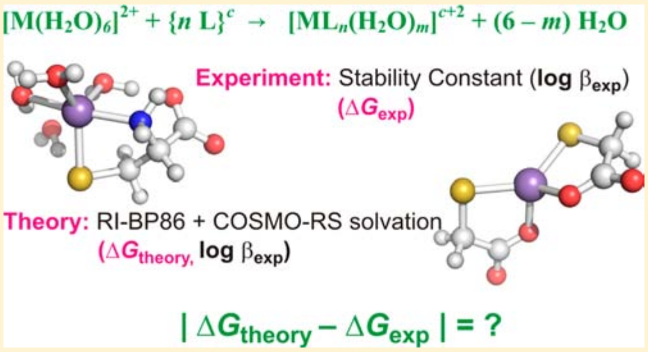
Inorganic Chemistry
Article
pubs.acs.org/IC

# Predicting the Stability Constants of Metal-Ion Complexes from First Principles

Ondrej Gutten and Lubomír Rulíšek*

Institute of Organic Chemistry and Biochemistry, Gilead Sciences Research Center & IOCB, Academy of Sciences of the Czech Republic, Flemingovo nám. 2, 166 10 Praha 6, Czech Republic

Ⓢ *Supporting Information*



**ABSTRACT:** The most important experimental quantity describing the thermodynamics of metal-ion binding with various (in)organic ligands, or biomolecules, is the stability constant of the complex ($\beta$). In principle, it can be calculated as the free-energy change associated with the metal-ion complexation, i.e., its uptake from the solution under standard conditions. Because this process is associated with the interactions of charged species, large values of interaction and solvation energies are in general involved. Using the standard thermodynamic cycle (in vacuo complexation and solvation/desolvation of the reference state and of the resulting complexes), one usually subtracts values of several hundreds of kilocalories per mole to obtain final results on the order of units or tens of kilocalories per mole. In this work, we use density functional theory and Møller–Plesset second-order perturbation theory calculations together with the conductor-like screening model for realistic solvation to calculate the stability constants of selected complexes—$[M(NH_3)_4]^{2+}$, $[M(NH_3)_4(H_2O)_2]^{2+}$, $[M(Imi)(H_2O)_5]^{2+}$, $[M(H_2O)_3(His)]^+$, $[M(H_2O)_4(Cys)]$, $[M(H_2O)_3(Cys)]$, $[M(CH_3COO)(H_2O)_3]^+$, $[M(CH_3COO)(H_2O)_5]^+$, $[M(SCH_2COO)_2]^{2-}$—with eight divalent metal ions ($Mn^{2+}$, $Fe^{2+}$, $Co^{2+}$, $Ni^{2+}$, $Cu^{2+}$, $Zn^{2+}$, $Cd^{2+}$, and $Hg^{2+}$). Using the currently available computational protocols, we show that it is possible to achieve a *relative* accuracy of 2–4 kcal·mol$^{-1}$ (1–3 orders of magnitude in $\beta$). However, because most of the computed values are affected by metal- and ligand-dependent systematic shifts, the accuracy of the "absolute" (uncorrected) values is generally lower. For metal-dependent systematic shifts, we propose the specific values to be used for the given metal ion and current protocol. At the same time, we argue that ligand-dependent shifts (which cannot be easily removed) do not influence the metal-ion selectivity of the particular site, and therefore it can be computed to within 2 kcal·mol$^{-1}$ average accuracy. Finally, a critical discussion is presented that aims at potential caveats that one may encounter in theoretical predictions of the stability constants and highlights the perspective that theoretical calculations may become both competitive and complementary tools to experimental measurements.

## 1. INTRODUCTION

Recent developments in both density functional theory (DFT) and ab initio wave function theory domains of computational chemistry, together with advances in the modeling of solvation effects,[1,2] resulted in the situation that theoretical calculations nowadays represent an integral part of many chemical and biochemical studies.[3] An appropriate selection and benchmarking of all components and methods necessary for accurate predictions of free-energy values for studied chemical processes, together with a careful setup of the model system (which is not always trivial), leads to high-quality computational data that complement and rival the experimental counterparts.

One of the challenges in computational (bio)chemistry is related to the quantitative treatment of metal-ion coordination in biomolecules, experimentally quantified by the stability constant ($\beta$) that is the equilibrium constant for the formation of a complex in solution.[4] In order to calculate these observable thermodynamic quantities from first principles, one usually applies a standard thermodynamic cycle consisting of the studied process in the gas phase (in this case, complexation of the ions with ligands) and (de)solvation of all of the species involved (i.e., of the complexed molecules vs free ligands and hydrated metal ions).[5] For ionic species, these elementary processes are usually associated with large energies of several hundreds of kilocalories per mole (the gas-phase association of the ion···neutral or ion···ion species and their solvation/desolvation energies) that almost cancel each other out to yield the final $\Delta G$ values of several kilocalories per mole (corresponding to dissociation or stability constants in the typical millimolar to femtomolar range).

However, what seems to be a small difference from the computational point of view and from the perspective of the large energy changes associated with the elementary processes is a markedly large value in chemical and biological systems.

**Received:** April 25, 2013
**Published:** September 3, 2013

Differences on the order of ∼5 kcal·mol$^{-1}$ govern many fundamental phenomena, such as metal-ion selectivity in biomolecules.[6−10] Thus, a deeper theoretical understanding of the metal-ion uptake by (bio)molecules may help us to answer fundamental questions, such as why nature selected various metal ions for performing specific functions.[11] Theoretical calculations can be viewed (and used) as a unique and complementary tool to well-established experimental methods[12−14] to correlate the calculated or experimental thermodynamics with the structural details.[15−18] Once, and only once, a satisfactory agreement between the computed and experimental data is obtained, decomposition of the total free-energy change, energy/structure mapping, or analysis of the electronic structure of the studied system can be done and provides us with the insights and concepts.[19,20]

Many systematic efforts to quantitatively assess the selectivity of metal binding by theoretical methods were described in the literature over the past 1.5 decades.[21−29] Recently, we presented a computational study[5] in which we critically evaluated the performance of the ab initio and DFT electronic structure methods in calculations of the energetics associated with metal-ion complexation on a set of five model $[MX_n]^{c+}$ complexes (M = $Fe^{2+}$, $Cu^{2+}$, $Zn^{2+}$, and $Cd^{2+}$; $n$ = 2 and 4−6) spanning four common coordination geometries. Reference values for the *gas-phase* complexation energies were obtained using the CCSD(T)/aug-cc-pVTZ method and compared with cheaper methods, such as DFT and RI-MP2. In the same study,[5] we applied two popular and presumably accurate solvation methods—conductor-like screening model for realistic solvation (COSMO-RS) and universal solvation model employing the full solute density (SMD)—to find out whether the calculated free-energy ($\Delta G$) changes associated with metal-ion complexation in solution match (or are at least in the range of) the experimental stability constants. The computational data highlighted several intricacies in theoretical predictions of the stability constants that may result in errors of several tens of kilocalories per mole in the final $\Delta G$ ($-RT \ln \beta$) values: (a) the covalent character of some metal−ligand bonds [e.g., copper(II) thiolate]; (b) various definitions of the reference state of some systems (e.g., Jahn−Teller unstable $[Cu(H_2O)_6]^{2+}$ vs $[Cu(H_2O)_4]^{2+}$); (c) the presence of the negatively charged ligand in the metal coordination sphere. A similar approach has been evaluated in a recent study by Delgmann and Schenk.[30] The investigation has confirmed that conventional solvation treatment by methods like the polarizable continuum model (PCM) or COSMO is insufficient. The application of more advanced methods (specifically COSMO-RS examined therein) in combination with appropriate quantum-chemical methods is essential to obtain good quantitative agreement with experimental data. Although a number of problematic points concerning COSMO-RS have been highlighted, in combination with careful analysis it stands as a very powerful tool for dealing with solvation effects.

