TARTU ÜLIKOOL

Füüsika - keemia teaduskond
Keemilise Füüsika Instituut

Ruslan Svetlitski

QSPR MODELLING OF LANTHANIDE-ORGANIC COMPLEX STABILITY CONSTANTS

Magistritöö
teoreetilise ja arvutikeemia erialal

Juhendaja: prof. Mati Karelson

Tatru 2004

Contents

1. Introduction **3**

2. Literature overview **6**

2.1 OSPR modeling 6

2.2 Principal component analysis (PCA) 7

3. Data set **9**

4. Meethodology **14**

4.1 QSPR modeling 14

4.2 Principal component analysis 14

5. Results and discussion **18**

6. Conclusions **30**

7. Kokkuvõte **31**

8. References and notes **32**

1. Introduction

Lanthanide chemistry has come under increasing study. In last few years, the experimentalists have sought to exploit their unique chemical and magnetic properties for variety of advanced materials, catalysis, and biomedical applications[1-4]. Especially rapidly increasing area is connected with luminescence properties of lanthanide-organic complexes. Although lanthanides are often referred to as rare earth elements, they are not scarce. In our technology-filled lives they are ubiquitous, especially where the emission of light is involved. The average man looks at lanthanide luminescence for at least 2 hours and 33 minutes a day. That is, if he only watches television. Computer screens too use lanthanide doped phosphors to convert cathode rays into visible information. The same type of materials is also found in fluorescent lighting.

Lanthanide ions are also applied as light generating and amplifying constituents in lasers (e.g. in the neodymium (III): YAG laser) and optical amplifiers (EDFA's: erbium (III) doped fiber amplifiers). Over the past few years, the latter have become a key component in optical telecommunications, where light instead of electricity carries information. The lanthanide-doped materials for these applications have thus far been inorganic glasses and crystals. Complexes of lanthanide ions with organic ligands hold the promise that for some applications they may provide alternatives. Scientists are working on polymers doped with lanthanide complexes for optical amplification, and the first prototypes of flat panel displays containing lanthanide complexes have seen the lights.

Currently, the most important use of luminescent lanthanide complexes is in medical diagnostics, where they are used to detect small amount of biomolecules that can tell about the physical state of patient. The well-known DELFIA (Dissociation

Enhanced Lanthanide Fluorescence Immunoassay) detection scheme of the Finnish company Wallac (now owned by PerkinElmer Lifesciences) is basis of large number of clinical tests, e.g. for testing newborn babies, in the diagnostics of many diseases and even to detect BSE (bovine spongiform encephalopathy, or mad cow disease).

The other area of using lanthanide complexes is magnetic resonance imaging (MRI) contrast agents[5]. The high-spin paramagnetism and long electronic time relaxation time of Gd^{3+} has made it pre-eminent among contrast agent for MRI[7,8]. Related complexes of Dy and Tm – with much stronger electronic relaxation times – are effective NMR shift reagents[9]. The controlled modulation of Lewis acidity across the series is allowing the development of complexes exhibiting phosphatase activity[10], while the redox activity of cerium, samarium, and europium may be expected to allow the development of further selective oxidants and reductants.

Lanthanide complexes in solution exhibit a well-defined luminescence that is characterized by narrow emission bands, large Stokes' shifts, and long excited-state lifetimes. Europium and terbium complexes possess excited-state lifetimes in aqueous solution of up to 5ms and emit in the red and green; they have been used as probes in fluoroimmunoassays[10,11] and show considerable promise in luminescence imaging and as sensors for certain bioactive ions[12]. The near-IR emission from the excited state Nd^{3+}, Yb^{3+}, and Er^{3+} is less long-lived, but complexes of these ions offer much promise as probes in vivo as tissue is relatively transparent to incident light with a wavelength around 1000 nm.

In the past few years several excellent reviews have appeared detailing aspects of the structure and solution dynamics of contrast agents [7,8,17], biomedical and NMR applications[7,8,14], complex design features[15], thermodynamic aspects of complex formation[7,16] , the development of luminescent lanthanide complexes operating in

aqueous media[10,11,17] , and the diagnostic and therapeutic uses of lanthanide – texaphyrin and – porphyrin complexes[18,19].

Compounds of gadolinium (Gd) are currently used as commercial MRI contrast agents. If the structure of potential contrast agents could be successfully determined, the relationship between chemical properties and relativity, which is current unclear, could be deducted and improved agents developed. Similarly, rational design of lanthanide complexes with important application in other areas could be pursued.

Chelation combined with solvent extraction is one of the most widely used techniques for preconcentration and separation of metal ions from aqueous samples for analytical purposes[20]. These solvent extraction procedures, however, are usually time consuming, especially for solids where leaching procedures are needed to release the metal ions before complexation and solvent extraction. The complex stabilities are very important for the development of new efficient methods of separation of lanthanides from solution. It is well known that the separability depends on the stability constants of the complexes formed[21].

The above-listed applications require the development of lanthanide chelates with carefully tailored chemical, structural and spectroscopic (or magnetic) properties. Thus, the aim of the present work is the development of predictive QSPR models of stability constants for lanthanide complexes with organic ligands. Such models enable to make reliable predictions of the stability constants for previously unknown complexes and to elucidate the structural factors determining the stability of complexes.

2. Literature overview

2.1 OSPR modeling

The beginning of QSPR dates back more then a century. In 1884 Mills developed a QSPR for prediction the melting points and boiling points of homologous series[22]. Similar pioneering work followed shortly after on quantitative structure activity relationships (QSAR) in studies of relationships between the potency of local anesthetics and oil/water partition coefficient[23], and between narcosis and chain length[24].

QSPR models are empirical equations, used for estimating various physical or thermodynamic properties of molecules. A QSPR model has the form

$$P = a + b \cdot D1 + c \cdot D2 + d \cdot D3 + \cdots, \qquad (1)$$

where P is the physical property of interest, a, b, c, . . . are regression coefficients, and D1, D2, D3, . . . are parameters derived from the molecular structure, so-called descriptors. A variety of different types of descriptors can be used[25]. The simplest types are constitutional or topological descriptors, such as the number of carbon atoms, and parameters describing types and order of the chemical bonds in the molecules. Various geometrical descriptors, including the principal moment of inertia, can also be used. The most important and also the most complicated descriptors, are electrostatic and quantum chemical descriptors. The electrostatic descriptors are parameters, which depend on the charge distribution within the molecule, including the dipole moment. An example of quantum chemically derived descriptors are the HOMO and LUMO energies[26].

2.2 Principal component analysis (PCA)

PCA[27,28] is relatively straightforward method for transforming a given set of data into principal component (PC) that are orthogonal (unrelated) to each other. In contrast to multiregression analysis (MRA), PCA requires no particular assumption about the underlying structure of the variables. Mester and Schwarz[29] studied the principal components of solvent ionicity using factor analysis and found a single principal component of ionicity, which is common to all the various operational charge definitions. Heberger and Lopata[30] performed PCA on experimental and calculated parameters of radical addition reactions to assess nucleophilicity and electrophilicity of radicals. Two principal components were extracted, accounting for electrophilic and nucleophilic properties of radicals.

The principal component model may be described by eq (2) where $\vec{x}_{ik}$ is the mean scaled value of the experimental quantities (variables) (scaling weights, w_k, transfer $\vec{x}_{ik}$ to unscaled data, $\vec{x}'_{ik} = w_k^{-1}\vec{x}_{ik}$); t_{ia} are scores; P_{ak} are loadings; e_{ik} are residuals; i is the chemical compound (object); k is the experimental measurement (variables); and a is the principal component.

$$x_{ik} = \vec{x}_{ik} + \sum_{a=1}^{A} t_{ia} P_{ak} + e_{ik} \qquad (2)$$

The number of PCs (scores) exiting in characteristic vector space is equal to, or less than, the number of variables in the data set. Each and every PC is orthogonal to all the other PCs. The first principal component is defined as that giving the largest contribution to the respective PCA of linear relationship exhibited in the data. The

second component may be viewed as the second best linear combination of variables that accounts for the maximum possible of the residual variance after the effect of the first component is removed from the data. Subsequent components are defined similarly until practically all the variance in the data is exhausted.

PCA allows the examination of set of characteristics (variables) of class of compounds (objects) to investigate the relations between them. It enables the identification of one, two, three, or more PCs derived from the characteristics for the compound examined. These components have defined values for each of the compounds (t_{1i}, t_{2i}, t_{3i}, the "scores") and are taken in certain proportions (p_{1k}, p_{2k}, p_{3k}, etc., the "loadings") for each type of characteristics. Graphical representations of these values, the "scores" plot for the compounds and the "loadings" plot for the characteristics, provide pictures that allow the recognition of systematic patterns that is otherwise difficult to deduce from the original data matrix.

Examples of some of applications of PC analysis in heterocyclic chemistry include investigations of (i) aromaticity[32,33,34,35] and of (ii) simultaneous dependence of S_N2 rates on alkyl group structure and leaving group nucleofugacity in nucleophilic displacements in which heterocycles act as leaving groups[36,37]. A multivariate statistical treatment was used for solvent characteristics where large numbers of solvents and many scales new dimensions to the problems generally investigated in LFERs.

3. Data set

The data set of experimental stability constants of complexes between lanthanide ions and structurally variable organic ligands was compiled from the literature (cf. references from Table 1). The 1:1 complex formation between a ligand L^n and a cation M^{3+} is given by

$$L^n + M^{3+} \leftrightarrow ML^{n+3}$$

For the QSPR model development, the logarithmic constants $\log K_1$ were used, where K_1 is defined as follows:

$$K_1 = \frac{[LnL^{n+3}]}{[Ln^{3+}][L^n]}$$

All stability constants correspond to aqueous solutions at the ionic force $\mu = 0.1$ and temperature 25 ^{0}C.

Table 1 gives the list of the 66 different organic ligands that were selected for the present QSPR study, each with 6 or more data points. Tables 2 and 3 include the additional information about the chemical structure of the organic ligands used. In Table 4, the 23 different metal descriptors that were used in present QSPR study are listed.

