Raos, *Acta Chim. Slov.* **2012**, *59*, 194–198.
9. A. Miličević, N. Raos, *J. Mol. Liquids* **2012**, *165*, 139–142.
10. A. Miličević, N. Raos, *Polyhedron* **2008**, *27*, 887–892.
11. A. Miličević, G. Branica, N. Raos, *Molecules* **2011**, *16*, 1103–1112.
12. A. Miličević, N. Raos, *Acta Chim. Slov.* **2013**, *60*, 120–123.
13. A. Miličević, N. Raos, *Chem Phys. Lett.* **2012**, *528*, 63–67.
14. C. H. Gammons, S. A. Wood, *Chem. Geol.* **2000**, *166*, 103–124.
15. S. A. Wood, D. J. Wesolowski, D. A. Palmer, *Chem. Geol.* **2000**, *167*, 231–253.
16. J. M. Cleveland, T. F. Rees, *Science* **1981**, *212*, 1506–1509.
17. W. P. Li, D. S. Ma, C. Higginbotham, T. Hoffman, A. R. Ketring, C.S. Cutler, S. S. Jurisson, *Nucl. Med. Biol.* **2001**, *28*, 145–154.
18. D. Parker, R. S. Dickins, H. Puschmann, C. Crossland, J. A. K. Howard, *Chem. Rev.* **2002**, *102*, 1977–2010.
19. S. N. Limaye, M. C. Saxena, *Can. J. Chem.* **1986**, *64*, 865–870.
20. C. Kremer, J. Torres, S. Domínguez, A. Mederos, *Coord. Chem. Rev.* **2005**, *249*, 567–590.
21. S. Aime, M. Botta, J. I. Bruce, D. Parker, V. Mainero, E. Terreno, *Chem. Commun.* **2001**, 115–116.
22. I. V. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. A. Palyulin, E. V. Radchenko, N. S. Zefirov, A. S. Makarenko, V. Y. Tanchuk, V. V. Prokopenko, *J. Comput. Aid. Mol. Des.* **2005**, *19*, 453–463.
23. VCCLAB, Virtual Computational Chemistry Laboratory, http://www.vcclab.org, 2005.
24. http://cactus.nci.nih.gov/services/translate/

25. L. B. Kier, L. H. Hall, *J. Pharm. Sci.* **1976**, *65*, 1806–1809.
26. L. B. Kier, L. H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.
27. L. B. Kier, L. H. Hall, Molecular Connectivity in Structure-Activity Analysis, Willey, New York, 1986.
28. Randić M, MATCH Commun. *Math. Comput. Chem.* **2008**, *59*, 5–124.
29. Lučić B, Trinajstić N, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121–132.
30. R. D. Shannon, C. D. Prewitt, *Acta Cryst. B26* **1970**, 1046–1048.
31. H. Irving, R. J. P. Williams, The stability of transition-metal complexes, Inorganic Chemistry Laboratory, Oxford, 1953.
32. R. D. Shannon, C. D. Prewitt, *Acta Cryst.* B25 **1969**, 925–946.
33. R. D. Shannon, *Acta Cryst. A32* **1976**, 751–767.
34. K. Popov, I. Pletnev, H. Wanner, A. Vendilo, The First International Proficiency Testing Conference, Sinia, Romania, 11th–13th October, 2007, pp. 324–334.

Povzetek

Za modeliranje stabilnosti lantanidnih kompleksov z amino kislinami smo uporabili 20 mono kompleksov La^{3+}, Ce^{3+}, Pr^{3+}, Nd^{3+} in Sm^{3+} z glicinom, alaninom, valinom in levcinom. Z uporabo kvadratnega modela, osnovanega na indeksu $^3\chi^v$, smo dobili $r = 0.978$ in S.E. = 0.08. Napovedno moč modela smo preverili z ločitvijo začetnega seta kompleksov na ($N = 15$) in ($N = 5$). Logaritem stabilnosti konstante, log K_1, kompleksa levcina z neodimijem (III) in za štiri komplekse s samarijem (III) smo napovedali z S.E. = 0.11.