JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

QSPR Prediction of the Stability Constants of Gadolinium(III) Complexes for Magnetic Resonance Imaging

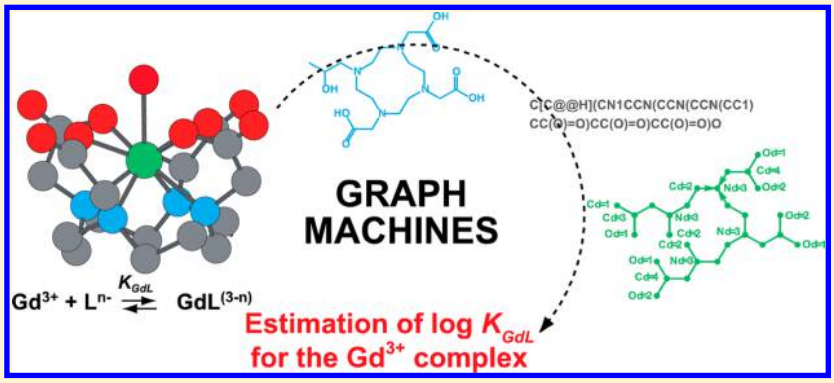
Fabienne Dioury,*[,†] Arthur Duprat,*[,‡,§] Gérard Dreyfus,[‡] Clotilde Ferroud,[†] and Janine Cossy[§]

[†]Laboratoire de Chimie moléculaire, génie des procédés chimiques et énergétiques (CMGPCE), Conservatoire national des arts et métiers (Cnam), 2 rue Conté, 75003 Paris, France

[‡]Signal Processing and Machine Learning (SIGMA) Lab, ESPCI ParisTech, 10 rue Vauquelin, 75005 Paris, France

[§]Laboratoire de Chimie Organique, ESPCI ParisTech, 10 rue Vauquelin, 75005 Paris, France

Supporting Information



ABSTRACT: Gadolinium(III) complexes constitute the largest class of compounds used as contrast agents for Magnetic Resonance Imaging (MRI). A quantitative structure−property relationship (QSPR) machine-learning based method is applied to predict the thermodynamic stability constants of these complexes (log K_{GdL}), a property commonly associated with the toxicity of such organometallic pharmaceuticals. In this approach, the log K_{GdL} value of each complex is predicted by a graph machine, a combination of parametrized functions that encodes the 2D structure of the ligand. The efficiency of the predictive model is estimated on an independent test set; in addition, the method is shown to be effective (i) for estimating the stability constants of uncharacterized, newly synthesized polyamino−polycarboxylic compounds and (ii) for providing independent log K_{GdL} estimations for complexants for which conflicting or questionable experimental data were reported. The exhaustive database of log K_{GdL} values for 158 complexants, reported for potential application as contrast agents for MRI and used in the present study, is available in the Supporting Information (122 primary literature sources).

INTRODUCTION

Magnetic resonance imaging (MRI) has evolved into a major noninvasive technique in medical diagnostics and biomedical research. To enhance the contrast between normal and diseased tissues, MRI examinations often require the administration of a paramagnetic contrast agent, among which Gd^{3+} complexes are prominent.[1,2] For the latter, the purpose of the ligands is to sequester the active cation and to form strong chelates that actually remain stable in the body and are excreted intact, thereby reducing significantly the toxicity of the free Gd^{3+} ion.[3] It is now well established that polyamino-polyacetate ligands, also called complexones, designed to satisfy the electronic demand of the cation, confer such a stability to the corresponding complexes, so that this class of compounds is the most widely used in clinical practice as contrast agents for MRI examinations.[4,5] In the most recent generation of contrast agents for functional and molecular imaging, the organic counterpart (the ligand) can append useful functionalities such as a targeting group that controls the biodistribution.[6] Such applications require more specific and more sensitive agents, so that a lot of effort is devoted to finding new gadolinium-based contrast agents with improved performance.

The high thermodynamic stability of a metal complex used as a pharmaceutical drug is commonly associated with a low toxicity,[7] as both cation and ligand binding abilities to endogenous substrates are neutralized. The thermodynamic stability constant, also called formation constant, of a gadolinium(III) complex (K_{therm} or K_{GdL} generally reported in log K unit) is useful to assess the amount of free Gd^{3+} or free ligand in a water solution. This constant thus appears to be crucial for the development of new contrast agents for MRI.[8] However, the experimental determination of these thermodynamic constants is long and tedious, so that the development of a computational predictive method would speed up the estimation of Gd^{3+} binding for newly designed ligands.

Received: April 16, 2014
Published: September 2, 2014

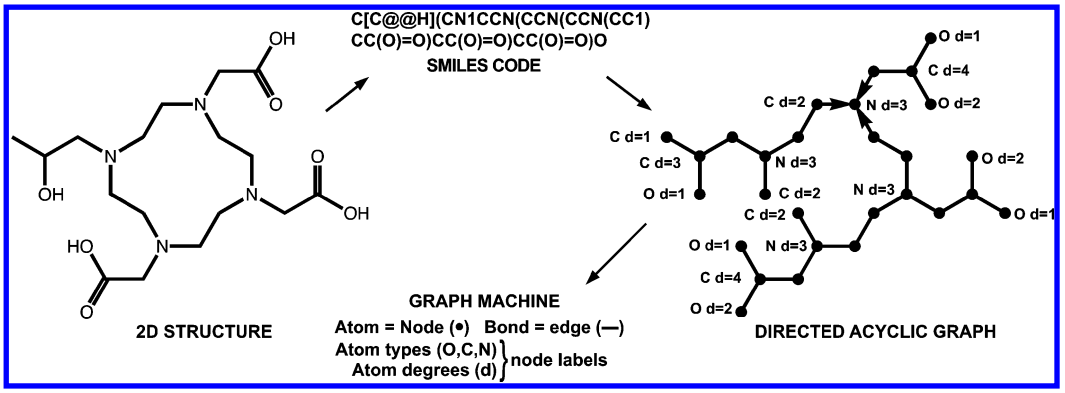


Figure 1. Encoding a molecule into a graph machine: the example of **HP-DO3A**.

Conventional QSPR predictive tools are most often based on relevant molecular descriptors that require the knowledge of 3D structures and are property-specific. Moreover, their design, computation and selection are currently a major burden in QSAR/QSPR applications. In this context, machine learning approaches that predict the properties of interest directly from the molecular structure, thereby exempting the designers from designing, computing, and selecting relevant molecular descriptors, are promising alternatives.

Graph machines are an example of such versatile methods: they perform prediction of properties or activities of molecules from their 2D structure encoded as a graph. In addition, as graph machines do not require any descriptor, a set of graph machines designed for predicting a property/activity of a set of molecules can be retrained for predicting other structure-dependent properties of the same molecules. By contrast, a standard (descriptor-based) regression model (linear, polynomial, neural net, etc.) that has been designed for predicting a given property must be redesigned for predicting another property, because descriptors are property-dependent.

A few computational tools were developed for the molecular modeling of Gd^{3+} complexes and for the correlation of their formation constant to a computed descriptor.[9] To our knowledge, a single predictive tool of log K_{GdL} was reported.[10] It was built with molecular descriptors computed on the basis of no more than 20 compounds representative of the main classes of polyamino-polyacidic ligands. This tool correlates the stability constants of Gd^{3+} complexes to the structures of the ligands; it was tested on a limited set of eight complexones.

For the present study, a database of stability constants for 158 Gd^{3+} complexes was constructed. It was used for training and testing predictive models of their stability, based on *graph machines*, a machine learning algorithm that provides predictions of molecular properties directly from the 2D structure of organic compounds without resorting to chemical descriptors. We show that the thermodynamic stability constants of Gd^{3+} complexes (log K_{GdL}) can be predicted from the 2D structure of the ligands alone: no prior knowledge of the structure of the complex, such as its overall charge or the number and nature of the coordinating sites, is required. Thus, chelation, which is essentially a 3D phenomenon, can be predicted from the information contained implicitly in the experimental data used for training and from the 2D structure of the ligand; loosely speaking, the training mechanism of graph machines can be said to provide a "projection" from the 2D structure to the 3D chelating mechanisms, given the available data. The fact that such a "projection" can be found from a relatively limited amount of data highlights the fact that, for the ligands considered in the present study, the 2D structure of the ligand is relevant for the prediction of the stability of the chelate.

The first section of the paper describes the model design and validation methodology. In the second section, the performances of the resulting model are assessed on an independent test set, and it is shown to be useful for assessing the validity of some reported experimental data, and for predicting the log K_{GdL} of new compounds. In addition, the database of the 158 Gd^{3+} complexes is provided as Supporting Information; it compiles the log K_{GdL} values together with the most important experimental information (analytical technique, medium, temperature), and it cites the primary literature sources that describe the experiments and their thermodynamic interpretation.

MATERIALS AND METHODS

Computational Methods. A graph machine is a composition of parametrized functions whose structure reflects the structure of the graph, so that the value taken by the function, after training, depends on the graph structure, and possibly on additional information.[11] In QSAR/QSPR applications, each node of the graph is a non-H atom, and each edge is a bond (single or multiple) between atoms. In order to take into account multiple bonds, the leaves of the graph contain the degree of each atom (i.e., the number of chemical bonds that bind it to the adjacent non-H atoms); the leaves also contain a label that indicates the nature of the atom and possibly additional data, such as stereochemical information. To handle acyclic graphs only, a minimum number of bonds are deleted if necessary, but the information about the existence of the deleted bonds is retained in the degrees of the adjacent nodes, present in the labels.

Figure 1 illustrates the steps that create the structure of a graph machine for QSAR/QSPR applications. The starting point is the SMILES description of the molecule of interest. It provides the planar representation, which is turned into an undirected, cyclic or acyclic, graph with appropriately labeled nodes. After deletion of edges and selection of the central node, a directed acyclic graph is constructed, where all paths in the graph end at the central node, chosen as described in ref 12.

The final step of the construction of a graph machine consists in postulating a parametrized function (termed *node function*), and implementing it at each node of the graph. The output of a node function is one of the inputs of the functions of the nodes to which it is linked by an oriented edge of the graph; since all

dx.doi.org/10.1021/ci500346w | J. Chem. Inf. Model. 2014, 54, 2718−2731

paths of the graph terminate at the central node, the output of that node (also termed output node) is the output of the graph machine. After training as described below, the output of the graph machine of a molecule is an estimation of the quantity of interest, that is, the log K_{GdL} of the chelate in the present study.

Node functions may be polynomials, neural networks, radial basis functions, etc. All node functions, except the output node functions, are identical within a graph machine and in all graph machines of the database; the output node function may be different from the other node functions, but all output node functions are identical in all graph machines of the database. Therefore, the number of parameters to be estimated during training is the number of parameters of the postulated node function and of the output node function. Although the parameters are not expected to have a straightforward physical or chemical interpretation, a graph machine is not a "black box": since its structure reflects the structure of the molecule, as shown in Figure 1, a graph machine is a "gray box" or "semi-physical" model. More detailed descriptions and didactic examples are provided in previous papers.[11,13,14]

In the present work, the node functions were chosen to be *neural networks*, a class of nonlinear functions that are known to be *universal approximators*, that is, to be able to approximate any sufficiently regular function in a bounded domain with arbitrary accuracy. Neural-network-based graph machines (also known as recursive neural networks[15]) have been described in detail with didactic examples,[11] as well as applications in QSAR;[13] a detailed mathematical analysis is provided by Scarselli et al.[16]

A neural network is a linear combination of nonlinear parametrized functions known as *hidden neurons*.[17] The complexity of a neural network is basically the number of its hidden neurons; therefore, the purpose of complexity selection for graph machines is to find the optimal number of hidden neurons given the available data, as described below.