The choice of a proper quantum-chemical method is very problematic. Although DFT is a popular choice, it is clear that none of the current functionals can present a final and universal answer for a wide range of properties. This is especially true for transition metals, which exhibit very diverse chemistry. For example, local-density approximation and generalized-gradient approximation functionals overstabilize low-spin states, although they can perform reasonably well in describing certain properties, e.g., geometries.[31] BP86 is of special interest to this study because of its involvement in the COSMO-RS protocol. Although it is considered to have a decent price/performance ratio, energetics is of limited accuracy, and it is always advisable to compare it with more accurate methods.[32]

The aim of this study is a careful analysis and inspection of all of the aforementioned caveats in the ab initio calculations of the stability constants performed on a much broader set of complexes with experimentally determined stability constants, using the set of eight biologically relevant divalent metal ions: $Mn^{2+}$, $Fe^{2+}$, $Co^{2+}$, $Ni^{2+}$, $Cu^{2+}$, $Zn^{2+}$, $Cd^{2+}$, and $Hg^{2+}$. Various details in adopted computational protocols are carefully analyzed and discussed, and it is concluded that while the accuracy of "absolute" values of binding free energies is still out of reach, the average accuracy of relative affinities of 2 kcal·mol$^{-1}$ might be, in principle, achievable.

## 2. METHODS

**2.1. Computational Details.** All calculations reported in this work were performed using the *TURBOMOLE 6.3* program. The quantum-chemical calculations were mostly performed using DFT, while in a few cases, the MP2 method has been employed. The geometry optimizations were carried out at the DFT level, employing the density-fitted (vide infra) BP86 functional (RI-BP86)[33a,34] and the def-TZVP basis set on all of the atoms.[35,36] All gas-phase structures represent true minima, based on a frequency calculation. The single-point energies were then calculated using the RI-BP86, RI-PBE,[37] BH-LYP,[33a−c] and B3LYP[33a−d,38] functionals or at the RI-MP2 level of theory. All DFT calculations using nonhybrid functionals (as well as MP2 calculations) were expedited by expanding the Coulomb integrals in an auxiliary basis set, using the resolution-of-identity (RI) approximation (density fitting).[39] A multipole-accelerated version of the RI algorithm was used for MP2 calculations. In most of the cases, the def2-TZVP basis set[35] was employed for all of the atoms, with two exceptions represented by the RI-PBE and RI-BP86 calculations, where def-TZVP was used. Grimme's D3 dispersion correction was used for all DFT calculations.[40] For Cd and Hg, small-core Stuttgart/Dresden pseudopotentials were applied to account for scalar relativistic effects.[41]

All metal ions were considered in their 2+ oxidation state. In all cases, high-spin states are assumed (with a couple of exceptions discussed in the text). The choice is justified by calculating energies of low-spin alternatives of selected systems in which one may expect the low-spin states to be closer in energy (Table S9 presented in the Supporting Information, SI). The exhaustive treatment of all of the spin-state splittings for all of the studied complexes is beyond the scope of this study; for iron(II) systems, the reader is referred, for example, to the above-mentioned work of Droghetti et al.[31]

Solvation (free) energies of all studied species were calculated using the COSMO-RS method,[42,43] as implemented in the COSMO*therm* program,[44] using the BP_TZVP_C30_1201.ctd parametrization file. The geometries in the solvent (water) were optimized using the COSMO method,[45] with COSMO radii of 2.0 Å for Mn−Zn, 2.2 Å for Cd, and 2.4 Å for Hg and $\varepsilon_r$ = 80.0. Quite some effort has been exerted to find the true minima; however, because of higher computational demands, this was not achieved in all of the cases. Single-point calculations used for the preparation of COSMO-RS calculations were done according to the recommended protocol, which includes RI-BP86/def-TZVP calculations with Grimme's D3 dispersion correction and uses $\varepsilon = \infty$ (ideal conductor) or 1 (vacuum) with the same radii as those described previously, vide supra. The Gibbs free energy (sometimes denoted free enthalpy) of a system with metal ion M and set of ligands $\{L_i\} \equiv \mathbf{L}$ (introduced in Figure 1) is defined as

$$G_{M,L}^{calc,\mu} = E_{el} + G_{solv} + E_{ZPVE} - RT \ln(q_{trans} q_{rot} q_{vib}) \quad (1a)$$

where $E_{el}$ is the in vacuo energy of the system, $G_{solv}$ is the solvation free energy, $E_{ZPVE}$ is the zero-point vibrational energy, and $-RT \ln(q_{trans}\, q_{rot}\, q_{vib})$ accounts for the entropic terms and the thermal correction to the enthalpy, obtainable from a frequency calculation and utilizing the ideal-gas approximation ($T$ = 298 K and $p$ = 2.48 MPa, which correspond to 1 M concentration).[46] As is discussed in more detail below, these latter

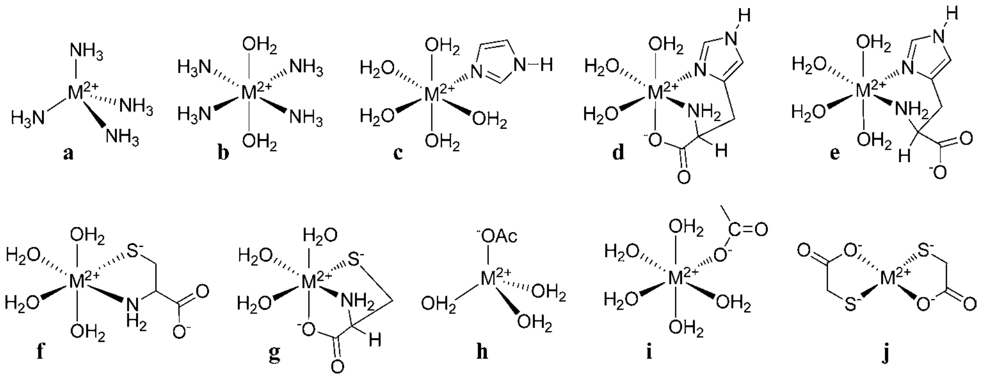


**Figure 1.** Model complexes: (a) $[\mathrm{M(NH_3)_4}]^{2+}$; (b) $[\mathrm{M(NH_3)_4(H_2O)_2}]^{2+}$; (c) $[\mathrm{M(Imi)(H_2O)_5}]^{2+}$; (d) $[\mathrm{M(His)(H_2O)_3}]^{+}$; (e) $[\mathrm{M(His)(H_2O)_4}]^{+}$; (f) $[\mathrm{M(Cys)(H_2O)_4}]$; (g) $[\mathrm{M(Cys)(H_2O)_3}]$; (h) $[\mathrm{M(CH_3COO)(H_2O)_3}]^{+}$; (i) $[\mathrm{M(CH_3COO)(H_2O)_5}]^{+}$; (j) $[\mathrm{M(SCH_2COO)_2}]^{2-}$.

two terms may sometimes present a nontrivial technical challenge. Therefore, we preferred to use the following equation as the practical (albeit not theoretically pure) solution and approximation to the free energy of the complex, using only a single, COSMO-optimized geometry:

$$G^{\mathrm{calc}}_{\mathrm{M,L}} = E_{\mathrm{el}} + G_{\mathrm{solv}} \qquad (1b)$$

Throughout the work, the values of $G^{\mathrm{calc}}_{\mathrm{M,L}}$ (based on eq 1b) and $E_{\mathrm{el}}$ obtained via the RI-BP86 method are reported, whereas $E_{\mathrm{el}}$ obtained via the RI-MP2 method and the values of $G^{\mathrm{calc},\mu}_{\mathrm{M,L}}$ (based on eq 1a) can be found in the SI (Tables S1−S6).

Experimental stability constants, $\beta_{\mathrm{M,L}}$, were converted to differences in the Gibbs free energy, using the formula

$$\Delta G^{\mathrm{exp}}_{\mathrm{M,L}} = -RT\ln K_{\mathrm{M,L}} \approx -RT\ln\beta_{\mathrm{M,L}} = -RT\ln\frac{[\mathrm{ML}_n]_\theta}{[\mathrm{M}]_\theta[\mathrm{L}]_\theta^{\,n}} \qquad (2)$$

where $K$ stands for the thermodynamical stability constant, $\beta$ for the experimental (concentration) stability constant, and $[\mathrm{X}]_\theta$ for the standard concentration.

