Charles University

Faculty of Science

Modeling of Chemical Properties of Nano- and Biostructures

Mgr. Ondrej Gutten

Metal-Ion Selectivity from Quantum-Chemical Perspective

Selektivita pro Ionty Kovů z Pohledu Kvantové Chemie

DOCTORAL THESIS

Supervisor: Lubomír Rulíšek

Prague, 2017

Prohlašuji, že jsem nepředložil práci ani její podstatnou část k získání jiného nebo stejného akademického titulu. Práci jsem zpracoval samostatně a uvedl všechny použité informační zdroje a literaturu.

Práce je postavena na čtyřech původních odborných článcích, kterých jsem spoluautorem. Můj podíl na prvoautorských článcích I, II a IV je větší část výpočetních experimentů a jejich analýzy. Mým podílem na článku III jsou výpočetní experimenty.

CONTENTS

Contents	Page No.

Acknowledgements

Abstract

Metal ions are a tempting tool for organisms thanks to the diversity of functions they have to offer, if they can be distinguished properly. Examining metal-ion selectivity computationally is challenging mainly due to complexity of electronic structure and solvation effects. A DFT-based protocol for predicting metal-ion selectivity of metal-binding systems was developed. The most essential part of the thesis is discussion of the magnitudes and sources of inherent errors, both for metal-ion complexes and small peptides. The thesis connects the work of four original papers. It includes computational and experimental benchmarks, a case-study validating the computational protocol for obtaining energetic and structural insights, and attempts applying the protocol to peptidic systems.

Abstrakt (CZ)

Ionty kovů jsou lákavým nástrojem pro organismy díky rozmanitosti funkcí, které mohou nabídnout, pokud jsou správně rozlišeny. Analýza selektivity pro ionty kovů je z výpočetního hlediska náročná, hlavně kvůli složitosti elektronické struktury a solvatačních efektů. Práce představuje protokol založený na teorii hustotního funkcionálu pro předpověď selektivity pro systémy vážoucí ionty kovů. Nejdůležitější částí práce je diskuse o velikostech a zdrojích chyb, a to jak pro komplexy iontů kovů, tak pro malé peptidy. Práce je postavena na čtyřech původných článcích. Obsahuje testování proti referenčním výpočetním a experimentálním datům, případovou studii ověřující protokol pro získání energetických a strukturálních poznatků a pokouší se aplikovat protokol na peptidové systémy.

Abbreviations

AAD Absolute Average Deviation

CFT Crystal-Field Theory

COSMO Conductor-like Screening Model

COSMO-RS Conductor-like Screening Model for Real Solvents

DFT Density-Functional Theory

GGA Generalized Gradient Approximation

HF Hartree-Fock

IW Irving-Williams

LFSE Ligand-Field Stabilization Energy

LFT Ligand-Field Theory

LSS Ligand-Specific Shift

MIS Metal-Induced Selectivity

MP2 Moller-Plesset 2nd order

MSS Metal-Specific Shift

GGA Generalized Gradient Approximation

QM Quantum-Mechanical

SIS Structure-Induced Selectivity

ZPVE Zero-Point Vibrational Energy

Metal-Ion Binding and Selectivity Concepts

Metal ions are the wizards of biochemistry. Despite their assignment to inorganic chemistry, they are essential for the organic to come to life - their unique abilities enable chemistry that is inaccessible to their organic counterparts.

The detailed list of roles that metal ions play in living organisms spans several scientific fields. This chapter does not attempt to cover it all. Instead, it should provide a basic overview of mechanisms relevant to metal-ion binding and the most important differences among them, with the focus on biochemical context.

Only a few prominent examples are presented in a hope of enticing the reader onto the journey of understanding the selectivity of metal ions.

1.1 Charge difference

The formal charge of an ion is expressed by its valence state. Often more than one valence state is available for metal ions.

In water, which is a powerful screening agent, charged species with low valence states are stable and semi-independent entities. The most common valence state of biologically relevant ions is I (Na^+, K^+, Cu^+) or II (Mg^{2+}, Ca^{2+}, Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}).

Higher valence states are associated with strongly bound ligands, either in catalytic centres (e.g. III and IV of manganese in the oxygen evolving center of photosystem II[1] or as oxoanions. For example, molybdenum enters the cell as a molybdate anion,[2] $MoO4^{2-}$; the oxygens may be exchanged for other ligands, but its VI oxidation state is maintained in all known cases.[3] Thus, its biological behaviour is more similar to e.g. sulphate anion than to fellow transition ions.[4]

The total charge of a binding site can be the determining factor driving its selectivity. For example, Li^+has been predicted to displace Mg^{2+}in binding sites with positive or neutral charge, but not in sites with negative charge. The stronger binding of a divalent cation competes with the smaller desolvation penalty of a monovalent cation.[5] The differences are pronounced in hydrophobic binding sites.

1.2 Effective Charge and Size

Although the total charge of two species may be identical, the locally concentrated charge (i.e. polar) of a ligand interacting at shorter distances with a smaller ion results in larger ionic contributions.

Ions with equal (formal) charge come in a variety of sizes. The ionic radius tends to decrease across the period, as the growing nuclear number contracts the electron clouds; e.g. r(Ca^{2+}) = 1.00 Å, while r(Zn^{2+}) = 0.74 Å. The ionic radius increases with periods (e.g. r(Mg^{2+}) = 0.72 Å). The relationship is somewhat more complicated going down the transition block.

Due to counteracting effect of relativistic contraction, these ions have similar sizes ($r(Cd^{2+})$ = 0.95 Å, $r(Hg^{2+})$ = 1.02 Å).[6]

In the case of some ion channels, the selectivity can be conferred by rigid construction of the pore. Rigidity is a complementary strategy to allostery (see Section 1.5), which ensures that no significant structural adaptation occurs upon binding of an ion and, thus, only the cognate ion can pass through favourably. This approach is utilized e.g. by $Na^+$and K^+channels. Changing the number of coordinating ligands and the size of the pore can ensure the selectivity for Na^+over K^+,[7] or vice versa.[8]

The channels can be optimized for the ion alone or its hydrated complex. Curiously, this can lead to a smaller channel being adapted to a larger ion (that is dehydrated), while a larger channel is optimized for an ion-water complex, such as in case of EEEE motifs that can be selective for both $Na^+$and Ca^{2+}.[9]

1.3 Geometry and Coordination Number

There are composite structural properties, encompassing cooperative factors of ligands (such as denticity) and ions. The most important of these is the (coordination) number and geometry of ligands. The geometry is particularly interesting for transition metal ions with partially filled d-orbitals. Understanding differences among their binding requires consideration of electronic structure.

1.3.1 Crystal and Ligand Field Theory

A first step in this direction is Crystal-Field Theory (CFT), which focuses on the ion by considering its electronic structure, specifically its d-orbitals, while thinking of ligands as point charges. The position of point charges represent the geometry of a complex. The differential interaction of the point charges with individual d-orbitals (each with a different angular dependence) accounts for geometrical preferences.

A textbook example is presented in Figure 1.1. of an For an octahedral

complex, originally 5-degenerate d-orbitals split into 2 levels of triple (t_2g - d_{xy}, d_{xz}, d_{yz}) and double (e_g - d_{z^2}, $d_{x^2y^2}$) degeneracy. The splitting can be rationalized by the position of orbital lobes. In the case of d_{xy}, d_{xz}, d_{yz} these point inbetween the ligands, resulting in lower electron repulsion than in the case of d_{z^2}, $d_{x^2y^2}$, where the orbital lobes point directly to the ligands and are, hence, higher in energy.
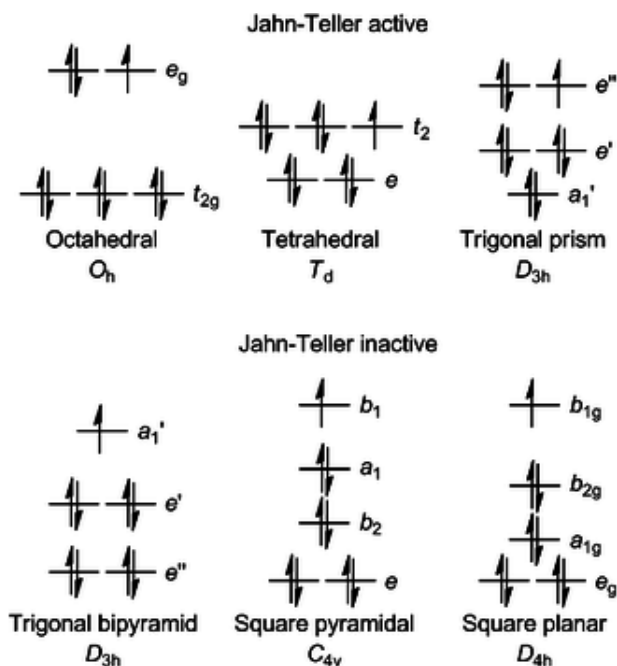
Figure 1.1: Splitting of d-orbitals in different geometries as predicted by CFT. Reprinted from Ref 10, Copyright 2013, with permission from The Royal Society of Chemistry.

Different geometries result in different ordering and energy gaps of individual orbitals. Ignoring differences in electron-electron repulsion, the energies of the (partially filled) d-orbitals show the amount of stabilization relative to a free ion. This difference is called Ligand-Field Stabilization Energy (LFSE) and it can help justify preferences of individual ions and geometries for each other.

An extension of CFT is the Ligand-Field Theory (LFT), which builds up on previous concepts by reconsidering ligands as orbitals, rather than just

point charges. This allows to distinguish different types of ligands by how their orbitals combine with those of the transition metal ion. The ligands are classified as σ-donors, π-donors and π-acceptors. The overall donating/accepting ability of a ligand is expressed by its position in spectrochemical series.

While the degeneracies obtained from LFT are not split any further, the energy gaps between individual manifolds are affected. This helps to rationalize e.g. the magnitude of the energy split between t2g and eg levels in Figure X and properties like spin states, excitation energies (color) of the complexes, or .

1.3.2 Coordination number

Coordination number is affected by size of both the ion and its ligands. The effect of open d-shells has been discussed above. However, even for d^0 or d^{10} the relationship to the coordination number is not straightforward.

For example, Mg^{2+} is only slightly smaller than Zn^{2+}, yet their most common observed coordination numbers are 6 and 4.[11] Even more surprisingly, Hg^{2+}, which is larger than Zn^{2+}, prefers coordination number of 2!

The subvalence structure of the ion adapts significantly upon binding of the first ligand.[12] This adaptation is more favourable than binding of extra ligands, overshadows the intuitive sterical argument. Its origin can be traced to importance of relativity for heavy metal complexes.[13]

1.4 Irwing-Williams Series

We limit, for a moment, our attention to transition metal ions in octahedral geometry. The stability constants of octahedral complexes follow a general trend:

$$Mn < Fe < Co < Ni < Cu > Zn$$

Although the differences vary, this trend is valid for a large range of ligands. This implies that the reason pertains to the properties of the metals.

This trend is usually justified in terms of ionic radius and LFSE. LFSE is the largest for Ni(II) and decreases again for Cu and Zn. The increased stability of Cu(II) is explained by the Jahn-Teller distortion of axial ligands.

However, this exaplanation cannot justify e.g. weak stability of Mg complexes, relative to Zn^{2+} with very similar ionic radius and identical (zero) LFSE. Moreover, the Irving-Williams (IW) series can be recognised in other geometries as well.

The charge accepting ability of an ion can be related to its ionization potential. The interaction can be intuitively rationalized as a partial repayment of the ionization cost by regaining fractional charge from the ligand.[14] , which hints at the importance of covalency. This link has been been confirmed by modern theoretical studies, either directly in terms of covalency[15] or in terms of charge transfer to the metal ion.[16]

The Irving Williams series has important implications for selectivity of metal ions. The challenge of correctly selecting the weaker binding ions and the strategies used are discussed in further sections.

1.5 Allostery

The immediate interaction of an ion with the binding site is in the spotlight of this work. In biological context there are, however, further concepts that are actively exploited in ion recognition. Although their detailed discussion is outside the scope of this writing, introducing them does serve to highlight the importance and limitations of affinity concepts introduced above.

The central theme of metal ion selectivity is the IW series. How can a generally weakly binding ion be selected against its competitors? The concepts of affinity described above does allow for limited perturbation of the IW series. However, the variation is not sufficient on its own and other concepts are exploited.

Allostery is an essential trick of metal ion homeostasis exploited in transporting, sensing, and chaperoning proteins.

Proteins could be considered impractically large, should binding of an ion be their only function. However, their complex structure allows for complex

structural changes as a result of binding the correct ion. In other words, **it is not the strength of binding but the fine structural details of the binding** unique to the cognate ion that elicit the response (signalling, ion release, ...). Allosteric response allows to shift the recognition from energetic grounds of the binding site to structural grounds of the entire protein. While this does solve the problem of recognition, it does not prevent non-cognate ions from binding to an active site.

A well-studied example of a protein binding non-cognate ion is a transcriptional regulator NikA. Other ions, such as Co^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+} bind to NikA as well, with affinities that observe the IW series. Only Cu^{2+}, which is the only one that binds in a similar geometry as Ni^{2+}, manage to induce the conformational changes necessary for the activity of the regulator.[17]

1.6   Abundance

Abundance of individual components affects the state of equilibrium. Maintaining the abundance of the metal ion is the very nature of homeostasis itself.

A particularly interesting observation is that the typical concentration of metal ions in a cell is inversely related to their position in IW series. Ions that bind less strongly are present in higher concentrations (e.g. Mg is found in mM concentrations) while free Cu is effectively absent.

Rather than defeating IW series in terms of absolute affinities, **cells match the baseline concentration of ions to the series itself**. Thus, even modest relative variations of affinity result in preferential binding of the given metal. This makes thermodynamically-driven assembly of metalloproteins a possibility in some cases.

This ingenious adaptation arises naturally from the IW series itself. The sensing (and efflux) proteins are subject to this trend as well. Strongly binding ions, such as e.g. copper, can activate respective sensors at lower concentrations, leaving only none or only a few to compete with the weaker binding ions, such as Mn^{2+} and Fe^{2+}.[18] Thus, both positive (optimization of binding site) and negative (elimination of competitors) selection is in play.

The homeostasis of ions does not intertwine only the abundance of ions, but also the abundance of sensors. The overall delicate balance thus only makes sense in the context of the entire cell and even species. This was demonstrated by overexpressing one of the sensors - e.g. doubling the amount of expressed Fur, an iron sensor, which shows only a weak preference for Fe(II) compared to Mn(II), resulting in Mn-induced iron response.[19] Similar affinity for both ions leaves the number of sensors as a determining factor of a mismetallation event. The balance thus may be upset not only by extreme ion fluctuations inside the cell, but also of its native sensors.

Specificity in Metal-ion Homeostasis in Biological Systems

Availability of metal ions has significant influence on the survival of an organism. Metal-ions are essentially "vitamins", as they cannot be synthesized from other precursors. The apparent concentrations necessary for a cell are orders of magnitude higher than what can be found in a "typical" environment. Thus, the organisms need to concentrate ions from the environment. On the other hand, very high concentrations can easily damage the delicate mechanisms of a cell due to unexpected, non-specific interactions. The amounts of metal ions need to be neither too low or too high, with the required amounts varying across different ions, compartments of a cell, and species.