As usual in statistical machine learning, the parameters of the node functions are estimated by minimizing the distance between the experimental values of log K_{GdL} of selected compounds that make up the *training set*, and the corresponding predictions provided by the graph machines. In the present work, the distance $J(\boldsymbol{\theta})$ is the usual least-squares cost function, that is, the sum, over all graph machines of the training set, of the squared prediction errors (eq 1)

$$J(\theta) = \sum_{i=1}^{N_T} [y_i - g_i(\theta)]^2 \tag{1}$$

where N_T is the number of ligands of the training set, y_i is the experimental value of log K_{GdL} of the i-th complex of the training set, $g_i(\boldsymbol{\theta})$ is the value of K_{GdL} predicted by the corresponding graph machine, and $\boldsymbol{\theta}$ is the vector of parameters (common to all node functions of all graph machines). Therefore, all graph machines are trained simultaneously.

The accuracy of the predictions performed on the examples of the training set is assessed by computing the root mean square training error (RMSTE; eq 2), where $\boldsymbol{\theta}_m$ is the vector of parameters of the node functions after training.

$$\text{RMSTE} = \sqrt{\frac{1}{N_T}\sum_{i=1}^{N_T} [y_i - g(\boldsymbol{\theta}_m)]^2} = \sqrt{\frac{J(\boldsymbol{\theta}_m)}{N_T}} \tag{2}$$

After training, when it is desired to predict the log K_{GdL} of a compound that is not present in the training set, its graph machine is constructed as explained above, and the node functions are assigned the parameters obtained by training, as the node functions are identical for all molecules.

In the present paper, the optimal complexity of the node functions, given the available data, was found by virtual leave-one-out, a nonlinear extension of the PRESS (predicted residual sum of squares) method.[14,18] The virtual leave-one-out prediction error of a ligand is a first-order approximation of the difference between the experimental value of log K_{GdL} and the prediction that would have been performed on that ligand if it had been withdrawn from the training set. The virtual leave-one-out score VLOOS (eq 3) is defined as the root-mean-square of the VLOO prediction errors

$$\text{VLOOS} = \sqrt{\frac{1}{N_T}\sum_{i=1}^{N_T} [y_i - g_i(\boldsymbol{\theta}_m{}^{-i})]^2} \tag{3}$$

where $g_i(\boldsymbol{\theta}_m{}^{-i})$ is a first-order approximation of the predicted value of log K_{GdL} of complex i provided by the i-th graph machine when the latter is not present in the training set, N_T is the number of ligands of the *training/validation set*, and y_i is the experimental value of log K_{GdL} of the i-th ligand of that set. Therefore, the score is an estimate of the generalization error of the model.

The assessment of the performance of the selected model is performed by applying it to a *test set* of complexes that are used neither for training nor for model selection. In addition, the model is used for estimating the stability constants of complexes with unknown or questionable experimental values (*application set*).

Database Construction. The development of a predictive tool based on graph machines exempts the model designer from finding and computing the structure-related quantities that are correlated to the property of interest and can be used as molecular descriptors. The only requirement is the availability of a database of molecules whose log K_{GdL} have been measured. In order to take advantage of the full power of the method, this database must sample the structural diversity of the ligands designed to complex Gd^{3+}.

Experimental investigations of the formation constants of metal complexes have been conducted by different analytical techniques for many decades. Specific compilations of reported data are available, and some computer databases are commercially available as well.[19−21] The problem encountered in compiling a collection of thermodynamic equilibrium data is that the data are difficult to compare because they come from different sources and were obtained in different experimental conditions.

In the case of Gd^{3+} complexes formed with polyamino-polycarboxylic acids, further difficulties in comparing the data are intrinsic to methods adapted to systems with very high equilibrium constants and high kinetic inertness. Generally, very long equilibration times are needed and, in the particular case of the macrocyclic compounds, a few days to several weeks are required to reach the equilibrium in aqueous solution at room temperature (3 weeks for the $[Gd(\mathbf{DOTA})]^-$ complex for instance Figure 2). The difficulties arising from the high values of stability constants are commonly resolved by the use of an auxiliary competing ligand, while the kinetic inertness can be overcome by performing discontinuous batch (out-of-cell) titrations.

Nevertheless, other sources of uncertainty remain; the example of $[Gd(\mathbf{DOTA})]^-$, the most studied complex of this

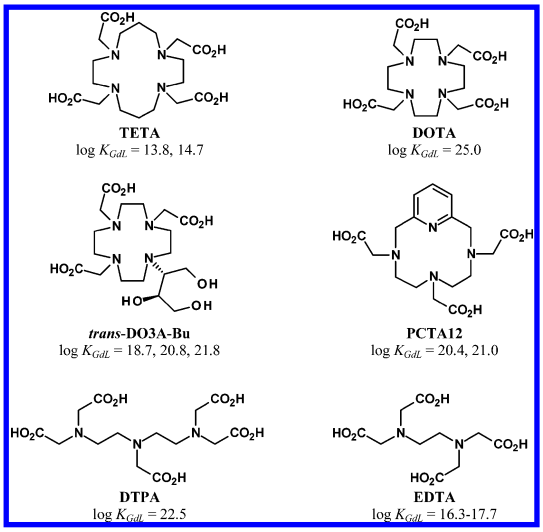


Figure 2. Some representative polyamino-polycarboxylic ligands for Gd^{3+} complexation and their experimental reported log K_{GdL} values.[22]

class, is typical of the difficulty of such experimental determinations: several values of its stability constant, obtained in different experiments performed at 25 °C by means of various techniques in quite different ionic media, have been reported, and range from log K_{GdL} = 22.1−28.0 (24% discrepancy).[22] One may explain the discrepancies as follows: (1) Specific experimental methodologies required for systems characterized by a high equilibrium constant, a high kinetic inertness, or both: out-of-cell and use of a competing ligand are sources of uncertainty. Moreover, one must make sure that measurements are done once all equilibria are reached including those of competing processes (see below). (2) Influence of the electrolyte: the high coordinating power (the denticity) of polyamino-polycarboxylic ligands makes them versatile ligands that are potentially able to coordinate all other metal cations of lower coordinating number. Among such competitive processes, the complexation of the alkaline cation of the electrolyte used to keep the ionic strength constant is one of the major cause of discrepancy in the reported values. Indeed, log K_{GdL} data have been most often determined in KCl or KNO_3 media (51% of the total values compiled), while K^+ and Na^+ are known to bind to complexones. In addition, such binding is ligand-dependent, as illustrated for the macrocyclic **DOTA** ligand (Figure 2) whose binding with Na^+ and K^+ is characterized by log K_{NaDOTA} = 4.4 and log K_{KDOTA} = 1.6, respectively, while it may be neglected for the homologous macrocycle **TETA** (Figure 2).[23] Another example of the electrolyte influence is found for ligand ***trans*-DO3A-Bu** (Figure 2) as Tóth et al.[24] showed that the stability constant ranges from log K_{GdL} = 18.7 to 21.8 (15% discrepancy) when NaCl, NMe_4Cl, or KCl is used as the electrolyte. To conclude with the electrolyte influence (nature and ionic strength), given that (i) only 23% of the log K_{GdL} data of our database have been determined in the expected more inert NR_4^+ medium, and (ii) investigations often failed to assess the influence of the ligand-electrolyte interaction, one may state that it is responsible for a part of the uncertainty in the published data. (3) Influence of the overall basicity of the ligand: apart from the interaction with the background electrolyte, the specific interactions with the H^+ cation, which are intrinsic to the ligand and can be quantified by the overall basicity of the latter (the stepwise protonation constants pK_a), are always taken into account and can be controlled by the pH of the medium. This, however, is the other major cause of the large range of log K_{GdL} values found in the literature as they are related to the set of pK_a values for which a similar variability is observed. (4) Finally, the formation of intermediate complexed species (protonated and hydroxo forms, polynuclear forms and complexes of stoichiometries higher than 1/1) may also cause additional inaccuracies in the measurements.

In practice, an important part of the experimental investigations in the domain aims at finding the technique and the conditions that minimize all of the competitive processes to postulate simpler models that are necessary for the processing of the experimental measurements.

In a recent IUPAC Technical Report, Anderegg et al. examined seven complexones, explained the causes of discrepancies between the reported log K_{ML} values, and categorized them according to their quality (reliability).[25] The conclusion is that one must remember that each reported piece of data should be taken with some reservation and examined carefully when building a datatset for which homogeneity is mandatory. This is especially important in the context of QSAR/QSPR, because the accuracy of the prediction made by a machine-learning based model cannot be better than the accuracy of the database used for training; therefore, an estimate of this experimental accuracy must be defined.

To build a large enough, reliable database, we decided to collect all the available experimental values of log K_{GdL} for the largest possible number of polyamino-polycarboxylic acids used as Gd^{3+} ligands for MRI applications. We chose to discard complexones with low denticity (glycine, iminodiacetic acids, nitrilotriacetic acids, ...) and with log K_{GdL} lower than 10, as such ligands are not suited for the complexation of cations with high coordinating number (8 or 9) such as Gd^{3+}, and are prone to give rise to numerous competitive binding processes. So, for all ligands, only data relative to equilibrium (eq 4) involving the complex $GdL^{(3-n)}$ of 1/1 stoichiometry were recorded.

$$Gd^{3+} + L^{n-} \overset{K_{GdL}}{\rightleftharpoons} GdL^{(3-n)} \text{ with } K_{GdL} = \frac{[GdL]}{[Gd][L]} \qquad (4)$$

Specific reviews dealing with Gd^{3+} chelates as MRI contrast agents are available.[4,26] They constitute an important source of information with a large number of log K_{GdL} data reported even though experimental aspects are sometimes omitted. On the basis of such sources, we decided to extract the data from the primary source of bibliography, that is, the original communications that describe the thermodynamic investigations. In addition to the log K_{GdL} values, we recorded the analytical method and experimental conditions in order to use data sets that are as reliable as possible. With such considerations in mind, a total of 158 polyamino-polycarboxylic ligands were compiled with 222 log K_{GdL} values distributed over 17 log K units ($10 < \log K_{GdL} < 27$) and coming from 111 articles published over the period 1953−2013 (see Supporting Information).

Concerning the chiral ligands, the log K_{GdL} values found in the literature are most frequently relative to compounds prepared as a mixture of stereoisomers. In the particular case of racemic mixtures, the composition is known, but investigations are done usually on ligands synthesized as an unqualified mixture of stereoisomers. As for the few ligands resulting from an asymmetric synthetic process and thus obtained as a well-defined single stereoisomer (12 out of the 50 chiral compounds), log K_{GdL} is available for this particular isomer

dx.doi.org/10.1021/ci500346w | J. Chem. Inf. Model. 2014, 54, 2718−2731

only. The unavailability of log K_{GdL} data for different stereoisomers of chiral ligands precludes the use of stereochemical features for the prediction of log K_{GdL} so that these particular ligands were encoded as stereochemically undefined structures (see for examples ***trans*-1,2-CDTA** and ***trans*-1,2-CPDTA** in Figure 3). Nevertheless, two chiral compounds

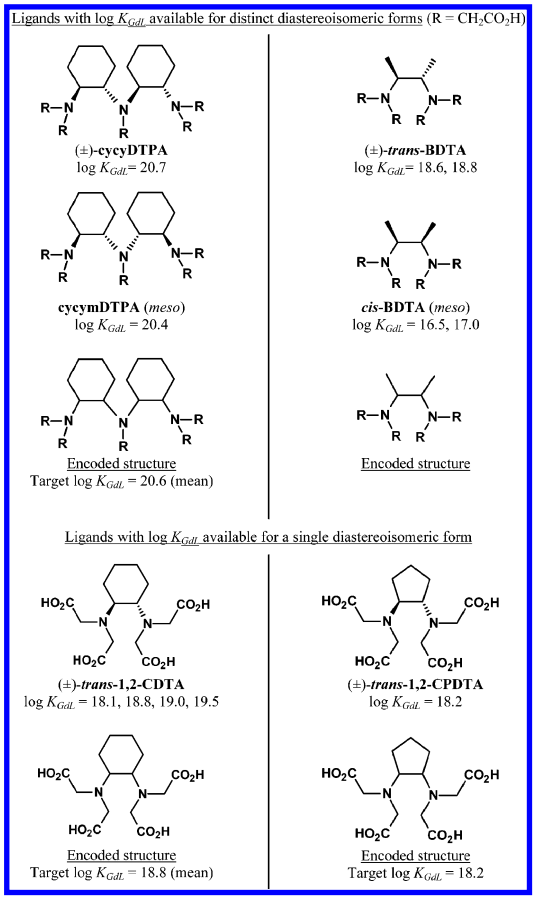


Figure 3. Chiral ligands for Gd^{3+} complexation and their experimental or target log K_{GdL} values.[22]

were found in the literature with log K_{GdL} values available for two distinct diastereoisomeric forms, the racemic mixture and the *meso* isomer: **cycy(m)DTPA** and **BDTA** (Figure 3). These compounds were included as well in our database as their virtual stereochemically undefined forms and associated with target log K_{GdL} values corresponding to the averaged experimental ones.