This is an approximation because the experimental values were obtained at generally nonzero ionic strengths and the measured concentration constants differ somewhat from the thermodynamical stability constants. The difference between the two, i.e., the dependence of the concentration constants on the ionic strength, was generally rather low in comparison to errors in the calculated values. Rigorously, the values could be extrapolated to $I = 0$, for example, with the help of Debye−Hückel theory to estimate the activity coefficients (or some of its extension, such as the Davies equation) using at least three experimentally determined values at varying ionic strength. Such data are not available in all cases, whereas in some other cases, multiple values originating in different sources can be found, and this questions the justifiability of such extrapolations to thermodynamical stability constants. Therefore, we prefer to use the experimental values pertinent to specific conditions in our comparison with the calculated values, thus assuming the activity coefficient to be equal to 1.

In silico, $\Delta G^{\mathrm{calc}}_{\mathrm{M,L}}$ was calculated as the difference of the Gibbs free energies, $G^{\mathrm{calc}}_{\mathrm{M,L}}$, as defined in eq 1, of the products and reactants in the general complexation reaction:

$$[\mathrm{M(H_2O)_6}]^{2+} + \{n\mathrm{L}\}^{c} \rightarrow [\mathrm{ML}_n(\mathrm{H_2O})_m]^{c+2} + (6-m)\mathrm{H_2O} \qquad (3)$$

where $c$ is the total formal charge of the ligands. The numbers of ligands, $n$, and $H_2O$ molecules, $m$, are specified in the corresponding tables.

**2.2. Model Systems.** The set of model systems comprised six complexes ($[\mathrm{M(NH_3)_4}]^{2+}$, $[\mathrm{M(Imi)}]^{2+}$, $[\mathrm{M(His)}]^{+}$, $[\mathrm{M(Cys)}]$, $[\mathrm{M(CH_3COO)}]^{+}$, and $[\mathrm{M(SCH_2COO)_2}]^{2-}$). These systems are simple enough to avoid serious problems with the correct geometry description with only a few binding modes to be tested. We have no structural information about the model systems apart from the number of ligands and their protonation state. This leaves an open question of how many $H_2O$ molecules should be explicitly included in the calculation. For some of the systems, we try two different values for this variable, which gives rise to the 10 systems that were studied, as depicted in Figure 1.

Including more $H_2O$ molecules than are needed to saturate the first coordination sphere (i.e., filling the second coordination sphere) might raise a concern about artificial hydrogen bonds being formed. This issue is very difficult to address because the exact hydrogen-bonding network around the complex is not known. Nevertheless, we carried out model calculations for the cysteinate complex (MD sampling). By comparing its complexes with 18 and 4 $H_2O$ molecules, we observed that similar hydrogen-bonding patterns are present. Probably the weakest correspondence has been found for $\mathrm{Hg^{II}}$. However, it is not clear whether 18 $H_2O$ molecules are sufficient to prevent potential artifacts or whether a missing hydrogen-bonding pattern is due to insufficient sampling (10 systems with 18 $H_2O$ molecules for each metal ion). These structures can be found in the SI.

The concentration constants for most (often for all) metal ions in the studied series were available and cover a wide range of values with $\log(\beta^{\mathrm{max}}_{\mathrm{L}}/\beta^{\mathrm{min}}_{\mathrm{L}}) = 1{-}35$, where $\beta^{\mathrm{max}}_{\mathrm{L}}$ and $\beta^{\mathrm{min}}_{\mathrm{L}}$ are the maximum (mostly in the $\mathrm{Hg^{2+}}$ complex) and minimum (mostly $\mathrm{Mn^{2+}}$) values of the concentration constants for a given set of ligands **L**.

## 3. RESULTS AND DISCUSSION

**3.1. Experimental Free Energies of Complexation.** In Table 1, we summarized the available experimental information on free-energy changes associated with the complexation of metal ions in the studied systems. The values of complexation free energies for a given metal M and set of ligands **L**, $\Delta G^{\mathrm{exp}}_{\mathrm{M,L}}$, are derived from the experimental values of $\log\beta_{\mathrm{M,L}}$ (via eq 2) obtained from Martell's tables.[47] We may observe the known general trends for 3d metal ions, conforming to Irving−Williams series of stability constants with $\mathrm{Ni^{2+}}$ and $\mathrm{Cu^{2+}}$ as the strongest binders, whereas the order of the other metal ions does vary somewhat (mainly Co, Zn, and Cd). However, the magnitude of these differences, $|\Delta G^{\mathrm{exp,max}}_{\mathrm{L}} - \Delta G^{\mathrm{exp,min}}_{\mathrm{L}}|_{(\mathrm{Mn^{2+}}\leftrightarrow\mathrm{Cd^{2+}})}$, varies significantly among the studied systems (from 1.1 kcal·mol$^{-1}$ for the $[\mathrm{M(CH_3COO)}]^{+}$ system to 15 kcal·mol$^{-1}$ for $[\mathrm{M(NH_3)_4}]^{2+}$). These magnitudes are even more pronounced if we include $\mathrm{Hg^{2+}}$ in the series, which has in all cases the highest binding affinity.

**3.2. Calculated Values of the Free Energies of Metal-Ion Complexation.** In Table 2, we summarize the calculated values of complexation free energies, $\Delta G^{\mathrm{calc}}_{\mathrm{M,L}}$, calculated according to eqs 1b and 3 using RI-BP86 for gas-phase electronic energies (the corresponding values of $\Delta G^{\mathrm{calc}}_{\mathrm{M,L}}$ obtained using RI-MP2 for the gas-phase electronic energies and $\Delta G^{\mathrm{calc},\mu}_{\mathrm{M,L}}$ obtained using eqs 1a and 3 using various methods—RI-BP86, RI-PBE, B3-LYP, BH-LYP, and RI-MP2—are listed in Tables S6 and S1−S5 in the SI, respectively). Representative equilibrium geometries of the

Table 1. Estimated Free Energies of Complexation (in kcal·mol$^{-1}$), $\Delta G^{\mathrm{exp}}_{\mathrm{M,L}}$, of the Studied Metal Ions As Calculated from Experimental Stability Constants[a]

| complex | $I$[b] | $\Delta G^{\mathrm{exp}}_{\mathrm{M,L}}$ [kcal·mol$^{-1}$] | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mn | Fe | Co | Ni | Cu | Zn | Cd | Hg |
| $[M(NH_3)_4]^{2+}$ | 2 | −2.3 | −4.5 | −7.6 | −11.1 | −17.6 | −13.2 | −10.1 | −26.1 |
| $[M(Imi)]^{2+}$ | 0.1 | −1.7 | | −3.3 | −4.1 | −5.7 | −3.5 | −3.7 | $-12.5^{h}$ |
| $[M(His)]^{+}$ | 0.1 | −4.5 | $-8.0^{c}$ | −9.4 | −11.8 | −13.9 | −8.9 | −7.7 | |
| $[M(Cys)]$ | 0.1 | −6.5 | $-9.0^{d}$ | −11.1 | −13.4 | | −12.4 | $-13.8^{g}$ | −19.7 |
| $[M(CH_3COO)]^{+}$ | 0 | −1.9 | −1.9 | −1.9 | −1.9 | −3.0 | −2.2 | −2.6 | −5.9 |
| $[M(SCH_2COO)_2]^{2-}$ | 0.1 | −10.3 | $-14.9^{e}$ | −16.6 | $-17.8^{f}$ | | −20.5 | | $-59.8^{i}$ |

[a] It is assumed that these represent thermodynamical equilibrium constants. Unless stated otherwise, the values are for $T$ = 298.15 K. [b] Ionic strength in mol·dm$^{-3}$. [c] $I$ = 3. [d] $T$ = 293.15 K. [e] $I$ = 0. [f] $T$ = 293.15 K. [g] $T$ = 310.15 K; $I$ = 0.15. [h] $I$ = 3. [i] $I$ = 1.