There are both multiple roles individual metal ions can fulfil and multiple ions that can fulfil a specific role. After all, if the desired effect is e.g. acidification of a water molecule, the important result is the hydroxyl anion, not the mechanism of achieving it. There are cases where the protein can effectively utilize more than one ion, not only *in vitro*, e.g. Co^{2+}-substituted carbonic anhydrase II,[20] but *in vivo* as well. There are multiple examples of Fe^{2+} replacing Mn^{2+},[21] both ions acting as interchangeable Lewis acids. The redox properties of the two ions are quite different. Under conditions of

oxidative stress, when iron presents a threat of undergoing deleterious Fenton chemistry, the bacteria can switch from Mn-dependent to Fe-dependent enzymes, such as super-oxide dismutase.[22]

The interchangeability is more of an exception than rule. Energetic "details" determine the efficiency of an enzyme, structural "details" of metal-ion binding can determine the difference between activation and non-responsiveness of a signalling protein. The importance of sensitive, selective, and effective metal-ion homeostasis is apparent.

The basic processes of metal ion homeostasis are presented in Figure 2.1. The picture is schematic and foreshadows strategies of ensuring proper levels of metal ions in a cell. The strategies are discussed in the following sections, with a focus on specificity and interplay for individual ions.

2.1 Sensing, Uptake and Efflux

Sensing At the very heart of the homeostasis lie the sensing mechanisms. These represent the cell's senses that allow it to make decisions about steering the other processes in order to maintain homeostasis. Sensors need to exist for every metal-ion relevant to the cell - which includes many non-native ions, such as Cd, Pb, Hg,[23] and others.

The recognition of cognate ion is achieved by adaptation of unique geometry, which drives the allosteric changes. The bound metals act as corepressors in case of metal-uptake regulating proteins, or as derepressors/activators in case of metal-efflux regulating proteins.[24]

Each sensor is expected to be specialized for its own cognate ion to the exclusion of all others.[25]

Monitoring of levels of native ions can be achieved through their intermediate products, such as haem sensors[26] or iron-sulfur cluster sensing proteins.[27]

Interestingly, metal ions can also be senesd by riboswitches, RNA sequences regulating RNA synthesis or translation.[28]
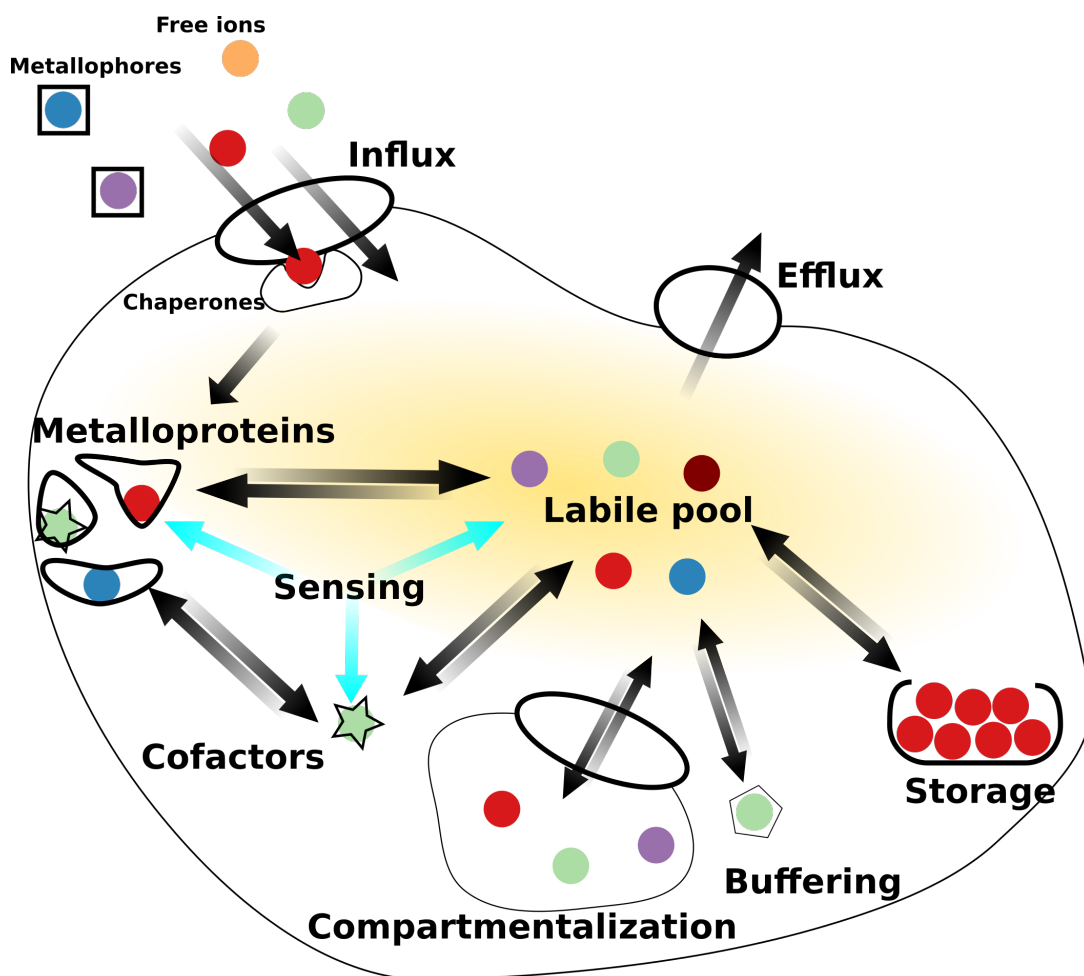
Figure 2.1: Schematic representation of major metal-ion processes of homeostasis

Uptake & Efflux The most principal tool in effecting metal ion homeostasis are membrane transporters. Due to metal ion charge they do not cross membrane spontaneously. The transport through membrane can occur passively through ion-channels, down the electrochemical gradient.

However, the desired concentrations are usually orders of magnitude higher than those found in a typical environment.[29] The import of ions is effected against the concentration gradient at the expense of energy or is coupled to a simultaneous, more favourale transport. Once this cost is paid, the impermeability of membranes allows holding up to the precious resource. This is

advantageous not only for achieving homeostasis but also for compartmentalization of intracellular space, which allows for more complicated setups.

Based on many structures of metal-binding domains, the selectivity of transporters is primarily dictated by a first-shell coordination sphere. For example, a Mn^{2+}-specific transporter might employ higher coordination numbers and prefer carboxylate ligands,[30] compared to a Zn-specific transporter.[31]

This adaptation of a binding site is necessary but not sufficient for required specifity. The first-shell preference is further amplified by allosteric response, which is required for the release of the ion from the binding site into the channel and across the membrane.[32]

Binding of non-cognate metals is not merely a laboratory exercise. A fascinating use of this concept is observed in host-pathogen interactions. Flooding a pathogen with zinc ions leads to inhibition of manganese uptake, which are essential for these organisms.[33] The zinc ions bind more strongly than the manganese ions, which results in a closed form of a protein, which does not effect the transport, effectivey clogging the manganese uptake pathway. Binding of cognate manganese, on the other hand, leaves the part of a protein flexible, which is necessary for the conformational change.[34] Manganese-starved cell is left vulnerable to oxidative stress.[35]

2.2   Internal Management

While controlling import and efflux of ions does provide basic functionality, a higher level of flexibility is beneficial to any living organism. First, protection from fluctuations of ion concentrations both outside and inside the cell is essential. Second, stockpiling in times of plenty can aid survival when resources become more scarce. Both aspects can be addressed by various forms of reversible depositing, which come at different levels of energetical cost, speed and ion affinity.

2.2.1 Bufferring

is achieved by maintaining presence of chelating agents (Lewis bases) that form labile complexes with the ions and proteins. Bufferring helps to maintain homeostasis without responding with potentially slow and expensive changes of cross-membrane transport.

The buffering is particularly important for ions high up IW series, as these can bind unspecifically and have deleterious effects. Bacterial cells have an overcapacity to bind zinc, by employing small thiol molecules[36] or metallothioneins.[37] This leaves virtually zero free ions, despite the total concentration being in almost milimolar concentrations.[29]

2.2.2 Storage

is a reversible removal of an ion from the labile pool. The best understood example is ferritin, which is a storage vessel for iron oxidized and deposited in form of ferric oxide. This speckle of rust can contain more than 2000 iron atoms.[38] To put this in perspective, an apparent iron concentration of 10^-4 M for a bacterial cell of volume 10^-14 l corresponds to a total quota of 6.10^5 iron atoms. Mere 300 copies of bacterioferritin can fully cover the requirements of such a cell.[29]

In a broader sense, storage includes the management of number of copies of functional metalloproteins. Apart from performing their function, these proteins also act as a "living" reservoir of a corresponding metal ion. Rather than maintaining the ion quota at fixed values, an organism can alter the demand for the ions.

For example, in case of iron deficiency, or under conditions of oxidative stress, an iron-dependent super-oxide dismutase may be exchanged for a manganese-dependent version.[22]

The breakdown of metalloproteins and their replacement with analogues that have different metal requirements can even be prioritized over modifying uptake/efflux, as that can, in some cases, entail costly construction/elimination of ion flux machinery.[39]

In an extreme cases of redirecting demand, organisms like *Lactobacillus acidophilus* dump the need for Fe altogether.[40]

2.2.3   Compartmentalization

allows maintaining different concentration arrangement in a single cell. In the broadest sense, it is what defines ion homeostasis - i.e. specific ion concentrations in a cell (as the essential compartment) that are different from those in the environment. In a less trivial example, compartmentalization can help deliver different cognate ions to proteins with otherwise identical binding sites, as is the case for a pair of Mn^{2+}- and Cu^{2+}-dependent proteins (opposite ends of IW series).[41]

Eukaryotic cells employ multiple cellular compartments that offer various adaptations. Some prominent examples might include

.

2.3   Indirect Metal Binding

Once a metal-ion is recognized, the recognition may shift to a metal-bound product. These can be transient binding agents (metallophores), delivery proteins (chaperones), or intermediate metal products (cofactors). These cases are an important part of metal ion homeostasis. However, they are not particularly interesting for understanding metal-ion recognition, as the problem is relayed from specificity of immediate interactions to those of a metal-bound complex.

Metallophores   are compounds capable of forming complexes with metal ions. Metallophores are an alternative to direct metal-ion uptake, which can have a couple of advantages.

First, metallophores can be deliberately produced and released by an organism scavenging the environment for the precious resource. Not limiting the binding process to peptide-like ligands can provide a better binding fit with both higher affinity and selectivity. Second, producing a signature met-

allophore gives advantage by masking it from other competitors and being the only one in town capable of recognizing the complex. The chemistry of iron binding complexes - siderophores - is particularly rich,[42] but metallophores have been described for a range of other ions.[43]

Chaperones are proteins that deliver metal-ion directly to another, final and functional recipient of a metal ion. Chaperones are known to participate in assembly of proteins that incorporate copper, nickel, and likely zinc.[24] all of which are high up Irwing-Williams series. These ions are kept at very low concentrations in a cell. The need for chaperones may be related to low availability of free form of these ions as well as propensity to potentially damaging interactions even with sites unoptimized for these proteins. While the chaperones are respectful of ion's binding preferences, the directionality of the ion transfer is governed by interaction of the chaperone with its partner (ion uptake complex, target protein).[44]

Quantities and Physico-Chemical Considerations

The ion homeostasis is fascinating for its complexity, ingenuity, interconnectivity. The first step in addressing this topic by quantum chemistry is translating these concepts into measurable and calculable physico-chemical quantities. Establishing these allows us to construct and test computational models, which is essential for drawing mechanistic conclusions.

3.1   Observable Quantities

To shift from qualitative to quantitative discussion we introduce the quantities useful for discussion of ion-ligand interactions.

The basic observable is a stability constant, β, defined as ratio of bound and unbound concentrations at given conditions.

$$\beta = \frac{a_{bound}}{a_{unbound}} \approx \frac{[ML]}{[M][L]} \tag{3.1}$$

where a stands for activity, and subscripts M and L represent the unbound metal ion and L a general set of ligands or a binding site, respectively. The latter part of the equation approximates the activity coefficient to be equal to 1. The set of ligands L constitutes some number of potentially dif-

ferent ligands. If sequential binding of these ligands is of interest, a step-wise binding constants (i.e. equilibrium of binding $(n+1)^{st}$ ligand to a complex of n ligands) may be of interest.

$$\beta_{n+1} = \frac{[Ml_{n+1}]}{[Ml_n][l]} \tag{3.2}$$

where l signifies a single general ligand, rather than a set of ligands. An overall stability constant is then simply a product of step-wise constants.

$$\beta = \prod_{i=1}^{n} \beta_i \tag{3.3}$$

Referring to $log\beta$ is more common, since stability constants span many orders of magnitude.

For discussion of metal ion selectivity, we are interested in preference of binding one ion over another. The relative stability constant for two ions is defined as a ratio between stability constants of the ions in question.

$$\beta_{M_1,M_2} = \frac{[M_1L][M_2]}{[M_2L][M_1]} \tag{3.4}$$

The relative stabilities tell us nothing about absolute stabilities. For example, the system can be highly selective, yet bind both ions very weakly.

3.2 Calculable Quantities

To relate these observable quantities to those we can compute we introduce the following relation

$$\Delta_r G = \Delta_r G^{\ominus} - RT \ln \frac{[ML]}{[M][L]} \tag{3.5}$$

$\Delta_r G$ stands for reaction Gibbs free energy, $\Delta_r G^{\ominus}$ stands for standard reaction Gibbs free energy, R is a gas constant (8.314 J K^{-1}mol^{-1}) and T is temperature.

A spontaneous reaction occurs if $\Delta_r G$ is negative. The direction of the reaction thus depends on the chemical properties of the species (expressed

collectively in the $\Delta_r G^\ominus$ term), as well as their relative amounts (expressed by the logarithmic term).

Calculation of the $\Delta_r G^\ominus$ term is the central subject of the computational efforts of the Papers I-IV.

Affinity and Allostery , discussed in Chapter 1 as separate strategies for ensuring selectivity, pertain to this term. However, the meaning of the "reaction" is slightly different in each case.

Affinity refers to the properties of the immediate binding site and the "reaction" is the process of binding of an ion. Allostery refers to the properties of an entire system (usually a protein), with the local details of binding at a site being amplified and translated into significant structural changes at a different part of a protein. The reaction is in this case not only the binding of an ion but the assorted structural change of a protein as well.

The importance of this selectivity mechanism is that the affinity of the binding site for a metal ion might be lower for the cognate metal ion but this is compensated by $\Delta_r G^\ominus$ of the structural change in the presence of the cognate ion only. The associated $\Delta_r G$ might still be more negative for binding of a non-cognate ion; but the binding of a cognate ion in conjunction with the structural change is unique to the cognate ion. This is the case of competitive inhibition of Mn uptake by Zn ions described in Section 2.1.

Abundance , also described in the previous chapter, refers to the logarithmic term in equation 3.5. Rather than optimizing the binding site and/or the whole protein, this strategy directs the relative stability by optimizing relative amounts of participating species. This relates both to the effective abundance of the ions (i.e available metal-ion concentrations that are inverse to the IW series) and the receptors (ligands - as exemplified by observed ross-talk upon overexpression of one of the sensors, Section 1.6).

Computational Considerations

Computational chemistry is the art of making approximations.

The primary interest of this thesis is the binding of metal ions to a set of ligands. These are represented explicitly in our calculations and should be represented as accurately as possible. Quantum chemistry offers a myriad of methods and their proper definition is left for the textbooks. Section 4.1 introduces only some of the concepts relevant for the discussion.