Database Partition into a Training/Validation Set, a Test Set, and an Application Set. As indicated in the Computational Methods section, model selection is a crucial step in the design of a machine-learning based model. Its purpose is to find the appropriate model complexity, given the available data, which provides the best generalization. In the following, we describe the procedure for partitioning the data set into (i) a training/validation set for model training and complexity selection, (ii) a test set for performance assessment, and (iii) an application set for demonstrating applications of the method to compounds with unknown or questionable log K_{GdL} values.

The reliability of the data present in both the training/validation and the test sets is critical to the success of the method. Consequently, compounds with several published log K_{GdL} values exhibiting large numerical discrepancies were discarded from these two sets and were natural candidates for the application one; the normalized range NR was defined as

$$\mathrm{NR} = 100\frac{\max(\log K_{GdL}) - \min(\log K_{GdL})}{\langle \log K_{GdL} \rangle} \qquad \text{(eq 5)}$$

where max(.), min(.), and ⟨.⟩ denote the maximum, minimum, and mean values found in the literature for the compound under investigation. Compounds with NR ≥ 25% were thus included into the application set. Nevertheless, special attention must be paid to the intensively studied **DOTA**, **DTPA**, and **EDTA** compounds (Figure 2) for which numerous log K_{GdL} with large discrepancies were reported: these ligands are the lead compounds for contrast agents that are commercially available or currently under clinical investigations. They are parent-structures of nearly all the ligands listed in our database, so that they were included in the training/validation set. With 10 published data and NR = 24%, **DOTA** was included in the training/validation set with the value found in an IUPAC technical report (log K_{GdL} = 25.0).[25] As for **DTPA**, with ten reported values and NR = 10%, the log K_{GdL} retained is that of Moeller et al. recommended in the IUPAC report (log K_{GdL} = 22.5).[25,27] In the case of **EDTA**, with seven reported values and a smaller NR (8%), we decided to retain the mean experimental value. For the other complexones with several reported log K_{GdL} values and a normalized range below 8%, the mean experimental value was retained as well, except in a few cases, discussed in the next paragraph, for which one of the reported experimental value was preferred to the mean one.

Assuming that data reported in a given publication are mutually consistent, some ligands were discarded from both the training/validation and test sets and included into the application set if one of the values seemed questionable. Such is the case of the three macrocyclic ligands **TETA**, **PEPA**, and **HEHA** (Figure 2 and Figure 4): Kodama et al. have studied these three ligands and reported log K_{GdL} values obtained by potentiometry in Na^+ electrolyte.[28] For **PEPA** and **HEHA**, no other data is available while **TETA** was also investigated by Clarke et al. by potentiometry in K^+ medium.[29] The latter was then examined by Anderegg et al.,[25] who recommended the value of Clarke. This led us to suspect the occurrence of a systematic error in Kodama's protocol, so that **PEPA** and **HEHA** were included in the application set, while **TETA** was included in the training/validation set with Clarke's value.

The case of **DO3A-L1** and **DO3A-L2** (Figure 5) is similar: as reference ligand, the authors also studied **DOTA** and reported a log K_{GdL} value well above that indicated in the IUPAC report (log K_{GdL} = 27.0 and 25.0 respectively)[25,30] so that values obtained in the same experimental conditions for these two **DOTA** derivatives were assumed unreliable and the ligands added to the application set.

For **PCTA14** and **PCTA12** (Figure 6), as well as for **EOB-DTPA** and **DTPA-BMEA** (Figure 7 and Figure 8), for which two log K_{GdL} were reported, the assumed mutual consistency of data obtained in a single research work led us to select the value coming from the publication that provided our training/validation or test set with the highest number of complexones.

For the two pyridine-containing cyclic ligands **PCTA14** and **PCTA12**, the values of log K_{GdL} retained were those found by Aime et al. (log K_{GdL} = 12.5 and 21.0 respectively) as, with **PCTA13** (Figure 6), the authors characterized and compared a total of three ligands in their work,[31] while Dioury et al.[32] and

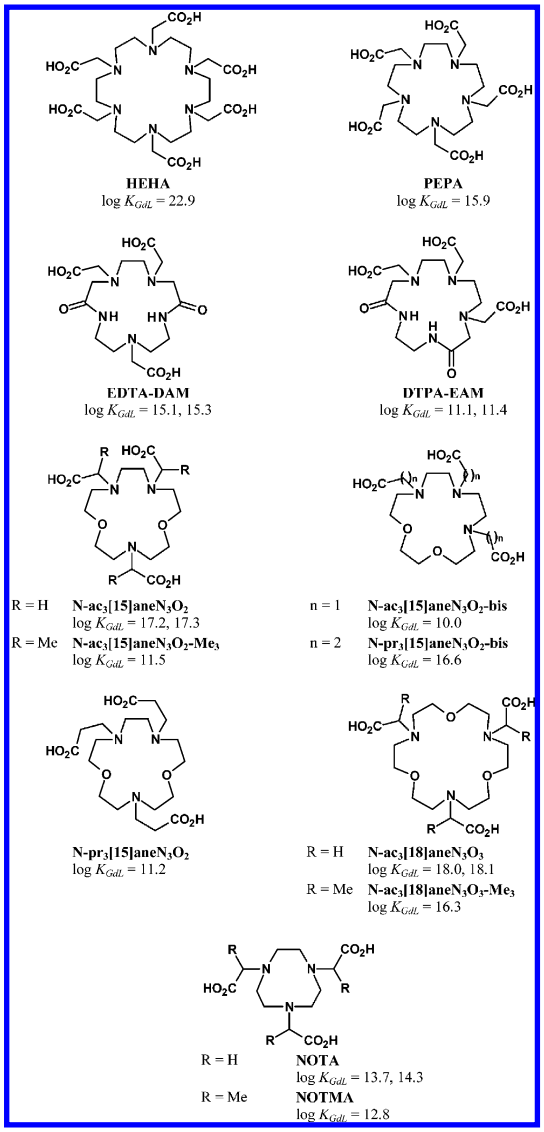


Figure 4. Some cyclic complexones and their experimental reported log K_{GdL} values.[22]

Tircsó et al.[33] reported the characterization of a single compound, **PCTA14** or **PCTA12** respectively. For **EOB-DTPA**, the value of Bianchi et al. was thus retained (log K_{GdL} = 22.8) as they investigated seven other complexones in similar conditions.[26] For **DTPA-BMEA**, the experimental value obtained by White et al. from potentiometric measurements (log K_{GdL} = 16.8)[34,35] was preferred to the value obtained previously by absorption spectroscopy (log K_{GdL} = 16.5),[36] because these authors characterized by potentiometry as well two other **DTPA** derivatives selected for the training/validation or test set (**DTPA-BMMEA**, **DTPA-BHMEA**, Figure 8).[34,35]

Moreover, by listing the values for the secondary *N,N″*-bisamides derived from **DTPA**, the values for **DTPA-BEA**, **DTPA-BHeptA**, and **DTPA-APD** (Figure 8), investigated as **DTPA-BMEA** by absorption spectroscopy, were found to be lower (log K_{GdL} = 15.3, 15.6, 15.3, respectively)[36,37] than the values generally obtained for this subclass of ligands (log K_{GdL} > 16). Therefore, data for the ligands studied under such experimental conditions were not retained for training, and **DTPA-BEA**, **DTPA-BHeptA**, and **DTPA-APD** were added to the application set.

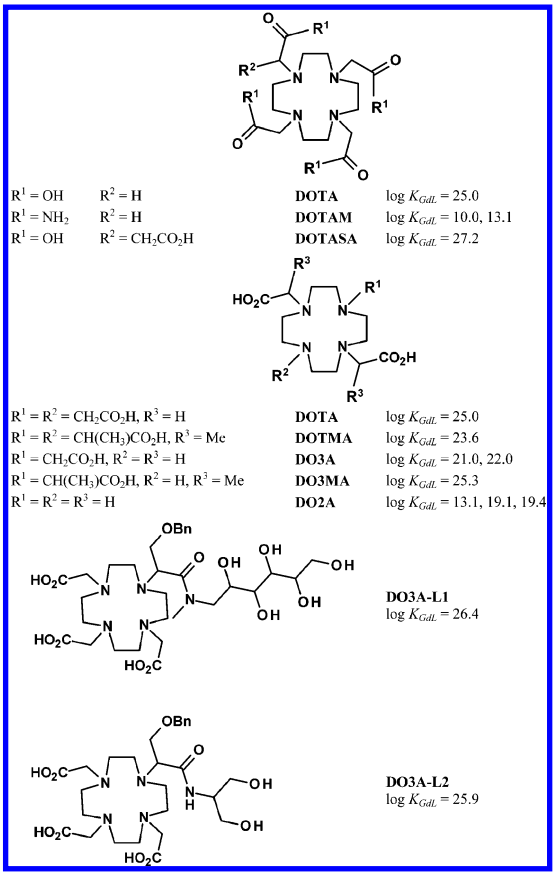


Figure 5. Some DOTA-like ligands and their experimental reported log K_{GdL} values.[22]

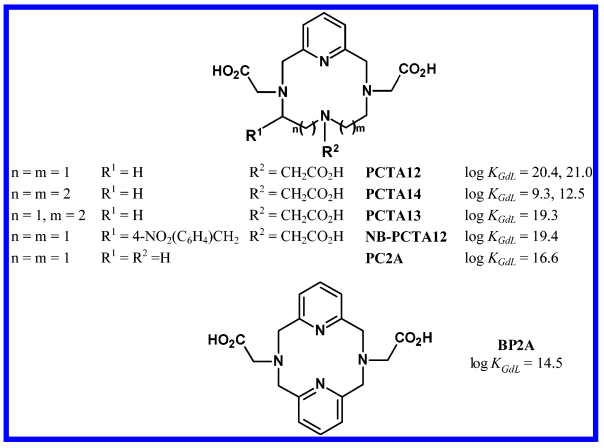


Figure 6. Pyridine-containing macrocyclic ligands and their experimental reported log K_{GdL} values.[22]

For **DOTAM** (Figure 5), Tircsó and Sherry argued that their measurements in a KCl background give a more realistic value than that previously obtained by Maumela et al. in a $NaNO_3$ medium (log K_{GdL} = 13.1 and 10.0 respectively)[38,39] so that the higher one was selected for the training process. For ***trans*-DO3A-Bu** (Figure 2) studied by a single group in three different media,[24] the higher data obtained in K^+ electrolyte (log K_{GdL} = 21.8) was selected since more than 50% of the ligands were characterized in similar KCl or KNO_3 media.

