DOI: 10.6060/mhc2012.121104s

New Approach for Accurate QSPR Modeling of Metal Complexation: Application to Stability Constants of Complexes of Lanthanide Ions Ln^{3+}, Ag^{+}, Zn^{2+}, Cd^{2+} and Hg^{2+} with Organic Ligands in Water

V. P. Solov'ev,[a@] A. Yu. Tsivadze,[a] and A. A. Varnek[b]

[a]*Institute of Physical Chemistry and Electrochemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation*
[b]*Laboratoire d'Infochimie, UMR 7177 CNRS, Université de Strasbourg, 67000 Strasbourg, France*
[@]*Corresponding author E-mail: solovev-vp@mail.ru*

In this paper, we propose a new definition of models applicability domain (AD) based on the selection of sufficient portion of individual QSPR models to be accepted for property prediction. Efficiency of this approach has been demonstrated in ensemble modeling of the stability constants logK of the 1:1 complexes of 17 lanthanide and transition metal ions (M) with various organic ligands (L) in water. The individual linear models based on substructural molecular fragment (SMF) descriptors were validated in 5-fold cross-validation procedure. Each test set compound was a subject of two previously developed ADs: fragment control and bounding box. After that, predictions for a given compound were considered reliable if the number of accepted models were larger than user defined portion of the total amount of selected individual models; otherwise the compound was discarded from the modeling. Application of this rule - "Quorum Control" – resulted in significant increase predictive performance of consensus models.

Keywords: Complexes with metal ions, organic ligands, QSPR modeling, stability constants, models applicability domain, multiple linear regression analysis, substructural molecular fragments.

Новый подход для точного QSPR моделирования комплексообразования металлов: применение к константам устойчивости комплексов лантанидных ионов Ln^{3+}, Ag^{+}, Zn^{2+}, Cd^{2+} и Hg^{2+} с органическими лигандами в воде

В. П. Соловьев,[a@] А. Ю. Цивадзе,[a] А. А. Варнек[b]

[a]*Институт физической химии и электрохимии Российской академии наук, 119991 Москва, Россия*
[b]*Laboratoire d'Infochimie, UMR 7177 CNRS, Université de Strasbourg, 67000 Страсбург, Франция*
[@]*E-mail: solovev-vp@mail.ru*

Предложено новое определение области применимости моделей (AD), основанное на выборе достаточной порции индивидуальных QSPR моделей, чтобы быть принятыми для предсказания свойства. Эффективность этого подхода продемонстрирована в использовании ансамбля моделей структура - свойство для оценки констант устойчивости $logK_{ML}$ комплексов 17 катионов лантанидов и переходных металлов (M) с разнообразными органическими лигандами (L) в воде. Прогностическая способность индивидуальных линейных моделей на основе субструктурных молекулярных фрагментов проверена процедурой внешнего пятикратного скользящего контроля. Для каждого тестируемого соединения применялись ранее разработанные методы AD: контроль новых фрагментов и границ числа вхождений фрагмента. Затем предсказание для данного соединения считалось надежным, если число принятых моделей было больше заданной пользователем части всех используемых индивидуальных моделей; иначе соединение исключалось из моделирования. Результатом применения этого правила "кворум контроля" было существенное увеличение прогностической способности консенсус моделей.

Ключевые слова: Металлокомплексы, органические лиганды, QSPR моделирование, константы устойчивости, область применимости моделей, множественный линейный регрессионный анализ, субструктурные молекулярные фрагменты.

 © ISUCT Publishing

Introduction

Binding of metal ions by organic ligands in solutions plays an important role in various workflows in industry[1-4] and biological processes.[4-7] Many efforts were done to design the ligands that selectively bind a given metal ion and allow to separate metal ions.[8-14] At present, a large amount of known experimental data concerning stability constants of metal-ligand complexes has been collected.[15-21] This opens an opportunity to develop quantitative structure - property relationships (QSPR) linking the stability constants with the structure of ligands which, in turn, can be used for computer-aided design of new metal binders.[22, 23]

To date, QSPR modeling of stability constants of the metal – ligand complexation was performed for alkali,[24-32] alkaline-earth,[32-38] rare-earth[39-42] and transition metal[23,34,36,37,40,43-45] ions. In many cases the practical application of the reported QSPR is complicated due to the lack of complete information about descriptors' calculations and details of machine-learning method implementation. In order to overcome this problem, we have developed the COMET (COmplexation of METals) software[39] which implements previously elaborated QSPR models of stability constants (log*K*) of the 1:1 (M:L) complexes of diverse organic ligands with alkaline-earth (Sr^{2+},[33,35] Ca^{2+}, Ba^{2+}, Mg^{2+}[33]), lanthanide (Ce^{3+}, Pr^{3+}, Nd^{3+}, Sm^{3+}, Eu^{3+}, Gd^{3+}, Tb^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, Tm^{3+}, Yb^{3+}, Lu^{3+})[39] and transition metal ions (Ag^{+},[40] Zn^{2+}, Cd^{2+} and Hg^{2+},[23] Mn^{2+}, Fe^{2+}, Y^{3+}, La^{3+}, Pb^{2+}, and UO_2^{2+}[43]) in water at 298 K and an ionic strength 0.1 M. All these models are based on substructural molecular fragments (SMF)[23,30,33] representing a subtype of the ISIDA descriptors.[46] SMF are subgraphs of molecular graph[30,47] whereas fragment occurrences are descriptor values. SMF descriptors are calculated solely from 2D chemical structures. It has been demonstrated that prediction performance of the models built on SMF is, at least, as good as that for the models involving molecular descriptors of different types,[48-50] E-state counts and E-state indices[40] and pharmacophore descriptors.[51] In our studies for alkaline-earth[33] and some heavy metal ions,[43] the root-mean squared error (*RMSE*) of predictions is comparable with experimental systematic errors.[15] For lanthanide ions,[39,52] Ag^{+},[40] Zn^{2+}, Cd^{2+}, and Hg^{2+},[23] *RMSE* of predictions of the log*K* values varies more widely: from 1.7 to 1.9 for transition metals and from 1.0 - 1.9 ($0.5 < \log K < 10$) to 2.5 - 3.7 ($\log K \geq 10$) for lanthanides.