Table 1. The organic ligands used in the QSPR treatment.

no.	ligand name	Ref.	no.	ligand name	Ref.	no.	ligand name	Ref.
1	IMDA	I-III	23	BIMDA	I	45	BCAM	I
2	Maleic acid	IV	24	DTPA	I	46	BCG	I
3	Acetate	III	25	EDTP	I, III	47	EDDM	I
4	α-Hydroxy-isobutyric acid	III	26	EEDTA	I, III	48	EDDS	I
5	4-Aminobenzoate	V	27	EGTA	I, IIII	49	EDDG	I
6	4-Hydroxybenzoate	V	28	HEDTA	I, III	50	DPDS	I
7	4-Nitrobenzoate	V	29	MEPDA	I, III	51	2-OPDTA	I
8	Acrylic acid	IV	30	MIMDA	I, III	52	OPDM	I
9	Methacrylic acid	IV	31	Nitrilotriacetic acid	I, III	53	OPDS	I
10	K22DAP	VI	32	PIMDA	I, III	54	OPDG	I
11	K22DA	VI	33	Glycolic acid	IIIX	55	EDDIP	I
12	K22DP	VI	34	Metoxyacetic acid	III	56	EDAP	I
13	K22MA	VI	35	Glyoxalic acid	III	57	EDTMP	I
14	K21DA	VI	36	a-Hydroxypropionic acid	III	58	OEAIP	I
15	EDTA	I,III	37	Picolinic acid	III	59	TEAIP	I
16	EDDA	I,III	38	Piperidin-2,6 dicarboxy acid	III	60	DETAIP	I
17	Malonic acid	VII	39	Glycine	III	61	OFIDA	I
18	4-dimethylaminobenzylidenepyruvate	VIII	40	Trimethylenediaminetetraacetic acid	I, III	62	KMIDA	I
19	4-dimethylaminocinnamylidenepyruvate	VIII	41	Triethylenetetraaminehexaacetic acid	I	63	DGL	I
20	Acetylacetone	III	42	CA	I			
21	1,2-Cyclohexylenedinitrilotetraacetic acid	I, III	43	IDS	I			
22	BENTA	I	44	BCA	I			

I. Kostromina, N.A.,1980. Complexing agents of rare earth metals, Nauka, Moscow; II .Hramov, V.P., 1974. Complexing agents of rare earth metals, Saratov university; III. Yatsimirskii, K.B., Kostromina, N.A., Sheka, Z. Davidenko, N. K., Kriss, E.E., Ermolenko, V.I., 1966. Chemistry of Complex Compounds of Rare Earth Elements. Naukova Dumka, Kiev (Russian Edition); IV. Panvushkin, V.T., Achrimenko, N.V., Khachatrian, A.S., 1998. Mixed-ligand complexes of three valent lanthanide ions with acetylacetone and some organic unsaturated acids. Polyhedron. 17, 3053-3058; V. Yun, S.J., Kang, S.K., Yun, S.S., 1999. Thermodynamics of complexation of lanthanides by some benzoic acid derivatives in aqueous solution. Thermochim. Acta. 333, 13 – 19; VI. Kim, J., Lee, C. N., Han, S. H., Suh, M. Y., 1997. Studies on complexation and solvent extraction of lanthanides in the presence of diaza-18 crown-6-di-isopropionic acid. Talanta, 45, 437-444; VII. Hirikawa, T., Hashimoto, Y., 1997. Simultaneous separation of yttrium and lanthanide ions by isotachophoresis. J. Chromatogr. A 772, 357-367; VIII. Pereira, N.C.S., Melios, C.B., Marques, R.N., Siqueira, O.S., De Morales, M., Molina, M., 1997. 4-Dimethylaminocinnamylidenepyruvic acid: synthesis, characterization and complexation with trivalent lanthanides, yttrium(III), scandium(III), thorium(IV) and uranium(VI) in aqueous solution. J. Alloys Compd. 297, 94-97;

Table 2. The ligand substructures

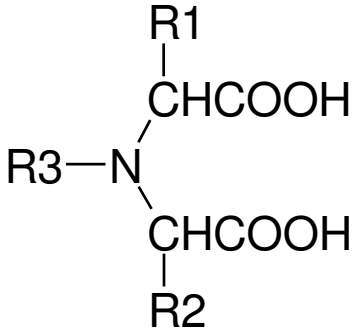
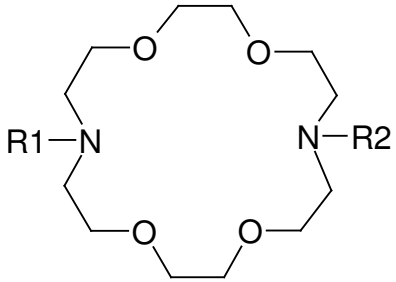
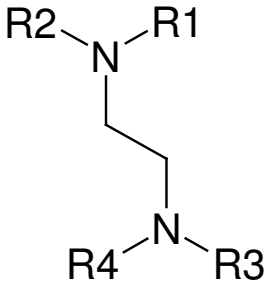
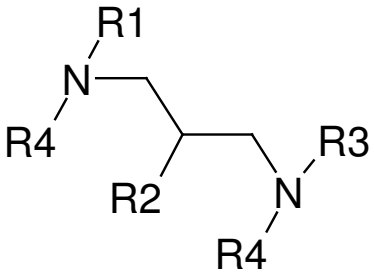
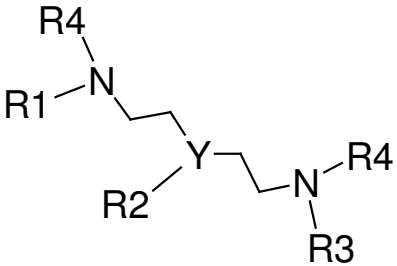
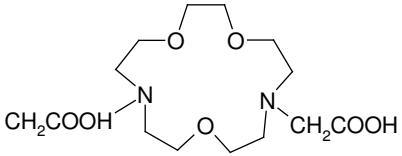
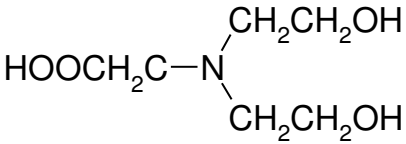
1	2	3	4
			
5	K21DA	DGL	
			

Table 3. The ligand structures

no.	Ligand	R1	R2	R3	R4	substructure
1	IMDA	-H	-H	-H	-	1
10	K22DAP	$-CH(CH_3)COOH$	$-CH(CH_3)COOH$	-	-	2
11	K22DA	$-CH_2COOH$	$-CH_2COOH$	-	-	2
12	K22DP	$-CH_2CH_2COOH$	$-CH_2CH_2COOH$	-	-	2
13	K22MA	-H	$-CH_2COOH$	-	-	2
15	EDTA	$-CH_2COOH$	$-CH_2COOH$	$-CH_2COOH$	$-CH_2COOH$	3
16	EDDA	$-CH_2COOH$	-H	$-CH_2COOH$	-H	3
22	BENTA	-H	-H	$-C_6H_4-HC(COOH)-$	-	1
23	BIMDA	-H	-H	$-CH_2-C_6H_4-$		1
24	DTPA	$-CH_2COOH$	$-CH_2COOH$	$-CH_2COOH$	$-CH_2COOH$	5 (Y=N)
25	EDTP	$-CH_2CH_2COOH$	$-CH_2CH_2COOH$	$-CH_2CH_2COOH$	$-CH_2CH_2COOH$	3
26	EEDTA	$-CH_2COOH$	-	$-CH_2COOH$	$-CH_2COOH$	(Y=O)
27	EGTA	$-CH_2COOH$	-	$-CH_2COOH$	$-CH_2COOH$	(Y= OCH_2CH_2O)
28	HEDTA	$-CH_2CH_2OH$	$-CH_2COOH$	$-CH_2COOH$	$-CH_2COOH$	4
29	MEPDA	-H	-H	$-CH_2-C_5H_3N-CH_3$	-	1
30	MIMDA	-H	-H	$-CH_3$	-	1
32	PIMDA	-H	-H	$-CH_2-C_5H_4N$	-	1
42	CA	$-CH_2COOH$	-H	-H	-	1
43	IDS	$-CH_2COOH$	$-CH_2COOH$	-H	-	1
44	BCA	-H	-H	$-CH(CH_2COOH)COOH$	-	1
45	BCAM	-H	-H	$-CH(COOH)COOH$	-	1
46	BCG	-H	-H	$-CH(CH_2CH_2COOH)COOH$	-	1
47	EDDM	-H	-H	$-CH(COOH)COOH$	-	3
48	EDDS	-H	-H	$-CH(CH_2COOH)COOH$	-	3
49	EDDG	-H	-H	$-CH(CH_2CH_2COOH)COOH$	-	3
50	DPDS	$-CH_2COOH$	-H	$-CH(CH_2COOH)COOH$	-H	4
51	2-OPDTA	$-CH_2COOH$	-OH	$-CH_2COOH$	$-CH_2COOH$	4
52	OPDM	$-CH(COOH)COOH$	-OH	$-CH(COOH)COOH$	-H	4
53	OPDS	$-CH(CH_2COOH)COOH$	-OH	$-CH(CH_2COOH)COOH$	-H	4
54	OPDG	$-CH(CH_2CH_2COOH)COOH$	-OH	$-CH(CH_2CH_2COOH)COOH$	-H	4
55	EDDIP	$(CH_3)_2C(PO_3H_2)$	-H	$(CH_3)_2C(PO_3H_2)$	-H	3
56	EDAP	$-CH_2PO_3H_2$	$-CH_2COOH$	$-CH_2PO_3H_2$	$-CH_2COOH$	3
57	EDTMP	$-CH_2PO_3H_2$	$-CH_2PO_3H_2$	$-CH_2PO_3H_2$	$-CH_2PO_3H_2$	3
58	OEAIP	$(CH_3)_2C(PO_3H_2)$	-	$(CH_3)_2C(PO_3H_2)$	$(CH_3)_2C(PO_3H_2)$	5 (Y=O)
59	TEAIP	$(CH_3)_2C(PO_3H_2)$	-	$(CH_3)_2C(PO_3H_2)$	$(CH_3)_2C(PO_3H_2)$	5 (Y=S)
60	DETAIP	$(CH_3)_2C(PO_3H_2)$	-H	$(CH_3)_2C(PO_3H_2)$	$(CH_3)_2C(PO_3H_2)$	5 (Y=N)
61	OFIDA	-H	-H	$HO-C_6H_4-$		1
62	KMIDA	-H	-H	$-CH_2-C_6H_3(CH_3)(OH)$		1

Table 4. The external descriptors for lanthanides.