Table 2. Calculated Values of Complexation Free Energies, $\Delta G^{\mathrm{calc}}_{\mathrm{M,L}}$ (kcal·mol$^{-1}$)[a]

| system | metal ion | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Mn | Fe | Co | Ni | Cu | Zn | Cd | Hg |
| $[M(NH_3)_4]^{2+}$ | 13.1 | 11.0 | 2.2 | $-1.8^{j}$ | −8.0 | −0.6 | 1.4 | −11.6 |
| $[M(NH_3)_4(H_2O)_2]^{2+}$ | $-8.4^{b}$ | $-11.3^{b}$ | $-14.4^{b}$ | $-17.9^{b}$ | $-28.5^{c}$ | $-14.9^{d}$ | $-17.3^{d}$ | $-27.7^{d}$ |
| | $-8.1^{b}$ | $-11.3^{b}$ | $-14.6^{b}$ | $-17.3^{b}$ | $-24.8^{b}$ | $-12.1^{b}$ | $-14.2^{b}$ | $-29.9^{d}$ |
| $[M(Imi)(H_2O)_5]^{2+}$ | −3.7 | $-4.6^{k}$ | −5.8 | −6.1 | −6.4 | −4.4 | −4.0 | −11.2 |
| $[M(H_2O)_3(His)]^{+}$ | 5.6 | 4.3 | 0.0 | −0.9 | −5.1 | 3.9 | 4.0 | $-5.2^{k}$ |
| $[M(H_2O)_4(His)]^{+}$ | $-11.8^{e}$ | $-12.9^{e}$ | $-11.0^{e}$ | $-17.9^{e}$ | $-23.9^{e}$ | $-13.4^{e}$ | $-13.2^{e}$ | $-21.4^{k}$ |
| | $-13.0^{f}$ | $-14.8^{f}$ | $-15.8^{f}$ | $-19.6^{f}$ | $-20.5^{f}$ | $-14.4^{f}$ | $-14.4^{f}$ | $-20.6^{k}$ |
| $[M(H_2O)_3(Cys)]$ | 13.8 | 12.1 | 6.4 | 9.1 | −1.6 | 12.9 | 6.4 | −13.9 |
| $[M(H_2O)_4(Cys)]$ | $4.1^{g}$ | $1.4^{g}$ | $0.6^{g}$ | $-0.4^{g}$ | $-14.4^{g,k}$ | $1.9^{g}$ | $-4.3^{g}$ | $-24.0^{g,j}$ |
| | $1.3^{e}$ | $1.7^{e}$ | $-2.9^{e}$ | $-0.6^{e}$ | $-4.5^{e,k}$ | $-2.6^{g}$ | $-6.9^{g}$ | $-22.1^{g,j}$ |
| | $1.3^{e}$ | $2.9^{e}$ | $-4.9^{e}$ | $-3.3^{e}$ | $-12.4^{e,k}$ | $0.4^{e}$ | $-6.7^{g}$ | $-22.6^{j}$ |
| | $6.1^{g}$ | $3.8^{g}$ | $-0.3^{g}$ | $3.5^{g}$ | $-12.5^{g,k}$ | $1.6^{g}$ | $-2.1^{g}$ | $-22.9^{j}$ |
| $[M(CH_3COO)(H2O)_3]^{+}$ | 30.1 | 31.9 | 27.5 | 35.8 | 25.0 | 28.7 | 27.3 | 22.7 |
| $[M(CH_3COO)(H2O)_5]^{+}$ | $3.7^{h}$ | $4.5^{h}$ | $2.8^{h}$ | $5.0^{h}$ | $0.0^{h}$ | $4.5^{h}$ | $2.8^{h}$ | $-1.0^{h}$ |
| | $4.6^{i}$ | $6.1^{i}$ | $6.4^{i}$ | $7.7^{i}$ | $-0.2^{i}$ | $7.4^{i}$ | $5.1^{i}$ | $0.3^{i}$ |
| $[M(SCH_2COO)_2]^{2-}$ | 35.6 | 31.0 | 25.7 | $40.7^{j}$ | $8.9^{k}$ | 28.0 | $21.5^{k}$ | −8.0 |

[a] Calculated using the RI-BP86/COSMO-RS protocol utilizing eqs 1b and 3. [b] Octahedral with $H_2O$ molecules in mutual cis (upper row) or trans (lower row) position. [c] Square pyramidal with $H_2O$ in the axial position; the other $H_2O$ is in the second sphere. [d] Tetrahedral; $H_2O$ molecules in the second sphere. [e] Tridentate binding mode. [f] $(N,N_{\mathrm{Imi}})$ bidentate binding mode. [g] (N,S) bidentate binding mode. [h] Syn binding mode of acetate. [i] Anti binding mode of acetate. [j] We think that these values of $\Delta G^{\mathrm{calc}}_{\mathrm{M,L}}$ might be incorrect and are excluded from further analysis for various reasons: (i) $[Ni(NH_3)_4]^{2+}$ and $[Ni(SCH_2COO)_2]^{2-}$ systems in square-planar geometry are low-spin complexes, as opposed to all other systems. $[Co(SCH_2COO)_2]^{2-}$ is potentially low-spin as well but is included in the analysis and assumed to have a high-spin ground state; (ii) the equilibrium geometry of $[Hg(H_2O)_m(Cys)]$ is entirely different from the geometries of other systems because of the preference of Hg for linear geometry. Although this trait is not unique to these systems, it still makes a direct comparison using the current protocol problematic. [k] Experimental stability constants are not available for these complexes.

studied complexes are depicted in Figure 2, whereas the complete set of all equilibrium geometries of studied complexes is deposited in the SI.

In an ideal case, each system would be represented by an ensemble of structures. However, a satisfactory sampling would require tens, and possibly hundreds, of conformations with at least two coordination spheres, which reaches way beyond affordable limits. Dealing with less sizable statistics is not reliable and, because of "steep" Boltzmann weights, is likely to be dominated by the most negative (i.e., the most stable) entry. We illustrate this by comparing Boltzmann-weighted averages of the overall free-energy change (i.e., calculated stability constant) for one of the systems—the $[M(H_2O)_x(Cys)]$ complex—with the same value obtained using the protocol utilized in this work (i.e., considering only one optimized structure). An ensemble of 10 structures with two (Table S7 in the SI) or one (Table S8 in the SI) coordination sphere of $H_2O$ molecules has been used in this comparison. Providing a larger bulk of water (18 $H_2O$ molecules in $[M(H_2O)_3(Cys)]\cdot 15H_2O$ system) introduces large "noise", and the ensemble is dominated by contribution from a single conformer. On the other hand, all conformers in a single coordination sphere ($[M(H_2O)_3(Cys)]$ system) are very similar, and the weights, as well as contributions, are almost identical.

For these reasons, we opt to represent the systems by a single structure, bearing fully in mind that systems with nonnegligible differences in conformational entropy will suffer systematic mistreatment. However, we do not expect this to be an issue for our chosen set of simple ligands. For some complexes, multiple conformations were considered as potential candidates. However, even a single binding mode is not easy to characterize by a single value of $\Delta G^{\mathrm{calc}}_{\mathrm{M,L}}$ (as can be demonstrated by a range of values obtained for, e.g., $[M(H_2O)_4(Cys)]$ system; Table 2). This is, in part, due to the lack of implementation of structure optimization with respect to our definition of $G^{\mathrm{calc}}_{\mathrm{M,L}}$ (eqs 1a and 1b), which includes two largely opposing terms (COSMO energy and a COSMO-RS correction). Only optimization with respect to the COSMO energy was available, whereas a rigorous optimization may provide different preferences of binding modes for more