There are two possible reasons of relatively poor prediction performance of the models. The first one is related to relatively poor quality of experimental data collected from various sources: log*K* values reported for the same equilibrium by different authors may differ.[15] Another reason could be related to some methodological problems: choice of descriptors, machine-learning methods and models applicability domain (AD). In this paper, we focus on the AD issue. We propose a new AD definition –"quorum control" – which significantly improves predictive performance. The efficiency of this approach has been demonstrated on previously studied datasets for metal complexation.

Here we report new QSPR ensemble models for the stability constant log*K* of the 1:1 (M:L) complexes of 13 lanthanide (Ce^{3+}, Pr^{3+}, Nd^{3+}, Sm^{3+}, Eu^{3+}, Gd^{3+}, Tb^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, Tm^{3+}, Yb^{3+}, Lu^{3+}) and 4 transition (Ag^{+}, Zn^{2+}, Cd^{2+}, Hg^{2+}) metal ions with sets of diverse organic molecules in aqueous solution at 298 K and an ionic strength 0.1 M. The models have been validated by external 5-fold cross-validation procedure. The root mean squared error (*RMSE*) of predictions is similar to systematic errors in experimental data. This is twice smaller compared to earlier reported models for which "quorum control" AD has not been applied.

Methods

Data Sets

The experimental stability constant (log*K*) values for the 1:1 (M:L) complexes of lanthanide (Ce^{3+}, Pr^{3+}, Nd^{3+}, Sm^{3+}, Eu^{3+}, Gd^{3+}, Tb^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, Tm^{3+}, Yb^{3+}, Lu^{3+}) and transition (Ag^{+}, Zn^{2+}, Cd^{2+}, Hg^{2+}) metal ions with diverse organic ligands in water were selected from the IUPAC Stability Constants Database (SC DB) (version 5.33, Academic Software)[15] at standard temperature 298 K and an ionic strength $I = 0.1$ M. Some log*K* values (around 15 %) were corrected to specified temperature and an ionic strength using the procedures included in SC DB.

2D structures of ligands, names of metal ions and corresponding log*K* values resulted from searching in SC DB were converted into Structure - Data Files (SDF) served as an input in the MLR module of the ISIDA (In Silico Design and Data Analysis) /QSPR program.[53] The data manager EdiSDF[35,46,54] was used to prepare data sets containing finally from 52 (Hg^{2+}) to 568 (Zn^{2+}) organic ligands (Figure 1). The log*K* values vary in the ranges from 0.6 – 1.8 to 17.9 – 24.7 (lanthanide ions), from 0.6 to 8.7 (Ag^{+}), from 0.1 to 21.3 (Zn^{2+}, Cd^{2+}) and from 4.9 to 28.5 (Hg^{2+}).

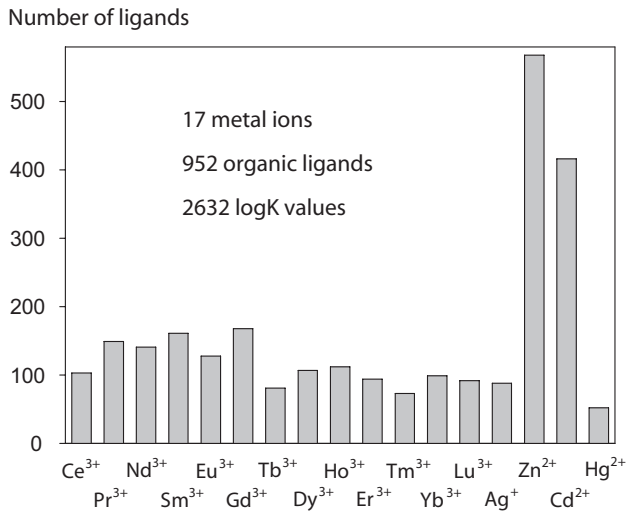
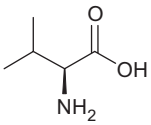
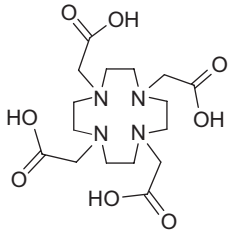
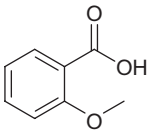
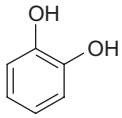
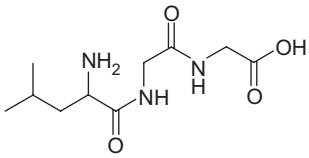
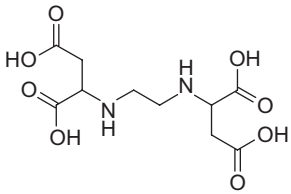
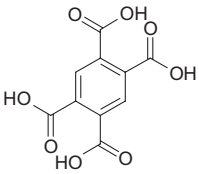
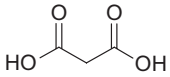


Figure 1. The number of ligands in individual datasets used in QSPR modeling of log*K*.