Notation	F01 [a]	F02 [b]	F03 [c]	F04 [d]	F05 [e]	F06 [f]	F07 [g]	F08 [h]	F09 [i]	F10 [j]	F11 [k]	F12 [l]
La	538.1	1067	1850.3	4819	0.207	138.906	57	2.74	22.39	3469	373.9	1.69
Ce	534.4	1050	1949	3547	0.209	140.116	58	2.7	20.69	3257	365	1.65
Pr	527	1020	2086	3761	0.21	140.908	59	2.67	20.8	3212	364	1.65
Nd	533.1	1040	2130	3900	0.215	144.24	60	2.64	20.59	3067	362.8	1.64
Sm	544.5	1070	2260	3990	0.224	150.36	62	2.59	19.98	1778	357.9	1.62
Eu	547.1	1085	2404	4120	0.227	151.964	63	2.56	28.97	1597	398.9	1.85
Gd	593.4	1170	1990	4250	0.235	157.25	64	2.54	19.9	3233	357.3	1.61
Tb	565.8	1110	2114	3839	0.237	158.925	65	2.51	19.3	3041	352.5	1.59
Dy	573	1130	2200	3990	0.243	162.5	66	2.49	19.01	2335	350.3	1.59
Ho	581	1140	2204	4100	0.246	164.93	67	2.47	18.74	2720	348.6	1.58
Er	589.3	1150	2194	4120	0.25	167.26	68	2.45	18.46	2510	346.8	1.57
Tm	596.7	1160	2285	4120	0.252	168.934	69	2.42	19.1	1950	344.7	1.56
Yb	603.4	1174.8	2417	4203	0.258	173.04	70	2.4	24.84	1467	388	1.74
Lu	523.5	1340	2022.3	4370	0.261	174.967	71	2.25	17.78	3315	343.5	-

[a]1st ionisation potential, $kJ \cdot mol^{-1}$. [b]2nd ionisation potential, $kJ \cdot mol^{-1}$. [c]3rd ionisation potential, $kJ \cdot mol^{-1}$. [d]4th ionisation potential, $kJ \cdot mol^{-1}$. [e]Atomic energy, ergs. [f]Atomic mass. [g]Atomic number. [h]Atomic radius, angstrom . [i]Atomic volume, $cm^3 \cdot mol^{-1}$. [j]Boiling point, ^0C. [k]Bond length in Me - Me, pm. [l]Bonding radius (Covalent radius), angstrom.

Table 4. The external descriptors for lanthanides (continued).

Notation	F13 [a]	F14 [b]	F15 [c]	F16 [d]	F17 [e]	F18 [f]	F19 [g]	F20 [h]	F21 [i]	F22 [j]	F23 [k]
La	6.146	0.0126	1.1	400	6.2	431	920	117.2	130	26.392	13.5
Ce	6.689	0.0115	1.12	350	5.5	423	795	115	128.3	26.622	11.4
Pr	6.64	0.0148	1.13	356	6.9	330	935	113	126.6	27.195	12.5
Nd	6.8	0.0157	1.14	328	7.1	285	1010	112.3	124.9	27.406	16.5
Sm	7.353	0.00956	1.17	207	8.6	175	1072	109.8	121.9	29.621	13.3
Eu	5.244	0.0112	1.2	175	9.2	175	822	108.9	120.6	27.657	13.9
Gd	7.901	0.00736	1.2	398	10	305	1311	107.8	119.3	37.111	10.6
Tb	8.219	0.00889	1.2	389	10.8	295	1360	106.3	118	28.607	11.1
Dy	8.551	0.0108	1.22	290	11.1	280	1412	105.2	116.7	28.113	10.7
Ho	8.795	0.0124	1.23	301	17	265	1470	104.1	115.5	27.213	16.2
Er	9.066	0.0117	1.24	317	19.9	285	1522	103	114.4	28.1	14.3
Tm	9.321	0.015	1.25	232	16.8	250	1545	102	113.4	27.029	16.8
Yb	6.57	0.0351	1.1	152	7.7	160	824	100.8	112.5	26.821	34.9
Lu	9.841	-	1.27	428	22	-	1656	-	111.7	26.245	16.4

[a]Density, $g \cdot cm^{-3}$. [b]Electrical conductivity, $10^6 \cdot cm^{-1} \cdot Ohm^{-1}$. [c]Electronegativity. [d]Enthalpy of atomisation $kJ \cdot mol^{-1}$. [e]Heat (Enthalpy) of fusion, $kJ \cdot mol^{-1}$. [f]Heat (Enthalpy) of vaporization, $kJ \cdot mol^{-1}$. [g]Melting point, ^0C. [h]Radius 6-coordinate, octahedral, ion (III), pm. [i]Radius 8-coordinate, ion (III), pm. [j]Specific heat, $J \cdot mol^{-1} \cdot K^{-1}$. [k]Thermal conductivity, $W \cdot m^{-1} \cdot K^{-1}$.

4. Meethodology

4.1 QSPR modeling

The geometrical structure of ligand molecules was optimized using the AM1[38] method within the MOPAC[39] program package. The geometry and other information from the output of quantum chemical calculations were inserted into the CODESSA[40] program, and descriptors for ligands were calculated. All these descriptors are derived solely from molecular structure and do not require experimental data to be calculated. Various data on physical properties were used as the descriptors for metals (Table 4). The CODESSA program was then used to find the best QSPR multilinear equations with 2, 3, or 4 descriptors depending on the size of the data set for a series of ligand complex with a given metal. Analogously, the QSPR equations were developed for a series of metal complexes with a given ligand. Both Heuristic and Best Multi-Linear correlation algorithms available in the CODESSA were used. The respective methodology has been described elsewhere[41]. The CODESSA program has already been successfully applied to correlate molecular structure with various properties including melting points[42], response factors[43], critical micelle concentrations[44,45], aqueous solubility of gases[46], glass transition temperatures of polymers[47], and solvent polarity scales[48].

4.2 Principal component analysis

A multivariate statistical treatment is particular suitable for determining the ligand characteristics.

Table 5. Initial matrix for principal component analysis

Ss\Ps	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tu	Yb	Lu
1	5.88	6.18	6.44	6.50	6.64	6.68	6.73	6.78	6.88	6.97	7.09	7.22	7.42	7.61
2	2.02	2.09	2.18	2.22	2.30	2.31	2.16	2.07	2.07	2.00	2.01	2.02	2.03	2.05
3	2.61	2.80	2.84	2.88	2.99	3.09	3.08	3.11	3.27	3.31	3.35	3.51	3.64	3.66
4	3.44	3.69	3.63	3.66	3.82	3.83	3.79	3.74	3.75	3.67	3.64	3.62	3.64	3.59
5	1.93	2.23	2.17	2.17	2.26	2.12	2.10	2.02	1.98	1.94	1.94	1.94	1.96	1.96
6	1.71	1.87	1.76	1.83	1.93	1.89	1.85	1.66	1.64	1.72	1.70	1.68	1.69	1.61
7	1.59	2.47	1.92	1.81	1.89	1.78	1.73	1.51	1.63	1.62	1.67	1.75	1.68	1.67
8	2.22	2.22	2.16	2.20	2.23	2.15	2.08	2.06	2.10	1.98	1.96	2.02	2.09	1.83
9	2.43	2.31	2.33	2.35	2.40	2.41	2.39	2.35	2.37	2.32	2.29	2.33	2.33	2.25
10	11.14	12.19	11.94	11.43	11.78	11.70	11.74	11.95	12.09	11.30	11.49	11.27	9.16	10.88
11	12.21	12.06	12.21	12.21	12.12	12.98	11.93	11.70	11.57	11.49	11.30	11.32	12.50	10.84
12	6.18	8.00	7.54	7.40	7.44	11.06	7.02	7.10	6.70	6.41	6.20	6.14	10.11	6.00
13	6.62	7.26	7.22	7.24	7.51	7.61	7.29	7.23	7.15	6.98	6.70	6.96	7.60	6.01
14	10.11	13.22	11.89	11.60	11.72	11.44	11.66	11.52	11.55	11.23	11.15	11.50	11.69	10.33
15	15.46	15.94	16.36	16.56	17.10	17.33	17.35	17.92	18.28	18.29	18.83	19.29	19.48	19.80
16	7.04	7.48	7.84	8.06	8.28	8.38	8.13	8.18	8.31	8.42	8.59	8.75	8.93	9.09
17	4.52	4.62	4.70	4.72	4.85	4.91	4.88	4.94	4.97	4.98	5.01	5.02	5.03	5.04
18	1.89	2.07	2.17	2.25	2.31	2.29	2.19	2.17	2.19	2.20	2.24	2.28	2.27	2.37
19	1.74	1.98	2.01	2.09	2.06	2.12	2.00	1.98	1.95	1.95	2.06	2.06	2.13	2.12
20	4.96	5.09	5.27	5.30	5.59	5.87	5.90	6.02	6.03	6.05	5.99	6.09	6.18	6.23
21	16.26	16.76	17.31	17.68	18.38	18.62	18.77	19.05	19.69	20.29	20.68	20.96	21.12	21.51
22	10.86	11.20	11.30	11.50	11.60	9.53	11.32	11.55	11.60	11.83	11.80	11.74	11.06	12.07
23	6.47	6.78	6.92	6.61	7.10	7.29	7.04	7.22	7.77	7.38	7.49	7.74	7.45	7.83
24	19.05	20.05	21.10	21.60	22.30	22.40	22.50	22.70	22.80	22.80	22.70	22.70	22.60	22.40
25	16.42	16.79	17.17	17.54	17.97	18.26	18.21	18.64	19.05	19.30	19.61	20.08	20.25	20.56
26	16.00	16.69	17.36	17.67	18.19	18.31	18.13	18.31	18.21	18.13	17.99	17.83	17.85	17.75
27	15.55	15.70	16.05	16.28	16.88	17.10	17.27	17.27	17.42	17.38	17.40	17.48	17.78	17.81
28	13.46	14.11	14.61	14.86	15.28	15.35	15.22	15.32	15.30	15.32	15.42	15.59	15.88	15.88
29	5.72	6.00	6.18	6.28	6.57	6.76	6.71	7.16	7.23	7.30	7.42	7.54	7.65	7.60
30	6.37	6.66	6.78	6.99	7.05	5.12	7.02	7.19	7.27	7.36	7.51	7.65	6.64	7.62
31	10.37	10.83	11.07	11.25	11.51	11.49	11.54	11.58	11.71	11.85	12.00	12.20	12.37	12.47
32	7.80	8.30	8.53	8.64	8.92	8.92	8.76	8.87	9.00	9.07	9.25	9.40	9.60	9.72

Table 5. Initial matrix for principal component analysis (continued).