Metal-ion complexes are often highly charged species that interact strongly with their environment. The problem of solvation is addressed in 4.2. While the solvent can be with some degree of success approximated by average behaviour, this is unlikely to be practical in the case of proteins, i.e. by neglecting parts of the protein that do not directly bind the ion.

Even in our our minimalistic model multiple rotational, translational, and vibrational states can be considered. Frequency analysis of a system also provides estimate of zero-point vibrational energy.

The three major terms mentioned above give rise to the following definition of free energy of a model system:

$$G = E_{gp} + G_{solv} + ZPVE - RT \ln Q \qquad (4.1)$$

where E_{gp} stands for gas-phase electronic energy G_{solv} stands for solvation free energy, $ZPVE$ stands for zero-point vibrational energy and $RTlnQ$ accounts for thermal corrections.

4.1 Quantum Chemistry

The aim of this work is to observe the combined behaviour of existing methods, rather than development of new ones. An in-depth description of quantum-chemical methods is considered outside the scope of this work. This section merely aims to introduce basic concepts that will be relevant for the later discussion.

4.1.1 Wave-Function Methods

Hartree-Fock (HF) method is the entrance gate to the world of quantum chemistry. An excellent step-by-step introduction to this method is presented elsewhere.[45]

The basic idea of the method is the expression of wave-function in terms of one-electron functions - orbitals - that do depend on each other, but only through the overall electron cloud formed by the other orbitals. The error introduced by this approximation is dubbed the dynamical correlation.

HF uses single Slater determinant for constructing wave-function. This is often insufficient. In transition metal systems this situation occurs commonly as a result of low-lying excited states. The difference between the proper description and a single-determinantal HF description is dubbed static correlation.

A lot of the interesting chemistry, unfortunately , happens thanks to these missing parts. The correlation is introduced by post-HF methods, typically in terms of the same one-electron functions. Such extensions are formally 'straightforward' but represent a rather impractical choice for describing dynamical correlation, which is inherently a multi-electron phenomenon, and converge relatively slowly.

A Born-Oppenheimer Hamiltonian uses only kinetic and Coulomb operators that have their classical interpretation. However, the requirement of anti-symmetric behaviour of fermionic wave-functions gives rise to a non-classical term - the exchange, K. The exchange term is a result of zero prob-

ability of two electrons having identical positions (and spin) at any given time. **The exchange term is inherently non-local.**

4.1.2   Density Functional Theory

Wave-function describes the system at hand completely. However, even in the approximate description by quasi-independent one-electron functions, it is a function of 3N spatial and N spin variables. Instead, electron density (a function of mere 3 spatial coordinates) can be used, without, in principle, losing any information;[46] a proper introduction to the formalism can be found elsewhere.

However, the connection between the wave-function and the density is of an unknown form. The challenge lies in reformulating the terms that contribute to energy as functionals of density - rather than functions of electron coordinates. The terms with unknown form are exchange, correlation, and kinetic energy. The world of DFT draws its richness from various attempts at expressing these terms.

Exchange and correlation are inherently non-local.[47] The DFT functionals are often classified based on the level of non-locality they employ. Local density approximations use only the value of density, while Generalized Gradient Approximation (GGA) functionals employ the first derivatives of density; meta-GGA functionals include first and second derivatives. All these functionals are called 'pure', as opposed to 'hybrid' functionals, which include some percentage of exact exchange from HF theory.

The list of functionals relevant for this work, along with the amount of exact exchange they use, is shown in Table 4.1.

The exchange functional should cancel out the Coulomb repulsion functional in one electron systems. The notorious failure of exchange functionals to do so results in self-interaction error. On the other hand, it is recognized that these functionals incorporate some static and dynamical correlation as well.[48]

Balancing DFT exchange functionals with HF exchange thus creates a

functional	BP86 PBE TPSS	B3-LYP	M06	MPWB1K	BH-LYP	M05-2X
exchange(%)	0	20	27	44	50	56

Table 4.1: DFT functionals used in Papers I-IV and the amount of HF exchange they use.

trade-off of reducing the self-interaction error at the expense of decreased correlation.

It may be expected that bonding in transition metal complexes will involve non-negligible amount of static correlation and, thus, smaller amounts of HF exchange would be desirable, relative to main group chemistry.[49] However, the relationship is more complicated, as evidenced by benchmark[50] and many case-studies.[51]

Clearly, there are multiple factors (such as the character of the systems and the quantities that are calculated) that play role and no single answer, without significant improvement of functionals in use, can be considered definitive.[52]

4.2   Solvation

While theoretical gas-phase chemistry has reached level of accuracy that is comparable to experiment, most of the interesting chemistry happens in solution. Solvent can have a profound effect on reactivity, energetics and mechanisms and cannot be ignored even for qualitative trends.

Explicit treatment of solvation should in principle provide the most realistic account of solvation. The resulting large number of degrees of freedom is treated by statistical averaging of the sampled space. Various flavours of molecular dynamics and Monte Carlo simulations offer effective ways of importance sampling in these multi-dimensional problems. While this has allowed for impressive advances, the problem of extensive sampling remains

a limiting factor.[53]

As a result, this approach is dominated by force-field and semiempirical quantum-mechanical energy functions, which often resort to functional descriptions that are too simplistic to describe the chemistry, especially in case of transition metals. Although ab-initio molecular dynamics (usually using Density-Functional Theory (DFT)) became feasible for simple systems and can provide promising results, e.g. in calculation of pKa[54] or analyzing solvent-metal interactions[55] , the application to metal-ligand binding is conceivable[56] but still rather impractical.

An attempt to locate middle ground between the two extremes are the multi-scale approaches, which combine QM an MM world by separating the system into layers that are described by increasingly more approximate methods.[57]

The inner layer consists of chemically active and electronically correlated parts and is treated by QM methods. The inner core is supported and polarized by outer layer that is typically treated on force-field level. Coupling of the two layers has to be carefully accounted for and raises several problems inherent to QM/MM methods. This approach is especially popular in enzyme mechanistic studies.

Some of the major problems of the aforementioned methods is the need for sampling and inappropriate representation of interactions between the system of interest and its surroundings. This is especially true for the water as a solvent, which has the ability to strongly polarize its solutes.

4.2.1 Polarizable-Continuum Models

The interaction of solute with solvent is traditionally split into several contributions. The effect of solvent is mainly electrostatic but other effects are relevant as well. These solvation energy is traditionally split to:

$$G = G_{el} + G_{dis} + G_{cav} + G_{rep} \tag{4.2}$$

G_{el} corresponds to electrostatic contribution, G_{dis} stands for dispersion interaction, G_{cav} represents the formation of a cavity within a solvent and

G_{rep} includes local changes to the structure of a solvent.[58]

This division is both phenomenological and situational. The separation is intuitive, although the terms are not strictly separable and measurable on their own. The most important is the electrostatic term and this fact is recognised by a class of approaches called Generalized Gradient Approximation (GGA). The remaining terms are usually estimated based on area of surface segments and atomic surface tensions with parameters fitted to numerous experimental data sets.[59] These models immerse the solute in homogeneous polarizable continuum. Both the continuum and the solute are polarized by the interaction.

The solute is first placed in a cavity, that is usually defined by scaled Van der Waals radii of the atoms of the molecule. The scaling (enlargement) is used to diminish problems with charge density reaching significant values outside of the cavity.

Outside of the cavity, we assert presence of a dielectric continuum by introducing solvent reaction field. The total electrostatic potential is subject to Poisson equation:

$$\nabla^2(V_{solute} + V_{rf}) = 0 \qquad (4.3)$$

The electrostatic potential comes from the solvent reaction field, V_{rf}, and the charge distribution of the solute, $V_{solute}(\vec{r}) = \int d\vec{r'} \frac{\sigma(\vec{r'})}{\vec{r}-\vec{r'}}$. The Poisson equation is a general relation that has to be valid regardless of the presence of the continuum.

The presence of a dielectric can be accounted for in different ways. An illustrative approach is to invoke the boundary condition that the electric field is continuous across the cavity surface. The electric field outside of the cavity is scaled by the dielectric constant, ϵ, of the solvent. In this way, the continuum represents the solvent by its macroscopic property.

$$\left(\frac{\partial V}{\partial n}\right)_{in} = \epsilon \left(\frac{\partial V}{\partial n}\right)_{out} \qquad (4.4)$$

where $\vec{n}$ is a unit vector pointing outside of the cavity.

A popular way of implementing these relations is by introducing apparent surface charge density on the surface of the cavity. The problem of complex

cavity shapes is solved by discretizing the surface to small fragments, which have their charge density represented by a single point charge. The magnitude of the charge is related by the above condition to the potential produced by solute charge density (obtained from Quantum-Mechanical (QM) calculation) and other surface charges. The set of self-referential linear equations can be solved iteratively to produce surface point charges, which are then used in the next QM cycle to polarize solute's charge density in a natural way.

Thus, the implementation requires merely a construction of a cavity, solving a set of linear equations and a mild increase in the number of SCF cycles.

4.2.2 Conductor-like Screening Model

There are numerous variations of the method outlined above. One worth mentioning in this work is the Conductor-like Screening Model (COSMO) approach, which derives the surface charges in a slightly different way.

The continuum in COSMO is a conductor. Instead of invoking the boundary condition the conductor implies zero electrostatic potential on the surface of the cavity. Thus, the surface charges are required to negate the potential of the solute charge density and other surface charges. The resulting charges are then scaled:[60]

$$f(\epsilon) = \frac{\epsilon - 1}{\epsilon + 0.5} \qquad (4.5)$$

which recovers the charges for a given value of dielectric constant. This approximative approach is sufficiently accurate even for low values of ϵ (2).[61] Thus, this approach is safe for water (with dielectric constant of ϵ 80).

The advantage of the above approach is its inherent simplicity. Moreover, long range polarization is included exactly (within the formulation of the model) without invoking ad hoc cut-offs or inhibitively large system sizes. The main problem is the simplistic representation of the solvent by a single macroscopic parameter - any structural details are ignored. Moreover, the model is inherently incapable of describing interactions that involves charge transfer between solvent and solute.

4.2.3 Conductor-like Screening Model for Real Solvents

The problems above are, to some extent, addressed by an extension of COSMO model called Conductor-like Screening Model for Real Solvents (COSMO-RS),[62].[61] The computational cost of this improvement is negligible and it is used heavily throughout this thesis.

An auxilliary quantity generated during the COSMO calculation is the set of charges located on a surface of the solute cavity. In a nutshell, COSMO-RS is a quantification of interaction of the surface segments of a solute with the surface segments of a solvent.

The starting point for COSMO-RS treatment are COSMO calculations for both the solute and the solvent. Thus, solvent and solute are treated at the same level of theory, while the COSMO model (and other PCM-like models as well) work with dielectric constant, ϵ, as the only reference to the nature of a solvent.

The segments and charges are assembled to a histogram, called a sigma-profile, shown in Figure 4.1.
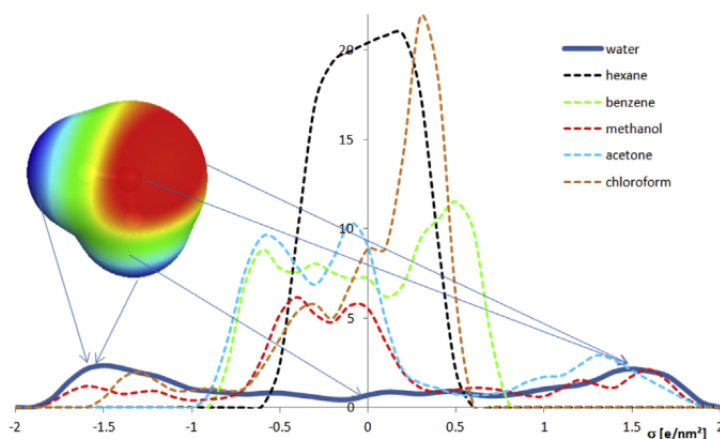

Fig. 1. σ-Surface of water and σ-profiles of water and 5 other solvents.

Figure 4.1: Examples of $\sigma-$profiles - charge density distributions. Reprinted from Ref 63, Copyright 2016, with permission from Elsevier.

The next step is the computation of chemical potential of a segment in a solvent environment. This chemical potential is related to energy of interaction with other segments, the relative frequency of the other segments (the σ-profile) and their own chemical potential.

$$\mu(\sigma) = kT \ln[\int d\sigma' p(\sigma') \exp{(E(\sigma, \sigma') - \mu(\sigma'))/kT}] \qquad (4.6)$$

with $\mu(\sigma)$ referring to the chemical potential of a segment with charge density σ, $p(\sigma)$ is the σ-profile of the solvent and $E(\sigma, \sigma')$ is the interaction energy of segments with respective charge densities σ and σ'.

There are several terms contributing to the energy of interaction of two segments. The major term corresponds to the electrostatic interaction of the charge densities of the two surfaces. So called hydrogen-bonding terms is treated separately. It is applied to sufficiently polar surfaces and accounts for charge transfer that is not included in the electrostatic term. Atom-specific parameters are used to reflect Van der Waals interaction between segments. COSMO-RS includes other terms that reflect structural details of the species studied. These are corrections that aim to treat problematic cases.

As in the case of calculation of apparent surface charges, this is a self-referential set of linear equations, which are very easy to solve. The result is a σ-potential, i.e. a chemical potential of a segment as a function of its charge density.

The chemical potential of the solute in a solvent is a weighted average of chemical potentials of its segments in the environment of a solvent.

Using only the auxilliary quantity of surface charge densities, a much more physical description of solute-solvent interaction has been obtained. The surface segments are treated as individual entities, which does not reflect the three-dimensional structure of the molecules and may lead to oversolvation. This is corrected for only partially.

CHAPTER 5

Optimization of Computational Protocol for Studying Metal-Ion Binding and Selectivity

Studying metal ion-ligand interactions is relevant and exciting not only for its biological implications. Moreover, contemporary computational chemistry has a potential to address these problems in a way that can provide insights both competitive and complementary to laboratory science. The specific challenges that have to be faced are put into perspective on the following pages. The details can be found in the attached Papers.

Systematic studies focusing on metal-ion selectivity[64,65,66] or on conformational exploration of ion-peptide systems[67,68] have been conducted before, but have avoided incorporation of solvation effect.

The main theme of this doctoral thesis is setting up a computational protocol that can address questions of metal ion-ligand and ion-peptide binding and selectivity in water environment in an effective, efficient and flexible manner.

For a computational chemist an observed error should not be a nuisance, but a core object of interest. Understanding the sources and the dependence

of errors on the decisions made throughout the computational experiment lies at the very heart of "applied" quantum chemistry.

Papers I and II focus on the most technical aspects of this quest and understanding the impact of individual choices on the ability of a computational protocol to reproduce absolute and relative stability constants of model complexes. Paper III attempts to further validate the computational protocol from a slightly different perspective, by looking at conformational free energies of copper-cyclam complexes. Paper IV attempts to transfer the computational protocol for use in metallopeptidic systems and explore to what extent is the selectivity determined by the immediate environment of the binding site.

5.1 Paper I & II

Understanding ion-ligand interactions quantitatively requires accurate description of these systems. There are, conceptually, two main challenges.

The first is to select a model that faithfully represents the system of interest. An obvious part of this challenge is ensuring the computer model corresponds with the physical reality - such as identifying the correct binding mode or relevant reference states. However, all models can be considered as a multi-scale representation of reality. The models will difer in where we draw the lines between individual layers and what we put inbetween the lines. This aspect becomes especially relevant when treating solvent species, deciding on what to consider explicitly, implicitly, and what to ignore altogether.