On the basis of the above criteria, 121 compounds were finally selected for the training/validation and the test sets. 93 compounds have a single reported log K_{GdL} value, and 28

dx.doi.org/10.1021/ci500346w | J. Chem. Inf. Model. 2014, 54, 2718−2731

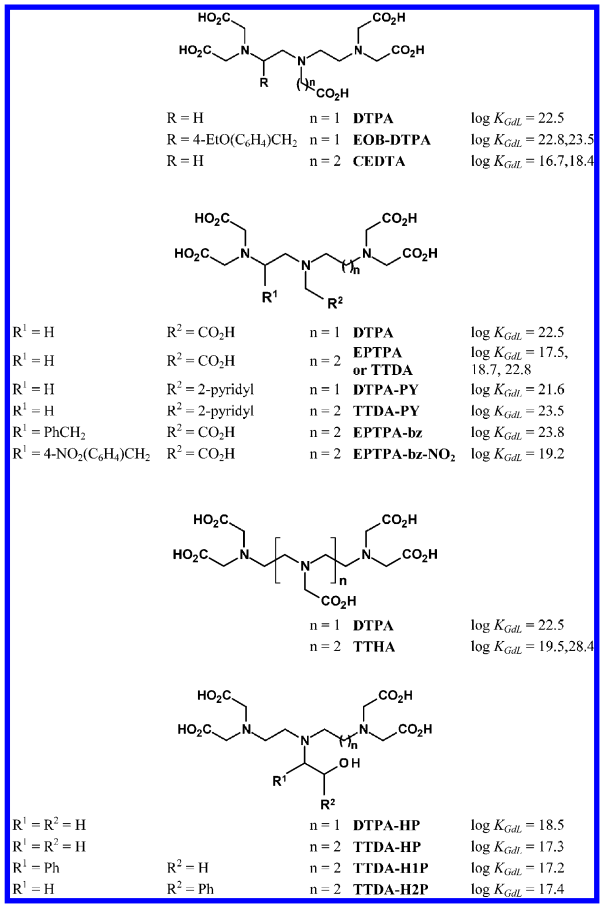


Figure 7. Some DTPA-like ligands and their experimental reported log K_{GdL} values.[22]

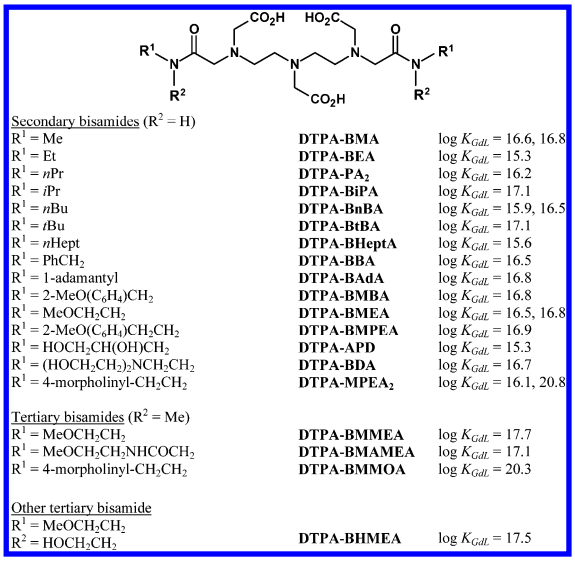


Figure 8. Some DTPA bisamides and their experimental reported log K_{GdL} values.[22]

compounds were characterized in distinct studies. For 19 out of these 28 compounds, the mean experimental value was retained, while, as discussed above, one specific experimental value was selected for each of the other 9 compounds. Their corresponding Gd^{3+} complexes have log K_{GdL} values ranging from 10.4 to 26.3, distributed as shown in Figure 9; 62% of the log K_{GdL} values are above 16.

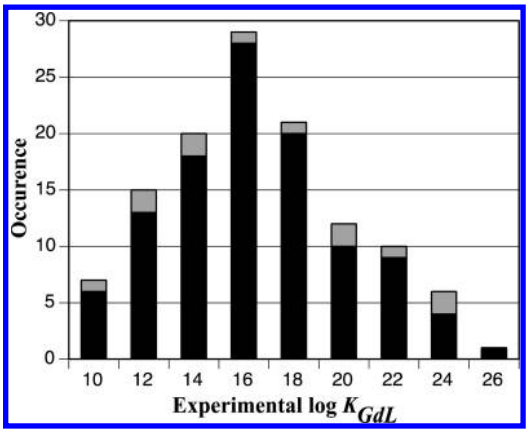


Figure 9. Histogram of log K_{GdL} of the 121 compounds in the training/validation set (black) and in the test set (gray).

Twelve of the above compounds were selected for the test set, based on the following criteria: (i) the values of their stability constants are uniformly distributed in the range shown in Figure 9, (ii) cyclic and linear ligands are in equal numbers, (iii) none of them are parent structures, and (iv) they all belong to different families.

As a result, the training/validation set contains 109 compounds. All 121 molecules are reported in Tables S1 and S2 of the Supporting Information.

The application set was first supplied with ligands discarded from the training/validation and test sets as discussed above. In addition, three kinds of molecules make up this 37 ligand data set: (i) existing ligands with no available log K_{GdL} value (Figure 10); (ii) ligands with several reported log K_{GdL} values that exhibit large normalized ranges, up to 37% for **DO2A** (Figure 5) and **TTHA** (Figure 7), larger than 25% for **EPTPA** (Figure 7), **DTPA-MPEA2** (Figure 8), and **ENDPDA** (Figure 11), and 10% for **CEDTA** (Figure 7). In such cases, the predictive method would help to make a choice between the reported values and to point to a suspected systematic error in some thermodynamic investigations; and (iii) ligands with a single and unreliable reported log K_{GdL} value; some considerations for ligands of that category were discussed above (see **HEHA**/**PEPA**, **DO3A-L1**/**DO3A-L2**, **DTPA-BEA**/**DTPA-BHeptA**/**DTPA-APD**). The case of the **EPTPA-bz** (Figure 7) illustrates the fact that critical comments led us to discard some values: Merbach et al. studied the parent-ligands **EPTPA** and **EPTPA-bz-NO$_2$** (Figure 7)[40] and claimed that the high stability found for **EPTPA-bz** by Wang et al. (log K_{GdL} = 23.8)[41] did not seem reasonable in comparison with the **DTPA** analogues. A similar comparison prompted us to do the same with **TTDA-PY** (Figure 7) as the single log K_{GdL} reported was in the same order of magnitude as the latter questionable compound (log K_{GdL} = 23.5).[42] Moreover, the value found for **TTDA-PY** is surprisingly larger than the value found previously for **DTPA-PY** (Figure 7), a **DTPA** derivative expected to form a strongest Gd^{3+} chelate than its homologous derivative (log K_{GdL} = 21.6).[43] It should be noted that, in this particular case, the other three ligands characterized by T.-H. Cheng et al. in the same work,[42] that is, **TTDA-HP**, **TTDA-H1P**, and **TTDA-H2P** (Figure 7) were kept in the training/validation set as the log K_{GdL} found (17.3, 17.2, 17.4 respectively) are consistent with the values they had reported earlier for the reference ligand **DTPA-HP** (log K_{GdL} = 18.5).[43]

dx.doi.org/10.1021/ci500346w | J. Chem. Inf. Model. 2014, 54, 2718−2731

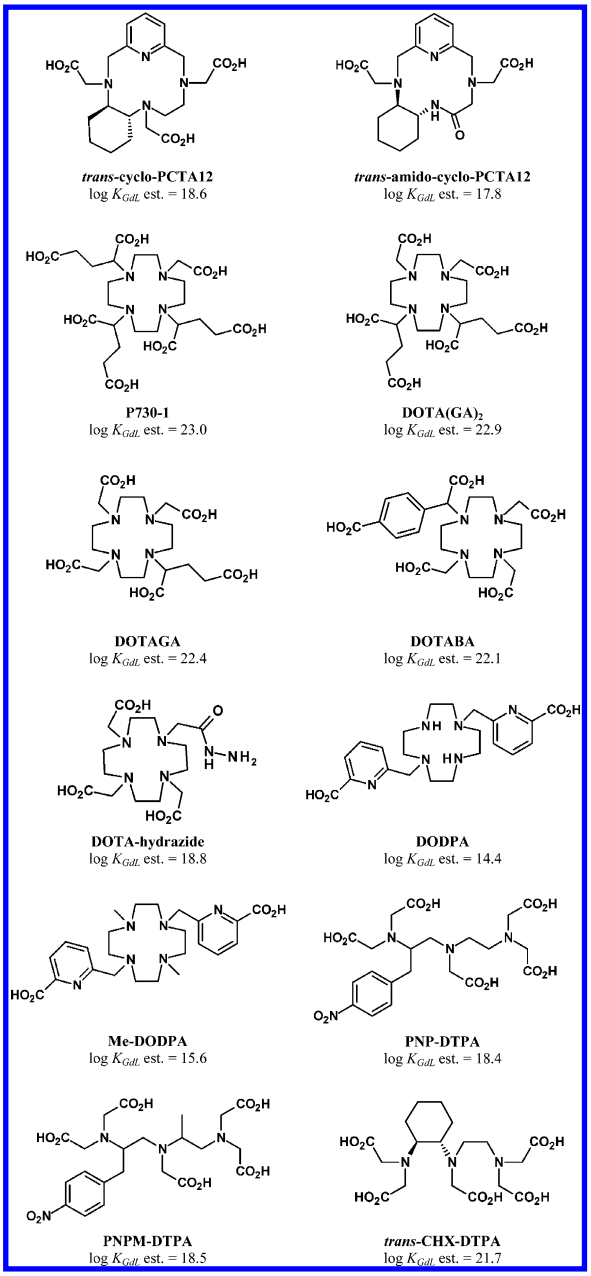


Figure 10. Potential MRI contrast agents with unknown experimental log K_{GdL}; the estimated log K_{GdL} value indicated is the value obtained in this work.

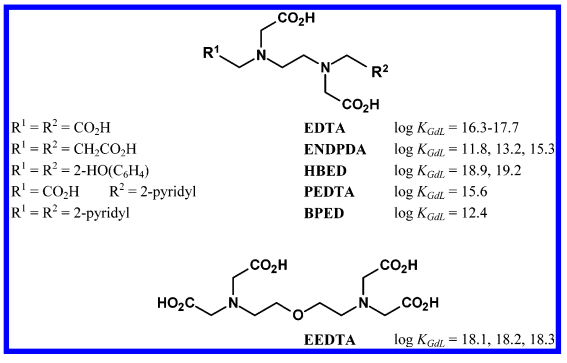


Figure 11. Some EDTA-like ligands and their experimental reported log K_{GdL} values.[22]

A similar analysis by structural similarity led us to assign the following compounds to the application set: **DTPA-EAM** as well as **N-ac$_3$[15]aneN$_3$O$_2$-bis** in comparison with their respective regioisomers **EDTA-DAM** and **N-ac$_3$[15]aneN$_3$O$_2$** (Figure 4), **DOTASA** and **DO3MA** comparatively to **DOTA** (Figure 5), **DTPA-BMMOA** compared to other tertiary bisamides derived from **DTPA** (Figure 8), **EEDTA** and **HBED** in comparison with **EDTA** (Figure 11). Ligand **N-pr$_3$[15]aneN$_3$O$_2$-bis** (Figure 4) was reported to give a complex of higher stability with gadolinium than the acetate analogous **N-ac$_3$[15]aneN$_3$O$_2$-bis** (log K_{GdL} = 16.6 vs 10.0).[44] These results are surprising as propionate pendant arms give rise to less favorable six-membered chelation rings in comparison with the five-membered rings given by the acetate ones so that it was also classified in the test data set for further evaluation. We decided to do the same with **N-pr$_3$[15]aneN$_3$O$_2$** (Figure 4) by considering the conflicting data reported for its regioisomer **N-pr$_3$[15]aneN$_3$O$_2$-bis** (log K_{GdL} = 11.2 and 16.6, respectively).[44,45] Several detailed Tables describing the content of the two sets can be found in the Supporting Information.