Ss\Ps	La	Ce	Pr	Nd	Sm	Eu	Gd	Tb	Dy	Ho	Er	Tu	Yb	Lu
33	2.55	2.69	2.78	2.89	2.91	2.93	2.79	2.82	2.92	2.99	3.00	3.06	3.13	3.15
34	2.02	2.06	2.07	2.10	2.13	2.12	2.06	2.06	2.05	2.07	2.08	2.08	2.08	2.09
35	2.36	2.39	2.44	2.48	2.55	2.50	2.49	2.52	2.56	2.58	2.60	2.61	2.65	2.68
36	2.60	2.76	2.85	2.87	2.88	2.95	2.89	2.90	3.01	3.02	3.16	3.19	3.23	3.27
37	3.54	3.74	3.85	3.88	4.06	4.07	4.06	4.15	4.22	4.22	4.26	4.36	4.43	4.45
38	5.30	5.70	5.80	5.90	6.10	6.10	6.10	6.20	6.30	6.30	6.40	6.50	6.60	6.70
39	3.20	3.40	3.60	3.60	4.00	4.20	4.00	4.20	4.10	4.20	4.20	4.30	4.40	4.40
40	11.23	11.75	11.91	12.36	13.08	13.49	13.73	14.10	14.67	14.84	15.15	15.39	15.94	15.84
41	22.94	23.77	23.71	23.78	23.66	23.28	23.45	23.55	23.74	23.55	23.40	23.23	22.97	22.94
42	9.86	9.61	10.00	10.26	10.31	10.26	10.32	10.29	10.67	10.63	10.64	10.77	10.91	10.91
43	11.62	11.72	11.79	11.86	11.90	11.93	11.86	11.82	11.93	11.90	11.93	11.95	11.96	11.89
44	11.05	11.08	10.86	11.12	11.27	11.28	11.54	11.64	11.75	11.80	11.95	12.12	12.03	11.88
45	11.26	11.30	11.43	11.62	11.72	11.88	11.75	11.85	12.09	11.97	11.84	12.34	12.11	12.13
46	9.49	10.21	10.39	10.43	10.52	10.60	10.74	10.77	10.77	11.51	11.57	11.87	11.90	11.92
47	10.98	11.58	12.00	12.27	12.68	12.69	12.58	12.53	12.54	12.53	12.45	12.40	12.46	12.45
48	12.07	12.81	13.28	13.41	13.77	13.83	13.67	13.72	13.55	13.75	13.75	14.00	14.11	14.15
49	6.81	7.49	7.71	7.94	8.30	8.48	8.61	8.38	9.06	9.24	9.37	9.59	9.82	10.00
50	8.72	8.90	9.40	9.54	9.84	10.02	9.95	10.46	10.56	10.76	10.90	11.09	11.16	11.12
51	15.40	17.30	17.50	17.68	19.62	19.66	19.90	19.86	20.44	20.42	20.78	21.16	20.30	20.20
52	9.88	10.29	10.50	10.59	10.83	10.65	10.57	11.00	10.93	11.11	11.41	11.24	11.05	11.12
53	9.35	9.60	9.91	10.29	11.05	11.27	11.12	11.56	11.90	11.83	12.00	12.12	11.97	12.01
54	7.21	7.36	7.83	7.93	8.44	8.56	8.12	8.81	9.03	9.10	9.30	9.52	9.66	9.40
55	10.13	10.58	10.13	11.59	12.56	12.44	12.27	12.18	12.89	12.11	13.39	13.34	13.92	13.37
56	15.60	14.00	16.51	17.56	17.12	17.80	17.35	19.10	17.45	19.02	20.12	20.96	17.65	17.68
57	20.15	20.27	21.00	21.47	22.39	22.40	21.80	21.52	21.80	21.85	21.62	21.71	22.62	22.70
58	11.06	13.29	11.48	11.04	12.02	18.38	13.37	12.42	12.04	11.44	11.42	10.78	12.76	13.36
59	10.01	10.02	10.43	10.91	11.80	11.35	11.80	11.46	11.80	12.04	12.01	12.42	12.90	13.03
60	9.89	11.20	11.62	12.01	14.01	16.28	12.68	13.89	13.60	13.61	13.50	13.40	13.30	13.35
61	11.20	11.08	11.31	11.20	12.60	12.05	12.00	12.42	13.20	12.50	12.89	13.15	10.65	12.43
62	12.61	12.94	12.82	12.70	12.84	11.90	12.05	12.90	13.10	13.00	13.01	13.02	11.71	11.79
63	4.84	5.22	5.44	5.57	5.74	5.70	5.59	5.53	5.49	5.43	5.39	5.43	5.45	5.46

The initial data set contained 14 lanthanides and 63 organic ligands. The matrix used is given in Table 5.

All missing values in this matrix was predicted with program CODESSA, as describe above, thus providing a complete matrix 63 x 14 suitable for a general principal component analysis of ligand properties. PCA was carried out using the SIMCA-P 9.0 program package with these 63 ligands as variables, each having 14 data point for the 14 metals (objects). Table 6 lists the percentage of variance covered by different components.

Table 6. Variance covered by the 10 components

component	% variance	tot. variance
1	64.3	64.3
2	15.2	79.5
3	7.2	86.7
4	4.6	91.3
5	2.7	94.0
6	2.2	96.2
7	1.3	97.5
8	0.9	98.4
9	0.7	99.1
10	0.3	99.4

The first principal component is responsible for 64.3% of variance, the second for 15.2%; the third for 7.2% and the fourth for 4.6%; these four components thus account for 91.3% of the total variance. The next three make up 1.3% - 2.7% each, so the total variance covered by seven components is 97.5 %. Thus it appears that four major orthogonal components determine how different functional groups of ligand take part in complex formation; probably some less important interactions are described by minor components.

5. Results and discussion

In Tables 5 and 6, the results of the QSPR treatment are summarized for the series of the organic ligands and the lanthanides, respectively. In the first column of Table 7, the ligands are listed in order given in Table 1, and the second column in both Tables 5 and 6 shows the number of experimental data points in the treatment, respectively. The coefficients of QSPR equations and the notations of the respective descriptors are given in the next columns, together with the *t*-test values. The natural value of the regression coefficient itself cannot be treated as an indicator of the importance of the descriptor in an equation as the absolute numeric values of the descriptors vary in a large range. Thus, the *t*-test value for each descriptor has been used instead for the purpose. The last three columns of Tables 5 and 6 show the statistical parameters of the QSPR equations: the squared correlation coefficients (R^2), the squared standard deviation (s^2), and the squared cross-validated correlation coefficients (R^2_{cv}). Most of the developed QSPR equations for ligands have satisfactory correlation coefficient; 58 out of 63 models for ligands have R^2 higher than 0.90 and only 1 model has $R^2 < 0.85$. In the case of QSPR equations for metals, 10 out of 14 models for metal have R^2 higher than 0.87 and no models have $R^2 < 0.84$.

The notations of the descriptors that were used in equations of Table 7 are listed in Table 4. The notations of the descriptors that were used in equations of Table 8 are presented in Table 9. All descriptors from Table 4 are different properties of lanthanides. The descriptors in Table 9 can be divided into six groups. The largest groups include the hydrogen bonding descriptors (6 descriptors), topological indices of the organic ligands (5), general electronic properties (5 descriptors) and bonding interactions (5 descriptors). In addition, descriptors reflecting the geometry and constitution (3) of ligands and partial surface areas (3) did appear in the QSPR models. The hydrogen bonding descriptors were involved 11 times, descriptors describing geometry and constitution of the ligands 10 times, the partial surface areas 10 times, electronic properties 9 times, descriptors describing the topology 8 times and bonding interaction descriptors 8 times.

Table 7. The QSPR models ($a_0+a_1d_1+a_2d_2+a_3d_3+a_4d_4$) on complex stability constants for organic ligands [a]

ligand	a_0	n [b]	a_1	d_1	*t*-test	a_2	d_2	*t*-test	a_3	d_3	*t*-test	a_4	d_4	*t*-test	R^2	s^2	R^2_{cv}
1	15.7	14	-3.26	F08	-25.7	-5.81E-07	F18	-3.17	7.80	F14	3.15	-1.48E-07	F04	-2.81	0.991	2.96E-03	0.948
2	3.36	13	-1.41E-06	F18	-9.06	9.36E-07	F02	3.89	9.72E-06	F17	-2.87	-0.0152	F23	-6.81	0.919	1.40E-03	0.746
3	3.93	14	-0.0212	F13	-1.71	-9.99E-07	F18	-4.76	-2.25E-06	F01	-3.25	7.19E-03	F23	-2.66	0.839	2.64E-03	0.607
4	-0.911	14	0.0659	F07	24.9	6.28E-08	F04	1.69	6.51	F14	3.67				0.988	1.54E-03	0.963
5	-3.82	10	1.16E-06	F18	-5.01	4.40E-02	F21	7.22	2.35E-06	F02	5.56	-4.07E-07	F04	-5.81	0.948	1.29E-03	0.744
6	-4.57	10	-1.99E-06	F18	-4.92	4.46E-02	F20	4.49	1.62E-06	F02	2.98	2.24E-04	F19	2.49	0.898	2.21E-03	0.675
7	-2.78	10	-1.03E-06	F18	-3.49	3.29E-02	F21	4.23	1.89E-06	F02	3.54	-3.07E-07	F04	-3.44	0.871	2.08E-03	0.565
8	1.33	9	- 7.82E-05	F10	5.43	-1.17E-05	F17	-7.09	-3.29E-01	F12	-3.58	1.52E-02	F20	8.33	0.988	2.12E-04	0.964
9	2.20	9	- 7.32E-06	F17	-14.82	-4.14	F14	-15.8	-2.85E-08	F18	-7.68	8.93E-08	F04	6.29	0.991	2.83E-05	0.953
10	16.5	8	- 5.51E-07	F02	-2.15	-5.87E-07	F04	-7.86	-0.117	F23	-13.3	-9.44E-07	F16	-2.58	0.992	2.91E-03	0.951
11	15.4	8	- 0.379	F13	-16.4	-1..93E-02	F23	-1.59	- 1.35E-06	F18	-3.81				0.987	5.39E-03	0.927
12	-5.54	8	7.34E-01	F09	24.1	-4.30E-06	F18	-14.7	7.71E-06	F02	17.3	-2.30E-06	F04	-24.54	0.999	1.10E-03	0.986
13	10.3	8	-4.20E-05	F17	-7.10	-4.15E-07	F04	-2.54	-36.2	F14	-3.34	2.05E-06	F18	- 2.81	0.988	6.59E-03	0.862
14	18.2	8	- 69.8	F14	-2.57	1.50E-06	F04	4.79							0.883	6.57E-02	0.683
15	-0.645	14	0.298	F07	34.9	-1.96E-07	F04	-1.51							0.991	1.91E-02	0.985
16	18.8	14	-3.64	F08	-15.4	-1.78E-06	F18	- 5.06	- 2.05E-07	F04	- 2.00				0.971	1.14E-02	0.877
17	9.50	14	-3.53E-02	F13	- 5.42	- 6.94	F14	- 6.79	- 3.96E-02	F20	- 22.3				0.990	3.44E-04	0.978
18	- 2.78	14	-1.07E-06	F18	- 6.36	1.24E-06	F02	7.03	- 1.61E-07	F04	- 3.56	-1.78E-06	F01	- 3.28	0.902	1.95E-03	0.908
19	2.78	14	-7.43E-07	F18	- 3.15	8.43E-07	F02	3.64	-1.77E-07	F23	- 3.02	3.60E-03	F01	2.25	0.809[1*]	3.17E-03	0.937
20	15.9	14	-1.69E-05	F17	-3.12	-2.05E-02	F23	-6.05	-9.34E-02	F21	-19.2	4.09E-03	F11	3.18	0.991	2.51E-03	0.982
21	45.9	13	5.19E-05	F17	4.66	2.61E-02	F23	3.83	-2.59E-01	F20	-21.5				0.997	1.11E-02	0.989
22	22.5	6	-6.70	F12	-24.8	44.0	F14	9.26	-2.10E-06	F16	-11.0				0.998	3.96E-04	0.971
23	2.81	9	3.74E-02	F06	12.4	-1.19E-01	F23	-5.40	-2.32E-06	F16	- 5.26	76.3	F14	4.35	0.984	3.84E-03	0.925
24	43.0	14	- 5.56E-06	F18	- 4.81	- 59.1	F14	- 3.93	-0.173	F20	- 8.73				0.935	0.112	0.873
25	1.82	14	2.95E-01	F07	41.2	-8.70E-07	F16	-2.76	-3.41E-06	F01	-2.83				0.996	8.76E-03	0.990
26	29.4	14	- 3.16	F08	- 5.37	-3.96E-07	F04	- 7.05	- 5.69E-06	F18	-4.49	- 5.40E-02	F23	-4.39	0.922	4.94E-02	0.617
27	36.5	14	-4.13E-05	F17	-3.67	-20.3	F01	-3.32	-1.57E-01	F21	-14.7				0.980	1.46E-02	0.961
28	28.4	14	- 3.17E-06	F18	- 7.44	- 3.53E-07	F04	- 3.05	-4.26	F08	-15.7	-8.97E-03	F23	- 1.38	0.979	1.36E-02	0.833
29	20.1	14	- 1.40E-07	F04	- 2.14	- 0.106	F21	-32.6							0.990	4.94E-03	0.953
30	20.8	9	-6.73E-07	F18	-2.79	23.9	F14	4.01	-8.57	F12	-16.6				0.991	1.73E-03	0.975
31	23.2	14	- 4.12	F08	-21.0	- 1.41E-06	F18	- 4.82	- 1.97E-07	F04	- 2.31				0.982	7.84E-03	0.964
32	19.6	14	- 3.62	F08	- 20.6	- 2.57E-07	F04	-3.31	- 1.34E-06	F16	- 5.11				0.980	6.64E-03	0.93
33	0.478	14	-6.38E-07	F18	- 4.84	2.42	F15	12.3	18.3	F14	11	- 9.16E-07	F01	-2.14	0.968	1.27E-03	0.921
34	0.275	13	- 5.82E-07	F18	- 7.96	- 5.77E-05	F04	- 4.56	6.15E-07	F02	5.23	0.0122	F21	6.00	0.906[2*]	1.20E-04	0.635