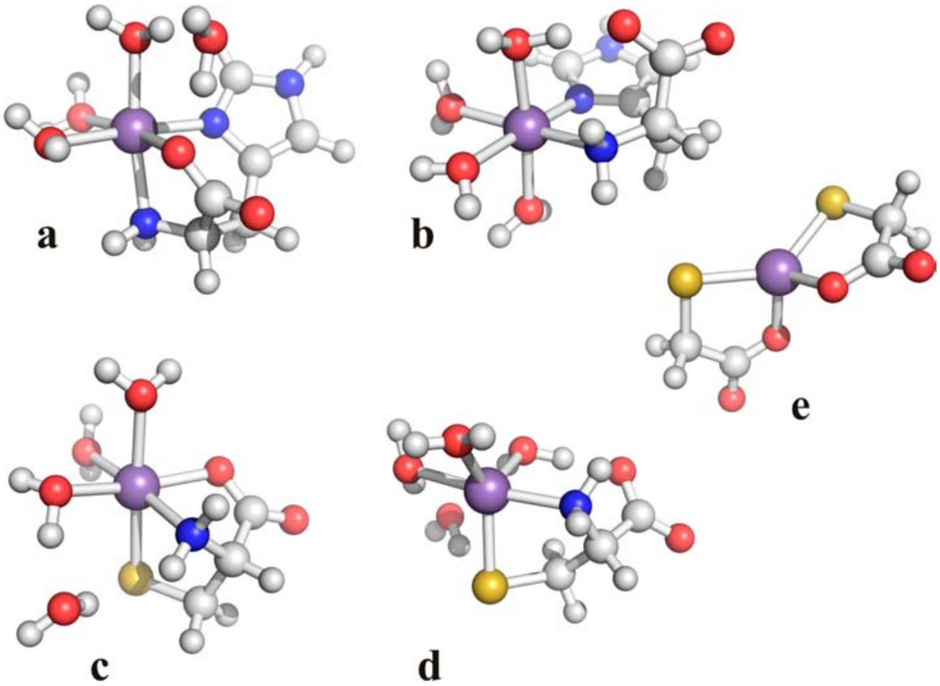


**Figure 2.** Representative equilibrium geometries for selected complexes studied in this work: (a) $[Mn(His)(H_2O)_4]^+$, (N,N,O) binding mode, one $H_2O$ in the second coordination sphere; (b) $[Mn(His)(H_2O)_4]^+$, (N,N) binding mode; (c) $[Mn(Cys)(H_2O)_4]$, (N,S,O) binding mode, one $H_2O$ in the second coordination sphere; (d) $[Mn(Cys)(H_2O)_3]$, (N,S) binding mode, trigonal-bipyramidal geometry, one $H_2O$ in the second coordination sphere; (e) $[Mn(SCH_2COO)_2]^{2-}$.

complex systems. Even if the global minimum with respect to $G_{M,L}^{calc}$ could be found, it may not necessarily represent the "true" structure of the real complex in solution. The reason for this is discussed in detail in section 3.3. In subsequent sections, the conformations with the lowest $G_{M,L}^{calc}$ are used as representatives of a given system.

The values of $\Delta G_{M,L}^{calc}$ for some of the systems—($[M(NH_3)_4(H_2O)_2]^{2+}$, $[M(Imi)(H_2O)_5]^{2+}$, $[M(H_2O)_4(His)]^+$, and $[M(CH_3COO)(H_2O)_5]^+$)—are in the vicinity of their experimental counterparts, but even for these systems, the error commonly exceeds 6 kcal·mol$^{-1}$ in either direction which is not satisfactory. For other systems, the deviations between theory and experiment are greater still.

It turns out that compelling information can be unveiled if we examine the differences between the experimental and calculated values, $\Delta\Delta G_{M,L}$, defined as

$$\Delta\Delta G_{M,L} = (\Delta G_{M,L}^{exp} - \Delta G_{M,L}^{calc}) \tag{4}$$

where $\Delta G_{M,L}^{exp}$ is defined by eq 2 and listed in Table 1 and $\Delta G_{M,L}^{calc}$ is obtained from calculation via eqs 1b and 3 and summarized in Table 2. Next, we show that a large part of this discrepancy can be identified and qualitatively predicted. To this end, we split $\Delta\Delta G_{M,L}$ into two contributions: the ligand-specific shift, $LSS_L$, and the metal-specific shift, $MSS_{M,L}$.

**3.3. Analysis of Ligand-Specific Shifts.** The first of these contributions, the ligand-specific shift, is constructed as an average of all values of $\Delta\Delta G_{M,L}$ for a given set of ligands, **L**, and denoted as $LSS_L$:

$$LSS_L = \sum_{M=\{Mn^{2+},...,Hg^{2+}\}}^{def} \frac{\Delta\Delta G_{M,L}}{n} \tag{5}$$

"def" indicates that the only allowed values of M are those for which $\Delta\Delta G_{M,L}$ is defined; i.e., both $\Delta G_{M,L}^{exp}$ and $\Delta G_{M,L}^{calc}$ are available. For a given method and a given set of ligands **L**, there is a single value of $LSS_L$ and it represents how far, on average, the calculated result is from the experimental one, i.e., an average error. The values of $LSS_L$ are listed in Tables 4 (last column) and S1−S6 in the SI.

The ligand-specific shifts, $LSS_L$, are significant, ranging from less than +8 to −50 kcal·mol$^{-1}$. The somewhat good correspondence between the experimental and calculated values of free energies of complexation for the four systems ($[M(NH_3)_4(H_2O)_2]^{2+}$, $[M(Imi)(H_2O)_5]^{2+}$, $[M(H_2O)_4(His)]^+$, and $[M(CH_3COO)(H_2O)_5]^+$) mentioned in section 3.2 is, in part, due to lower values of their ligand-specific shifts (+6.3, +1.0, +7.4, and −5.4 kcal·mol$^{-1}$, respectively). $LSS_L$ consists of two major contributions. One originates in the ZPVE − $RT \ln Q$ term that is neglected in eq 1b (included in eq 1a) and is especially notable in cases where there is a change in the number of species upon complexation (e.g., $[M(NH_3)_4]^{2+}$, $[M(CH_3COO)(H2O)_3]^+$, $[M(H_2O)_3(Cys)]$, and $[M(SCH_2COO)_2]^{2-}$). Comparing Table 2 (based on eq 1b) with Table S1 in the SI (values obtained by using eq 1a) shows that if this contribution is included, the magnitude of the overall error significantly decreases.

The ZPVE − $RT \ln Q$ term is commonly estimated by invoking the ideal gas and rigid-rotor/harmonic-oscillator approximation, which requires a gas-phase structure of the system that needs to be obtained in addition to the COSMO-optimized structure. However, this may introduce an error that increases the more the two structures diverge from each other. An alternative approach is to utilize these approximations for a COSMO-optimized structure. However, besides being accompanied by technical complications, it is not rigorous[48] and, again, introduces a systematic error that is difficult to control.

The other source of error is a certain bias of the adopted protocol for solvating the charged ligands in their unbound and bound states. The magnitude of this error largely corresponds to

Table 3. Contribution of Protocol Bias to Ligand-Specific Shifts, $LSS_L$ (kcal·mol$^{-1}$)

| acid−conjugate base | $\Delta G^{exp}$ | $\Delta G^{calc}$ | $\Delta\Delta G^{exp/calc}$ | complexation | $LSS_L$[a] |
|---|---|---|---|---|---|
| $NH_3 \rightarrow NH_4^+$ | −12.6 | −17.3 | 4.7 | $[M(H_2O)_6]^{2+} + 4NH_3 \rightarrow [M(NH_3)_4(H_2O)_2]^{2+} + 2H_2O$ | 2.8 |
| Imi → ImiH$^+$ | −9.5 | −12.6 | 3.1 | $[M(H_2O)_6]^{2+}$ + Imi → $[M(Imi)(H_2O)_5]^{2+} + H_2O$ | 0.3 |
| His$^-$ → HisH$_2^{2+}$ | −23.0 | 27.1 | 4.1 | $[M(H_2O)_6]^{2+}$ + His$^-$ → $[M(His)(H_2O)_3]^{2+} + 3H_2O$ | 3.0 |
| Cys$^{2-}$ → Cys$^+$ | −27.8 | −23.2 | −4.6 | $[M(H_2O)_6]^{2+}$ + Cys$^{2-}$ → $[M(Cys)(H_2O)_3] + 3H_2O$ | −8.0 |
| $CH_3COO^- \rightarrow CH_3COOH$ | −6.5 | −4.6 | −1.9 | $[M(H_2O)_6]^{2+}$ + Ac$^-$ → $[M(Ac)_4(H_2O)_5]^+ + H_2O$ | −8.6 |
| $SCH_2COO^{2-} \rightarrow HSCH_2COOH$ | −19.3 | −9.7 | −9.6 | $[M(H_2O)_6]^{2+}$ + 2tg$^{2-}$ → $[M(tg)_2]^{2-} + 6H_2O$[b] | −13.5 |

[a] Calculated using the RI-BP86/COSMO-RS protocol based on eqs 1a and 3; i.e., the terms ZPVE and *RT* ln *Q* are included. Full data, from which the $LSS_L$ value has been obtained, can be found in Table S1 in the SI. [b] tg stands for the thioglycolate ligand, $(SCH_2COO)^{2-}$.

the type of group (carboxylic acid, thiolate, amine nitrogen, etc.). Apparently, complexation energies for acetate, thioglycolate, and cysteine are not negative enough (even if the ZPVE − *RT* ln *Q* term is included), implying too favorable solvation of the unbound ligands. For the other systems (ammonia, imidazole, and histidine), the values are, on the contrary, too negative, indicative of overstabilization of bound imidazole and amine N atoms over neutral unbound states. In the case of histidine, which experiences the opposing effects of carboxyl O and N atoms, the latter prevails.