If several values of the stability constant log*K* were available for a particular ligand (Table 1), for selections we followed the recommendations of IUPAC;[55] in some cases the most recent data or the data consistent with respect to different experimental methods were chosen. It should be noted that discrepancies in experimental log*K* values reported for the same equilibrium by different authors may attain rather high values (till 2.3 log*K* unites, see Table 1).

Studied molecules include crown-ethers, thia-, and aza-crowns with neutral and acidic lariat groups, cryptands; derivatives of carboxylic and polycarboxylic acids; polyamines, (thio)ethers, various amino acids and aminocarboxylates; derivatives of phosphonous, (di)phosphoric, (di)phosphonic and phosphinic acids; cyclic and acyclic polydentate ligands with the terminal carboxy and phosphoryl groups separated by various cyclic or acyclic spacers;

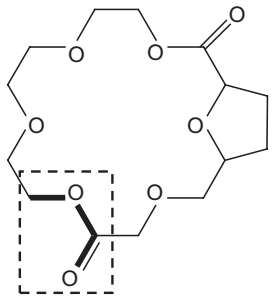
Table 1. The log*K* values for the 1:1 (M:L) complexes of several studied ligands with lanthanide ions in water at temperature 298 K and an ionic strength 0.1 M. The demonstration of a discrepancy in the experimental log*K* values.

ligand	metal ion	log*K*	ref.	ligand	metal ion	log*K*	ref.
	Ce^{3+}	3.96	[56]		Gd^{3+}	24.67	[58]
		5.02	[57]			27.00	[59]
	Pr^{3+}	1.96	[60]		Dy^{3+}	9.12	[62]
		1.14	[61]			11.34	[63]
	Nd^{3+}	3.27	[64]		Ho^{3+}	13.10	[66]
		1.75	[65]			13.60	[67]
	Eu^{3+}	5.81	[68]		Er^{3+}	4.89	[70]
		4.86	[69]			4.42	[71]

various (di)sulfonic acids; ternary amines with phosphono and carboxy groups; mono- and dipodands of ternary amines; amide, phenol, glucose, imidazole, adenosine, inosine, uridine, uracil, cytidine, thymidine, thymidine, adenine and guanosine derivatives; pyridines; purine, phenanthroline, hydrazide derivatives, *etc*.

Descriptors

SMF[23,30,54,72] as subgraphs of molecular graphs of the ligands were used as descriptors in QSPR models. Molecules were represented with implicit hydrogen atoms. Two classes of the SMF descriptors were generated: shortest topological paths with explicit representation of atoms and bonds, and terminal groups as shortest paths but defined by length and explicit identification of terminal atoms and bonds[23,30,33,43] (Figure 2).



a	b
C-O-C=O	C-[4]=O
C-O-C	C-[3]-C
C-O	C-[2]- O
O-C=O	O-[3]=O
C-O	C-[2]- O
C=O	C=[2]=O

Figure 2. Two classes of SMF: shortest topological paths with explicit representation of atoms and bonds (a), and terminal groups as shortest paths defined by length and explicit identification of terminal atoms and bonds (b). The SMF types IAB(2-4) and IAB(2-4)t are shown.

Single, double and triple bonds were considered different in acyclic and cyclic non-aromatic motifs. For every class of the sequences, the minimal ($2 \le n_{min} \le 4$) and maximal ($6 \le n_{max} \le 15$) numbers of constituent atoms are defined. The sequences include all intermediate shortest paths with *n* atoms: $n_{min} \le n \le n_{max}$. 60 types of the sequences of two classes have been generated varying the values of n_{min} and n_{max}. SMF descriptors of each particular type were used as an initial descriptors' pool in QSPR modeling to build several QSPR models using different variable selection technique.

Models Building and Validation

QSPR modeling was performed using Multiple Linear Regression Analysis (MLR) of the ISIDA/QSPR program[53] (Figure 3) with combined forward and backward stepwise variable selection techniques.[23,30,33,46,73] MLR is applied to build linear relationships between independent variables (SMF descriptors: X_i, i =1, 2,…) and a dependent variable (here target property $Y = \log K$): $Y = c_0 + \Sigma c_i X_i$, where every descriptor value (SMF count x_{ij}, j = 1, 2,…, n; here n is the number of ligands) is associated with observed property value (y_j, j = 1, 2,…, n), c_i is descriptor contribution, and c_0 is the independent term which is omitted in a part of models. The Singular Value Decomposition method is used to fit contributions c_i and to minimize the sum of squared residuals which are squared differences between the property values calculated by the model ($y_{j,calc}$) and observed values ($y_{j,exp}$) in the training set. The program can generate more than 25,000 MLR models; each of them corresponds to particular type of the SMF descriptors and MLR equation ($c_0 = 0$ or $c_0 \neq 0$) and applied variable selection technique. Here, three sub-algorithms FVS-1, FVS-2 and FVS-3 for forward stepwise variable selection [33] and the algorithm for backward stepwise variable selection[28] have been applied. The efficiency

of the FVS procedure was compared with an implementation of Genetic Algorithm[33,51] on the QSPR modeling of antifilarial and different types of anti-HIV activities. The results show similar predictive performance of computationally expensive GA-based approaches and FVS calculations. The leave-one-out (LOO) cross–validation correlation coefficient Q served as a criterion of model selection: the acceptable models were characterized by $Q^2 > 0.5$.