Table 7. The QSPR models ($a_0+a_1d_1+a_2d_2+a_3d_3+a_4d_4$) on complex stability constants for organic ligands [a] (continued)

ligand	a_0	n [b]	a_1	d_1	*t*-test	a_2	d_2	*t*-test	a_3	d_3	*t*-test	a_4	d_4	*t*-test	R^2	s^2	R^2_{cv}
35	1.19	14	6.70E-07	F02	14.9	3.79E-07	F03	17.3	-1.19E-02	F09	-9.22	1.70	F14	3.11	0.990	1.24E-04	0.966
36	1.88	14	2.01	F15	3.56	-7.85E-8	F04	- 2.25	- 0.0112	F20	- 1.184	16.9	F14	4.48	0.974	1.31E-03	0.897
37	9.53	14	- 7.01E-08	F04	- 2.54	- 0.0471	F20	-27.5	- 3.23E-05	F10	- 2.46				0.990	8.70E-04	0.960
38	14.1	14	-2.69	F08	-35.8	-2.13E-07	F04	-6.54	-8.53E-07	F18	-7.62				0.993	1.15E-03	0.985
39	15.6	13	- 0.128	F13	- 5.34	-17.9	F14	- 4.85	- 0.0983	F20	- 15.4				0.977	4.47E-03	0.962
40	-5.91	9	8.32E-07	F03	8.24	-2.32E-07	F04	-5.17	80.8	F05	75.3				0.999	1.70E-03	0.999
41	29.8	12	- 0.0587	F07	- 4.08	- 7.40E-07	F04	- 4.31	- 1.20E-06	F18	- 1.66	6.53E-04	F19	2.88	0.900	0.0162	0.734
42	8.33	14	7.93E-04	F19	6.67	24.6	F14	5.65	-1.90E-06	F18	-7.64	1.18E-06	F02	2.53	0.975	5.56E-03	0.938
43	10.8	14	1.12	F15	8.35	-5.02E-08	F04	-2.04	-6.43E-07	F18	-7.47	15.34	F14	4.48	0.952	6.47E-04	0.791
44	6.93	14	1.53E-06	F02	3.23	1.24E-06	F03	7.29	4.05E-04	F19	3.27				0.924	0.0101	0.865
45	6.23	14	3.71E-06	F01	3.59	-2.17E-03	F11	-1.46	0.0254	F06	10.2				0.970	6.64E-03	0.947
46	18.7	14	5.93E-05	F17	4.17	0.0423	F23	4.74	-0.0666	F20	-4.31	-4.76E-07	F04	-3.57	0.978	0.0164	0.929
47	11.7	14	-4.91E-06	F18	-10.3	-5.33E-07	F04	-3.93	4.20E-06	F02	8.21	-0.0317	F23	-4.38	0.949	0.0169	0.516
48	12.2	14	-3.99E-06	F18	-9.40	4.69E-06	F02	6.02	-7.27E-07	F04	-5.27	1.67E-05	F17	1.52	0.959	0.0178	0.661
49	26.9	14	-5.80E-07	F16	-1.13	-0.169	F20	-20.2							0.976	2.44E-02	0.959
50	21.9	14	2.96E-04	F19	2.51	7.19E-07	F03	3.71	-1.27E-01	F20	-15.8				0.994	5.22E-03	0.979
51	-20.4	8	14.2	F15	32.2	3.71E-05	F01	42.9	3.13E-06	F03	27.6	-1.16E-06	F04	-19.5	0.999	2.94E-03	0.953
52	22.3	14	-1.93E-07	F04	-2.33	-7.70E-02	F20	-10.2	3.16E-05	F17	4.80	-2.52E-06	F02	-4.62	0.978	5.57E-03	0.958
53	31.3	14	-1.73E-06	F16	-4.67	-0.161	F21	-30.6	-28.5	F14	-5.77				0.991	0.0111	0.983
54	-2.87	14	1.36E-01	F07	15.6	-3.25E-02	F09	-2.57	1.60E-06	F03	6.20				0.988	9.84E-03	0.978
55	21.9	13	-8.10E-04	F10	-2.53	4.18E-06	F02	1.32	-4.13E-02	F09	-0.745	-9.50E-02	F21	-1.71	0.924	1.56E-01	0.859
56	14.1	13	1.12E-02	F19	6.88	1.67E-01	F23	4.27	-1.92E-05	F02	-4.66	4.43E-01	F09	4.02	0.878	3.87E-01	0.748
57	21.3	10	5.51	F05	2.29	0.442	F22	3.70	-7.91E-06	F18	-14.6				0.981	0.0123	0.955
58	-46.0	6	29.5	F12	3.35	-170	F14	-8.81	-2.81E-06	F04	-3.94	2.14E-05	F02	6.29	0.997	0.0178	0.999
59	28.0	6	-6.08	F08	-11.9	-3.12E-06	F18	-3.55							0.980	0.0346	0.956
60	18.0	10	2.15E-03	F19	3.62	-1.06E-05	F18	-4.34	-1.16E-06	F04	-1.63				0.923 [e]	0.166	0.759
61	12.6	7	2.06E-03	F19	6.29	-7.16E-04	F10	-6.47	-0.0751	F23	-2.16				0.966	0.0369	0.823
62	18.8	7	-125	F09	-51.5	-0.0918	F22	-105	-0.0386	F23	-29.9	-1.12E-06	F16	-25.9	0.999	4.13E-05	0.998
63	-4.71	13	-3.95E-06	F18	-7.52	-5.09E-07	F04	-5.96	5.00E-04	F19	4.03	4.21E-06	F02	4.97	0.934[3*]	5.40E-03	0.310

[a] the numeration of ligands corresponds to Table 1.
[b] number of data points in the set.
[1*] R^2 = 0.809 was obtained using five descriptors (2.78 -7.43E-07 x F18 + 8.43E-07 x F02 – 1.77E-07 x F04 + 3.60E-03 x F23 – 1.47E-06 x F01)
[2*] R^2 = 0.843 was obtained using five descriptors (0.275 - 5.82E-07 x F18 - 5.77E-05 x F04 + 6.15E-07 x F02 + 1.22E-02 x F21 + 5.29E-06 x F17)
[3*] R^2 = 0.934 was obtained using five descriptors (-4.71 - 3.95E-06 x F18 - 5.09E-07 x F04 + 5.00E-04 x F19 + 4.21E-06 x F02 + 6.75E-02 x F02)

Table 8. The QSPR models ($a_0+a_1d_1+a_2d_2+a_3d_3+a_4d_4$) on complex stability constants for lanthanides

metal	a_0	n [a]	a_1	d_1	*t*-test	a_2	d_2	*t*-test	a_3	d_3	*t*-test	a_4	d_4	*t*-test	R^2	s^2	R^2_{cv}
La	-6.28	76	0.00159	G01	10.9	133	H02	7.22	5.90	P02	4.74	1.26	E02	3.07	0.846	4.10	0.825
Ce	11.0	53	0.00262	G01	14.8	40.0	G02	3.80	175	H04	4.83	-5.36	T04	-5.64	0.889	2.65	0.863
Pr	8.16	58	0.101	T05	13.5	45.1	G02	5.00	90.9	P03	6.68	-3.49	B02	-3.29	0.910	2.80	0.893
Nd	- 1.82	74	0.194	G03	8.63	16.3	H03	10.5	8.47	P02	6.68	-0.909	B03	- 3.18	0.879	3.66	0.857
Sm	15.38	70	0.00157	G01	9.22	-4.27	B04	-3.07	131	H02	6.38	3.99	P02	3.27	0.863	4.43	0.843
Eu	-4.65	57	0.00210	G01	10.6	18.9	H01	3.53	5.19	P01	3.58	1.97	E02	3.98	0.897	4.03	0.878
Gd	-9.51	70	0.242	G03	9.98	1.17	E05	2.45	8.40	P02	5.96	394	H05	8.12	0.864	4.73	0.841
Tb	-6.59	58	90.1	H06	7.89	3.79	T02	7.98	3.88	P01	2.58	1.73	E05	3.28	0.873	4.27	0.846
Dy	-0.479	63	0.208	G03	7.77	17.2	H03	9.44	9.68	P02	6.46	-1.33	B03	-3.57	0.881	4.30	0.855
Ho	-28.4	57	5.91	T01	14.3	2.16	E05	3.93	11.8	T03	6.91	-458	E04	-3.48	0.897	3.75	0.880
Er	32.5	63	0.203	G03	8.35	17.4	H03	9.43	10.3	P01	6.76	-10.8	B01	-4.29	0.883	4.05	0.859
Tm	-35.6	45	6.05	T01	11.46	225	H02	8.35	-452	E03	-2.28	1.85	B01	2.86	0.910	3.91	0.888
Yb	23.9	59	4.80	T02	6.77	2.74	E01	6.05	1191	P03	7.46	-699	E03	-4.01	0.880	4.99	0.860
Lu	0.0203	58	4.24	T02	7.94	158	H02	4.80	1.52	B05	4.78	-1.46	B03	-2.93	0.860	4.97	0.835

[a] number of data points in the set.