The other challenge, not truly separable from the first one, is accurate quantum-chemical description of the chosen model. The outcome of this description should be computable quantities that are comparable to benchmarks and/or experimental values.

The field of quantum chemistry has advanced to a point where description of gas-phase low-temperature isolated systems combats the accuracy of description of these systems by experiment. Focusing on purely computational benchmarks is attractive due to elimination of a number of factors, allowing direct and clean comparison of accurate benchmark methods with computationally cheaper methods. Data obtained through CCSD(T), considered a golden standard in the realm of single-reference systems, represent accessible benchmarks.

Ultimately, however, our interest lies in metal ion complexes in solvent environment, as only these are relevant for undestanding the mechanisms of binding and selectivity in biological systems or for most of laboratory chemistry. Experimental data for these systems represent the more relevant benchmarks. It remains, however, suspect to numerous factors such as approximate treatment of solvent, ionic strength, pairing the model with physical reality, non-zero temperature effects, and experimental deviations.

Both benchmarks can provide valuable insights. Hence, there is no reason not to be greedy. The gas-phase benchmarks are discussed in Section

5.1.1. Complex analysis with reference to experimental stability constants is addressed in 5.1.2.

Establishing and understanding the strengths and weaknesses of a computational protocol for metal ion complexing systems is the main objective of Papers I and II. Both requirements, i.e. choosing a proper model and its accurate description, are addressed.

5.1.1 Gas-Phase Benchmarks

Work in Paper I   In Paper I, five model systems are considered. Each system has 2-6 simple ligands. The set of ligands includes water, ammonia, acetate, methylthiolate, imidazol or phenol. These are bound to each of four divalent ions - Cd, Zn, Cu, Fe. Reference systems with corresponding number of water molecules (only) are built as well.

Single-point energies are calculated for these systems by a wide range of methods. CCSD(T) is used as a reference. Hartree-Fock, MP2, SOS-MP2, SCS-MP2 are included as representatives of wave-function methods. A variety of, generally computationally cheaper, DFT functionals represent majority of the tested methods.

Apart from single-point calculations, a subset of the systems is tested for sensitivity of equilibrium structure to the choice of method used for optimization.

Gas-Phase Energies   CCSD(T) is a single reference method that builds up on Hartree-Fock wave-function. This is a safe assumption in case of closed-shell systems, such as complexes with Cd^{2+} and Zn^{2+} ions. For some of the open-shell systems, containing Fe^{2+} and Cu^{2+}, there are indications that use of single-reference methods is less warranted.

The quantity of interest is the energy of interaction with a given set of ligands, defined by equation 5.1. The aim of the benchmark is to identify methods capable of reproducing the CCSD(T) value of this quantity as closely as possible but at significantly lower computational cost.

System	[CdL$_4$]	[CuL$_2$]$^+$
CCSD(T)	375.8	342.6
MP2	375.1	320.5
PBE+D3	384.3	337.6
B3-LYP+D3	377.9	334.8
TPSS+D3	382.0	333.6
M06	380.2	341.1
BH-LYP	370.5	314.2
M05-2X	375.3	316.9
MPWB1K	376.2	320.4

Table 5.1: Interaction energies [kcal.mol^{-1}] of selected complexes. The Cd^{2+}complex is a representative example, while the Cu^{2+}complex is a problematic case.

$$[M(H_2O)_n]^{2+} + nL^c \rightarrow [ML_n]^{c+2} + 6H_2O \qquad (5.1)$$
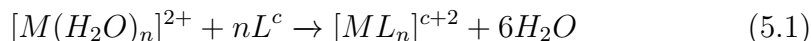
Out of the tested methods **the best agreement is provided by Moller-Plesset 2nd order (MP2)** in combination with a triple-ζ basis set, **with the errors within 1.5 kcal.mol^{-1}**. This is an extraordinary result, as the computational price is comparable to hybrid DFT calculations. The success, however, comes at the price of questionable robustness, as evidenced by a couple of cases where the method fails spectacularly. An example of a typical and of a problematic case is presented in Table 5.1.

The problematic cases are all open-shell species and all contain a metal-thiolate bond. Copper systems seem to be especially problematic in this regard. The problem is reflected in increased errors, difficulties in converging the wave-function and unwarranted energy differences of some virtually identical structures. A partially multi-reference character or presence of intruder states might be responsible for these anomalies. Without resolving the specific case, it is noted that MP2 appears as a very attractive, although not fully reliably method for the problem of interest.

DFT methods show a wide range of accuracy that is, in general, inferior to that of MP2. The average absolute errors are in the range of several

units to more than 10 kcal.mol^{-1}. A vague trend that can be recognized is improved performance of functionals with higher amount of exact exchange. Indeed, the top three functionals all include more than 40% of exact exchange, while the worst three were pure functionals (i.e. with 0 % exact exchange). Among the worst performing was also BP86, which will be relevant for later discussion. Notably, the cases precarious for the wave-function methods were handled by DFT without increased difficulty.

It appears that a DFT functional with high amount of exchange can obtain interaction energies of these systems slightly more robustly although at cost of decreased accuracy compared to MP2. Both approaches can be considered satisfactory.

For discussion of selectivity, differences of interaction energies for different metal ions are relevant. The error in this doubly differential quantity is significantly less pronounced, typically in units of kcal.mol^{-1}. This error appears to be of similar magnitude even for methods with poor performance on the absolute scale.

Structure Optimization The investigation of structure dependence on the method used for optimization reveals few remarkable differences. Employing o one method for both the optimization and any further calculations is to be preferred. Expectedly, the **requirements are less stringent for obtaining the geometry** than they are for calculation of interaction energies.

Even with one method, there is a danger of obtaining multiple minima even for rather simple systems. Before drawing conclusions from differences in energy of two complexes with different metal ions, it is advisable to swap the ions and restart the geometry optimizations.

Dispersion   is often not properly included in most DFT functionals. This deficiency is typically remedied by empirical correction. Its contribution to a binding energy will depend on the size and nature of ligands, but can be expected to be within several units of kcal.mol^{-1}. If the reference state is similar to the target complex (i.e. equal number of ligands, similar character

of ligands), the dispersion correction is unlikely to contribute significantly to the total value of interaction energy. However, including dispersion correction is still essential for optimization of structures.

5.1.2 Experimental Benchmarks

Work in Paper I and II The systems subjected to gas-phase benchmarks were also considered in solution. Moreover, complexes of small ligands with a series of divalent metal ions (Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, Hg^{2+}), for which the experimental stability constants were available, were built. The systems are presented in Figure 5.1. Different binding modes, amount of hydration and reference states are considered.

Calculations that include in-solvent optimizations, gas-phase single-point energies, free energy of solvatation, thermal corrections and zero-point vibrational energies are performed.
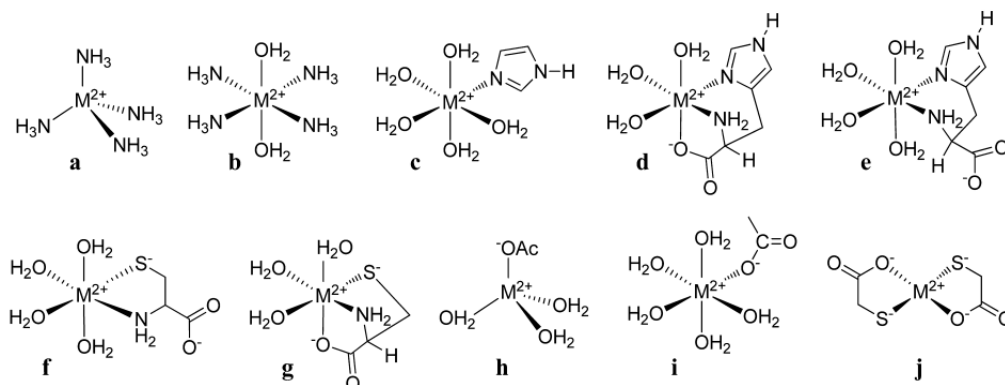

Figure 5.1: Systems studied in Paper II.

Reference states The previous analysis assumed equal number of ligands in the target complex and the reference state. It is observed that the good performance of some methods benefits from cancellation of errors - an interaction with a ligand is compared to the interaction with a water molecule in the reference state.

A more natural reference state is tested instead:

$$[M(H_2O)_6]^{2+} + nL^c \rightarrow [ML_n(H_2O)_m]^{c+2} + (6{-}m)H_2O \qquad (5.2)$$
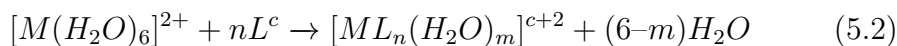
With the hexahydrated reference state the error accumulates and its magnitude can be partially related to the difference in the number of ligands between the complex and the reference state. This highlights the importance of correctly representing the systems of interest.

Implicit vs Explicit Solvation   Inevitably, this leads to the question of which solvent molecules to represent explicitly. In the case of an ideal solvent model this choice would be inconsequential. However, the accumulated error mentioned above shows that this is not the case. Removal of an explicit solvent molecule is not properly compensated by introduction of implicit solvent, or its advanced variants like COSMO-RS or SMD (used in Papers I and II). Apparently, there is a systematic misrepresentation of the first-shell solvent molecules.

Is one representation more justifiable than the other? What are the possible discriminating factors? The answers can be obtained by considering the nature of our solvent models. Even though both COSMO-RS and SMD go beyond polarizable continuum models, they build up on them and inherit some of the inherent limitations. Specifically, the solvent is represented by a macroscopic descriptor of a bulk solvent and can only account for solute-solvent charge transfer, at best, parametrically.

This is known to cause problems with charged or even just highly polar species, where this description meets its limits. The assumed bulk solvent molecule becomes qualitatively different to the actual strongly interacting solvent molecule. For these reasons, a model that prevents direct contact of an ion with the continuum is preferable. This is especially relevant for the doubly charged metal ion, but similar reasoning pertains to (charged) ligands.

It is advisable to include all first-shell ligands explicitly. For reference states of metal ions this usually implies hexahydrated octahedral complex. A

System	Average error
$[M(NH_3)_4(H_2O)_2]^{2+}$	7.6
$[M(Imi)(H_2O)_5]^{2+}$	1.3
$[M(H_2O)_3(His)]^+$	2.2
$[M(H_2O)_3(Cys)]$	17.6
$[M(CH_3COO)(H_2O)_5]^+$	12.4
$[M(SCH_2COO)_2]^{2-}$	9.9

Table 5.2: Errors [kcal.mol^{-1}] from MP2 + COSMO-RS calculations of binding free energies for selected complexes.

known complication is the cuprous ion, Cu(II). Due to Jahn-Teller distortion the axial ligands are bound less strongly or not at all. A hexaydrated Cu(II) complex optimized in GGA ends up with one of the water molecules floating away to the second solvation shell. This not only exposes the metal ion to the continuum, but also introduces an explicit water molecule to the second solvation shell. A reference state with five water molecules instead faces the potential accumulative error mentioned above. Either way, this leads to systems that are qualitatively different to other metal ions. An analogous problem is presented by many Cu(II) systems and is related to the problem of comparing different structures mentioned in previous section.

Using hexaligated complexes on both sides appears to diminish the differences among the methods. However, this may be caused by consideration of simple ligands (due to availability of experiental data), where only a few ligands differ from the water molecule. Hence, we do not infer any conclusions from this observation.

In-Depth Error Analysis The computational protocol so far consists of MP2 for gas-phase energies, COSMO-RS for solvation, and standard frequency analysis for the estimation of thermal contributions and Zero-Point Vibrational Energy (ZPVE). Table 5.2 shows the errors for computation of binding free energies using this computational setup.

Even with this setup the errors are in units to tens of kcal.mol^{-1}from

the experimental values. We have to conclude that **the presented com-
putational protocol is NOT capable of reproducing experimental
stability constants of metal ion complexes**.

In order to understand the source of the error we start with a naive
assumption. We assume that there are two major contributions to the error
(i.e. the diference between the experimental and calculated binding free
energies) - a ligand-specific shift (LSS) and a metal-specific shift (MSS).

$$\Delta\Delta G_{M,L} = (\Delta G_{M,L}^{exp} - \Delta G_{M,L}^{calc}) = LSS_L + MSS_{M,L}$$

The difference between this assumption and actual values is incorporated
into MSS. Thus, the MSS is both metal- (M) and ligand- (L) dependent.

Assuming our definition of free energy is in principle exact there are three
possible sources of error. The errors stemming from gas-phase energy have
been discussed previously. Most of the following discussion elaborates on
issues related to solvation free energy.

We refrain from analysing contribution of ZPVE and thermal corrections
to the total error. We do, however, note that this term exhibits low variability
across different metal ions and only becomes notable, although still of minor
importance, when the binding modes are not identical.

Ligand-Specific Shift   The Ligand-Specific Shift (LSS) is defined as:

$$LSS_L = \sum_{M}^{n} \frac{\Delta\Delta G_{M,L}}{n} \tag{5.3}$$

The processes of ligand binding a proton and binding to a metal ion
are qualitatively similar, as both constitute charged Lewis acids. Thus, we
expect similar trends for the LSS_L and pKa of a corresponding lingand. This
comparison is shown in 5.3.

Moreover, the ligands are simple systems that are unlikely to suffer from
insufficient description of wave function or thermal contributions. This allows
us to attribute all of the error in these calculations to solvation and relate
the conclusions to the behaviour of the solvation model for the metal ion

System	LSS	pKa error
$[\text{M(NH}_3)_4(\text{H}_2\text{O})_2]^{2+}$	2.8	4.7
$[\text{M(Imi)(H}_2\text{O})_5]^{2+}$	0.3	3.1
$[\text{M(H}_2\text{O})_3(\text{His})]^{+}$	3.0	4.1
$[\text{M(H}_2\text{O})_3(\text{Cys})]$	−8.0	−4.6
$[\text{M(CH}_3\text{COO)(H}_2\text{O})_5]^{+}$	−8.6	−1.9
$[\text{M(SCH}_2\text{COO})_2]^{2-}$	−13.5	−9.6

Table 5.3: Ligand-Specific Shifts [kcal.mol^{-1}] obtained from BP86 + COSMO-RS calculations of binding free energies for selected complexes and error in calculation of pKa of corresponding ligands.

complexes. The pKa and β error are not quantitatively identical, since the LSS quantity includes errors in all three terms (i.e. also gas-phase energy, ZPVE, and thermal corrections, not just solvation) and also includes errors common to all studied metals.

Just as in case of metal ions, including the first solvation shell might be beneficial for, especially charged, ligands as well. However, the first solvation shell of a ligand might not be defined as clearly as in the case of a metal ion.

An important and related point is that the error of pKa prediction is related to the continuum treatment of ligands. Each functional group is associated with different error, implying systematic favouring or disfavouring of exposure of some groups to the continuum. Comparing binding free energies of different binding modes is, thus, subject to this bias.

Metal-Specific Shift   Metal-Specific Shift (MSS) is defined as:

$$MSS_{M,L} = \Delta\Delta G_{M,L} - LSS_L \tag{5.4}$$

The magnitude of MSS is fairly small, which means, that the differences in binding free energies across the metal ion series, i.e. selectivities, are reproduced surprisingly well.