The log$K$ values were predicted by consensus models (CMs). One consensus model combines predictions issued from a multitude of individual models originated from different types of the SMF descriptors and variable selection algorithms.[23,28,35,73,74] Thus for every compound from the test set, the *target* property is computed as an arithmetic mean of values obtained by individual models excluding those leading to outlying values according to Tompson's rule.[75] If a test compound is identified as being outside an applicability domain (AD) of individual model, the prediction by given model for a given compound is not included in CM.

In order to validate CM, the external 5-fold cross validation (5-CV) was applied.[40,74] In this procedure, an entire dataset is divided in 5 non-overlapping pairs of training and test sets. Predictions are prepared for all molecules (n) of the initial dataset, since each of them belongs to one of the test sets. The descriptor selection and model acceptance procedures were performed only on the training folds. Predictive performance of CM has been estimated using coefficient of determination (R_0^2) and root-mean squared error ($RMSE$) for a combination of all five test sets

$$R_0^2 = 1 - \sum_{i=1}^{n}\left(Y_{\exp,i} - Y_{pred,i}\right)^2 \Big/ \sum_{i=1}^{n}\left(Y_{\exp,i} - \langle Y \rangle_{\exp}\right)^2$$

$$RMSE = \left[\sum_{i=1}^{n}\left(Y_{\exp,i} - Y_{pred,i}\right)^2 \Big/ n\right]^{1/2}$$

where Yexp and Ypred are, respectively, experimental and predicted values of the stability constant logK.

Models Applicability Domain Definitions

Ensemble of three approaches for model AD has been applied. Two of them - Bounding Box and Fragment Control – have been reported earlier.[46] The bounding box method considers as AD a multi-dimension descriptor space confined by minimal and maximal values of counts of SMF descriptors involved in an individual model. Fragment control rejects a prediction for a test compound containing SMF fragments which don't occur in the initial SMF pool generated for the training set.[23,33,43]

Here, we introduce "Quorum Control", a new AD definition which discards a test compound if CM includes less than n_Q % of the total number of selected individual models ("no consensus without quorum"). Here n_Q = 15 - 20 % was used. Application of the "Quorum Control" AD results in discarding from the initial dataset molecules containing "rare" fragments, *e.g.*, occurring in less than 3 molecules. Another speaking, this new AD withdraws the molecules with statistically insignificant fragments. The role of the approach is to discard the molecules for which predictions are considered as unreliable. This leads to iterative CMs building with reduction of the modeling sets in 1.5 - 2 times. Finally, 952 organic ligands and 2632 log$K$ values were involved in the modeling. Separately, this includes 17 datasets of 103 (Ce^{3+}), 149 (Pr^{3+}), 141 (Nd^{3+}), 161 (Sm^{3+}), 128 (Eu^{3+}), 168 (Gd^{3+}), 81 (Tb^{3+}), 107 (Dy^{3+}), 112 (Ho^{3+}), 94 (Er^{3+}), 73 (Tm^{3+}), 99 (Yb^{3+}), 92 (Lu^{3+}), 88 (Ag^{+}), 568 (Zn^{2+}), 416 (Cd^{2+}) and 52 (Hg^{2+}) organic ligands (Figure 1).

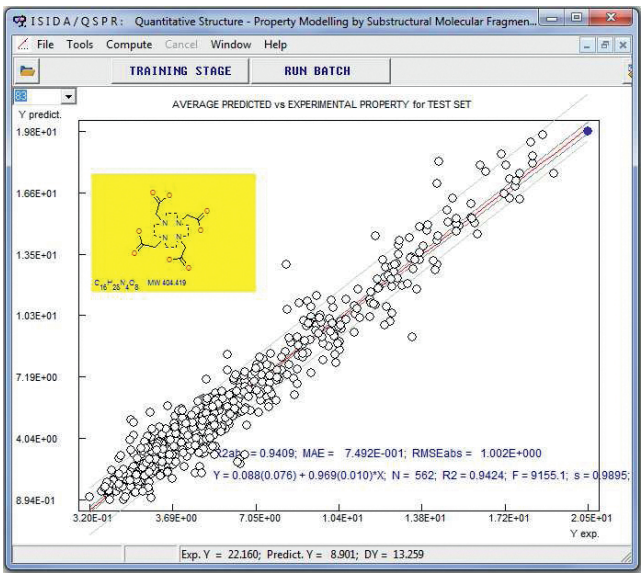


Figure 3. The interface of the ISIDA/QSPR program shows predicted versus experimental log$K$: highlighted ligand for selected data point for the Zn^{2+} complexation.