Table 9. Descriptors used in the QSPR models for lanthanides

	Descriptor name
	Bonding Interactions
B01	Max coulombic interaction for bond H-C
B02	Min coulombic interaction for bond H-C
B03	Max coulombic interaction for bond C-C
B04	Min coulombic interaction for bond C-C
B05	Number of double bonds
	Partial Surface Areas
P01	Square root of Charged Surface Area (MOPAC PC) for atom C
P02	Square root of Charged Surface Area for atom C
P03	Square root of Partial Surface Area for atom O
	Geometrical/Constitutional
G01	Gravitation index (all atoms' pairs)
G02	Relative number of N atoms
G03	Shadow plane YZ
	Topological
T01	Average Complementary Information content (order 1)
T02	Average Complementary Information content (order 2)
T03	Average Information content (order 0)
T04	Average Information content (order 1)
T05	Complementary Information content (order 2)
	Electronic Properties
E01	HOMO-1 energy
E02	LUMO energy
E03	Max 1-electron react. index for atom O
E04	Min 1-electron react. index for atom O
E05	Tot hybridization comp. of the molecular dipole
	Hydrogen Bonding
H01	HA dependent HDSA-1/TMSA (Zefirov PC)
H02	HA dependent HDSA-2/TMSA (Zefirov PC)
H03	HACA-2/SQRT(TMSA) (MOPAC PC)
H04	H-donors FCPSA (version 2)
H05	HACA-2/TMSA (MOPAC PC)
H06	HA dependent HDCA-2/SQRT(TMSA) (Zefirov PC)

This distribution of molecular descriptors in QSPR models indicates that the bidentate complex formation with the lanthanide ions is predominantly determined by the hydrogen-bonding related properties, geometrical and even topological structure of the ligands. The descriptors reflecting the charge distribution in the ligands and the related electrostatic interactions have smaller contributions.

In the case of the correlations with the lanthanide (metal) descriptors, the most important contribution is given by the successive ionization potentials of the metals. Those descriptors appear altogether 65 times, of which 32 cases involve the fourth ionization potential of the metal, i.e. the ionization potential of the Ln^{3+} ion. Another group of the descriptors of substantial importance includes the heats of vaporization (30 times) and fusion (10 times) of the metals. In principle, these descriptors (physical properties) depend on the London forces between the metal atoms and may thus reflect similar non-covalent interactions in the complexes.

The descriptors in each model are given in Tables 5 and 6 in order of the (absolute) *t*-test values. In this way, the most significant descriptors for each model are in the d1 column (Tables 5 and 6). If two or several metals or organic ligands have the same most significant descriptors, it follows the complex stability for those metals or organic ligands should depend predominantly on the same chemical parameter or effect.

Notably, the overall fitness of the QSPR models with metals as variables is excellent (Fig. 1). Thus, the prediction of $\log K_1$ in cases when the QSPR equation is known for a given organic ligand would be very reliable. On the other hand, the predictions from the QSPR models with ligands as variables are less precise (Figure 2). This is, however, not unexpected bearing in mind large structural variability of the organic ligands used.

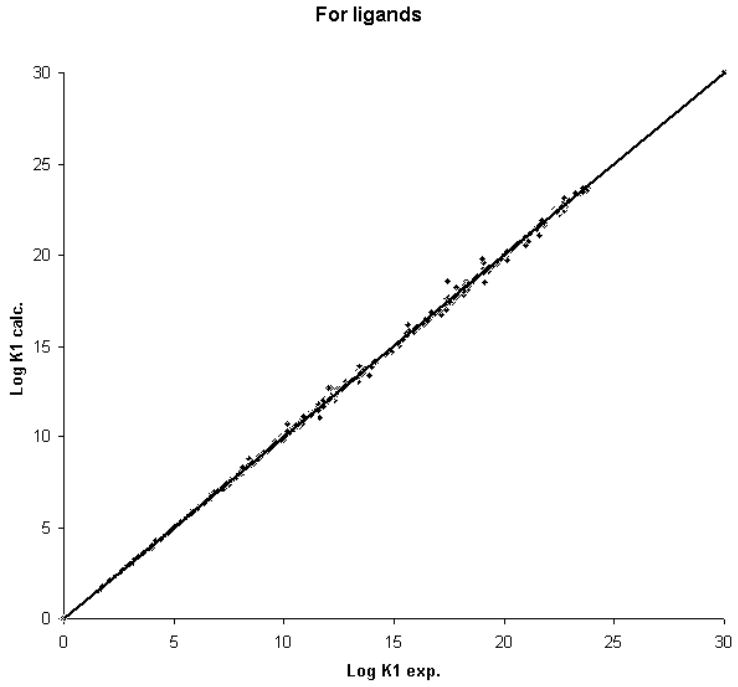


Figure 1. Correlation between the experimental and predicted data from QSPR models for single ligands (metals variable). $R^2 = 0.999$

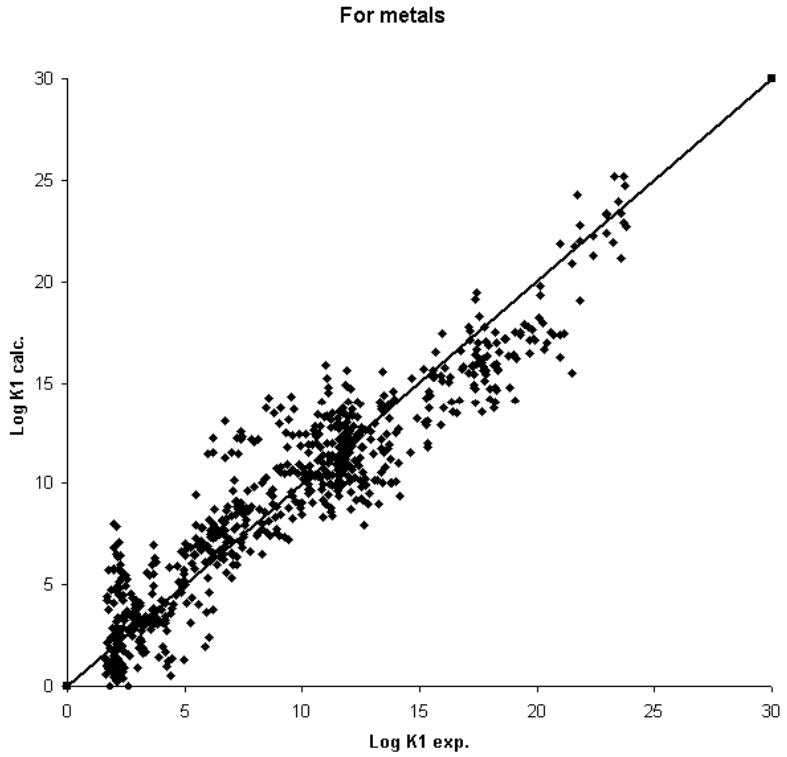


Figure 2. Correlation between the experimental and predicted data from QSPR models for single metals (ligands variable). $R^2 = 0.878$

A significant correlation was found between the predictions of unknown $logK_1$ values, proceeding from the QSPR equations for the ligands and for the metals, respectively (R^2 = 0.6, Figure 3).

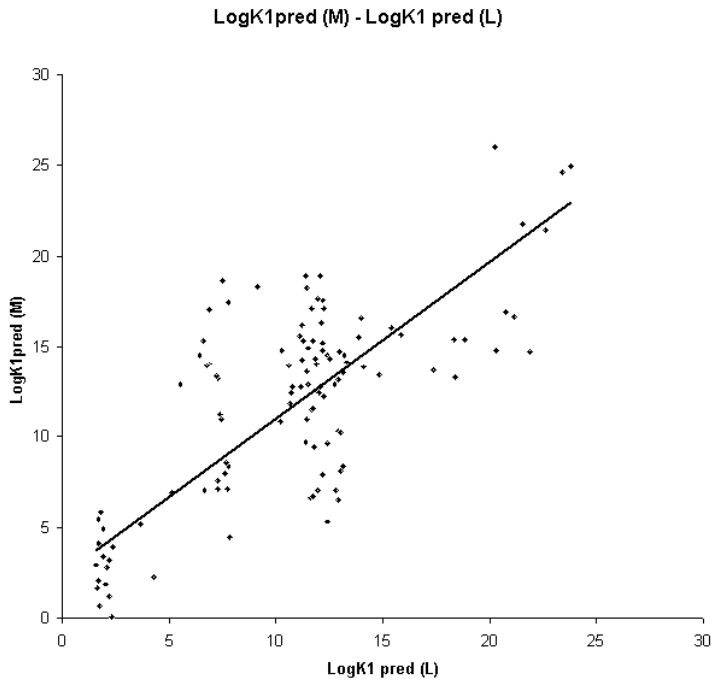


Figure 3. Correlation between the predicted data for unknown complexes from the QSPR models for metals and ligands as variable, respectively. R^2 = 0.597

PCA grouped ligand by they property. Table 10 presents the loadings of the ligands in the first fourth principal components. The total variance in each scale covered by the first fourth components is also given in the final column. (I) for 43 of the 63 ligands, the four components describe 90% or more of their variance; (II) another twelve ligands are described rather well with 80-89% of variance; (III) a further five ligands are described with 70-79% of variance; and just 3 ligands poorly described with only 60-69% of their variance accounted for by the four main components.

Figure 4 shows the loadings of second PCA component plotted against the loadings of the first component. Labels are defined as in Table 1.