It is also quite invariant, which is reflected by small root-mean-squares of

System	RMS$_L$	RMS$_L^C$
[M(NH$_3$)$_4$(H$_2$O)$_2$]$^{2+}$	2.7	1.7
[M(Imi)(H$_2$O)$_5$]$^{2+}$	1.3	1.9
[M(H$_2$O)$_3$(His)]$^+$	1.5	0.8
[M(H$_2$O)$_3$(Cys)]	2.6	1.5
[M(CH$_3$COO)(H$_2$O)$_5$]$^+$	1.3	0.8
[M(SCH$_2$COO)$_2$]$^{2-}$	3.6	2.4

Table 5.4: Variation of Metal-Specific Shifts [kcal.mol^{-1}] obtained from BP86 + COSMO-RS calculations of binding free energies for selected complexes, before (RMS$_L$) and after(RMS$_L^C$) empirical correction.

MSS$_{M,L}$ for a specified ligand, L, see Table 5.4. MSS averaged over all tested systems can be applied as an empirical correction.

The RMS$_L^C$ thus represents error in reproducing selectivity of a given complex after applying this correction - the value is in all but one case below 2 kcal.mol^{-1}, which can be considered satisfactory.

It is worth mentioning that the error is largest for the problematic Cu(II) ion (2.5 kcal.mol^{-1}). This suggests that the analysis of reference states was well founded and the magnitude of MSS for each metal is related to the uncertainty in defining its reference hydrated state.

Selectivity The choice of method for calculation of relative (rather than absolute) stability constants does not seem to be as critical. Considering relative binding free energies benefits significantly from cancellation of errors.

The results in 5.4 are calculated with BP86, which was among the worst performers in the benchmark calculations done in Paper I. When comparing the methods in their ability to reproduce differences in binding free energies, the BP86 method is better than others by a small margin. This apparent discrepancy has two possible sources.

First, the nature of error is likely similar for all metal ions studied, as mentioned above. This observation is, naturally, system dependent. Among the studied complexes, the largest error is observed for the (SCH2COO)$^{2-}$ ligands. While there are multiple sources of error possible, we recall to the

thiolate ligands presenting a challenge for many of the tested methods. The increased error in this case may be due to related reasons.

Second, the solvation method used, COSMO-RS, is parametrized for BP86. Although metal ion complexes are not a natural system for this solvation method, it is likely that the performance of BP86 is aided, to some extent, by the parametrization.

The appeal of using BP86 as a method of choice is two-fold. First, it is a pure functional, which warrants a lower computational cost. Second, thanks to the setup of computational protocol a single single-point calculation is sufficient, instead of three calculations required for combining COSMO-RS with other methods. The limitations of BP86 for systems of more complicated electronic structure, such as presence of thiolate ligands, have to be borne in mind at all times.

As mentioned previously, calculation of ZPVE and thermal corrections is a costly and burdened with systematic errors, due to problems of estimating these terms for species in solvent. Their contribution to the accuracy of relative stability constants is negligible when comparing identical binding modes. Even for different binding modes this contribution is comparable to the expected error of these computations. Avoiding the frequency analysis altogether, thus, appears as a reasonable choice.

5.1.3 The Computational Protocol for Selectivity

The proposed version of the computational protocol thus consists of BP86 optimization and single-point calculation, combined with COSMO-RS. Frequency analysis is omitted.

It is counterintuitive that the minimalistic choice for the protocol, i.e. in-solvent optimization followed by a single single-point calculation, could be sufficient for reproducing selectivity properties of metal-ion complexes. Moreover, it appears to be an optimal choice. I hope I have convinced the reader that this choice is not made of convenience or serendipity and that the reasons behind it are well motivated and justified.

There are limitations of the chosen approach. Most importantly, these

are possibly insufficient description of the electronic structure in case of more complicated systems (such as those containing thiolate ligands) and bias in comparing different conformations stemming both from the inherent properties of the solvation method and omission of frequency analysis.

Still, after applying a justifiable empirical correction, the **error in reproducing selectivity of studied complexes was reduced below 2 kcal.mol^{-1}**.

Identifying these limitations allows awareness and accounting for problematic cases by using more reliable wave-function or DFT methods, inclusion of frequency analysis and at least qualitative analysis of solvation bias.

5.2 Paper III

The computational protocol seems to work well in reproducing relative stability constants. Some level of uncertainty, however, remains in comparison of different conformations. The study presented in Paper III offered an opportunity to further test its capabilities, both in describing structural and energetic properties of metal-ion complexes.

   The systems in question (see Figure 5.2) were designed to bind ^{64}Cu quickly and easily; the resulting complex to be examined for a potential use in nuclear medicine. The systems are composed of a cyclam ring and a variety of phosphonate and phosphinate pendants.
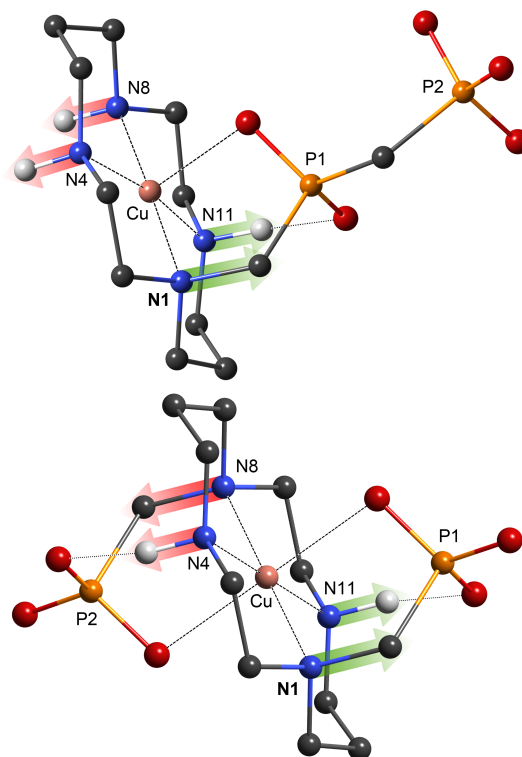

Figure 5.2: Phosphonate complexes L^2 and te2P

   The system of interest, L^2, exhibited unusual kinetic and thermodynamic behaviour. The kinetics of complexation suggested a three-step process. The

first step includes formation of an out-of-cage coordination of copper by the pendant arm. The second step transfer of copper from out-of-cage to in-cage complex of unknown structure. The last step results in formation of one of the major cyclam conformers, namely conformer I (see 5.3.

Spectral characteristics of the intermediate complex suggest that it is one of the major cyclam conformers as well (see Figure 5.3). However, due to its transient nature, its identification by experimental methods remained elusive. Could the proposed computational help identify the structure of this intermediate?

The three-step process, mentioned above, results in formation of conformer I. Under different experimental conditions, this conformer further transforms to conformer III. This conversion is not unusual, as conformer III is the most stable for many cyclam complexes. However, the conversion is usually quantitative, while for system L^2 a mixture of conformers I and III is observed. Can this anomaly be justified by the proposed computational protocol?

Work in Paper III   Given at least one pendant arm attached to one of the nitrogens, there are 8 possible conformers defined by chirality of protonated nitrogen atoms.

All of these conformers were built for the three systems in question and subjected to the DFT + COSMO-RS protocol. Rather than examining interaction with different ions, as was the case in previous Papers, relative energies of conformers of individual systems were obtained and examined.

5.2.1   Explaining Kinetic Behaviour

It is, in principle, not possible to justify kinetics of a process by merely looking at energies of local minima. However, the conversion of the intermediate complex to conformer I is pH dependent and is proposed to be catalyzed by a hydroxide anion. The intermediate complex is, at the same time, believed to be one of the major cyclam conformers. Thus, the transition state should be related to the intermediate complex, differing by the interconversion of
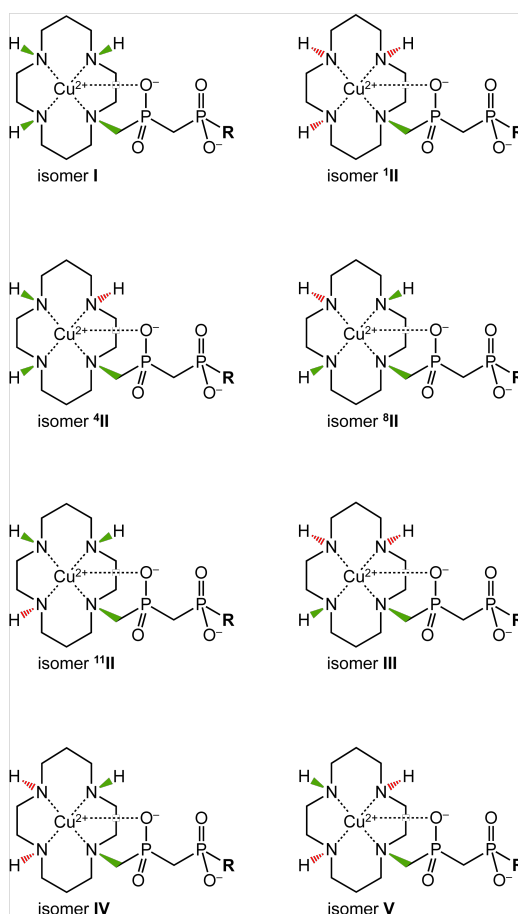
Figure 5.3: Major conformers of cyclam complexes.

the nitrogen's chirality. The energetical cost of the transition state comes on the top of the conformational energy of the intermediate complex. Thus, the conformational energy is a discriminating, albeit not decisive, factor that can contribute to the identification of the intermediate complex.

Two low-energy candidates for the intermediate complex are [4]II and [11]II conformers, with conformational free energies 1.4 and 4.0 kcal.mol^{-1}(relative to the conformer I), respectively.

The [4]II complex conformer is expected to have similar probabilities of converting to I and III conformers; a single flip of chirality required in each case. However, only conversion to I is observed. Such behaviour seems to

be more in line with the [11]II conformer, for which only the conversion to I is natural. This is supported by the structure of [11]II, in which the pendant arm stabilizes the two hydrogens not undergoing a conversion. These observations collectively point to the [11]II conformer as being the elusive intermediate complex.

No intermediate complexes were observed for the other examined system te2P. In agreement with these observations, the conformational free energies of these systems are relatively higher than for L^2.

The overall complexation process of L^2, as proposed in Paper III, is shown in Figure 5.4.
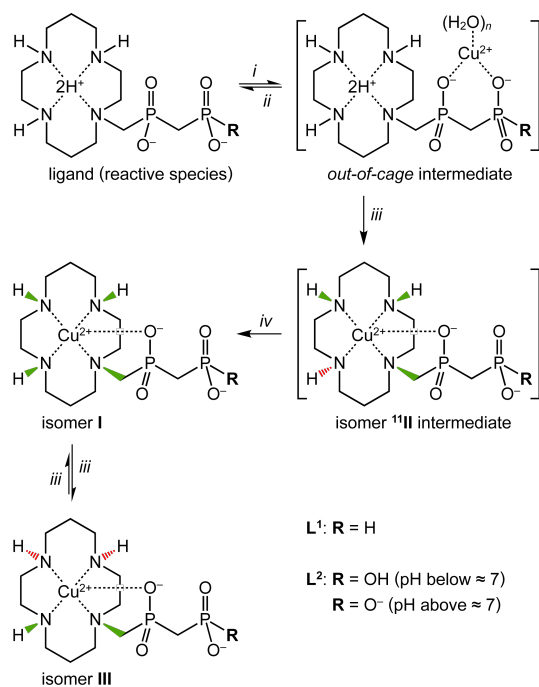
Figure 5.4: Proposed intermediates of the complexation process of L^2

5.2.2 Explaining Thermodynamic Behaviour

A more natural question to address is the formation of thermodynamic mixture of I and III conformers of the L^2 system. No mixture was observed for the te2P system, which yields conformer III exclusively.

The observed ratio of I/III conformers of in the thermodynamically stable mixture of L^2 suggests conformational free energy difference of ca. 0.5 kcal.mol^{-1}. No mixture is observed for the te2P system, implying significantly larger difference.

In agreement with these observations, the lowest conformer is correctly identified to be III. For L^2 system, the I conformer is only 1.4 kcal.mol^{-1}above conformer III, while it is 5.4 kcal.mol^{-1}higher for the te2P system. The difference is likely caused by the stabilization of I conformer by the longer pendant arm of L^2.

5.2.3 Conclusions

The computational protocol was applied to a case-study of copper-cyclam complexes. The conformational energies and structures obtained from the protocol were used to aid identification of the intermediate complex and to justify the formation of equilibrium mixture. Although the question posed in this study did not relate to metal-ion selectivity or the conformational energies of different binding modes (see Paper II), the agreement of the results with all observations further validates the computational protocol.

5.3 Paper IV

In order to understand metal ion selectivity in biological context it is necessary to make a step towards more complex systems. Paper IV focuses on peptidic systems, as these are abundant and well explored, but it is likely that the conclusions are not system-specific and can be transferred e.g. to interaction of nucleotides and DNA with ions.

To what extent is the computational protocol transferable to metallopeptides? There are at least two obvious problems to be faced.

First, it is not practical to represent the entire peptide by quantum mechanical description. Even modestly sized peptides are computationally prohibitive. There is an obvious need for choosing an approximate representation of the metal-binding peptide. The definition of a ligand thus becomes unclear. Cutting bonds along non-polar (essentially C-C) bonds, similar to the practice of QM/MM calculations, is a reasonable solution. But how far from the binding site should the cut be made?

Second problem is much more fundamental. Even in simple metal-ion complexes the approximate description of outer shells is problematic. This is despite the fact that water as a solvent is extensively studied and chemically rather simple. Peptides exhibit immense variability and have potentially very complex dynamics that depend on the details of the whole large molecule. What level of accuracy can be hoped for if only simplified model is considered? Where could such a model be useful?

Work in Paper IV   In Paper IV, six amino acid sequences were selected. These were either predicted or observed to bind metal ions with their side chains. The peptides are referred to based on the amino acid codes of the putative binding residues, and are presented in Figure 5.5.

Four models representing the binding sites were examined. The models differed in distance of the truncation site from the binding atom and capped by a hydrogen atom. The truncation included only the binding atom (TINY), one methyl group (SMALL), the entire side chain (ALPHA), or the entire amino acid (FULL_AA).

47
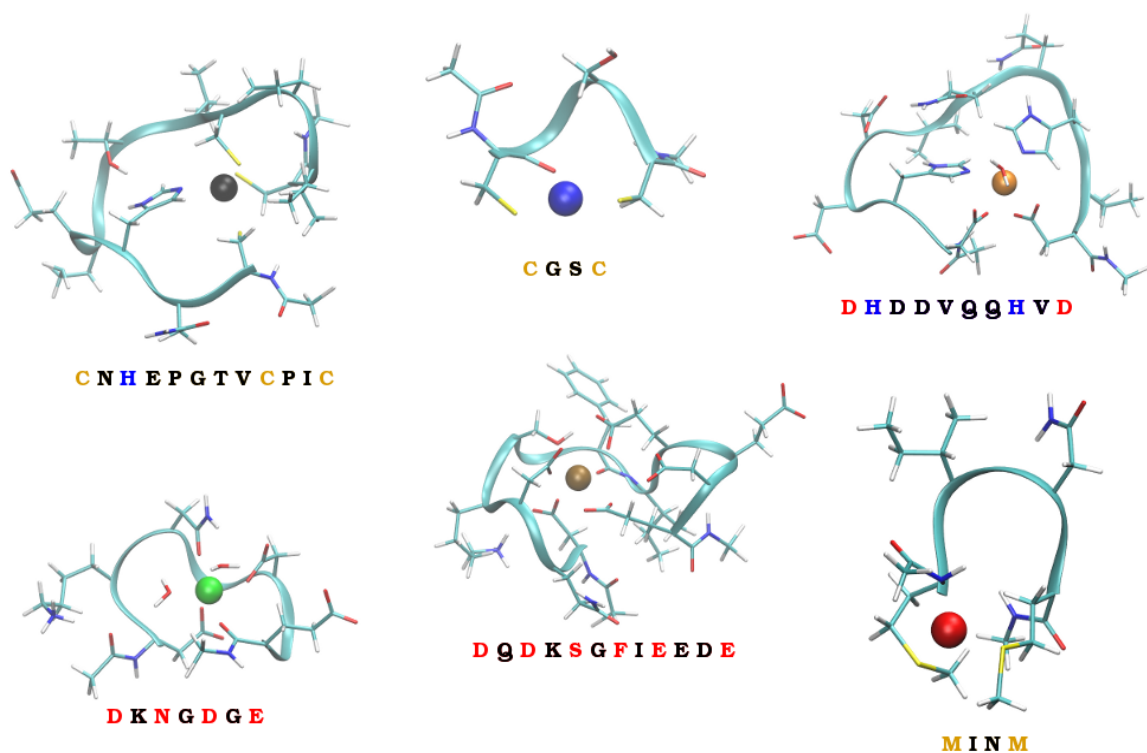
Figure 5.5: Six peptide systems studied in Paper IV.