Complexity Selection Results. To select models that can generalize satisfactorily, graph machines of increasing complexities were trained. For each node function complexity, 2000 models were trained with different random initial parameter values (mean training time of the five hidden neuron model selected: 0.6 s for 150-epoch training on a quad-core i7−2600 @ 3.6 GHz; for more details on implementation see ref 46) and the mean of the ten smallest RMSTEs (eq 2) was computed (Table 1). Table 1 also contains the values of VLOOS obtained

Table 1. Estimation of the Quality of Training and of Prediction for Increasing Graph Machine Complexity

no. of hidden neurons	2	3	4	5	6	7
RMSTE[a]	1.73	1.07	0.63	0.36	0.14	0.03
VLOOS[b]	2.12	1.52	1.12	0.95	0.82	1.31

[a]Average of the RMSTEs of the 10 models (out of 2000) having the smallest RMSTEs for the 109 compounds of the training/validation set. [b]VLOO scores averaged over the 20 models (out of 2000) having the smallest VLOO scores for the 109 compounds of the training/validation set.

from (eq 3) where $g_i(\boldsymbol{\theta}_m^{-i})$ is the mean VLOO prediction of the log K_{GdL} of complex i provided by the 20 models out of 2000 that have the smallest VLOO scores.

As expected, Table 1 shows that the RMSTE decreases monotonically as the complexity (number of hidden neurons) of the node functions increases. By contrast, the VLOO score first decreases, goes through a minimum, and starts increasing; the VLOO score is minimum for 6 hidden neurons, but the difference between the VLOO scores obtained with 5 and 6 hidden neurons is not very large. In such a situation, the recommended practice is to select the model with the smallest complexity. In order to have a more quantitative assessment, the normalized standard deviation of the leverages[17] was computed for models with five and six hidden neurons

$$\sigma_n = \sqrt{\frac{N}{p(n-p)}\sum_{k=1}^{N}\left(h_{kk} - \frac{p}{N}\right)^2}$$

where h_{kk} is the leverage of example k, N is the number of training examples, and p is the number of parameters of the model. The leverages have the following properties

dx.doi.org/10.1021/ci500346w | J. Chem. Inf. Model. 2014, 54, 2718−2731

$$0 \le h_{kk} \le 1 \text{ and } \sum_{k=1}^{N} h_{kk} = p$$

which show that the leverage of an example can be interpreted as the proportion of the total number of parameters p that is used by the model to fit that example. Therefore, models that exhibit high leverages for some examples are very likely to overfit these examples, while an example whose leverage is close to zero has a very small influence on the model. A model such that all leverages are equal to p/N is very unlikely to overfit the data since all examples have the same influence on the model. The normalized standard deviation of the leverages is equal to 0 if all leverages are equal (smallest risk of overfitting), and it is equal to 1 if p leverages are equal to 1 and all others are equal to zero (worst risk of overfitting). Therefore, if two models have similar VLOO scores, the model whose leverages have the smaller normalized standard deviation should be selected; in the present case, the mean value of the normalized standard deviations of the leverages of the 20 five-hidden-neuron models with the smallest VLOOS was found to be smaller than that of six-hidden-neuron models. Thus, five hidden neurons was the selected complexity.

In Figure 12, the leave-one-out estimates of log K_{GdL} of all compounds of the training/validation set (gray filled circle) and the estimates of log K_{GdL} of compounds of the test set (black filled square) are plotted against the experimental values (these values are reported in Table S4 of the Supporting Information).

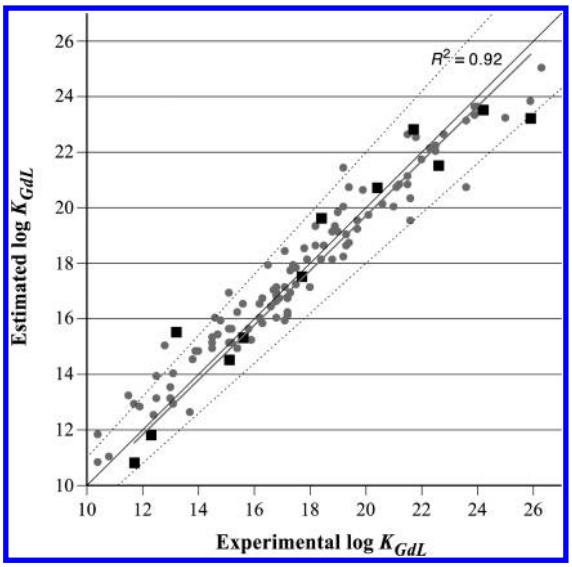


Figure 12. Virtual leave-one-out estimates of log K_{GdL} for the 109 compounds of the training/validation set (gray filled circle), and estimates for the 12 compounds of the test set (black filled square) vs experimental values of log K_{GdL}. Graph machines whose node functions are neural networks with five hidden neurons were used. VLOO estimates are the mean of the VLOO estimates provided by the 20 models with the smallest VLOO scores. For consistency, estimates for the test are the mean values predicted by the same models. The coefficient of determination R^2 has the same value (within three digits) for the training/validation set and for the test set. The regression line for the VLOO estimates and the bisector are not distinguishable.

The prediction error made by a statistical model should always be compared to the experimental error. Unfortunately, most log K_{GdL} values are reported without any estimation of the experimental uncertainty. A rough estimate thereof was derived from 10 values available for log K_{GdDOTA}, 10 values available for log K_{GdDTPA}, and 7 values available for log K_{GdEDTA} which have a normalized range of 24%, 10% and 8% respectively: as a guideline, an accuracy of ±10% was thus considered a realistic goal for the predicted values of log K_{GdL}. Figure 12 shows that all test ligands but one are located between the two dashed lines that display the ±10% error range.

As a baseline test (suggested by one of the reviewers), our method was compared to the "naïve" predictor, which predicts that the log K_{GdL} of a ligand is just the mean of the log K_{GdL} of the ligands that belong to the same family. This applies only to ligands whose family has a significant number of members in the available data set. We considered families with more than five members: DOTA-like (8 ligands), DO3A-like (12 ligands), PCTA-like (6 ligands), NOTA-like (7 ligands), cyclic DTPA-like (5 ligands), aza-crown ethers (9 ligands), DTPA-like (35 ligands), EDTA-like (13 ligands), and homologous EDTA-like (6 ligands). For those 101 ligands, the mean squared prediction error of the naïve predictor was found to be 4.85, while the VLOO mean squared error of the graph machines was 0.81. Therefore, graph machines provide a reduction of RMSE by a factor of 2.5 with respect to the naïve predictor.

In view of the above results, the model was applied to the molecules of the application set, as discussed in the following section.

APPLICATION SET: RESULTS AND DISCUSSION

The predictive method described above can be useful: (i) to provide a first estimate of the stability constants for uncharacterized and for newly designed, hence hitherto nonsynthesized, complexones and (ii) to provide an independent estimation of log K_{GdL}, either when reported experimental results exhibit large numerical discrepancies or when single measurements are deemed questionable in view of values reported for similar compounds.

First log K_{GdL} Estimations for Uncharacterized Complexes. We have found 12 examples of reported complexes prepared as potential MRI contrast agents with no published thermodynamic stability constants: **cyclo-PCTA12**,[47] **amido-cyclo-PCTA12**,[48] **P730-1**,[49] **DOTA(GA)$_2$**,[50] **DOTA-GA**,[50,51] **DOTABA**,[51] **DOTA-hydrazide**,[52] **DODPA** and **MeDODPA**,[53] **PNP-DTPA** and **PNPM-DTPA**,[54] and **CHX-DTPA**[55] (Figure 10). The estimated stability constants (log K_{GdL} est.) are reported in Figure 10.

The predicted stability constants for the 12-membered macrocycles **cyclo-PCTA12** and **amido-cyclo-PCTA12** are consistent with that measured for the parent **PCTA12** as well as for the other three pyridine-containing macrocycles of similar cavity size, namely **NB-PCTA12**, **BP2A** and **PC2A** (Figure 6). For **cyclo-PCTA12**, a vicinal di-*C*-substituted **PCTA12** derivative, the estimated log K_{GdL} = 18.6 is lower than that of the parent ligand **PCTA12** (Figure 6). One can assume that the *trans* cyclohexyl subunit rigidifies the skeleton and that this entropic effect is not favorable for the complexation. A similar trend is observed for the linear series of **DTPA** derivatives as **cycyDTPA** and **cycymDTPA** (Figure 3) give Gd^{3+} complexes with less stability than that obtained with the parent ligand **DTPA** (20.7, and 20.4 respectively, versus 22.5). For the ***trans*-amido-cyclo-PCTA12**, the smaller estimated value log K_{GdL} = 17.8 can be rationalized by the effect of the substitution of an effective acetate subunit by an intracyclic carbonyl element that may be less favorable for the cationic complexation.

dx.doi.org/10.1021/ci500346w | J. Chem. Inf. Model. 2014, 54, 2718−2731

The predicted values for the four **DOTA** analogues bearing an appended arm on at least one of the acetate subunit, **P730-1**, **DOTA(GA)$_2$**, **DOTAGA**, and **DOTABA**, by ranging from 22.1 to 23.0, were found to be smaller than that of the parent **DOTA**. This is satisfactory, since **DOTA** is currently known as the best ligand for the gadolinium complexation with a log K_{GdL} ≈ 25.0, and since substitutions on acetate pendant arms are prone to generate steric hindrance that may disturb the metal complexation.

As for the **DOTA-hydrazide** ligand, the estimated log K_{GdL} = 18.8 is consistent with the assumed lower coordinating ability of a hydrazide subunit compared to that of a carboxylic one. A similar trend was observed for the **DOTAM** and **DOTA** ligands (Figure 5) where the replacement of the four carboxylic subunits by carboxamide ones led to a dramatic loss of stability of the corresponding gadolinium complex (log K_{GdL} =13.1 and 25.0, respectively).[25,38]

The predicted values for **DODPA** and **Me-DODPA** (14.4 and 15.6, respectively) are consistent with that estimated for **DO2A** (log K_{GdL} = 14.4, Table 2). It may be noted that, on such scaffold, a picolinic acid subunit appeared to be as effective as a carboxylic one in terms of gadolinium chelation ability. Moreover, these values are in agreement with those measured for other hexadentate 12-membered tetraazamacrocyclic ligands like **BP2A** and **PC2A** (log K_{GdL} = 14.5 and 16.6, respectively).

Table 2. Experimental and Estimated log K_{GdL} Values for Complexes with Conflicting Reported Data

ligand	experimental log K_{GdL}	refs	estimated log K_{GdL}
DO2A (Figure 5)	13.1, 19.1, 19.4	57, 56	14.4
TTHA (Figure 7)	19.5, 28.4	58, 59	23.3
CEDTA (Figure 7)	16.7, 18.4	60	18.7
EPTPA (Figure 7)	17.5, 18.7, 22.8	40, 61	16.5
DTPA-MPEA$_2$ (Figure 8)	16.1, 20.8	62, 36	16.6
ENDPDA (Figure 11)	11.8, 13.2, 15.3	63, 58	16.0
BDTA (Figure 3)	18.6[a], 18.8[a]	64−66	18.5[c]
	16.5[b], 17[b]	67, 64	

[a]Racemic *trans* diastereoisomer used for the log K_{GdL} evaluation. [b]Ligand stereochemically well-defined: *meso* diastereoisomer used for the log K_{GdL} evaluation. [c]Estimated value for a stereochemically undefined form (see Database Construction section).