Table 10. Loadings of the four principal components

Ligands number	PC1	PC2	PC3	PC4	R
1	0.153	-0.023	-0.045	0.095	0.993
2	-0.042	0.281	0.031	-0.010	0.837
3	0.034	0.271	0.162	-0.115	0.910
4	0.151	-0.043	-0.071	0.073	0.985
5	-0.077	0.228	0.140	0.213	0.959
6	-0.074	0.243	0.044	0.073	0.813
7	-0.073	0.101	0.089	0.391	0.788
8	-0.124	0.136	-0.001	-0.019	0.800
9	-0.082	0.115	-0.077	-0.380	0.844
10	-0.066	0.062	0.340	-0.144	0.799
11	-0.079	0.212	-0.214	-0.057	0.899
12	-0.005	0.245	-0.256	0.036	0.880
13	-0.026	0.252	0.003	-0.057	0.646
14	-0.034	0.161	0.174	0.257	0.624
15	0.154	-0.053	-0.040	0.031	0.992
16	0.153	0.037	-0.033	0.086	0.991
17	0.155	0.019	0.018	-0.068	0.989
18	0.125	0.145	0.050	0.151	0.912
19	0.101	0.177	-0.031	0.266	0.927
20	0.151	0.005	-0.015	-0.113	0.963
21	0.154	-0.053	-0.028	0.019	0.994
22	0.058	-0.182	0.276	0.176	0.886
23	0.143	-0.028	0.031	-0.047	0.847
24	0.141	0.091	0.110	-0.112	0.971
25	0.154	-0.052	-0.045	0.027	0.996
26	0.117	0.161	0.138	-0.156	0.957
27	0.153	0.017	-0.018	-0.095	0.974
28	0.151	0.078	0.008	0.020	0.987
29	0.154	-0.038	-0.010	-0.031	0.977
30	0.072	-0.186	0.272	0.127	0.923
31	0.155	-0.005	-0.023	0.079	0.998
32	0.154	0.017	-0.029	0.109	0.996
33	0.148	0.026	-0.044	0.135	0.954
34	0.059	0.235	0.024	0.142	0.731
35	0.153	-0.020	-0.019	0.066	0.971
36	0.150	-0.026	-0.047	0.130	0.975
37	0.156	-0.001	-0.009	0.026	0.991
38	0.156	0.006	0.003	0.067	0.994
39	0.152	0.047	-0.024	-0.073	0.979
40	0.154	-0.045	-0.039	-0.014	0.981
41	-0.047	0.121	0.375	0.053	0.877
42	0.149	-0.048	-0.049	-0.030	0.939
43	0.139	0.117	0.040	-0.012	0.921
44	0.141	-0.095	-0.017	-0.055	0.904
45	0.148	-0.003	-0.009	-0.083	0.912
46	0.148	-0.047	-0.024	0.135	0.962
47	0.124	0.167	0.114	-0.084	0.968
48	0.145	0.106	0.028	0.055	0.975
49	0.154	-0.023	-0.026	0.056	0.983
50	0.155	-0.038	-0.009	-0.027	0.988
51	0.147	0.047	0.094	-0.082	0.958
52	0.147	-0.014	0.119	0.016	0.936
53	0.153	-0.002	0.028	-0.112	0.989
54	0.153	-0.025	-0.020	-0.030	0.964
55	0.147	0.029	-0.043	-0.026	0.895
56	0.119	-0.038	0.095	-0.188	0.727
57	0.131	0.131	-0.082	-0.014	0.895
58	0.011	0.216	-0.204	-0.091	0.664
59	0.152	-0.009	-0.053	0.016	0.947
60	0.107	0.191	0.004	-0.189	0.913
61	0.088	-0.032	0.263	-0.265	0.839
62	-0.038	-0.088	0.374	-0.081	0.788
63	0.080	0.249	0.133	-0.010	0.932

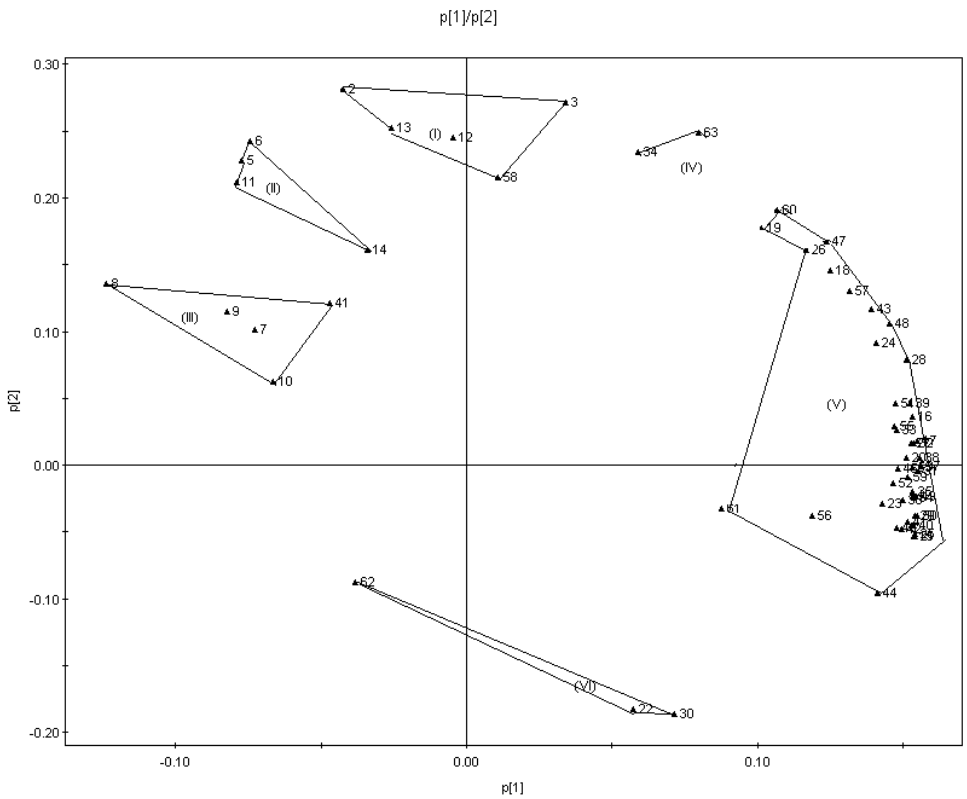


Figure 4 Loadings of the second PCA component plotted versus the loadings of the first component.

(I) The five ligands have small negative or positive loadings for the first component and large positive loadings for the second component. These ligands have or only one functional group, which takes part in complex formation, if more, the metal can interact with only one group at time.

(II) The four ligands have small to medium negative loadings for the first component and medium to large positive loadings for the second component. These ligands have two functional groups, with good complex formation ability. The bonding between functional groups and metal ion is strong.

(III) The five ligands have medium to large negative loadings for the first component and small to medium positive loadings for second component. These ligands have two or more functional groups, but with different ability to give

bonds with metals, or between some of these group is intermolecular bonds and they didn't take part in complex formation.

(IV) The two ligands have medium positive loadings for the first component and medium to large positive loadings for the second component. These ligands have in the main structure heteroatoms, which can participate in complex formation, but with weak bonds.

(V) This large group of ligands has large positive loadings for the first component and medium positive to negative loadings for the second component. These ligands have in structure heteroatoms, which can give strong bond with metal, and two or more functional groups with good complex formation ability.

(VI) The three ligands have medium positive to negative loadings for the first component and large negative loadings for the second component. In these ligands the nitrogen, which can participate in complex formation, is partially blocked.

The scores of the first four principal components for the 14 metals are presented in Table 11.

Table 11. Scores of first four principal components

Notation	PC1	PC2	PC3	PC4
La	-12.963	-5.543	-3.106	-2.049
Ce	-9.652	0.467	1.034	3.106
Pr	-6.249	0.424	0.854	1.368
Nd	-4.150	1.587	0.875	1.436
Sm	-0.684	4.235	2.100	0.168
Eu	0.073	6.896	-2.984	-1.975
Gd	-0.627	1.513	0.297	-0.935
Tb	1.086	-0.333	1.446	-1.941
Dy	2.709	-0.722	1.988	-1.988
Ho	3.578	-1.793	1.188	-0.769
Er	5.007	-2.265	1.316	0.008
Tu	6.493	-2.182	1.043	-0.030
Yb	6.894	0.717	-4.592	1.602
Lu	8.485	-3.002	-1.460	1.999

Figure 5 presents the plot of the scores of the second component against the scores of the first component.

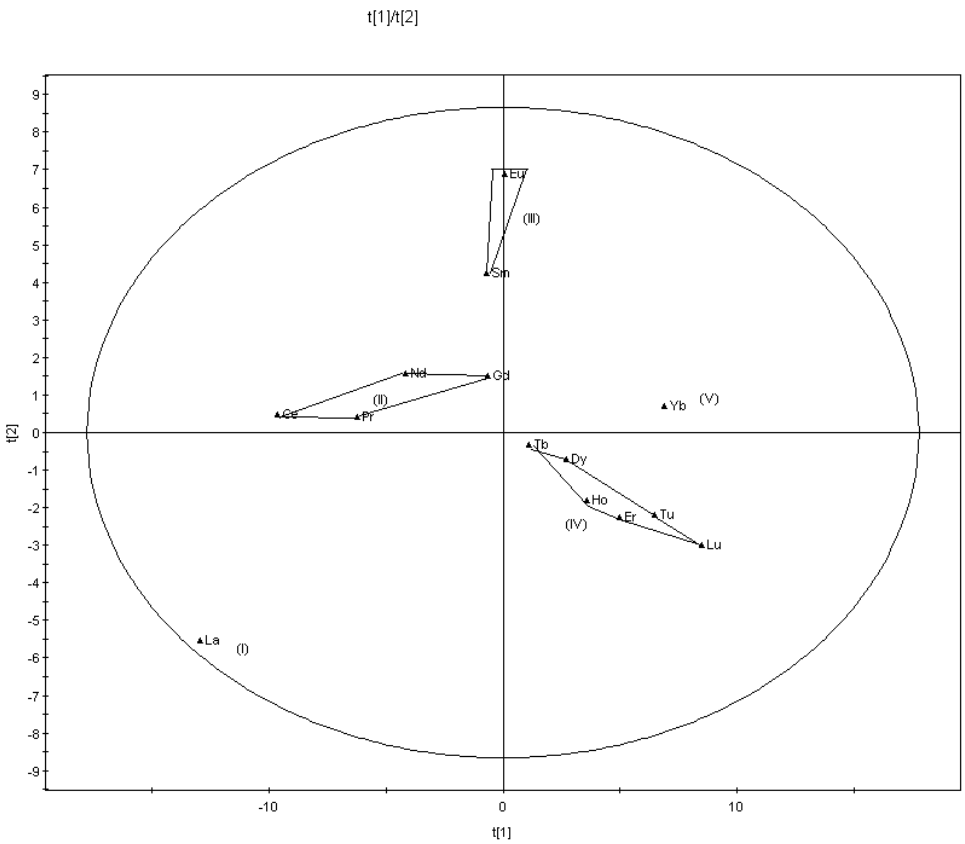


Figure 5. Plot of the scores of second component versus the scores of the first component.

The PCA score grouping directly reflects the properties of metal. The metals with similar physical properties get into one group.

6. Conclusions

We have demonstrated that the theoretical molecular descriptors can been successfully applied in the development of predictive QSPR models for the stability constants of lanthanide (III) – organic complexes. These constants are also well correlated with various physical properties of lanthanide metals used as the descriptors characterizing the metal ions in the series of data for a constant organic ligand. A satisfactory correlation was found between the stability constants for previously unmeasured complexes predicted from the QSPR equations for a constant ligand and a constant metal ion, respectively.

Present results of PCA analysis allow classify and group metal and ligands in to separately groups.