Each of the metal ions from the series examined previously - Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, Hg^{2+}- was inserted into each peptide and each model.

Each system was optimized. Subsequently, the metal ion of each system was substituted with all other metal ions and subjected to the computational protocol (without optimization).

Several substitutions of side-chains with and without optimization were performed for selected systems in order to systematically study the various effects that could affect the effect of the system not included in the models.

5.3.1 The Thermodynamic Cycle

The relative free energy is a state quantity. For the purpose of analysis we split the binding of different ion into two steps. First, we exchange the metal-ion - without optimizing the system. The differences in binding free energy of this step are addressed as MSS. The second step is optimization of the system. Differences in binding free energy of this step are addressed as LSS. The process is shown schematically in Figure 5.6.
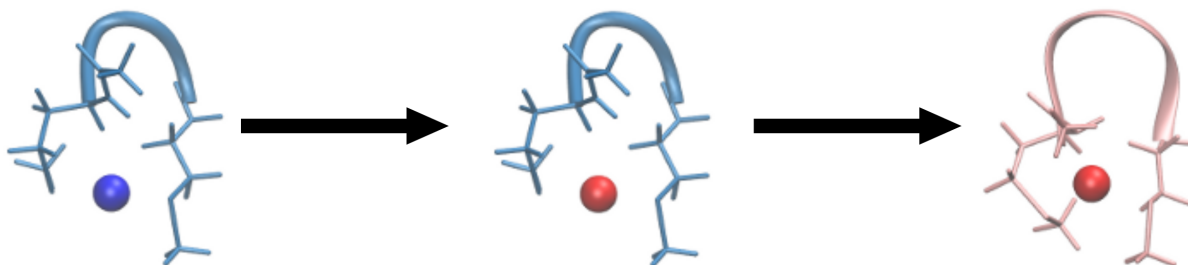
Figure 5.6: Two-step exchange of metal-ion

This split allows us to isolate the effect of outer-shells (i.e. those parts of the peptide not included in the model) on binding of the metal-ion from the overall process of structural adaptation of a peptide. The idea is similar to any thermodynamic cycle. The focus of the discussion in on the Metal-Induced Selectivity (MIS), because addressing the structure-induced selectivity is better suited for e.g. QM/MM shcemes, which are outside the scope of this thesis.

5.3.2 Metal-Induced Selectivity

The Absolute Average Deviation (AAD), i.e. the distance from the mean, of the following quantity is studied:

			System			
model	CC	MM	DHHD	DNDO	CHCC	DDSOEE
TINY	2.2	1.8	3.8	2.5	0.6	0.8
SMALL	0.9	0.8	1.5	0.4	0.3	0.5
ALPHA	0.4	0.5	0.4	0.1	0.3	0.4
FULL_AA	0.2	0.1	0.1	0.1	0.1	0.2

Table 5.5: A representative example of AAD values.

$$G^{2ndshell}_{M,L,model} = G^{peptide}_{M,L} - G^{model}_{M,L} \qquad (5.5)$$

This quantity isolates the interaction of the outer shells with the metal ion. The average (in AAD) is taken over all metal ions - in line with the idea of MIS. Note that the structure of peptide (and the model) is held frozen for all metal ions.

The AAD values for a representative example are shown in Table 5.5 and are generally below 1 kcal.mol^{-1}, even for SMALL models.

Error Trends Across Systems   AAD$_{L,model}$ varies across the systems studied. Three conceptual sources of this variation have been identified. The list is likely incomplete but it does cover most of the error of MIS observed and should prove useful for future application.

Hardness   Hardness of the binding ligand is an important determinant. Softer ligands are more sensitive to the truncation than harder ligands such as carboxylates. Thus, ALPHA models are recommended for thiolates (cysteine side-chains of CC and CHCC) and thiolethers (methionine side-chains of MM), while SMALL models are sufficient for most other cases.

We note that complexity of the thiolate-metal bond, shown repeatedly to be difficult to address, is not relevant in this case - the computational method is the same for all terms in equation 5.5 and, thus, the error reflects only the effect of outer shell.

Shielding   There are, potentially, other parts of the peptide that significantly affect outer-shell selectivity that cannot be easily identified and/or included in the model. The most obvious case is the **presence of charged ligands** in the vicinity of the binding site.

As observed previously, including the first-shell ligands is necessary but doable, because the presence of solvent molecules can be easily surmised in the case of metal-ion complexes. In a peptide, the binding site and the binding partners are defined by the peptide. Fewer ligands result in exposure of the ion to the continuum in a model. Thus, **coordination number** is an important determinant, with the higher number favouring proper shielding of the ion from the rest of the peptide and, thus, lower AAD.

Although the coordination number cannot be accounted for without expanding the model beyond the first-shell ligands, it does hint at the measure of the outer-shell selectivity that can be expected from the model.

Hydrogen Bonds   Other than explicitly charged moieties close to the binding site, strongly interacting partners - such as hydrogen bonds - may play role as well. Their influence is relatively small, typically only few tenths of kcal.mol^{-1}. In the case of ideal hydrogen bonds and strong acceptors their importance may increase.

In conclusion, the MIS revolves around the measure of interaction of the model with the parts of the system not included in the model. Maximising the shielding of the ion from the rest of the peptide helps to improve the selectivity description of a model.

5.3.3   Structure-Induced Selectivity

The metal-induced selectivity is only a thought experiment. Connecting its relevance to the structural adaptation of a model is a required to justify its meaningfulness.

The effect of structure adaptation of the binding site - coupled with the structural changes of an entire peptide - is the basic principle of allostery

and can be arbitrarily large.

An argument can be made that this does not invalidate the previous analysis. First, there are systems, among those studied, where the structural adaptation has negligible effect on relative binding free energies, i.e. the local minima for a pair of ions is identical. Thus, it is conceivable that for potentially rigid peptides the first-shell models capture the selectivity properties very well.

Apart from these ideal cases, there are other ways of accounting for the structural change. The content of this analysis is essentially identical to the art of proper selection of the QM system in QM/MM studies. The truncation criterion revolves, in this case, around selectivity properties. Thus, it is natural to try to apply the conclusions to a multi-scale model. Structure-Induced Selectivity (SIS) would thus be accounted for by e.g. point charges and MM description using well established methods.

5.3.4 Conclusions

The computational protocol described in Papers I and II in conjunction with proper selection of a model is a useful tool for addressing selectivity of metallopeptides.

The error in reproducing selectivity properties is enlarged, compared to structurally simpler metal-ion complexes, by ca. 1 kcal.mol^{-1}. The true magnitude of the error depends on details of the binding site and its environment.

The above conclusion is only valid under an optimistic assumption of a rigid peptide. Thus, it should serve only as a proof of concept that invites to couple the protocol with QM/MM schemes.

5.4 Conclusions and Outlook

The thesis attempts to show that the computational protocol at hand is well suited to address problems of metal-ion selectivity and that it is economic and flexible enough to allow modifications for more problematic cases. In its basic form it consists of BP86 coupled with COSMO-RS solvation model. Further work may be required to address the issues raised in this thesis and to couple the protocol with QM/MM techniques for investigation of metallopeptidic systems.

Despite the best efforts, it seems that calculation of absolute stability constants is out of reach of contemporary "single-point" methods. While a variety of quantum-chemical methods are available, it is the solvation that still proves to be the most challenging aspect.

On the other hand, the computation of relative stability constants (relevant for examining metal-ion selectivity properties) is much more accessible. It is the cancellation of errors in multiple aspects that warrants the use of cheaper methods.

This is partly due to different requirements for calculation of - seemingly similar - absolute and relative binding free energies. Moreover, the gas-phase energy and solvation free energy are correlated quantities, which presents a hurdle in benchmarking the methods separately. Ultimately, it is their combination that is relevant. Thus, the BP86 functional, that performs poorly in gas-phase benchmarks, is much more attractive in combination with a solvation method parametrized for this functional.

Similarly, the ZPVE and thermal corrections obtained from frequency analysis make a major contribution to the values of absolute stability constants, but a negligible one to the values of relative stability constants. Once the objective shifts to the latter, it is advisable to omit this expensive part of the calculation.

Representation of a system by a single structure is justifiable only for relatively simple systems. Even in these cases, identifying the structure may not be intuitive and conformational sampling is advised. The bias for exposure of some functional groups to the continuum makes the evaluation of

such sampling problematic and requires further work.

While the first-shell ligands are, naturally, the most important determinants of selectivity, the outer shells can have non-negligible contribution as well. Although the use of the protocol for metallopeptidic systems has been conceptually demonstrated, coupling of the protocol with QM/MM schemes is necessary for practical use.

With a reliable computational protocol at hand, it would be interesting to explore the factors of metal-ion selectivity systematically. While the subject has been studied extensively for at least a century, resulting in many qualitative and semi-quantitative concepts, a universal protocol and unbiased systematic approach is certainly attractive.

This could result in quantification of importance of individual factors, their cooperative effects, and, possibly, identification of yet unknown concepts. Along these lines, a work on construction of a database of metal-ion complexes composed of ligands that represent amino acid side-chains is under way. The database will contain complexes exhaustively combining different metal ions, ligands, and coordination geometries. The database was the ultimate motivation behind the presented Papers. Unfortunately, I have not managed to compile its results within this thesis and I invite the reader to stay tuned for the exciting study that goes beyond technical benchmarking and protocol validation.

Bibliography

[1] Vera Krewald, Marius Retegan, Nicholas Cox, Johannes Messinger, Wolfgang Lubitz, Serena DeBeer, Frank Neese, and Dimitrios A. Pantazis. Metal oxidation states in biological water splitting. *Chemical Science*, 6(3):1676–1695, February 2015.

[2] Douglas C. Rees, Robert P. Gunsalus, Sabina Rech, and Yonglin Hu. Crystal structure of the molybdate binding protein ModA. *Nature Structural and Molecular Biology*, 4(9):703, September 1997.

[3] Lalla Aicha Ba, Mandy Doering, Torsten Burkholz, and Claus Jacob. Metal trafficking: From maintaining the metal homeostasis to future drug design. *Metallomics*, 1(4):292, 2009.

[4] Todor Dudev and Carmay Lim. Oxyanion Selectivity in Sulfate and Molybdate Transport Proteins: An ab Initio/CDM Study. *Journal of the American Chemical Society*, 126(33):10296–10305, August 2004.

[5] Todor Dudev and Carmay Lim. Competition between Li+ and Mg2+ in Metalloproteins. Implications for Lithium Therapy. *Journal of the American Chemical Society*, 133(24):9506–9515, June 2011.

[6] R. D. Shannon. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(5):751–767, September 1976.

[7] Todor Dudev and Carmay Lim. Factors Governing the Na+ vs K+ Selectivity in Sodium Ion Channels. *Journal of the American Chemical Society*, 132(7):2321–2332, February 2010.

[8] Todor Dudev and Carmay Lim. Determinants of K+ vs Na+ Selectivity in Potassium Channels. *Journal of the American Chemical Society*, 131(23):8092–8101, June 2009.

[9] Todor Dudev and Carmay Lim. Why voltage-gated Ca 2+ and bacterial Na + channels with the same EEEE motif in their selectivity filters confer opposite metal selectivity. *Physical Chemistry Chemical Physics*, 14(36):12451–12456, 2012.

[10] Malcolm A. Halcrow. Jahn–Teller distortions in transition metal compounds, and their importance in functional molecular and inorganic materials. *Chemical Society Reviews*, 42(4):1784–1795, January 2013.

[11] Minko Dudev, Jonathan Wang, Todor Dudev, and Carmay Lim. Factors Governing the Metal Coordination Number in Metal Complexes from Cambridge Structural Database Analyses. *The Journal of Physical Chemistry B*, 110(4):1889–1895, February 2006.

[12] Robin Chaudret, Julia Contreras-Garcia, Mickaël Delcey, Olivier Parisel, Weitao Yang, and Jean-Philip Piquemal. Revisiting H2O Nucleation around Au+ and Hg2+: The Peculiar "Pseudo-Soft" Character of the Gold Cation. *Journal of Chemical Theory and Computation*, 10(5):1900–1909, May 2014.

[13] Xiao-Gen Xiong, Wen-Hua Xu, Jun Li, and Pekka Pyykkö. Aspects of bonding in small gold clusters. *International Journal of Mass Spectrometry*, 354-355(Supplement C):15–18, November 2013.

[14] Irving and Williams. The stability of Transition-Metal complexes. *Journal of Chemical Society*, pages 3192–3210, 1953.

[15] Serge I. Gorelsky, Lipika Basumallick, Josh Vura-Weis, Ritimukta Sarangi, Keith O. Hodgson, Britt Hedman, Kiyoshi Fujisawa, and Ed-

ward I. Solomon. Spectroscopic and DFT Investigation of [M{HB(3,5-iPr2pz)3}(SC6F5)] (M = Mn, Fe, Co, Ni, Cu, and Zn) Model Complexes: Periodic Trends in Metal-Thiolate Bonding. *Inorganic Chemistry*, 44(14):4947–4960, July 2005.

[16] Pradeep R. Varadwaj, Arpita Varadwaj, and Bih-Yaw Jin. Ligand(s)-to-metal charge transfer as a factor controlling the equilibrium constants of late first-row transition metal complexes: Revealing the Irving–Williams thermodynamical series. *Physical Chemistry Chemical Physics*, 17(2):805–811, 2015.

[17] Christine M. Phillips, Eric R. Schreiter, Yayi Guo, Sheila C. Wang, Deborah B. Zamble, and Catherine L. Drennan. Structural Basis of the Metal Specificity for Nickel Regulatory Protein NikR,. *Biochemistry*, 47(7):1938–1946, February 2008.

[18] Kevin J. Waldron and Nigel J. Robinson. How do bacterial cells ensure that metalloproteins get the correct metal? *Nature Reviews Microbiology*, 7(1):nrmicro2057, January 2009.

[19] Zhen Ma, Melinda J. Faulkner, and John D. Helmann. Origins of Specificity and Crosstalk in Metal Ion Sensing by Bacillus subtilis Fur. *Molecular microbiology*, 86(5):1144–1155, December 2012.

[20] K. Håkansson, A. Wehnert, and A. Liljas. X-ray analysis of metal-substituted human carbonic anhydrase II derivatives. *Acta crystallographica. Section D, Biological crystallography*, 50(Pt 1):93–100, January 1994.