Because the complexation process can be viewed as the reverse of proton exchange with $H_2O$ molecules (i.e., a proton rather than a metal ion acts as the Lewis acid), we tried to justify the validity of the above hypothesis by calculating the p$K_a$ values (or, more precisely, $\Delta G_{protonation}$ values, which are also applicable in the case of multiple protonation sites) of studied ligands.

As can be seen from the results presented in Table 3, the ligand-dependent shifts, $LSS_L$, qualitatively mimic the deviations between theoretically predicted and experimental p$K_a$'s (i.e., $\Delta\Delta G^{exp/calc}$), and they may serve as the "zero-order" estimates of the expected $LSS_L$ values. It must be emphasized that this correlation does not include the number of ligands and uses only an average of values (viz. definition of $LSS_L$ in eq 5) over a series of metal ions. It is, by no means, meant to be quantitative, but it does demonstrate the major contributions to $LSS_L$, i.e., the protocol bias and the ZPVE − *RT* ln *Q* term.

The nonnegligible magnitudes of these systematic deviations (Table 3) warn that extra care needs to be taken when comparing two conformations for which the difference between their respective $\Delta G_{M,L}^{calc}$ values is in the direction of the "protocol bias". For example, in complexes with histidine, the (N,N$_{Imi}$) binding mode seems to be preferred over the tridentate mode, but whether this might be due to the bias of the protocol used or it is indeed the preferred binding mode found in solution is difficult to conclude unambiguously. On the other hand, for cysteine, the (N,S) mode is always preferred over the (N,O) mode (data not shown) in spite of the understabilization of the bound thiolate group, which leaves more confidence in concluding that this is indeed the preferred binding mode.

The description and analysis of both components of $LSS_L$ remain largely qualitative. Still, we feel that comprehending the nature of $LSS_L$ (or, in general, understanding systematic deviations in protocols used for calculations of solvation Gibbs free energies) is important because it has significant implications in various calculations of the thermodynamic properties of molecules [stability constants, reduction potentials, or acidity constants (p$K_a$'s)].

A better quantitative insight into the $LSS_L$ values might be obtained by comparing the calculated solvation energies with experimental data. This approach, however, is only partially applicable because even if the solvation values for ligands and metal ions are all available, they certainly are not available for the resulting complex.

**3.4. Analysis of Metal-Specific Shifts.** If one is, however, focused on the selectivity of a particular site for a given metal ion, the ligand-specific shifts, $LSS_L$, can be disregarded because they do not affect the relative affinities of a series of metal ions for the particular site. We define the metal-specific shifts as

$$MSS_{M,L} = \Delta\Delta G_{M,L} - LSS_L \tag{6}$$

In other words, it is a measure of how the predicted values deviate from the actual relative affinities. In order to quantify how systematic the shifts are, we use two types of root mean squares (RMSs) of these $RMS_{M,L}$ values: denoted as $RMS_L$ and $RMS_M$. The former, $RMS_L$, is calculated from the values of $RMS_{M,L}$ for one specific set of ligands **L** and all possible metal ions:

$$RMS_L \overset{def}{=} \sqrt{\sum_{M=\{Mn^{2+},...,Hg^{2+}\}} \frac{(MSS_{M,L})^2}{n}} \tag{7}$$

A low value of $RMS_L$ implies that the relative affinities of metal ions for a given set of ligands **L** are reproduced well. The latter, $RMS_M$, is analogously defined for one specific metal ion and all possible sets of ligands **L**:

$$RMS_M \overset{def}{=} \sqrt{\sum_{L=\{[M(NH_3)_4]^{2+},...,[M(SCH_2COO)_2]^{2-}\}} \frac{(MSS_{M,L})^2}{n}} \tag{8}$$

A low value of $RMS_M$ implies that the affinity of a given metal ion M is reproduced with a similar error for various systems.

First, we focus our attention on the values of $RMS_L$ in Table 4. Admittedly, some of the systems ($[M(Imi)]^{2+}$ and $[M(CH_3COO)]^+$) have a smaller range of binding free energies (cf., $\Delta G_{M,L}^{exp}$ in Table 1), which probably also contributes to the lower variance of $MSS_{M,L}$ and, hence, lower $RMS_L$. However, this is not the sole reason for the lower values of $RMS_L$ because these also remain quite low for more selective systems ([M(Cys)] and $[M(His)]^+$). Additionally, a large part of $RMS_L$ is often due to one or two outlying values, while the other values are much smaller.

Systems that differ in the number of water ligands do possess some similarity in the values of individual metal-specific shifts but are not entirely equivalent in this respect. While this can be indicative of one of the systems being a better representation of an actual species in solvent, it has to be borne in mind that the search for local minima is not consistent across the metal-ion series, and this can easily be a more influential factor than the geometry of ligands around a metal ion.

The values in Table 2 were calculated using RI-BP86 and eqs 1b and 3, although in our previous work,[5] we mildly advocated for use of the RI-MP2 method for calculating the gas-phase interaction (complexation) energies for the complexes of divalent

Table 4. Calculated Values of Metal-Specific Shifts and Related Statistics ($kcal\cdot mol^{-1}$)[a]

| system | $MSS_{M,L}$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mn | Fe | Co | Ni | Cu | Zn | Cd | Hg | $RMS_L$ | $LSS_L$ |
| $[M(NH_3)_4]^{2+}$ | −2.8 | −2.8 | 2.9 | excl.[b] | 3.1 | 0.1 | 1.2 | −1.8 | 2.5 | −12.7 |
| $[M(NH_3)_4(H_2O)_2]^{2+}$ | −0.3 | 0.6 | 0.7 | 0.5 | 4.6 | −4.5 | 0.9 | −2.5 | 2.7 | 6.3 |
| $[M(Imi)(H_2O)_5]^{2+}$ | 1.0 | | 1.5 | 1.0 | −0.3 | −0.1 | −0.7 | −2.3 | 1.3 | 1.0 |
| $[M(H_2O)_3(His)]^+$ | 0.8 | −1.4 | 1.5 | −0.1 | 2.0 | −1.9 | −0.9 | | 1.5 | −10.9 |
| $[M(H_2O)_4(His)]^+$ | 1.1 | −0.6 | −1.0 | 0.5 | 2.7 | −1.9 | −0.7 | | 1.5 | 7.4 |
| $[M(H_2O)_3(Cys)]$ | 0.9 | 0.0 | 3.6 | −1.3 | | −4.2 | 1.0 | excl.[b] | 2.6 | −21.1 |
| $[M(H_2O)_4(Cys)]$ | 0.8 | −1.9 | 2.3 | −1.5 | | −1.3 | 1.6 | excl.[b] | 1.8 | −8.5 |
| $[M(CH_3COO)(H2O)_3]^+$ | −0.7 | −2.5 | 1.9 | −6.4 | 3.3 | 0.4 | 1.3 | 2.7 | 3.2 | −31.3 |
| $[M(CH_3COO)(H2O)_5]^+$ | −0.2 | −0.9 | 0.8 | −1.5 | 2.4 | −1.2 | 0.0 | 0.5 | 1.3 | −5.4 |
| $[M(SCH_2COO)_2]^{2-}$ | 1.5 | 1.3 | 4.9 | excl.[b] | | −1.4 | | −4.7 | 3.6 | −46.9 |
| $MSS_{M,avg}$[c] | 0.2 | −0.9 | 1.9 | −1.1 | 2.5 | −1.6 | 0.4 | −1.3 | | |
| $RMS_M$ | 1.3 | 1.4 | 1.6 | 2.3 | 1.5 | 1.7 | 1.0 | 2.6 | | |

[a] The protocol used was RI-BP86/COSMO-RS. Only the most negative value of $\Delta G_{M,L}^{calc}$ for each of the systems is listed, and only those for which experimental data are available are used for calculation of the LSS and RMS quantities. The empty fields are due to missing experimental data. [b] "excl." denotes results that were excluded from analysis; Table 2. [c] Arithmetic average of $MSS_{M,L}$ values over all systems (over all values of "L").

metal ions and ligands with N, S, and O donor atoms [considering the CCSD(T) calculations as the reference]. In contrast, the RI-BP86 functional belonged to the least satisfying methods investigated. Surprisingly, the composite RI-MP2/COSMO-RS protocol (i.e., RI-MP2 values used for the gas-phase complexation energies and the RI-BP86/def-TZVP method used for calculation of the solvation energies within the COSMO-RS framework) is comparable to the presented RI-BP86/COSMO-RS values (data can be found in the SI, Tables S2 and S6, for values obtained using eqs 1a, 1b, and 3). The same essentially holds true for the three other functionals (RI-PBE, BHLYP, and B3LYP; data in the SI, Tables S3−S5) that were shown to be superior (BH-LYP, B3-LYP) or on par (RI-PBE) to RI-BP86 in the gas phase [with respect to the CCSD(T) reference].