Results and Discussion

1800 individual structure - property models were built for every metal ion in 5-CV procedure, only the most robust models ($Q^2 > 0.5$) were selected to include in CMs. Obtained CMs demonstrate a reasonable predictive ability of log$K$ in 5-CV procedure: $RMSE$ values vary from 0.52 (Ag^{+}) to 1.10 (Dy^{3+}) with the exception of 1.39 (Hg^{2+}) (Figures 4 and 5). $RMSE$ value is essentially higher for Hg^{2+} than for the rest of metal ions as a result of moderate data set of diverse ligands. Squared determination coefficient R_0^2 changes from 0.900 (Ag^{+}) to 0.977 (Gd^{3+}).

"Quorum Control" guided QSPR models perform better than previously reported models[23,39,40,52] for all studied metal ions. In whole, $RMSE$ values of predictions obtained in this work are twice lower than those for the earlier reported models[23,39,40,52] (Figure 4) and they are close to experimental systematic errors (see Table 1).

Different individual models involve several common fragments whose contributions into log$K$ for particular metal slightly vary from one model to another one. For instance, in

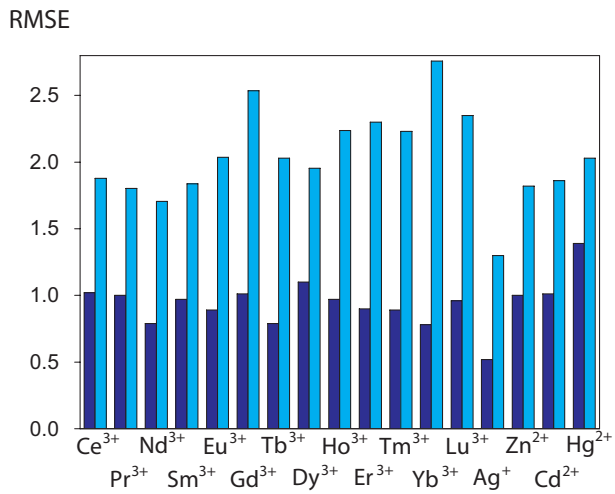


Figure 4. $RMSE$ of the log$K$ predictions obtained in 5-CV procedure: this work (*dark blue*) and from refs[23,40,52] for lanthanide ions,[52] Ag^{+},[40] Zn^{2+}, Cd^{2+}, and Hg^{2+}[23] (*blue*).