[21] Joseph A. Cotruvo and JoAnne Stubbe. Metallation and mismetallation of iron and manganese proteins in vitro and in vivo: The class I ribonucleotide reductases as a case study. *Metallomics : integrated biometal science*, 4(10):1020–1036, October 2012.

[22] J. A. Fee. Regulation of sod genes in Escherichia coli: Relevance to superoxide dismutase function. *Molecular Microbiology*, 5(11):2599–2610, November 1991.

[23] Alison G. Tebo, Lars Hemmingsen, and Vincent L. Pecoraro. Variable primary coordination environments of Cd(II) binding to three helix bundles provide a pathway for rapid metal exchange. *Metallomics : integrated biometal science*, 7(12):1555–1561, December 2015.

[24] Zhen Ma, Faith E. Jacobsen, and David P. Giedroc. Coordination Chemistry of Bacterial Metal Transport and Sensing. *Chemical Reviews*, 109(10):4644–4681, October 2009.

[25] Mario A. Pennella and David P. Giedroc. Structural Determinants of Metal Selectivity in Prokaryotic Metal-responsive Transcriptional Regulators. *Biometals*, 18(4):413–428, August 2005.

[26] Gesine Rudolph, Geo Semini, Felix Hauser, Andrea Lindemann, Markus Friberg, Hauke Hennecke, and Hans-Martin Fischer. The Iron Control Element, Acting in Positive and Negative Control of Iron-Regulated Bradyrhizobium japonicum Genes, Is a Target for the Irr Protein. *Journal of Bacteriology*, 188(2):733–744, January 2006.

[27] Senapathy Rajagopalan, Sarah J. Teter, Petrus H. Zwart, Richard G. Brennan, Kevin J. Phillips, and Patricia J. Kiley. Studies of IscR reveal a unique mechanism for metal-dependent regulation of DNA binding specificity. *Nature structural & molecular biology*, 20(6):740–747, June 2013.

[28] Michael J. Cromie, Yixin Shi, Tammy Latifi, and Eduardo A. Groisman. An RNA Sensor for Intracellular Mg2+. *Cell*, 125(1):71–84, April 2006.

[29] Caryn E. Outten, O'Halloran, and Thomas V. Femtomolar Sensitivity of Metalloregulatory Proteins Controlling Zinc Homeostasis. *Science*, 292(5526):2488–2492, June 2001.

[30] Valeria Rukhman, Rina Anati, Meira Melamed-Frank, and Noam Adir. The MntC Crystal Structure Suggests that Import of Mn2+ in Cyanobacteria is Redox Controlled. *Journal of Molecular Biology*, 348(4):961–969, May 2005.

[31] Soojay Banerjee, Baoxian Wei, Maitrayee Bhattacharyya-Pakrasi, Himadri B. Pakrasi, and Thomas J. Smith. Structural Determinants of Metal Specificity in the Zinc Transport Protein ZnuA from Synechocystis 6803. *Journal of Molecular Biology*, 333(5):1061–1069, November 2003.

[32] Elizabeth L. Borths, Kaspar P. Locher, Allen T. Lee, and Douglas C. Rees. The structure of Escherichia coli BtuF and binding to its cognate ATP binding cassette transporter. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):16642–16647, December 2002.

[33] Bart A. Eijkelkamp, Jacqueline R. Morey, Miranda P. Ween, Cheryllynn Y. Ong, Alastair G. McEwan, James C. Paton, and Christopher A. McDevitt. Extracellular Zinc Competitively Inhibits Manganese Uptake and Compromises Oxidative Stress Management in Streptococcus pneumoniae. *PLOS ONE*, 9(2):e89427, 18-Feb-2014.

[34] Rafael M. Couñago, Miranda P. Ween, Stephanie L. Begg, Megha Bajaj, Johannes Zuegg, Megan L. O'Mara, Matthew A. Cooper, Alastair G. McEwan, James C. Paton, Bostjan Kobe, and Christopher A. McDevitt. Imperfect coordination chemistry facilitates metal ion release in the Psa permease. *Nature Chemical Biology*, 10(1):nchembio.1382, November 2013.

[35] Abiodun D. Ogunniyi, Layla K. Mahdi, Michael P. Jennings, Alastair G. McEwan, Christopher A. McDevitt, Mark B. Van der Hoek, Christopher J. Bagley, Peter Hoffmann, Katherine A. Gould, and James C. Paton. Central Role of Manganese in Regulation of Stress Responses, Physiology, and Metabolism in Streptococcus pneumoniae. *Journal of Bacteriology*, 192(17):4489–4497, September 2010.

[36] Zhen Ma, Pete Chandrangsu, Tyler C. Helmann, Adisak Romsang, Ahmed Gaballa, and John D. Helmann. Bacillithiol is a major buffer of the labile zinc pool in Bacillus subtilis. *Molecular microbiology*, 94(4):756–770, November 2014.

[37] Robert A. Colvin, William R. Holmes, Charles P. Fontaine, and Wolfgang Maret. Cytosolic zinc buffering and muffling: Their role in intracellular zinc homeostasis. *Metallomics*, 2(5):306–317, May 2010.

[38] S. C. Andrews. Iron storage in bacteria. *Advances in Microbial Physiology*, 40:281–351, 1998.

[39] Jung-Ho Shin and John D. Helmann. Molecular logic of the Zur-regulated zinc deprivation response in Bacillus subtilis. *Nature Communications*, 7, August 2016.

[40] David B. Sabine and Jacqueline Vaselekos. Trace Element Requirements of Lactobacillus acidophilus. *Nature*, 214(5087):520, April 1967.

[41] Brian Reale, Christopher Dennison, Conrad Bessant, Joe Gray, Katsuko Sato, Kevin J. Waldron, Mark J. Banfield, Nigel J. Robinson, Steve Tottey, Susan J. Firbank, and Timothy R. Cheek. Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature*, 455(7216):1138, October 2008.

[42] Manjula Sritharan. Iron Homeostasis in Mycobacterium tuberculosis: Mechanistic Insights into Siderophore-Mediated Iron Uptake. *Journal of Bacteriology*, 198(18):2399–2409, September 2016.

[43] Stephan M. Kraemer, Owen W. Duckworth, James M. Harrington, and Walter D. C. Schenkeveld. Metallophores and Trace Metal Biogeochemistry. *Aquatic Geochemistry*, 21(2-4):159–195, July 2015.

[44] Antonio Rosato, Francesca Cantini, Isabella C. Felli, Ivano Bertini, Leonardo Gonnelli, Lucia Banci, Nick Hadjiliadis, Petros Voulgaris, and Roberta Pierattelli. The Atx1-Ccc2 complex is a metal-mediated protein-protein interaction. *Nature Chemical Biology*, 2(7):367, July 2006.

[45] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, Mineola, N.Y, revised ed. edition edition, July 1996.

[46] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Physical Review*, 136(3B):B864–B871, November 1964.

[47] John P. Perdew, Matthias Ernzerhof, Aleš Zupan, and Kieron Burke. Nonlocality of the density functional for exchange and correlation: Physical origins and chemical consequences. *The Journal of Chemical Physics*, 108(4):1522–1531, January 1998.

[48] Axel D. Becke. A real-space model of nondynamical correlation. *The Journal of Chemical Physics*, 119(6):2972–2977, July 2003.

[49] Nathan E. Schultz, Yan Zhao, and Donald G. Truhlar. Density Functionals for Inorganometallic and Organometallic Chemistry. *The Journal of Physical Chemistry A*, 109(49):11127–11143, December 2005.

[50] Filipp Furche and John P. Perdew. The performance of semilocal and hybrid density functionals in 3d transition-metal chemistry. *The Journal of Chemical Physics*, 124(4):044103, January 2006.

[51] Raquel Rios-Font, Mariona Sodupe, Luis Rodríguez-Santiago, and Peter R. Taylor. The Role of Exact Exchange in the Description of Cu2+-(H2O)n (n = 1-6) Complexes by Means of DFT Methods. *The Journal of Physical Chemistry A*, 114(40):10857–10863, October 2010.

[52] Jeremy N. Harvey. On the accuracy of density functional theory in transition metal chemistry. *Annual Reports Section "C" (Physical Chemistry)*, 102(0):203–226, 2006.

[53] Eric Paquet and Herna L. Viktor. Molecular Dynamics, Monte Carlo Simulations, and Langevin Dynamics: A Computational Review. https://www.hindawi.com/journals/bmri/2015/183918/, 2015.

[54] Anil Kumar Tummanapelli and Sukumaran Vasudevan. Estimating successive pKa values of polyprotic acids from ab initio molecular dynamics using metadynamics: The dissociation of phthalic acid and its isomers. *Physical Chemistry Chemical Physics*, 17(9):6383–6388, February 2015.

[55] Emilie Cauët, Stuart Bogatko, John H. Weare, John L. Fulton, Gregory K. Schenter, and Eric J. Bylaska. Structure and dynamics of the hydration shells of the Zn2+ ion from ab initio molecular dynamics and combined ab initio and classical molecular dynamics simulations. *The Journal of Chemical Physics*, 132(19):194502, May 2010.

[56] Aleksandar Y. Mehandzhiyski, Enrico Riccardi, Titus S. van Erp, Thuat T. Trinh, and Brian A. Grimes. Ab Initio Molecular Dynamics Study on the Interactions between Carboxylate Ions and Metal Ions in Water. *The Journal of Physical Chemistry B*, 119(33):10710–10719, August 2015.

[57] Marc W. van der Kamp and Adrian J. Mulholland. Combined Quantum Mechanics/Molecular Mechanics (QM/MM) Methods in Computational Enzymology. *Biochemistry*, 52(16):2708–2728, April 2013.

[58] Jacopo Tomasi, Benedetta Mennucci, and Roberto Cammi. Quantum Mechanical Continuum Solvation Models. *Chemical Reviews*, 105(8):2999–3094, August 2005.

[59] Aleksandr V. Marenich, Christopher J. Cramer, and Donald G. Truhlar. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *The Journal of Physical Chemistry B*, 113(18):6378–6396, May 2009.

[60] Andreas Klamt. The COSMO and COSMO-RS solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):699–709, September 2011.

[61] Andreas Klamt, Volker Jonas, Thorsten Bürger, and John C. W. Lohrenz. Refinement and Parametrization of COSMO-RS. *The Journal of Physical Chemistry A*, 102(26):5074–5085, June 1998.

[62] Andreas Klamt. Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena. *The Journal of Physical Chemistry*, 99(7):2224–2235, February 1995.

[63] Andreas Klamt. COSMO-RS for aqueous solvation and interfaces. *Fluid Phase Equilibria*, 407(Supplement C):152–158, January 2016.

[64] Lubomír Rulíšek and Zdeněk Havlas. Theoretical Studies of Metal Ion Selectivity. 1. DFT Calculations of Interaction Energies of Amino Acid Side Chains with Selected Transition Metal Ions (Co2+, Ni2+, Cu2+, Zn2+, Cd2+, and Hg2+). *Journal of the American Chemical Society*, 122(42):10428–10439, October 2000.

[65] Lubomír Rulíšek and Zdeněk Havlas. Theoretical Studies of Metal Ion Selectivity.† 2. DFT Calculations of Complexation Energies of Selected Transition Metal Ions (Co2+, Ni2+, Cu2+, Zn2+, Cd2+, and Hg2+) in Metal-Binding Sites of Metalloproteins. *The Journal of Physical Chemistry A*, 106(15):3855–3866, April 2002.

[66] Lubomír Rulíšek and Zdeněk Havlas. Theoretical Studies of Metal Ion Selectivity. 3. A Theoretical Design of the Most Specific Combinations of Functional Groups Representing Amino Acid Side Chains for the Selected Metal Ions (Co2+, Ni2+, Cu2+, Zn2+, Cd2+, and Hg2+). *The Journal of Physical Chemistry B*, 107(10):2376–2385, March 2003.

[67] Matti Ropo, Markus Schneider, Carsten Baldauf, and Volker Blum. First-principles data set of 45,892 isolated and cation-coordinated conformers of 20 proteinogenic amino acids. *Scientific Data*, 3:160009, February 2016.

[68] M. Ropo, V. Blum, and C. Baldauf. Trends for isolated amino acids and dipeptides: Conformation, divalent ion binding, and remarkable similarity of binding to calcium and lead. *Scientific Reports*, 6:35772, November 2016.

Papers

Paper I

Ondrej Gutten, Ivana Beššeová, and Lubomír Rulíšek. Interaction of Metal Ions with Biomolecular Ligands: How Accurate Are Calculated Free Energies Associated with Metal Ion Complexation? *The Journal of Physical Chemistry A*, 115(41):11394–11402, October 2011.

Paper II

Ondrej Gutten and Lubomír Rulíšek. Predicting the Stability Constants of Metal-Ion Complexes from First Principles. *Inorganic Chemistry*, 52(18):10347–10355, September 2013.

Paper III

Tomáš David, Vojtěch Kubíček, Ondrej Gutten, Přemysl Lubal, Jan Kotek, Hans-Jűrgen Pietzsch, Lubomír Rulíšek, and Petr Hermann. Cyclam Derivatives with a Bis(phosphinate) or a Phosphinato–Phosphonate Pendant Arm: Ligands for Fast and Efficient Copper(II) Complexation for Nuclear Medical Applications. Inorganic Chemistry, 54(24):11751–11766, December 2015.

Paper IV

Ondrej Gutten and Lubomír Rulíšek. How simple is too simple? Computational perspective on importance of second-shell environment for metal-ion selectivity. *Phys. Chem. Chem. Phys.*, 17(22):14393–14404, 2015.

Other Publications

Daniel Bím, Ondrej Gutten, Jakub Chalupský, Martin Srnec, and Lubomír Rulíšek. TEORETICKÁ BIOANORGANICKÁ CHEMIE A SPEKTROSKOPIE. Chem. Listy, (110):354–364, January 2016.

Ondrej Gutten, Daniel Bím, Jan Řezáč, and Lubomír Rulíšek. Macrocycle Conformational Sampling by DFT-D3/COSMO-RS Methodology. Journal of Chemical Information and Modeling, Just accepted, 2017.

Jan Řezáč, Daniel Bím, Ondrej Gutten, and Lubomír Rulíšek. Towards Accurate Conformational Energies of Smaller Peptides and Medium-Sized Macrocycles: MPCONF196 Benchmark Energy Data Set. Submitted, 2017.

Interaction of Metal Ions with Biomolecular Ligands: How Accurate Are Calculated Free Energies Associated with Metal Ion Complexation?