In order to provide some arguments in favor of the observed performance of RI-BP86, we may recall that the presented computational scheme of the calculation of $G$ (see eqs 1a and 1b) includes terms of which the gas-phase complexation energies, $E_{el}$, and solvation energies, $G_{solv}$, are the most important ones. They almost cancel out to yield the final values of $G$ on the order of units or tens of kilocalories per mole. It is worth mentioning that neither of these two contributions alone contains accurate information about the absolute or relative complexation energies. In the COSMO-RS protocol, the solvation energy is obtained from the gas-phase energy and COSMO single-point ($\varepsilon = \infty$) calculations using the functional for which COSMO-RS has been parametrized, i.e., the RI-BP86 functional. If we use RI-BP86 (with the same basis set) for the gas-phase energies as well, this value cancels out and is eliminated from the final $G$ of a given species. Hence, only the RI-BP86 energy of a system in a conductor-like environment ($\varepsilon = \infty$), as described by COSMO theory and a COSMO-RS correction to the nonideal-conductor behavior of the solvent, is present in the final value of $G$.

It is possible that the COSMO(-RS) RI-BP86 energy is free of the taint present in the gas-phase RI-BP86 calculations or that this is compensated for in the COSMO-RS scheme or it is at least not too variable across the studied metal ions. It should also be remembered that the evaluation can be skewed when the systems studied are not represented well. One of the obvious candidates is the $[M(CH_3COO)(H_2O)_3]^+$ system, which has a consistently low reproducibility of relative affinities. Either way, RI-BP86 appears as a competitive choice to the investigated alternatives (see the SI, Tables S1−S6). Apart from its considerably lower computational cost, it has another valuable property, as discussed in the following paragraph. Similar conclusions concerning a good price/performance ratio for BP86 for transition-metal complexes were also formulated by Furche and Perdew.[32]

### 3.5. Elimination of Bias for Individual Metal Ions.

We turn our attention to the values of $RMS_M$ presented in Table 4 (i.e., root mean square of metal-specific shifts for a given metal, as defined by eq 8). These can be loosely interpreted as a bias of the adopted protocol for a given metal ion. An encouraging finding is that, in the case of RI-BP86, by calibration of the described protocol a large part of this bias could be eliminated. Thus, the new value is calculated as follows:

$$\Delta G_{M,L}^{C} = \Delta G_{M,L}^{calc} + MSS_{M,avg} + LSS_L \tag{9}$$

where $\Delta G_{M,L}^{calc}$ is the free energy of binding calculated as before, $MSS_{M,avg}$'s are calibration values obtained in a fashion described below, i.e., one value for each metal ion. The last term, $LSS_L$, is a ligand-specific shift, which is unknown for reasons discussed in section 3.3. We use the exact values here (obtained from experimental values) to highlight the increased precision of the obtained *relative* values, which can be seen from a comparison of Tables 1 and 5.

This calibration is doable thanks to a relatively small variation of the metal-specific shift, $RMS_{M,L}$, across different systems, i.e., low values of $RMS_M$. Interestingly, these values are lowest for RI-BP86, whereas the RI-MP2, BHLYP, B3LYP, and PBE methods are trailing behind in this sense in almost all cases.

Although calibration can be done in a number of ways, the quality of which will always depend on the set of systems chosen, the results should not vary fundamentally. This statement is based on relatively low values of $RMS_M$ for all M for a diverse group of ligands presented, and these are assumed to remain low even if we included other systems. To minimize the influence of any one of the systems on the calibration, we use average values of MSS (listed in Table 4), denoted as $RMS_{M,avg}$.

Values of $\Delta G_{M,L}^{C}$ are presented in Table 5 and can be directly compared to experimental values from Table 1. The contribution of the calibration can be assessed by comparing the $RMS_L$ values of calibrated ($RMS_L^C$) and uncalibrated ($RMS_L^{orig}$) protocols. The calibration improves the prediction of selectivity in almost all cases, with a single exception of the $[M(Imi)(H_2O)_5]^{2+}$ system. Although this specific calibration is certainly not optimal, it is

Table 5. Calculated Values of Complexation Free Energies after Removal of Metal-Specific Bias $\Delta G^{C}_{M,L}$ (eq 9) and the Corresponding $RMS_L$ (kcal·mol$^{-1}$)[a]

| | $\Delta G^{C}_{M,L}$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| complex | Mn | Fe | Co | Ni | Cu | Zn | Cd | Hg | $RMS^{C}_{L}$ | $RMS^{orig}_{L}$ |
| $[M(NH_3)_4]^{2+}$ | 0.6 | −2.6 | −8.6 | excl.[b] | −18.1 | −14.9 | −10.9 | −25.7 | 1.7 | 2.5 |
| $[M(NH_3)_4(H_2O)_2]^{2+}$ | −1.9 | −6.0 | −6.4 | −12.7 | −19.7 | −10.3 | −10.6 | −25.0 | 1.7 | 2.7 |
| $[M(Imi)(H_2O)_5]^{2+}$ | −2.5 | −4.5[c] | −2.9 | −6.2 | −2.9 | −5.1 | −2.6 | −11.6 | 1.9 | 1.3 |
| $[M(H_2O)_3(His)]^{+}$ | −5.1 | −7.5 | −9.0 | −12.9 | −13.4 | −8.6 | −6.4 | −17.5[c] | 0.8 | 1.5 |
| $[M(H_2O)_4(His)]^{+}$ | −5.5 | −8.4 | −6.5 | −13.4 | −14.0 | −8.6 | −6.6 | −15.4[c] | 1.5 | 1.5 |
| $[M(H_2O)_3(Cys)]$ | −7.2 | −10.0 | −12.8 | −13.2 | −20.2[c] | −9.9 | −14.4 | excl.[b] | 1.5 | 2.6 |
| $[M(H_2O)_4(Cys)]$ | −7.1 | −8.1 | −11.6 | −13.0 | −20.4[c] | −12.8 | −15.0 | excl.[b] | 0.7 | 1.8 |
| $[M(CH_3COO)(H2O)_3]^{+}$ | −1.1 | −0.4 | −1.9 | 3.4 | −3.8 | −4.2 | −3.5 | −9.9 | 2.6 | 3.2 |
| $[M(CH_3COO)(H2O)_5]^{+}$ | −1.6 | −1.9 | −0.8 | −1.5 | −2.9 | −2.6 | −2.2 | −7.8 | 0.8 | 1.3 |
| $[M(SCH_2COO)_2]^{2-}$ | −11.3 | −16.8 | −19.3 | excl.[b] | −35.4[c] | −20.5 | −25.0[c] | −56.2 | 2.4 | 3.3 |

[a] The protocol used was RI-BP86/COSMO-RS. [b] "excl." denotes results that were excluded from analysis; Table 2. [c] Experimental stability constants were not available for these complexes.

conceptually simple, avoids parametrization, and as such is a significant improvement that gets the majority of the absolute values of precision of the relative metal-ion affinity below the threshold of 2 kcal·mol$^{-1}$, or even 1 kcal·mol$^{-1}$.