The predictions for the three *C*-functionalized DTPA-derivatives **PNP-DTPA**, **PNPM-DTPA**, and **CHX-DTPA** follow the trend observed by other authors for the complexation of the lanthanoïd yttrium(III): the stability constants for the *p*-nitrobenzyl substituted complexes tended to increase with additional substituents on the carbon backbone, while the highest stability constant was observed for the cyclohexyl derivative **CHX-DTPA**.[54] Besides, the calculated value for **CHX-DTPA** (21.7) proved to be in agreement with those measured for the two dicyclohexyl DTPA-derivatives **cycyDTPA** and **cycymDTPA** (20.7 and 20.4, respectively, Figure 3). Moreover, by comparison with the parent ligand **DTPA** (Figure 7), the *trans* cyclohexyl subunit is still not favorable for the gadolinium complexation: a similar trend was observed previously for **cyclo-PCTA12** and the parent ligand **PCTA12** (see above).

Independent log K_{GdL} Estimation for Conflicting Experimental Data. In the case of ligands characterized by different research groups and/or by different analytical methods that provided inconsistent or conflicting data, the computed data delivers an independent evaluation that can help the chemist to make a choice between the experimental values and to reveal hidden systematic errors in an analytical protocol. Seven examples are given in Table 2.

The computed log K_{GdL} = 14.4 for **DO2A** (Figure 5) provides a value closer to the lowest of the three experimental ones; however, the two higher values reported (log K_{GdL} ≈ 19)[56] seemed to be inconsistent with values of the stability constant reported for **DOTA** and **DO3A** (Figure 5) and with the effect of the loss of an acetic acid pendant arm on the stability of the corresponding gadolinium complexes. Moreover, with such consideration in mind, the computed value for **DO2A** is consistent with the reported one for the analogous rigidified pyridine-containing tetraazamacrocycle **BP2A** (14.5, Figure 6).

The case of **TTHA** (Figure 7) is similar: in comparison with the octadentate **DTPA** (Figure 7) and **DOTA** (Figure 5), which are currently considered to be the best chelating agents for Gd^{3+} with log K_{GdL} ≈ 22.5 and 25.0, respectively, the lower as well as the higher values reported for **TTHA** seemed unreasonable. Consequently, the decadentate **TTHA**, able to totally satisfy the electronic demand of Gd^{3+} known to have a coordination number of 8 or 9,[4] was expected to give a 1/1 gadolinium complex of similar stability. It should be noted that none of these experimental values reported had been retained in Anderegg's IUPAC report.[25] The estimated value log K_{GdL} = 23.3 thus appears to be in agreement with the stability of the two reference ligands.

The predicted values of the stability constants of both **CEDTA** and **EPTPA** ligands (Figure 7) can be compared to that reported for the parent **DTPA** (Figure 7): these two homologous derivatives are expected to give gadolinium complexes of lower stability due to the replacement of one of the five-membered intramolecular chelation ring formed on complexation by a less favorable six-membered one. For **CEDTA**, the predicted value matches the highest experimental one while for **EPTPA**, the calculated value is closer to the lowest of the three experimental ones. Moreover, the comparison of the computed values for **CEDTA** and **EPTPA** highlights a remarkable difference in the structural modulation of the **DTPA** skeleton as the replacement of an ethylene bridge connecting two *N*-atoms by a propylene one seems more damaging than the enlargement of one *N*-appended arm from acetate to propionate.

In the case of **DTPA-MPEA$_2$** (Figure 8), the computed value of log K_{GdL} = 16.6 permits to select the most reasonable of the two measured values (log K_{GdL} = 16.1);[62] these values are in agreement with those collected for 11 other examples of secondary DTPA-bisamides found in the literature, all being characterized by a log K_{GdL} ≈ 16−17 (**DTPA-BMA**, **DTPA-PA2**, **DTPA-BiPA**, **DTPA-BnBA**, **DTPA-BtBA**, **DTPA-BBA**, **DTPA-BAdA**, **DTPA-BMBA**, **DTPA-BMEA**, **DTPA-BMPEA**, and **DTPA-BDA** (Figure 8).

As for **ENDPDA** (Figure 11), the computed log K_{GdL} = 16.0 suggests that the higher experimental log K_{GdL} = 15.3 is the most reliable. These values are consistent with that found for the parent **EDTA** (Figure 11; log K_{GdL} ≈ 17.2) and with the expected decrease in stability because of the formation of the less favorable six-membered chelation rings with propionate pendant arms instead of the five-membered rings with acetate ones.

The case of the chiral ligand **BDTA** (Figure 3) is interesting because it highlights the fundamental importance of the construction of the training/validation set. It should be noted that the predicted log K_{GdL} = 18.5 is very close to the experimental determination for the *trans* stereoisomer. This result can be explained if one considers the learned structural feature for similar chiral ligands bearing two vicinal asymmetric centers, namely, **cycyDTPA** and **cycymDTPA**, **CPDTA**, and **CDTA** (Figure 3). As discussed above, stereochemical features were not taken into account in the present study, so that the chiral compounds were encoded as stereochemically undefined structures. However, for the few ligands available, the experimental values have always been obtained with stereoisomers with a *trans* relative configuration so that it could be expected that the modeling method considers each ligand bearing two vicinal asymmetric centers as its *trans* stereoisomer.

New log K_{GdL} Estimation for Ligands with Questionable Reported Stability Constant. The reasons that led us to classify some data as poorly reliable were discussed above. These cases are presented in Table 3.

Table 3. Experimental and Estimated log K_{GdL} Values for Complexes with Questionable Reported Data

ligand	experimental log K_{GdL}	refs	estimated log K_{GdL}
HEHA (Figure 4)	22.9	28	19.9
PEPA (Figure 4)	15.9	28	18.5
DTPA-EAM (Figure 4)	11.1, 11.4	68, 69	14.4
N-ac$_3$[15]aneN$_3$O$_2$-bis (Figure 4)	10.0	44	17.4
N-pr$_3$[15]aneN$_3$O$_2$-bis (Figure 4)	16.6	44	15.9
N-pr$_3$[15]aneN$_3$O$_2$ (Figure 4)	11.2	45	14.4
DOTASA (Figure 5)[a]	27.2	70	22.3
DO3A-L1 (Figure 5)[a]	26.4	30	22.9
DO3A- L2 (Figure 5)[a]	25.9	30	22.9
***(R,R,R)*-DO3MA** (Figure 5)[b]	25.3	71	16.8
***(S)*-EPTPA-bz** (Figure 7)[b]	23.8	41	19.6
TTDA-PY (Figure 7)	23.5	42	14.8
DTPA-BEA (Figure 8)	15.3	36	16.7
DTPA-BHeptA (Figure 8)	15.6	36	16.2
DTPA-APD (Figure 8)[c]	15.3	37	16.8
DTPA-BMMOA (Figure 8)	20.3	62	17.9
EEDTA (Figure 11)	18.1, 18.2, 18.3	72, 21	15.6
HBED (Figure 11)	18.9, 19.2	73, 19	17.1

[a]Ligand stereochemically undefined; probable mixture of stereoisomers used for the log K_{GdL} evaluation. [b]Ligand stereochemically well-defined used for the log K_{GdL} evaluation. [c]Mixture of stereoisomers used for the log K_{GdL} estimation.

The estimated values for **HEHA** and **PEPA** (Figure 4) follow the experimental trend as the higher stability is found for the dodecadentate macrocyclic ligand **HEHA**. Moreover, the estimated value log K_{GdL} = 18.5 for **PEPA** is in agreement with the experimental one reported for the oxygenated analogous **N-ac$_3$[15]aneN$_3$O$_2$** (Figure 4) as a higher stability is expected when two *O*-centers are replaced by two aminoacetic subunits of higher chelating potential.

The estimated log K_{GdL} for the four 15-membered macrocyclic ligands **DTPA-EAM**, **N-ac$_3$[15]aneN$_3$O$_2$-bis**, **N-pr$_3$[15]aneN$_3$O$_2$-bis**, and **N-pr$_3$[15]aneN$_3$O$_2$** (14.4, 17.4, 15.9, and 14.4, respectively) are consistent together, as well as with the reported log K_{GdL} found for regioisomeric analogs **EDTA-DAM** and **N-ac$_3$[15]aneN$_3$O$_2$** (Figure 4). For **DTPA-EAM**, the estimated value is higher than the experimental determination, albeit consistent with the reported log K_{GdL} for its regioisomer **EDTA-DAM** (log K_{GdL} = 14.4 and 15.1−15.3, respectively). The value predicted for **N-ac$_3$[15]-aneN$_3$O$_2$-bis** is far higher than the experimental one but consistent as well with the value reported for its regioisomer **N-ac$_3$[15]aneN$_3$O$_2$** (log K_{GdL} = 17.4 and 17.2−17.3, respectively). As for the estimated value for **N-pr$_3$[15]aneN$_3$O$_2$-bis** and for **N-pr$_3$[15]aneN$_3$O$_2$** (log K_{GdL} = 15.9 and 14.4, respectively) they are, as expected, lower than that found for the respective acetate derivatives. Therefore, from these results, one can conclude that the two isomeric 15-membered macrocyclic skeletons, that is, **[15]aneN$_3$O$_2$**/**[15]aneN$_3$O$_2$-bis**, and **DTPA-EAM**/**EDTA-DAM** give rise to complexes of similar stability with Gd^{3+}.

Concerning the **DOTA** derivatives **DOTASA**, **DO3A-L1**, and **DO3A-L2** (Figure 5) the estimated log $K_{GdL} \approx 22-23$ is consistent with values pertaining to ligands with high number of chelating subunits, which can fully satisfy the electronic demand of Gd^{3+}, and whose efficiency is modulated by steric hindrance because of appended arms. Therefore, the corresponding Gd^{3+} complexes are expected to be slightly less stable than the complex formed with the parent **DOTA** ligand (log $K_{GdL} \approx 25$).

As for **DO3MA**, a trimethylated derivative of **DO3A** (Figure 5), the estimated log K_{GdL} = 16.8 seems more reasonable than the surprisingly high experimental provided value: comparing the effect of methyl substitution on the acetate pendant arms of several macrocyclic ligands of our training/validation set, namely **NOTA** vs **NOTMA**, **DOTA** vs **DOTMA**, **N-ac$_3$[15]-aneN$_3$O$_2$** vs **N-ac$_3$[15]aneN$_3$O$_2$-Me$_3$**, and **N-ac$_3$[18]aneN$_3$O$_3$** vs **N-ac$_3$[18]aneN$_3$O$_3$-Me$_3$** (Figures 4 and 5), it may be concluded that such appendix is always damaging for the stability of the corresponding Gd^{3+} complex while the extent of the damage depends on the macrocyclic scaffold.

The log K_{GdL} = 19.6 estimated for **EPTPA-bz** (Figure 7), which is lower than the reported value log K_{GdL} = 23.8[41] that Merbach et al. deemed unreasonable,[40] is less surprising when compared with data reported for **EPTPA** (also named **TTDA**) and **EPTPA-bz-NO$_2$** (Figure 7). It is also consistent with the lower log K_{GdL} = 14.8 estimated for **TTDA-PY** (Figure 7) as the replacement of one acetate subunit by a 2-pyridinylmethyl one had proved detrimental for the gadolinium complexation if one takes into account the experimental log K_{GdL} differences for **DTPA** vs **DTPA-PY** (Figure 7), and for **EDTA** vs **PEDTA** and **BPED** (Figure 11).

The suspected underestimated experimental log K_{GdL} for the three secondary DTPA-bisamides **DTPA-BEA**, **DTPA-BHeptA**, and **DTPA-APD** (Figure 8) have predicted log K_{GdL} = 16.7, 16.2, and 16.8, respectively; they are in agreement with the value estimated previously for **DTPA-MPEA$_2$** (log K_{GdL} = 16.6, Table 2) together with values reported for others secondary bisamides (Figure 8).

As for the suspected overestimated log K_{GdL} measured for the tertiary DTPA-bisamide **DTPA-BMMOA** (Figure 8), the predicted log K_{GdL} = 17.9 is in good agreement with the two other *N,N″*-dimethylated tertiary DTPA-bisamides found in the literature **DTPA-BMAMEA** and **DTPA-BMMEA** (Figure 8).