7. Kokkuvõte

Käesoleva töö eesmärgiks oli näidata, et teoreetilisi molekulaardeskriptoreid võib edukalt rakendada lantanoid (III) orgaaniliste komplekside püsivuskonstantide ennustamiseks QSPR mudelidega. Need konstandid korreleeruvad hästi ka lantanoidide mitmete füüsikaliste omadustega, mida kasutati metalliioone iseloomustavate deskriptoritena konstantse orgaanilise ligandiga andmeseeriates. Konstantse ligandi ja konstantse metalliiooni jaoks saadud QSPR võrrandite abil samade komplekside jaoks ennustatud püsivuskonstantide väärtused olid rahuldavas kooskõlas.

Andmete põhikomponentide analüüsi (PCA) tulemused võimaldasid klassifitseerida metalle ja ligande erinevatesse rühmadesse nende kompleksimoodustamise võime järgi.

8. References and notes

1. 1 Alexander, V. Design and Synthesis of Macro cyclic Ligands and Their Complexes of Lanthanides and Actinides. *Chem. Rev.* **1995**, *95*, 273-343.

2. Gun'ko, Y. K.; Edelman, F. T, Lanthanide and Actinides. Annual Survey of their Organometallic Chemistry Covering the Year 1994. *Coord. Chem. Rev.* **1996,** *156*, 1-89.

3. Hogerheide, M. P.; Boersma, J.; van Koten, G. Intramolecular Coordination in Group 3 and Lanthanide Chemistry. An Overview. *Coord. Chem. Rev.* **1996,** *155*, 87-126.

4. Cotton, S, A.; Aspects of the Lanthanide-Carbon σ-Bond. *Coord. Chem. Rev.* **1997,** *160*, 93-127.)

5. Laufer, R. B. Paramagnetic Metal complexes as Water Proton Relaxation Agents for MRI Imaging: Theory and Design. *Chem. Rev.* **1987,** *87*, 901-927.

6. Parke, D.; Dickins, R. S.; Puschmann, H.; Crossland, C.; Howard, J. A. K. Being exited by lanthanide Coordination complexes: Aqua Species, Chirality, Exited-State Chemistry, and Exange Dymanics. *Chem. Rev*. **2002**, *102*, 1977-2010.

7. Merbach, A. E.; Toth, E. Eds. The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging. Willey: New York. 2001.

8. Caravan, P.; Ellison, J. J.; McMurray, T. J.; Lauffer, R. B. Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications. *Chem. Rev*. **1999**, *99*, 2263-2352.

9. Peters, J. A.; Huskens, J.; Raber, D. J. Lanthanide induced shifts and relaxation rate enhancements. *Prog. Nucl. Magn. Reson. Spectrosc*. **1996**, *28*, 283-350.

10. Komiyama, M.; Takeda, N.; Shigekawa, H. *Chem. Commun.* **1999**, 1443.

11. Hemmila, I. K. Applications of Fluorescence in Immunoassays. Willey: New York. 1991

12. Mathis, G. Probing molecular interactions with homogeneous techniques based on rare earth cryptates and fluorescence energy transfer. *Clin. Chem.* **1995**, *41,* 1391-1397.

13. Parker, D. Luminescent lanthanide sensors for pH, pO2 and selected anions. *Coord. Chem. Rev.* **2000**, *205*, 109-130.

14. Aime, S.; Botta, M.; Casellato, U.; Tamburini, S.; Vigato, P. A. NMR Evidence for Interconversion between Two Enantiomeric Forms of Macrocyclic Schiff Base Lanthanide(III) Complexes Through Reversible Ring Contraction and Expansion. *Inorg. Chem*. **1995**, *34*, 5825-5831.

15. Alexander, V. Design and Synthesis of Macrocyclic Ligands and Their Complexes of Lanthanides and Actinides. *Chem. Rev.* **1995**, *95*, 273-342.

16. Bianchi, A.; Calabi, L.; Corona, F.; Fontana, S.; Losi, P.; Maiocchi, A., Paleari, L., Valtancoli, B. *Coord. Chem. Rev.* **2000**, *204*, 309.

17. Toth, E.; Burai, L.; Merbach, A. E. Similarities and differences between the isoelectronic GdIII and EuII complexes with regard to MRI contrast agent applications. *Coord. Chem. Rev.* **2001**, *216*, 363-382.

18. Mody, T. D.; Sessler, J. L. In Supramolecular Technology. Reinhoudt. D. N., Ed.: Wiley: New York, 1999: Chapter 7. pp 245-293.

19. Mody, T. D.; Sessler, J. L. *Porphyrins Pthalocyanines*. **2001,** *5*, 134.

20. C.M. Wai, Preconcentration of Trace Elements by SolventExtraction in Preconcentration Techniques for Trace Elements, CRC Press, Boca Raton, FL, **1991**, Ch. 4, pp. 111-119.

21. Hirokawa, T., Kiso, Y. Complex- forming Equilibria in Isotachophoresis. VI. Simulation of Isotachophoretic Equilibria of lanthanoids and Determination of Mobilities and Stability Constants of Acetate and ∝-Hydroxyisobutyrate complexes. *J. Chromatogr.* **1984**, *312*, 11-29.

22. Mills, E. J.; *Philosophical Magazine 1884*, *17*, 173-187.

23. Meyer, H. Zur Theorie der Alkoholnarkose. *Arch. Exper. Pathol. Pharmakol.* **1899**, *42*, 109-118

24. Overton, E. Studien über die Narkose zugleich ein Beitrag zur allgemeinen Pharmacologie; Verlag Gusav Fischer: Jena (German), **1901**; pp. 141

25. Karelson, M. Molecular Descriptors in QSAR/QSPR. John Wiley & Sons, New York, 2000.

26. Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027-1043.

27. Draper, N.; Smith, M. In *Applied Regression Analysis*, 2nd ed.; Wiley-Interscience: New York, **1981**.

28. Mather, P. M. In Computational Methods of Multivariate Analysis in Physical Geography; Wiley: London, 1976

29. Meister, P. M.; Schwarz, W. H. E. Principal Component of Iinicity. *J. Phys. Chem.* **1994**, *98,* 8245-8252

30. Heberger, K.; Lopata, A. Assessment of Nucleophilicity and Electrophilicity of Radicals, and Palr and Enthalpy Effects on Radical Addition Reactions. *J. Org. Chem.* **1998**, *63*, 8646-8653

31. Musumarra, G.; Bruno, M.; Katritzky, A. R.; Sakizadeh, K.; Alunni, S.; Clementi, S. Kinetics and Mechanism of Nucleophilic Displacements with

Heterocycles as Leaving Groups. Part 19. Chemometric Investigation of the Simultaneous Dependence of S_N2 Rates on Alkyl Groups Structure and Leaving Group Nucleofugacity. *J. Chem Soc., Perkin Trans. 2* **1985**, 1887-189.

32. Katritzky, A. R.; Barczynski, P.; Musumarra, G. Pisano, D.; Szafran, M. Aromaticity as a Quantitative Concept. 1. A statistical Demonstration of the Orthogonality of "Classical" and "Magnetic" Aromaticity in Five- and Six-Membered Heterocycles. *J. Am. Chem. Soc*. **1989**, *111*, 7-15.

33. Katritzky, A. R.; Feygelman, V.; Musumarra, G.; Barczynski, P.; Szafran, M. Aromaticity as a Quantitative Concept. 2. Sixteen Familiar Five- and Six-Membered Monocyclic Heterocycles. *J. Prakt. Chem*. **1990**, *332*, 853-869.

34. Katritzky, A. R.; Feygelman, V.; Musumarra, G.; Barczynski, P.; Szafran, M. Aromaticity as a Quantitative Concept. 3. Benzo-fused Five- and Six-Membered Heterocycles. *J. Prakt. Chem*. **1990**, *332*, 885-897

35. Katrizk, A. R.; Karelson, M.; Sild, S.; Krygowski, T. M.; Jug, K.; Aromaticity as a Quantitative Concept. 7. Aromaticity Reaffirmed as a Multidimensional Characteristics. *J. Org. Chem.* **1998**, *63*, 5228-5231.

36. Sjöström, M.; Wold, W. Linear Free Energy Relationships. Local Empirical Rules-Or Fundamental Laws of Chemistry. *Acta Chem. Schand., Ser. B* **1981**, *35*, 537-554.

37. Albano, C.; Wold, S. Multivariate Analysis of Solvolysis kinetic Data: an Empirical Classification Paralleling Charge Delocalization in the Transition State. *J. Chem. Soc., Prekin Trans. 2* **1980**, 1447-1451.

38. Dewar, M.J.S., Zoebisch, E. G., Healy, E. F., Stewart, J. J. P. AM1: A new General Purpose Quantum Mechanical Model. *J. Am. Chem. Soc*. **1985** *107*, 3902-3909.

39. Stewart, J.J.P. *MOPAC Manual* 6.0; QCE No 455, Seiler F. J. Research Laboratory, **1990**.

40. Katritzky, A.R., Lobanov, V.S., Karelson, M. *CODESSA Reference Manual Version 2.0*; Gainesville, **1994**.

41. Katritzky, A.R., Lobanov, V.S., Karelson, M. QSPR: Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc Rev.* **1995***,* 279-287.

42. Katritzky, A.R., Maran, U., Karelson, M., Lobanov, V.S. Prediction of Melting points for the Substituted Benzenes: A QSPR Approach. *J. Chem. Inf. Comput. Sci.* **1997***, 37*, 913-919.

43. Katritzky, A.R., Ignatchenko, E.S., Barcock, R.A., Lobanov, V.S., Karelson, M. Prediction of Gas Chromatographic Retention Times and Response Factors Using a General Quantitative Structure-Property Relationship Treatment. *Anal. Chem.* **1994**, *66*, 1799-1807.

44. Huibers, P.D., Lobanov, V.S., Katritzky, A.R., Shah, D.O., Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure-Property Relationship Approach. 1. Nonionic Surfactants. *Langmuir* **1996**, *12*, 1462-1470.

45. Huibers, P.D., Lobanov, V.S., Katritzky, A.R., Shah, D.O., Karelson, M. Prediction of Critical Micelle Concentration Using a Quantitative Structure-Property Relationship Approach 2. Anionic Surfactants. *J. Colloid Interface Sci.* **1997**, *187*, 113-120.

46. Katritzky, A.R., Mu, L., Karelson, M. A QSPR Study of the Solubility of Gases and Vapors in Water. *J. Chem. Inf. Comput. Sci.* **1996**,. *36,* 1162-1168.

47. Katritzky, A.R., Rachwal, P., Law, K.W., Karelson, M., Lobanov, V.S. Prediction of polymer Glass Transition Temperatures Using a General Quantitative Structure-Property Relationship Treatment. *J. Chem. Inf. Comput.* **1996**, *Sci. 36*, 879-884.

48. Katritzky, A.R., Tamm, T., Wang, Y., Sild, S., Karelson, M. QSPR Treatment of Solvent Scales. *J. Chem. Inf. Comput, Sci.* **1999**, *39*, 684-691.