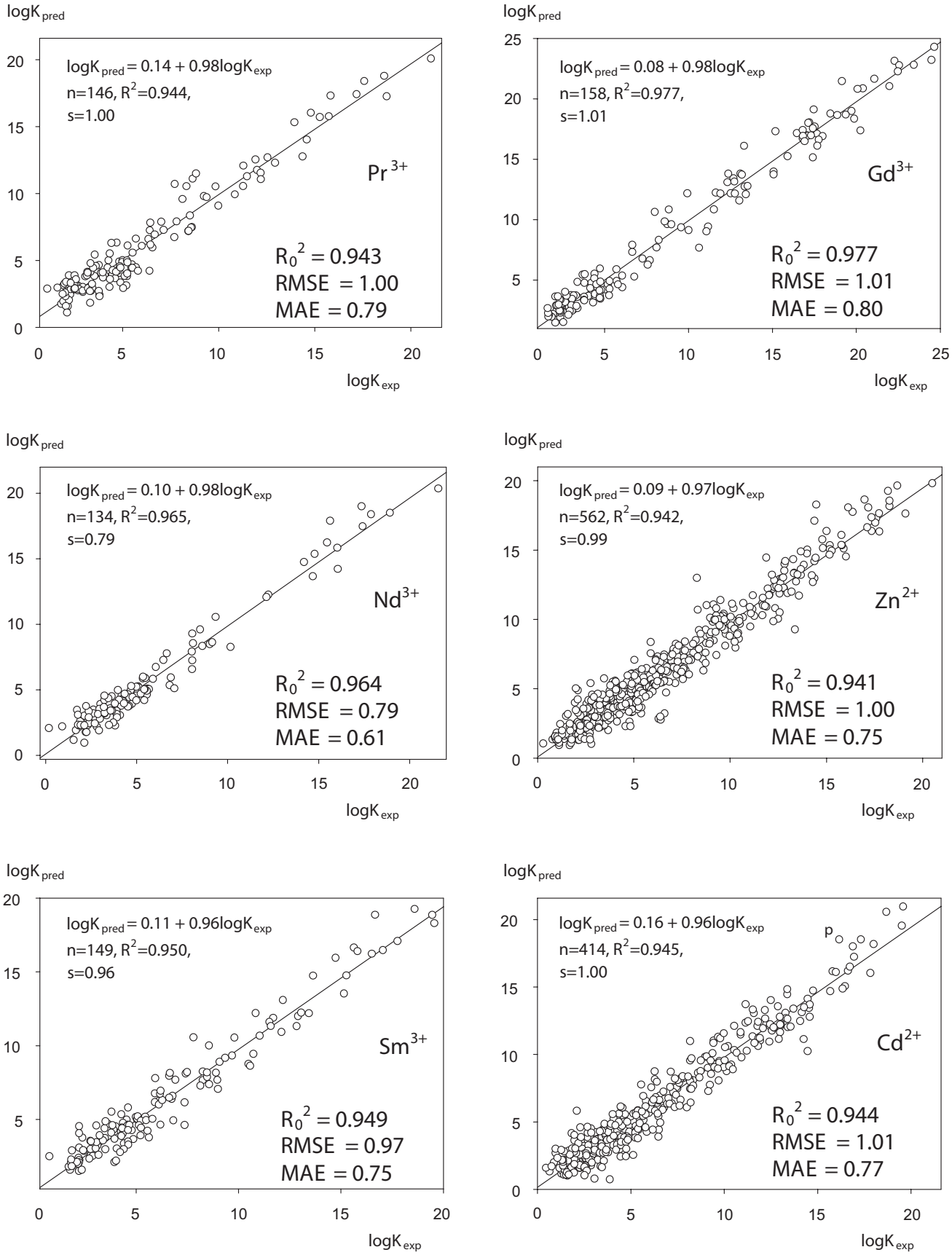


Figure 5. Predictive performance of the models built on 6 largest datasets for Pr^{3+}, Nd^{3+} , Sm^{3+}, Gd^{3+}, Zn^{2+}, and Cd^{2+}. The plots show the correlations between predicted (in 5-fold cross validation) versus experimental log*K* values.

the complexes of lanthanide ions, shortest topological paths O=C-C.N.C-C=O of 2,6-dicarboxypiperidyl-*N*-ethanoic acid, N-C-C=O occurred in 2-aminoacids and different acyclic and macrocyclic ligands with $NCH_2CO(OH)$ group(s), C.C.O.C.C.N-C-P=O in different aza-macrocycles with $CH_2PO(OH)_2$ pendant group(s) contribute each about 2.5 log*K* units. (Here C.N, O.C, and C.C are fragments with single bond in cycles). Information about molecular moieties with high positive or negative contributions into log*K* could be particularly useful for the design of new metal binders.

Conclusions

In order to improve the predictive performance of QSPR models prepared with the ISIDA QSPR program, a new definition of models applicability domain (AD) has been proposed. This approach is closely related to the ensemble modeling technique in which many individual models (instead of one sole) are generated for one training set, followed by their application on the test set compounds. Some of these individual models may not be accepted because of the Bounding Box and Fragment Control AD approaches. More models are discarded, smaller predictive performance is expected. Thus, Quorum Control, a new AD definition imposes a threshold for the acceptance rate of individual models leading to reliable consensus.

Efficiency of this approach has been demonstrated in ensemble modeling of the stability constants log*K* of the 1:1 complexes of 17 lanthanide and transition metal ions with various organic ligands in water. The individual linear models based on substructural molecular fragment descriptors were validated in 5-fold cross-validation procedure. It has been shown that application Quorum Control AD significantly increases predicted performance of the models. Thus, the root-mean squared error of predictions *RMSE* varies from 0.52 (Ag^+) to 1.39 (Hg^{2+}), which is similar to systematic experimental errors.

Acknowledgment. We thank CNRS France, the French Embassy in Russia and the Russian Foundation for Basic Research (project no. 09-03-93106) for support.