Finally, for the two **EDTA** derivatives, **EEDTA** and **HBED** (Figure 11), with suspected overestimated log $K_{GdL} \approx 18-19$ in

comparison with that of the parent ligand, the lower estimated values log K_{GdL} = 15.6 and 17.1, respectively, appear more relevant. Moreover, on such scaffold, in terms of gadolinium complexation ability, it may be noted that (i) the 2-hydroxyphenylmethyl subunits act as the acetate ones; (ii) the elongation of the central chain connecting the two *N*-atoms is detrimental.

CONCLUSION

The QSPR predictive method described in the present work makes use of a large database of 158 polyamino-polycarboxylic complexants of Gd^{3+} designed for application as MRI contrast agents. The efficiency of this approach is demonstrated by the prediction of log K_{GdL} values for new compounds, linear or cyclic, that are consistent with the stability constant values of reference ligands. Moreover, the models can provide independent values that are of interest for complexants with conflicting or questionable experimental data. It may be useful as well for chemists to assess the effect of structural modulation, therefore aiding the rational design of improved MRI contrast agents. Finally, one can take advantage of the versatility of our computational approach based on graph machines: once built, the graph machines can serve, subject to the availability of an appropriate database, to develop other QSAR/QSPR predictive tools for all activity/property related to the same set of compounds. In the particular case of polyamino-polycarboxylic ligands, new predictive models of other properties of interest such as the relaxivity of the gadolinium complexes, or the stability constants of complexes formed with other metallic cations for medical purposes, are under investigation.

ASSOCIATED CONTENT

Supporting Information

Structures of the ligands, experimental log K_{GdL} values, experimental conditions of the thermodynamic measurements, references relative to the 158 polyamino-polycarboxylic ligands used in that work, measured and estimated data of log K_{GdL} for the 109 molecules of the training/validation set and the 12 molecules of the test set, and SMILES notations and names of the 158 Gd^{3+} chelating agents. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: fabienne.dioury@cnam.fr.
*E-mail: arthur.duprat@espci.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank Dr. Jean-Luc Ploix, consultant, for designing and developing the Python libraries and graphical interface used in this study.

REFERENCES

(1) Bellin, M.-F. MR contrast agents, The old and the new. *Eur. J. Radiol.* **2006**, *60*, 314−323.
(2) Yan, G.-P.; Robinson, L.; Hogg, P. Magnetic resonance imaging contrast agents: Overview and perspectives. *Radiography* **2007**, *13*, e5−e19.
(3) Idée, J.-M.; Port, M.; Raynal, I.; Schaefer, M.; Le Greneur, S.; Corot, C. Clinical and biological consequences of transmetallation induced by contrast agents for magnetic resonance imaging: A review. *Fundam. Clin. Pharmacol.* **2006**, *20*, 563−576.
(4) Caravan, P.; Ellison, J. J.; McMurry, T. J.; Lauffer, R. B. Gadolinium(III) chelates as MRI contrast agents: Structure, dynamics, and applications. *Chem. Rev.* **1999**, *99*, 2293−2352.
(5) Rohrer, M.; Bauer, H.; Mintorovitch, J.; Requardt, M.; Weinmann, H.-J. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest. Radiol.* **2005**, *40*, 715−724.
(6) Jacques, V.; Desreux, J. F. New classes of MRI contrast agents. *Top. Curr. Chem.* **2002**, *221*, 123−164.
(7) Cacheris, W. P.; Quay, S. C.; Rocklage, S. M. The relationship between thermodynamics and the toxicity of gadolinium complexes. *Magn. Reson. Imaging* **1990**, *8*, 467−481.
(8) Morcos, S. K. Extracellular gadolinium contrast agents: Differences in stability. *Eur. J. Radiol.* **2008**, *66*, 175−179.
(9) Reichert, D. E.; Hancock, R. D.; Welch, M. J. Molecular mechanics investigation of gadolinium(III) complexes. *Inorg. Chem.* **1996**, *35*, 7013−7020 and references therein..
(10) Qi, Y.-H.; Zhang, Q.-Y.; Xu, L. Correlation analysis of the structures and stability constants of gadolinium(III) complexes. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1471−1475.
(11) Goulon-Sigwalt-Abram, A.; Duprat, A.; Dreyfus, G. From Hopfield nets to recursive networks to graph machines: Numerical machine learning for structured data. *Theor. Comput. Sci.* **2005**, *344*, 298−334.
(12) Jochum, C.; Gasteiger, J. Canonical numbering and constitutional symmetry. *J. Chem. Inf. Comput. Sci.* **1977**, *17*, 113−117.
(13) Goulon, A.; Picot, T.; Duprat, A.; Dreyfus, G. Predicting activities without computing descriptors: Graph machines for QSAR. *SAR QSAR Environ. Res.* **2007**, *18*, 141−153.
(14) Goulon-Sigwalt-Abram, A.; Duprat, A.; Dreyfus, G. In *Unconventional Computation*; Springer: Berlin, 2006; Vol. *4135*, pp 1−19.
(15) Baldi, P.; Pollastri, G. The principled design of large-scale recursive neural network architectures - DAG-RNNs and the protein structure prediction problem. *J. Machine Learning Res.* **2003**, *4*, 575−602.
(16) Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; Monfardini, G. Computational capabilities of graph neural networks. *IEEE Trans. Neural Networks* **2009**, *20*, 81−102.
(17) See for instance: Dreyfus, G. In *Neural Networks: Methodology and Applications*; Springer: Berlin, 2005.
(18) Monari, G.; Dreyfus, G. Local overfitting control via leverages. *Neural Comput.* **2002**, *14*, 1481−1506.
(19) Pettit, L. D.; Powell, K. J.; *Academic Software Stability Constants Database*, version 4.0; IUPAC and Academic Software: Yorks, U.K., 1999.
(20) Smith, R. M.; Martell, A. E.; Motekaitis, R. J.; *Texas A&M University Critically Selected Stability Constants of Metal Complexes Database*, Version 8.0; NIST: College Station, TX, 2004.
(21) Martell, A. E.; Smith, R. M. In *Critical Stability Constants*. Plenum: New York, 1974; Vol. *1*.
(22) log K_{GdL} values and corresponding references given in the Supporting Information.
(23) Delgado, R.; Frausto da Silva, J. J. R. Metal complexes of cyclic tetraazatetraacetic acids. *Talanta* **1982**, *29*, 815−822.
(24) Tóth, É.; Király, R.; Platzek, J.; Radüchel, B.; Brücher, E. Equilibrium and kinetic studies on complexes of 10-[2,3-dihydroxy-(1-hydroxymethyl)-propyl]-1,4,7,10-tetraazacyclododecane-1,4,7-triacetate. *Inorg. Chim. Acta* **1996**, *249*, 191−199.
(25) Anderegg, G.; Arnaud-Neu, F.; Delgado, R.; Felcman, J.; Popov, K. Critical evaluation of stability constants of metal complexes of complexones for biomedical and environmental applications. *Pure Appl. Chem.* **2005**, *77*, 1445−1495.
(26) Bianchi, A.; Calabi, L.; Corana, F.; Fontana, S.; Losi, P.; Maiocchi, A.; Paleari, L.; Valtancoli, B. Thermodynamic and structural properties of Gd(III) complexes with polyamino-polycarboxylic