## 4. CONCLUSIONS

Throughout this work, we tried to give a complete account of our efforts aiming at quantitatively accurate predictions of the stability constants of metal ions in (bio)inorganic systems, using the modern methods of computational chemistry. Together with a careful benchmarking of quantum-chemical methods to obtain accurate gas-phase complexation energies carried out in our previous study,[5] this leads us to think that the computational protocol used in this study represents the current state-of-the-art of computational chemistry to treat the problem at hand (ab initio predictions of the "absolute values" of the stability constants). The only ingredient missing might be the extensive conformational sampling of all of the studied complexes, as has been mentioned in the discussion. However, the increase of computational demands to address this problem would be formidable, and we are not aware of a standardized protocol that would enable one to treat large sets of various complexes on equal footing.

Looking at the results presented in Table 2, one may conclude that straightforward application of the presented protocol leads only to a mediocre agreement between the experimental and theoretical stability constants at best and that this problem cannot be handled properly by contemporary theoretical chemistry. However, careful analysis of the trends in computed stability constants and systematic errors present therein enabled us to suggest a computationally sound and robust protocol for estimating the *relative* affinities of metal ions for the formation of complexes with ligands to within ~2 kcal·mol$^{-1}$ average accuracy (after removal of the systematic metal-specific shifts). This requires a single geometry that represents the structure in the water (solvent) environment. Using these, only a single RI-BP86 calculation with COSMO of product complex is required, as long as decomposition of the free energy into individual steps of the thermodynamic cycle is not desired; COSMO-RS solvation energy calculation is also required but comes at practically no computational cost. A large part of the protocol's taint is eliminated by simple calibration. The choice of the particular calibration values for removal of the metal-dependent shifts is based on a set of structurally simple systems, and its details do not fundamentally influence the results obtained.

The fitness of the method used for electronic structure calculation is significantly altered when it is to be combined with solvation energies calculated using the COSMO-RS protocol. A seemingly inappropriate (as judged by the accuracy of the gas-phase interaction energies) RI-BP86 functional performs, in conjunction with COSMO-RS, better than other methods not only in the accuracy of the relative affinities but also in the consistency of behavior in a wide range of systems of diverse character.

Admittedly, the protocol has a notable weak point. The question of structural representation of the system (e.g., the coordination of $H_2O$ molecules) is not easily addressed because the protocol is unable to reliably predict a correct conformation because of their inconsistent treatment. Partly, at least, this is due to the nonzero charge of the systems. However, there is no simple dependence of the magnitude of error, nor obvious remedy, available, leaving a direct comparison of the affinity of metal ions for different ligands still elusive to computational treatment, whereas the metal-ion selectivity for the particular site can be addressed reasonably well.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

Equilibrium geometries of all of the molecules studied. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: rulisek@uochb.cas.cz.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The project was supported by the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic (Project RVO: 61388963).

## ■ REFERENCES

(1) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760−768.

(2) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. *Acc. Chem. Res.* **2009**, *42*, 489−492.

(3) Rokob, T. A.; Srnec, M.; Rulíšek, L. *Dalton Trans.* **2012**, *41*, 5754−5768.

(4) Beck, M. T.; Nagypal, I. *Complex Equilibria: Stability Constants*; Halsted Press: New York, 1989.

(5) Gutten, O.; Beššeová, I.; Rulíšek, L. *J. Phys. Chem. A* **2011**, *115*, 11394−11402.

(6) Sigel, R. K. O.; Sigel, H. *Acc. Chem. Res.* **2010**, *43*, 974−984.

(7) Dudev, T.; Lim, C. *Chem. Rev.* **2003**, *103*, 773−787.

(8) Glusker, J. P. *Adv. Protein Chem.* **1991**, *42*, 1−76.

(9) Rulíšek, L.; Vondrášek, J. *J. Inorg. Biochem.* **1998**, *71*, 115−127.

(10) Lee, K. H.; Matzapetakis, M.; Mitra, S.; Marsh, E. N. G.; Pecoraro, V. L. *J. Am. Chem. Soc.* **2004**, *126*, 9178−9179.

(11) Williams, R. J. P. *Biometals* **2007**, *20*, 107−112.

(12) *Handbook of Metal−Ligand Interactions in Biological Fluids*; Berthon, G., Ed.; Marcel Dekker: New York, 1995; Vol. 2.

(13) Peacock, A. F. A.; Hemmingsen, L.; Pecoraro, V. L. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 16566−16571.

(14) Sudhir, P.-R.; Wu, H.-F.; Zhou, Z.-C. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1517−1521.

(15) Dudev, T.; Lim, C. *Annu. Rev. Biophys.* **2008**, *37*, 97−116.

(16) Kuppuraj, G.; Dudev, M.; Lim, C. *J. Phys. Chem. B* **2009**, *113*, 2952−2960.

(17) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(18) Ryde, U. *Curr. Opin. Chem. Biol.* **2003**, *7*, 136−142.

(19) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 1253−1272.

(20) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; John Wiley & Sons, Inc.: New York, 1997.

(21) Rulíšek, L.; Havlas, Z. *J. Phys. Chem. A* **1999**, *103*, 1634−1639.

(22) Rulíšek, L.; Havlas, Z. *J. Chem. Phys.* **2000**, *112*, 149−157.

(23) Rulíšek, L.; Havlas, Z. *J. Am. Chem. Soc.* **2000**, *122*, 10428−10439.

(24) Rulíšek, L.; Havlas, Z. *J. Phys. Chem. A* **2002**, *106*, 3855−3866.

(25) Rulíšek, L.; Havlas, Z. *Int. J. Quantum Chem.* **2003**, *91*, 504−510.

(26) Rulíšek, L.; Havlas, Z. *J. Phys. Chem. B* **2003**, *107*, 2376−2385.

(27) Kožíšek, M.; Svatoš, A.; Buděšínský, M.; Muck, A.; Bauer, M. C.; Kotrba, P.; Ruml, T.; Havlas, Z.; Linse, S.; Rulíšek, L. *Chem.—Eur. J.* **2008**, *14*, 7836−7846.

(28) Dudev, T.; Lim, C. *J. Phys. Chem. B* **2001**, *105*, 10709−10714.

(29) Dudev, T.; Lim, C. *Acc. Chem. Res.* **2007**, *40*, 85−93.

(30) Delgmann, P.; Schenk, S. *J. Comput. Chem.* **2012**, *33*, 1304−1320.

(31) Droghetti, A.; Alfé, D.; Sanvito, S. *J. Chem. Phys.* **2012**, *137*, 124303.

(32) Furche, F.; Perdew, J. P. *J. Chem. Phys.* **2006**, *124*, 044103.

(33) (a) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098−3100. (b) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785−789. (c) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652. (d) Stephens, P. J.; Devlin, F. J.; Frisch, M. J.; Chabalowski, C. F. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(34) Perdew, J. P. *Phys. Rev. B* **1986**, *33*, 8822−8824.

(35) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.

(36) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829−5835.

(37) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868.

(38) Hertwig, R. H.; Koch, W. *Chem. Phys. Lett.* **1997**, *268*, 345−351.

(39) Eichkorn, K.; Treutler, O.; Öhm, H.; Häser, M.; Ahlrichs, R. *Chem. Phys. Lett.* **1995**, *240*, 283−290.

(40) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(41) Andrae, D.; Haussermann, U.; Dolg, M.; Stoll, H.; Preuss, H. *Theor. Chim. Acta* **1990**, *77*, 123−141.

(42) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224−2235.

(43) Klamt, A.; Jonas, V.; Buerger, T.; Lohrenz, J. C. W. *J. Phys. Chem.* **1998**, *102*, 5074−5085.

(44) Eckert, F.; Klamt, A. *COSMOtherm*, version C3.0, release 12.01; COSMOlogic GmbH & Co. KG: Leverkusen Germany, 2011.

(45) Klamt, A.; Schuurmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.

(46) Jensen, F. *Introduction to Computational Chemistry*; John Wiley & Sons: New York, 1999.

(47) Martell, A. E.; Smith, R. M. *Critical Stability Constants*; Plenum Press, New York, 1974−1989.

(48) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. *Acc. Chem. Res.* **2009**, *42*, 489−492.