References

1. *Comprehensive Coordination Chemistry II. Applications of Coordination Chemistry*. (Ward M. D., Ed.). San Diego: Elsevier, **2003**.
2. Duca G. *Homogeneous Catalysis with Metal Complexes. Fundamentals and Applications. Springer Series in Chemical Physics*. Berlin, Heidelberg: Springer, **2012**. Vol. 102.
3. Kumar S., Dhar D.N., Saxena P. N. *J. Sci. Ind. Res.* **2009**, *68*, 181-187.
4. Tretyakov Y.D., Martynenko L.I., Grigoryev A.N., Tsivadze A.Y. *Neorganicheskaya Khimiya. Khimiya Elementov. 1.* [*Inorganic Chemistry. Chemistry of Elements. Book 1]*. Moscow: Khimia, **2001** (in Russ.).
5. *Comprehensive Coordination Chemistry II. Bio-coordination Chemistry*. (Que L.J., Tolman W.B., Eds.). San Diego: Elsevier, **2003**.
6. Bhattacharya P.K. *Metal Ions in Biochemistry*. Harrow, U. K.: Alpha Science International, **2005**.
7. *Metal Ions in Biological Systems. Vol 37. Manganese and Its Role in Biological Processes*. (Sigel A., Sigel H., Eds.). New York: CRC Press, **2000**.
8. Drahoš B., Kotek J., Císařová I., Hermann P., Helm L., Lukeš I., Tóth E. *Inorg. Chem.* **2011**, *50*, 12785-12801.
9. Svobodová I., Lubal P., Plutnar J., Havličková J., Kotek J., Hermann P., Lukeš I. *Dalton Trans.* **2006**, 5184-5197.
10. Bazzicalupi C., Bencini A., Bianchi A., Borsari L., Danesi A., Giorgi C., Lodeiro C., Mariani P., Pina F., Santarelli S., Tamayo A., Valtancoli B. *Dalton Trans.* **2006**, 4000-4010.
11. Hancock R.D. *Analyst* **1997**, *122*, 51R-58R.
12. Hancock R.D., Martell A.E. *Chem. Rev.* **1989**, *89*, 1875-1914.
13. Martell A.E., Hancock R.D., Motekaitis R.J. *Coord. Chem. Rev.* **1994**, *133*, 39-65.
14. Dimmock P.W., Warwick P., Robbins R.A. *Analyst* **1995**, *120*, 2159-2170.
15. *IUPAC Stability Constants Database*. version 5.33, 2012. http://www.acadsoft.co.uk/ (accessed Sept. 20, 2012).
16. *NIST Critically Selected Stability Constants of Metal Complexes Database*. version 8.0, 2004. http://www.nist.gov/srd/nist46.cfm (accessed May 14, 2012).
17. *Joint Expert Speciation System. A powerful research tool for modelling chemical speciation in complex environments*. version Oct. 3, 2011, 1985 - 2012. http://jess.murdoch.edu.au/jess_home.htm (accessed Sept. 20, 2012).
18. Martell A.E., Smith R.M. *Critical Stability Constants*. New York: Plenum Press, **1989**. Vol. 6.
19. Christensen J.J., Izatt R.M. *Handbook of Metal Ligand Heats and Related Thermodynamic Quantities*. New York: Marcel Dekker Inc., **1983**.
20. Izatt R.M., Pawlak K., Bradshaw J.S., Bruening R.L. *Chem. Rev.* **1991**, *91*, 1721-2085.
21. Solov'ev V.P., Vnuk E.A., Strakhova N.N., Raevsky O.A. *Thermodynamics of Complexation of the Macrocyclic Polyethers with Salts of Alkali and Alkaline-Earth Metals*. Moscow: VINITI, **1991** (in Russ.).
22. Varnek A., Solov'ev V. In: *Ion Exchange and Solvent Extraction, A Series of Advances*. (Sengupta A. K., Moyer B. A., Eds.). Boca Raton: CRC Press, Taylor and Francis Group, **2009**, Vol. 19, p. 319-358.
23. Solov'ev V., Sukhno I., Buzko V., Polushin A., Marcou G., Tsivadze A., Varnek A. *J. Incl. Phenom. Macrocycl. Chem.* **2012**, *72*, 309-321.
24. Daraei H., Irandoust M., Ghasemi J.B., Kurdian A.R. *J. Incl. Phenom. Macrocycl. Chem.* **2012**, *72*, 423-435.
25. Li Y., Su L., Zhang X., Huang X., Zhai H. *Chemom. Intell. Lab. Syst.* **2011**, *105*, 106-113.
26. Ghasemi J.B., Ahmadi S., Ayati M. *Macroheterocycles* **2010**, *3*, 234-242.
27. Ghasemi J., Saaidpour S. *J. Incl. Phenom. Macrocycl. Chem.* **2008**, *60*, 339-351.
28. Solov'ev V.P., Varnek A.A. *Rus. Chem. Bull.* **2004**, *53*, 1434-1445.
29. Varnek A.A., Wipff G., Solov'ev V.P., Solotnov A.F. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 812-829.
30. Solov'ev V.P., Varnek A.A., Wipff G. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 847-858.
31. Gakh A.A., Sumpter B.G., Noid D.W., Sachleben R.A., Moyer B.A. *J. Incl. Phenom. Mol. Recognit. Chem.* **1997**, *27*, 201-213.
32. Shi Z.G., McCullough E.A. *J. Incl. Phenom. Mol. Recognit. Chem.* **1994**, *18*, 9-26.
33. Solov'ev V.P., Kireeva N., Tsivadze A.Y., Varnek A. *J. Incl. Phenom. Macrocycl. Chem.*, **2012**, publ. online 13.06.12, DOI: 10.1007/s10847-10012-10185-x.
34. Cabaniss S. E. *Environ. Sci. Technol.* **2008**, *42*, 5210-5216.
35. Solov'ev V.P., Kireeva N.V., Tsivadze A.Y., Varnek A.A. *J. Struct. Chem.* **2006**, *47*, 298-311.