ligands: basic compounds for the development of MRI contrast agents. *Coord. Chem. Rev.* **2000**, *204*, 309−393.
(27) Moeller, T.; Thompson, L. C. Observations on the rare earths—LXXV(1): The stabilities of diethylenetriaminepentaacetic acid chelates. *J. Inorg. Nucl. Chem.* **1962**, *24*, 499−510.
(28) Kodama, M.; Koike, T.; Mahatma, A. B.; Kimura, E. Thermodynamic and kinetic studies of lanthanide complexes of 1,4,7,10,13-pentaazacyclopentadecane-*N*,*N*′,*N*″,*N*‴,*N*⁗-pentaacetic acid and 1,4,7,10,13,16-hexaazacyclooctadecane-*N*,*N*′,*N*″,*N*‴,*N*⁗,*N*′′′′′-hexaacetic acid. *Inorg. Chem.* **1991**, *30*, 1270−1273.
(29) Clarke, E. T.; Martell, A. E. Stabilities of trivalent metal−ion complexes of the tetraacetate derivatives of 12-membered, 13-membered, and 14-membered tetraazamacrocycles. *Inorg. Chim. Acta* **1991**, *190*, 37−46.
(30) Aime, S.; Anelli, P. L.; Botta, M.; Fedeli, F.; Grandi, M.; Paoli, P.; Uggeri, F. Synthesis, characterization, and 1/T1 NMRD profiles of gadolinium(III) complexes of monoamide derivatives of DOTA-like ligands. X-ray structure of the 10-[2-[[2-hydroxy-1-(hydroxymethyl)-ethyl]amino]-1-[(phenylmethoxy)methyl]-2-oxoethyl]-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid-gadolinium(III) complex. *Inorg. Chem.* **1992**, *31*, 2422−2428.
(31) Aime, S.; Botta, M.; Crich, S. G.; Giovenzana, G. B.; Jommi, G.; Pagliarin, R.; Sisti, M. Synthesis and NMR studies of three pyridine-containing triaza macrocyclic triacetate ligands and their complexes with lanthanide ions. *Inorg. Chem.* **1997**, *36*, 2992−3000.
(32) Dioury, F.; Sylvestre, I.; Siaugue, J.-M.; Wintgens, V.; Ferroud, C.; Favre-Réguillon, A.; Foos, J.; Guy, A. Regioselectively N-functionalised 14-membered azapyridinomacrocycles bearing trialkanoic acid side chains as ligands for lanthanide ions. *Eur. J. Org. Chem.* **2004**, 4424−4436.
(33) Tircsó, G.; Kovacs, Z.; Sherry, A. D. Equilibrium and formation/dissociation kinetics of some Ln(III)PCTA complexes. *Inorg. Chem.* **2006**, *45*, 9269−9280.
(34) Imura, H.; Choppin, G. R.; Cacheris, W. P.; de Learie, L. A.; Dunn, T. J.; White, D. H. Thermodynamics and NMR studies of DTPA-bis(methoxyethylamide) and its derivatives. Protonation and complexation with Ln(III). *Inorg. Chim. Acta* **1997**, *258*, 227−236.
(35) White, D. H.; de Learie, L. A.; Moore, D. A.; Wallace, R. A.; Dunn, T. J.; Cacheris, W. P.; Imura, H.; Choppin, G. R. The thermodynamics of complexation of lanthanide(III) DTPA-bisamide complexes and their implication for stability and solution structure. *Invest. Radiol.* **1991**, *26* (Suppl. 1), S226−S228.
(36) White, D. H.; Rajagopalan, R.; Kuan, K. T.; Lin, Y.; Wallace, R. A.; Rogic, M. M.; Bosworth, M. E.; Robbins, M. S.; Ralston, W. H.; Adams, M. D.; Dunn, T. J. Synthesis and characterization of nonionic paramagnetic metal complexes as potential magnetic resonance imaging contrast agents. *Invest. Radiol.* **1990**, *25* (Suppl. 1), S56−S57.
(37) Sherry, A. D.; Cacheris, W. P.; Kuan, K. T. Stability constants for Gd^{3+} binding to model DTPA-conjugates and DTPA-proteins: Implications for their use as magnetic resonance contrast agents. *Magn. Reson. Med.* **1988**, *8*, 180−190.
(38) Pasha, A.; Tircsó, G.; Benyó, E. T.; Brücher, E.; Sherry, A. D. Synthesis and characterization of DOTA-$(amide)_4$ derivatives: Equilibrium and kinetic behavior of their lanthanide(III) complexes. *Eur. J. Inorg. Chem.* **2007**, 4340−4349.
(39) Maumela, H.; Hancock, R. D.; Carlton, L.; Reibenspies, J. H.; Wainwright, K. P. The amide oxygen as a donor group. Metal ion complexing properties of tetra-*N*-acetamide substituted cyclen: A crystallographic, NMR, molecular mechanics, and thermodynamic study. *J. Am. Chem. Soc.* **1995**, *117*, 6698−6707.
(40) Laus, S.; Ruloff, R.; Tóth, É.; Merbach, A. E. Gd(III) complexes with fast water exchange and high thermodynamic stability: potential building blocks for high-relaxivity MRI contrast agents. *Chem.—Eur. J.* **2003**, *9*, 3555−3566.
(41) Cheng, T.-H.; Lee, T.-M.; Ou, M.-H.; Li, C.-R.; Liu, G.-C.; Wang, Y.-M. Thermodynamic stability and physicochemical characterization of ligand (4*S*)-4-benzyl-3,6,10-tris(carboxymethyl)-3,6,10-triazadodecanedioic acid (H_5[(*S*)-4-Bz-ttda]) and its complexes formed with lanthanides, calcium(II), zinc(II), and copper(II) ions. *Helv. Chim. Acta* **2002**, *85*, 1033−1050.
(42) Cheng, T.-H.; Wang, Y.-M.; Lin, K.-T.; Liu, G.-C. Synthesis of four derivatives of 3,6,10-tri(carboxymethyl)-3,6,10-triazadodecanedioic acid, the stabilities of their complexes with Ca(II), Cu(II), Zn(II) and lanthanide(III) and water-exchange investigations of Gd(III) chelates. *J. Chem. Soc., Dalton Trans.* **2001**, 3357−3366.
(43) Cheng, T.-H.; Wang, Y.-M.; Lee, W.-T.; Liu, G.-C. Synthesis of two *N*′-2-pyridylmethyl and *N*′-2-hydroxypropyl derivatives of diethylenetriaminepentaacetic acid and the stabilities of their complexes with Ln^{3+}, Ca^{2+}, Cu^{2+} and Zn^{2+}. *Polyhedron* **2000**, *19*, 2027−2037.
(44) Hong, C.; Choi, Y.; Choppin, G. R. The stability constant of transition and lanthanide metal ions complexes with 15 membered macrocyclic azacrown ligands. *J. Korean Chem. Soc.* **2004**, *48*, 577−582.
(45) Choi, K.-Y.; Lee, Y.-I.; Kil, H.-S.; Kim, D.-W.; Chung, Y.-S.; Kim, C.-S.; Hong, C.-P.; Sim, W. Potentiometry of the dioxa−triaza macrocyclic complexes as receptors for first-row transition and lanthanide metals. *Microchem. J.* **1996**, *53*, 180−187.
(46) Duprat, A.; Ploix, J.-L.; Dioury, F.; Dreyfus, G. Toward big data in QSAR/QSPR. Presented at the IEEE Workshop on Machine Learning for Signal Processing, Reims, France, 2014.
(47) Dioury, F.; Sambou, S.; Guéné, E.; Sabatou, M.; Ferroud, C.; Guy, A.; Port, M. Synthesis of a tricyclic tetraazatriacetic ligand for gadolinium(III) as potential contrast agent for MRI. *Tetrahedron* **2007**, *63*, 204−214.
(48) Dioury, F.; Ferroud, C.; Guy, A.; Port, M. Synthesis of an hexadentate tricyclic tetraazadiacetic ligand as precursor for MRI contrast enhancement agents. *Tetrahedron* **2009**, *65*, 7573−7579.
(49) Meyer, D.; Port, M.; Rousseaux, O.; Simonot, C. Metal chelates of macrocyclic polyaminocarboxylic derivatives and their use for diagnostic imaging. EP0922700, 1999.
(50) Kielar, F.; Tei, L.; Terreno, E.; Botta, M. Large relaxivity enhancement of paramagnetic lipid nanoparticles by restricting the local motions of the Gd(III) chelates. *J. Am. Chem. Soc.* **2010**, *132*, 7836−7837.
(51) Henig, J.; Tóth, E.; Engelmann, J.; Gottschalk, S.; Mayer, H. A. Macrocyclic Gd^{3+} chelates attached to a silsesquioxane core as potential magnetic resonance imaging contrast agents: Synthesis, physicochemical characterization, and stability studies. *Inorg. Chem.* **2010**, *49*, 6124−6138.
(52) Fuge, F.; Weiler, M.; Gätjens, J.; Lammers, T.; Kiessling, F. Comparison and systematic optimization of synthetic protocols for DOTA−hydrazide generation. *Tetrahedron Lett.* **2013**, *54*, 918−920.
(53) Rodriguez-Rodriguez, A.; Esteban-Gomez, D.; de Blas, A.; Rodriguez-Blas, T.; Fekete, M.; Botta, M.; Tripier, R.; Platas-Iglesias, C. Lanthanide(III) complexes with ligands derived from a cyclen framework containing pyridinecarboxylate pendants. The effect of steric hindrance on the hydration number. *Inorg. Chem.* **2012**, *51*, 2509−2521.
(54) McMurry, T. J.; Pippin, C. G.; Wu, C.; Deal, K. A.; Brechbiel, M. W.; Mirzadeh, S.; Gansow, O. A. Physical parameters and biological stability of yttrium(III) diethylenetriaminepentaacetic acid derivative conjugates. *J. Med. Chem.* **1998**, *41*, 3546−3549.
(55) Brechbiel, M. W.; Gansow, O. A.; Pippin, C. G.; Rogers, R. D.; Planalp, R. P. Preparation of the novel chelating agent *N*-(2-aminoethyl)-*trans*-1,2-diaminocyclohexane-*N*,*N*′,*N*″-pentaacetic acid (H_5CyDTPA), A preorganized analogue of diethylenetriaminepentaacetic acid (H_5DTPA), and the structures of $Bi^{III}(CyDTPA)^{2-}$ and $Bi^{III}(H_2DTPA)$ complexes. *Inorg. Chem.* **1996**, *35*, 6343−6348.
(56) Huskens, J.; Torres, D. A.; Kovacs, Z.; André, J. P.; Geraldes, C. F. G. C.; Sherry, A. D. Alkaline earth metal and lanthanide(III) complexes of ligands based upon 1,4,7,10-tetraazacyclododecane-1,7-bis(acetic acid). *Inorg. Chem.* **1997**, *36*, 1495−1503.
(57) Chang, C. A.; Chen, Y.-H.; Chen, H.-Y.; Shieh, F.-K. Capillary electrophoresis, potentiometric and laser excited luminescence studies of lanthanide(III) complexes of 1,7-dicarboxymethyl-1,4,7,10-tetraazacyclododecane (DO2A). *J. Chem. Soc., Dalton Trans.* **1998**, 3243−3248.

(58) Garg, A. K.; Madhavan, A.; Garg, V.; Malik, W. U. Polarographic studies on rare earth complexes with polyaminopolycarboxylic acids. *Indian J. Chem., Sect. A: Inorg., Phys., Theor. Anal.* **1981**, *20A*, 994−997.
(59) Sosnovsky, G.; Maheswara Rao, N. U. Gadolinium, neodymium, praseodymium, thulium and ytterbium complexes as potential contrast enhancing agents for NMR imaging. *Eur. J. Med. Chem.* **1988**, *23*, 517−522.
(60) Vasil'eva, V. F.; Lavrova, O. Y.; Dyatlova, N. M.; Yashunskii, V. G. Synthesis and complex-forming properties of *N*′-(β-carboxyethyl)-diethylenetriamine-*N*,*N*,*N*″,*N*″-tetraacetic acid. *J. Gen. Chem. USSR* **1966**, *36*, 688−692.
(61) Wang, Y.-M.; Lee, C.-H.; Liu, G.-C.; Sheu, R.-S. Synthesis and complexation of Gd^{3+}, Ca^{2+}, Cu^{2+} and Zn^{2+} by 3,6,10-tri-(carboxymethyl)-3,6,10-triazadodecanedioic acid. *J. Chem. Soc., Dalton Trans.* **1998**, 4113−4118.
(62) Periasamy, M.; White, D.; DeLearie, L.; Moore, D.; Wallace, R.; Lin, W.; Dunn, J.; Hirth, W.; Cacheris, W.; Pilcher, G.; Galen, K.; Hynes, M.; Bosworth, M.; Lin, H.; Adams, M. The synthesis and screening of nonionic gadolinium(III) DTPA-bisamide complexes as magnetic resonance imaging contrast agents. *Invest. Radiol.* **1991**, *26* (Suppl. 1), S217−S220.
(63) Rizkalla, E. N.; Niu, C.; Choppin, G. R. Thermodynamics and nuclear magnetic resonance studies of lanthanide complexation by ethylenediamine-*NN*′-diacetate-*NN*′-di-3-propionate. *Inorg. Chim. Acta* **1988**, *146*, 135−142.
(64) Irving, H. M. N. H.; Sharpe, K. Complexes of meso- and DL-2,3-diaminobutane-*N*,*N*,*N*′,*N*′-tetraacetic acid with rare earth cations. *J. Inorg. Nucl. Chem.* **1971**, *33*, 217−231.
(65) Novak, V.; Svicekova, M.; Majer, J. New complexanes. VI. Determination of the stability constants of the lanthanide complexes of racemic 2,3-diaminobutane-*N*,*N*,*N*′,*N*′-tetraacetic acid by polarographic investigation of the exchange equilibria. *Chem. Zvesti* **1966**, *20*, 252−260.
(66) Voloder, K.; Simeon, V.; Weber, O. A. Physicochemical study of complexes between lanthanides and EDTA derivatives. *Arh. Hig. Rada Toksikol.* **1968**, *19*, 47−53.
(67) Novak, V.; Majer, J.; Svicekova, M. New complexons. III. Polarographic determination of the stability constants of the lanthanide complexes of meso-2,3-diaminobutane-*N*,*N*,*N*,*N*′-tetraacetic acid. *Chem. Zvesti* **1965**, *19*, 817−825.
(68) Carvalho, J. F.; Kim, S. H.; Chang, C. A. Synthesis and metal complex selectivity of macrocyclic DTPA and EDTA bis(amide) ligands. *Inorg. Chem.* **1992**, *31*, 4065−4068.
(69) Frey, S. T.; Chang, C. A.; Carvalho, J. F.; Varadarajan, A.; Schultze, L. M.; Pounds, K. L.; Horrocks, W. D. Characterization of lanthanide complexes with a series of amide-based macrocycles, potential MRI contrast agents, using Eu^{3+} luminescence spectroscopy and molecular mechanics. *Inorg. Chem.* **1994**, *33*, 2882−2889.
(70) André, J. P.; Brücher, E.; Király, R.; Carvalho, R. A.; Mäcke, H.; Geraldes, C. F. G. C. H_5dotasa (=(αRS)-α-(carboxymethyl)-1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid), An asymmetrical derivative of H_4dota (= 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid) substituted at one acetate pendant arm: ^{1}H-NMR and potentiometric studies of the ligand and its lanthanide(III) complexes. *Helv. Chim. Acta* **2005**, *88*, 633−646.
(71) Kang, S. I.; Ranganathan, R. S.; Emswiler, J. E.; Kumar, K.; Gougoutas, J. Z.; Malley, M. F.; Tweedle, M. F. Synthesis, characterization, and crystal structure of the gadolinium(III) chelate of (1R,4R,7R)-α,α′,α″-trimethyl-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid (DO3MA). *Inorg. Chem.* **1993**, *32*, 2912−2918.
(72) Mackey, J. L.; Hiller, M. A.; Powell, J. E. Rare earth chelate stability constants of some aminopolycarboxylic acids. *J. Phys. Chem.* **1962**, *66*, 311−314.
(73) Ma, R.; Motekaitis, R. J.; Martell, A. E. Stability of metal ion complexes of *N*,*N*′-bis(2-hydroxybenzyl)ethylenediamine-*N*,*N*′-diacetic acid. *Inorg. Chim. Acta* **1994**, *224*, 151−155.