36. Toropov A.A., Toropova A.P., Nesterova A.I., Nabiev O.M. *Russ. J. Coord. Chem.* **2004**, *30*, 611–617.
37. Toropov A.A., Toropova A.P. *Russ. J. Coord. Chem.* **2002**, *28*, 877-880.
38. Raevskii O.A., Sapegin A.M., Chistyakov V.V., Solov'ev V.P., Zefirov N.S. *Koord. Khim.* **1990**, *16*, 1175-1184 (in Russ.).
39. Varnek A., Fourches D., Kireeva N., Klimchuk O., Marcou G., Tsivadze A., Solov'ev V. *Radiochim. Acta* **2008**, *96*, 505-511.
40. Tetko I.V., Solov'ev V.P., Antonov A.V., Yao X.J., Fan B.T., Hoonakker F., Fourches D., Lachiche N., Varnek A. *J. Chem. Inf. Model.* **2006**, *46*, 808-819.
41. Svetlitski R., Lomaka A., Karelson M. *Separat. Sci. Technol.* **2006**, *41*, 197-216.
42. Qi Y.-H., Zhang Q.-Y., Xu L. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1471-1475.
43. Solov'ev V., Marcou G., Tsivadze A.Y., Varnek A. *Ind. Eng. Chem. Res.* **2012**, *51*, 13482-13489.
44. Raos N., Miličević A. *Arch. Ind. Hyg. Toxicol.* **2009**, *60*, 123-128.
45. Grgas B., Nikolić S., Paulić N., Raos N. *Croat. Chem. Acta* **1999**, *72*, 885-895.
46. Varnek A., Fourches D., Horvath D., Klimchuk O., Gaudin C., Vayer P., Solov'ev V., Hoonakker F., Tetko I.V., Marcou G. *Curr. Comput.-Aided Drug Des.* **2008**, *4*, 191-198.
47. Baskin I., Varnek A. In: *Chemoinformatics Approaches to Virtual Screening*. (Varnek A., Tropsha A., Eds.). Cambridge: RSC Publishing, **2008**, p. 1-43.
48. Katritzky A.R., Fara D.C., Yang H., Karelson M., Suzuki T., Solov'ev V.P., Varnek A. *J. Chem. Inf. Comp. Sci.* **2004**, *44*, 529-541.
49. Varnek A., Fourches D., Sieffert N., Solov'ev V.P., Hill C., Lecomte M. *J. Solv. Extr. Ion. Exch.* **2007**, *25*, 1-26.
50. Varnek A., Fourches D., Solov'ev V., Klimchuk O., Ouadi A., Billard I. *Solv. Extr. Ion Exch.* **2007**, *25*, 433-462.
51. Horvath D., Bonachera F., Solov'ev V., Gaudin C., Varnek A. *J. Chem. Inf. Model.* **2007**, *47*, 927-939.
52. Kireeva N. *Modèles QSPR multiples pour les constants de stabilité de complexes métaux – ligands organiques en solution et de points de fusion de liquides ioniques. Thesis*. Strasbourg: Université de Strasbourg, **2009**.
53. *ISIDA (In Silico Design and Data Analysis) program*. version 5.76, 2008-2012. http://infochim.u-strasbg.fr/spip.php?rubrique53 or http://vpsolovev.ru/programs/ (accessed 09.11.12).
54. Varnek A., Fourches D., Hoonakker F., Solov'ev V.P. *J. Comput. Aid. Mol. Des.* **2005**, *19*, 693-703.
55. Arnaud-Neu F., Delgado R., Chaves S. *Pure Appl. Chem.* **2003**, *75*, 71-102.
56. Batyaev I., Fogileva R. *Zh. Neorg. Khim.* **1974**, *19*, 670-673 (in Russ.).
57. Sekhon B.S., Chopra S.L. *Thermochim. Acta* **1973**, *7*, 151-157.
58. Bianchi A., Calabi L., Ferrini L., Losi P., Uggeri F., Valtancoli B. *Inorg. Chim. Acta* **1996**, *249*, 13-15.
59. Aime S., Anelli P. L., Botta M., Fedeli F., Grandi M., Paoli P., Uggeri F. *Inorg. Chem.* **1992**, *31*, 2422-2428.
60. Choppin G.R., Liu Q., Rizkalla E.N. *Inorg. Chim. Acta* **1988**, *145*, 309-314.
61. Makushova G.N., Ternovaya T.V., Kostromina N.A. *Koord. Khim.* **1981**, *7*, 372-376 (in Russ.).
62. Narayana V.G., Swamy S.J., Lingaiah P. *Indian J. Chem.* **1986**, *25A*, 491-493.
63. Pakhomova D.V., Kumok V.N., Serebrennikov V.V. *Zh. Neorg. Khim.* **1969**, *14*, 1434-1435 (in Russ.).
64. Sandhu R.S., Gandhi J.S., Kumar R. *Thermochim. Acta* **1981**, *47*, 117-119.
65. Feige P., Mocker D., Dreyer R., Münze R. *J. Inorg. Nucl. Chem.* **1973**, *35*, 3269-3275.
66. Martynenko L.I., Muratova N.M., Borisova A.P. *Zh. Neorg. Khim.* **1980**, *25*, 713-716 (in Russ.).
67. Babich V.A., Gorelov I.P. *Zh. Anal. Khim.* **1971**, *26*, 1832 (in Russ.).
68. Wang Z.-M., van de Burgt L.J., Choppin G.R. *Inorg. Chim. Acta* **1999**, *293*, 167-177.
69. Choppin G.R., Rizkalla E.N., El-ansi T. A., Dadgar A. *J. Coord. Chem.* **1994**, *31*, 297-304.
70. Degischer G., Choppin G.R. *J. Inorg. Nucl. Chem.* **1972**, *34*, 2823-2830.
71. Powell J.E., Farrell J.L., Neillie W.F.S., Russell R. *J. Inorg. Nucl. Chem.* **1968**, *30*, 2223-2231.
72. Solov'ev V., Oprisiu I., Marcou G., Varnek A. *Ind. Eng. Chem. Res.* **2011**, *50*, 14162-14167.
73. Varnek A., Solov'ev V.P. *Comb. Chem. High Throughput Screening* **2005**, *8*, 403-416.
74. Varnek A., Kireeva N., Tetko I.V., Baskin I.I., Solov'ev V.P. *J. Chem. Inf. Model.* **2007**, *47*, 1111-1122.
75. Muller P.H., Neumann P., Storm R. *Tafeln der mathematischen Statistik*. Leipzig: VEB Fachbuchverlag, **1979**.

Received 23.11.2012
Accepted 14.12.2012