Ondrej Gutten, Ivana Beššeová, 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

S *Supporting Information*

ABSTRACT: To address fundamental questions in bioinorganic chemistry, such as metal ion selectivity, accurate computational protocols for both the gas-phase association of metal—ligand complexes and solvation/desolvation energies of the species involved are needed. In this work, we attempt to critically evaluate the performance of the ab initio and DFT electronic structure methods available and recent solvation models in calculations of the energetics associated with metal ion complexation. On the example of five model complexes ([$M^{II}(CH_3S)(H_2O)$]$^+$, [$M^{II}(H_2O)_2$-$(H_2S)(NH_3)$]$^{2+}$, [$M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)$], [$M^{II}(H_2O)_3(SH)(CH_3COO)(Im)$], [$M^{II}(H_2S)(H_2O)(CH_3COO)$-$(PhOH)(Im)$]$^+$ in typical coordination geometries) and four metal ions (Fe^{2+}, Cu^{2+}, Zn^{2+}, and Cd^{2+}; representing open- and closed-shell and the first- and second-row transition metal elements), we provide reference values for the gas-phase complexation energies, as presumably obtained using the CCSD(T)/aug-cc-pVTZ method, and compare them with cheaper methods, such as DFT and RI-MP2, that can be used for large-scale calculations. We also discuss two possible definitions of interaction energies underlying the theoretically predicted metal-ion selectivity and the effect of geometry optimization on these values. Finally, popular solvation models, such as COSMO-RS and SMD, are used to demonstrate whether quantum chemical calculations can provide the overall free enthalpy (ΔG) changes in the range of the expected experimental values for the model complexes or match the experimental stability constants in the case of three complexes for which the experimental data exist. The data presented highlight several intricacies in the theoretical predictions of the experimental stability constants: the covalent character of some metal—ligand bonds (e.g., Cu(II)—thiolate) causing larger errors in the gas-phase complexation energies, inaccuracies in the treatment of solvation of the charged species, and difficulties in the definition of the reference state for Jahn—Teller unstable systems (e.g., [Cu(H$_2$O)$_6$]$^{2+}$). Although the agreement between the experimental (as derived from the stability constants) and calculated values is often within 5 kcal·mol^{-1}, in more complicated cases, it may exceed 15 kcal·mol^{-1}. Therefore, extreme caution must be exercised in assessing the subtle issues of metal ion selectivity quantitatively.
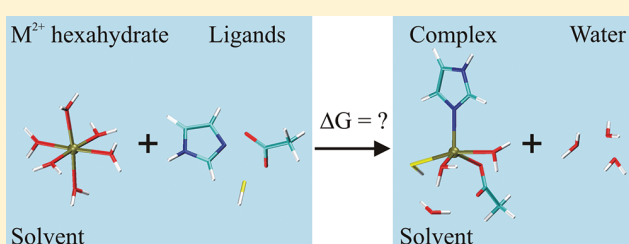
1. INTRODUCTION

In the past decade, computational chemistry has made significant progress that was fuelled by the method development, accumulated experience, and enormous increase in the available computational power. It has become an indispensable and integral part of many studies addressing the fundamental principles of chemical reactivity and molecular interactions, which are the underlying principles in most of chemistry and biology.

Owing to the consistent and enduring work of many groups, including the fundamental contributions of Hobza and co-workers,[1-3] the contemporary quantum chemical methods and computational protocols have almost converged to the situation where we can calculate the energies associated with the weak noncovalent intermolecular interactions with an accuracy rivaling the experimental values.[1] The differences of a few tenths of kcal·mol^{-1} with respect to the most accurate

experimental data can be obtained for small model complexes, such as dimers of nucleic acids.[4] Moreover, cheaper and reasonably accurate methods, such as DFT+D (DFT method including empirical dispersion) have been developed[5-7] and calibrated on the standard S22 (ref 6) and S66 data sets[8] (benchmark values of the interaction energies for the weakly bound molecular complexes). They can be conveniently used in quantum chemical calculations for systems containing several hundred atoms and, as such, are optimal tools in addressing many problems in computational drug design.[9]

The situation is quite different in the field of ionic interactions, mostly concerning metal ion coordination in biomolecules. Here, the standard thermodynamic cycle consisting of the free energy (or free enthalpy) differences in the gas-phase complexation of the ions and ligands and the difference in the solvation free energies of the complexed and free ligands is usually associated with large energies of several hundred $kcal \cdot mol^{-1}$ (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 ΔG values of several $kcal \cdot mol^{-1}$.

Still, small differences in the free energies of the complexation of metal ions in biomolecules (corresponding to experimental stability constants) govern many fundamental phenomena, such as metal ion selectivity in biomolecules.[10−14] Fully understanding these factors may assist in answering fundamental questions, such as why nature selected various metal ions for performing specific functions.[15]

Although various experimental methods to determine the overall thermodynamics and stoichiometry of metal uptake by biological systems exist,[16−18] theoretical calculations represent a unique and complementary tool to correlate the thermodynamics with the structural details.[19−22] Moreover, the calculations may lead, in principle, to a quantitative assessment of various energetic (and entropic) terms comprising the overall free energy (or free enthalpy) value,[23,24] which is a necessary prerequisite for obtaining insight and (presumably) control over the metal-ion selectivity and a possible design of the new functionalities into the peptide sequences.[25]

There have been systematic efforts to explore the selectivity of metal binding by computational chemistry.[26−34] However, to the best of our knowledge, there is not a computational method or protocol available that would enable us to predict these properties reliably with an accuracy challenging the experimental thermodynamic values acquired in condensed phase, i.e., within an accuracy of 1−2 $kcal \cdot mol^{-1}$ in the free enthalpy (ΔG) value in solution.

In our previous studies, the combination of the DFT-(B3LYP) method with medium-sized basis sets, such as Pople's 6-311+G(d,p) was used to draw the qualitative or semiquantitative trends in the binding of selected (Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, and Hg^{2+}) metal ions.[26,28] It was shown that for hydrated ions of Zn^{2+} and Ni^{2+} the relative energies accompanying the substitution of one water ligand for methanol, ammonia, or methanethiol are reasonably well predicted by the B3LYP functional in comparison with the reference QCISD(T) calculations, using mostly 6-311G(d) basis sets,[26] and that DFT methods can be also applied to high-spin octahedral Co^{2+} complexes with degenerate or near-degenerate ground electronic states.[27] As a convenient and affordable strategy, the B3LYP/6-311G+(d,p)//B3LYP/6-31+G(d) protocol was used to assess the affinity of simple functional groups representing amino acid side chains, measured as the above-mentioned energy of the substitution of one functional group for the water molecule in the first coordination sphere of the metal ion.[28] Also, the cooperative effect associated with the nonadditivity of the simultaneous binding of two functional groups to the metal ion (replacing two water molecules in the reference perhydrated complex) was investigated.[29] These attempts resulted in the proposal of specific combinations of metal-binding residues and the de novo design of specific metal-binding peptides,[30,31] including their synthesis and experimental quantification of their thermodynamic properties.[32] However, the experimentally determined overall affinity and

metal-ion selectivity of the designed peptides was not, in many cases, what had been predicted by theoretical calculations. This can be attributed to the approximations adopted in a fairly complex protocol of which the most crucial part is the evaluation of the complexation (interaction) energies of metal ions with their first-sphere ligands.

A similar computational strategy (DFT(B3LYP) with medium basis set calculations) was adopted by Dudev and Lim in their systematic investigation of the binding affinities of selected metal ions to the model binding sites corresponding to their protein counterparts.[33−35] Most of their efforts involved Mg^{2+}, Ca^{2+}, and Zn^{2+} ions in various protein environments, addressing questions related to the protonation states of protic metal-binding residues or various coordination modes of specific natural or non-natural[36] residues. Nevertheless, as mentioned above, all of these computational approaches critically depend on the accuracy of the quantum chemical methods used[37−39] and on the accuracy of the computed solvation and entropic terms to correlate the data with their experimental thermodynamic counterparts.

The aim of this study is the careful evaluation and bench-marking of the ab initio and density functional methods for the calculation of the interaction (complexation) energies of four representative metal ions, open-shell Fe^{2+} and Cu^{2+}, and closed-shell Zn^{2+} and Cd^{2+}, with model metal-binding sites. It also includes the critical assessment of the geometry effects on the calculated values and an attempt to address the solvation and entropic effects associated with metal binding in biomolecules. These efforts shall bring us closer to addressing the intriguing questions of metal-ion selectivity and affinity with quantitative accuracy by means of theoretical calculations.

2. METHODS

2.1. Computational Details. All of the calculations reported in this work were performed using the TURBOMOLE 6.3 program. The quantum chemical calculations were performed using the density functional theory (DFT) and correlated ab initio methods. The geometry optimizations of the systems studied in sections 3.2, 3.3, and 3.4 were carried out at the DFT level, employing the Perdew−Burke−Ernzerhof (PBE) functional.[40] The systems of section 3.1 were optimized using the RI-MP2[41] method (density-fitted second-order Møller−Plesset perturbation theory) or at the DFT level, employing one of the following functionals: PBE, B3LYP, or M06. The DFT(PBE) calculations were also expedited by expanding the Coulomb integrals in an auxiliary basis set, using the resolution-of-identity (RI-J) approximation (density fitting).[42] For all of the geometry optimizations, the def-SV(P) basis set was employed on all the atoms.[43,44]

The single-point DFT energies were calculated using the PBE,[40] B3LYP,[45,46] TPSS,[47] M06,[48] M052X,[49] BHLYP,[45a−c] and MPWB1K[50] functionals. For most of these calculations the def2-TZVP basis set was employed on all the atoms.[43] In few of the DFT calculations, either cc-pVTZ was employed for all the atoms or a combination of 6-31G* for light atoms and def2-TZVP for the metal ions. The ab initio reference energies were calculated using the CCSD(T) method (UCCSD(T) in case of open-shell systems). In addition to the above-mentioned basis sets, the correlation consistent aug-cc-pVDZ,[51] aug-cc-VTZ,[52] and aug-cc-pVQZ[50] basis sets were used (the last one only for the single point RI-MP2 calculations). To allow for solvation effects, the conductor-like screening model (COSMO) method,[53] an
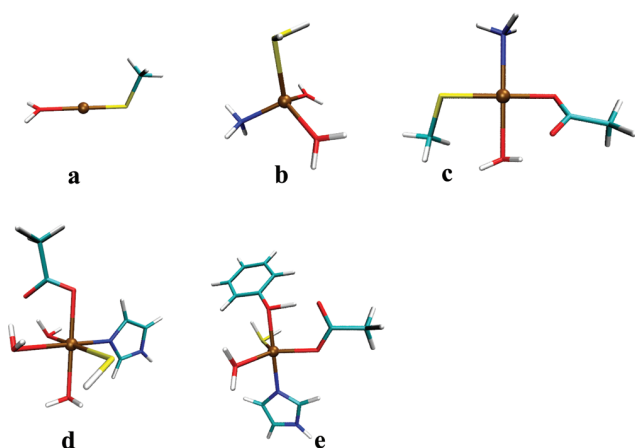
Figure 1. Model complexes: (a) $[M^{II}(CH_3S)(H_2O)]^+$ in linear coordination geometry; (b) $M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ in tetrahedral coordination geometry; (c) $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ in square-planar coordination geometry; (d) $[M^{II}(H_2O)_3(SH)$-$(CH_3COO)(Im)]$ in octahedral coordination geometry; (e) $[M^{II}$-$(H_2S)(H_2O)(CH_3COO)(PhOH)(Im)]^+$ in trigonal bipyramidal coordination geometry.

improved COSMO-RS,[54,55] and SMD[56] (with the default parameters) were used as solvation models. The COSMO radii used for the studied ions were 2.0 Å for Mn and Zn, 2.2 Å for Cd, and 2.4 Å for Hg. The Gibbs free energy was then calculated as the sum of the following contributions:

$$G = E_{el} + G_{solv} + E_{ZPE} - RT \ln(q_{trans}q_{rot}q_{vib}) \qquad (1)$$

where E_{el} is the in vacuo energy of the system, G_{solv} is the solvation free energy, E_{ZPE} is the zero-point energy, and $-RT \ln(q_{trans}q_{rot}q_{vib})$ accounts for the entropic terms and the thermal correction to the enthalpy obtained from a frequency calculation at 298 K and 1 atm using the ideal-gas approximation.[57]

Moreover, the correction of $(1.9 \cdot \Delta n)$ kcal·mol^{-1} (corresponding to the difference between the concentration of the ideal gas at 298 K and 1 atm and its 1 mol·L^{-1} concentration) has been applied for the reactions in which the number of moles (Δn) changed.

2.2. Model Systems. The basic set included five model systems that should cover a wide variety of metal−ligand interactions encountered in the metal-binding sites (Figure 1).

The choice of the ligands and their combinations is rather arbitrary. We aimed to include both charged and uncharged ligands, all three metal-binding atoms occurring in proteins (O, N, S) and a range of ligand size: from H$_2$O, H$_2$S up to PhOH.

Given the set of the ligands, all three open-shell metal ions were considered in their high-spin states ($S = 1/2$ for Cu^{2+}, $S = 2$ for Fe^{2+}, and $S = 5/2$ for Mn^{2+}), which should be their ground electronic states.

2.3. Interaction Energies and Gibbs Energies (Free Enthalpies). In previous works,[26,28,29,31] the authors have introduced the following equation to evaluate the metal ion selectivity of the given site quantitatively. The interaction energy has been defined as

$$\Delta E_{int}'(L_n) = E([ML_n]^c) - E([M(H_2O)_n]^{2+})$$
$$- E(Bq_M L_n^{(c-2)}) + E(Bq_M(H_2O)_n) \qquad (2)$$

where (L_n) represents the set of ligands, $\{L_i\}$, Bq$_M$ stands for the ghost atom representing the metal ion M^{2+}, and c is the charge of the

$[ML_n]$ complex. $\Delta E_{int}'$ is therefore an interaction energy of metal ions with the preorganized set of ligands and includes a correction for the larger part of the basis set superposition error (BSSE) with respect to the arbitrarily chosen reference state: perhydrated M^{2+} complex in a given coordination geometry. More precisely, it is the energy needed for (or acquired in) the transfer of the ion from the reference state to the preorganized (L_n) site.

This equation works well for the relative trends in the computed interaction (complexation) energies. It also accounts for interligand interaction, which can be desirable in cases where the complex metal-binding site in the biomolecule is represented by the first-sphere ligands and, therefore, some of these interligand interactions might not exist in the protein.

However, for the purpose of the current study, there are two drawbacks with the usage of eq 2. First, it cannot be related, in a direct way, to the experimental stability constants. Second, in the case of ligands with a formal negative charge (e.g., carboxylates, thiolates), the $E(Bq_M L_n^{(c-2)})$ term is quite sensitive to the molecular geometry and small changes in the geometry can result in differences of several kcal·mol^{-1} for the $(E(Bq_M L_n^{(c-2)}) - E(Bq_M(H_2O)_n))$ term owing to the Coulomb repulsion of two negatively charged ligands.

Therefore, we present interaction energies and free enthalpies corresponding to the overall process in solution (or gas phase):

$$\{L_1 + \ ... \ + L_n\}^{c-2} + [M(H_2O)_m]^{2+} \rightarrow [ML_n]^c + mH_2O \qquad (3)$$
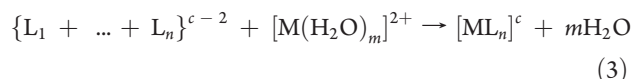
where the $\{L_i\}^{c-2}$ notation is used to denote the overall charge of the (noninteracting) ligands; $m = n$ for the benchmarking purposes (molecular geometries and molecular energies; sections 3.1 and 3.2. below) or $m = 6$ for assessments of the correlation between the calculated and experimental stability constants (to refer to the standard reference state of the studied ions in solution, presumably their hexahydrates)[58] Correspondingly, ΔE_{int} (or ΔG_{int}) is defined as

$$\Delta E_{int}(L_n) = E([ML_n]^c) + mE((H_2O))$$
$$- E([M(H_2O)_m]^{2+}) - (\Sigma_i E(L_i))^{c-2} \qquad (4)$$

and is used throughout the paper, if not explicitly mentioned otherwise.

3. RESULTS AND DISCUSSION

3.1. Calibrating the Molecular Geometry Optimizations: Performance of the RI-MP2 and DFT Methods. Concerning the choice of computational protocol, it should be borne in mind that the requirements for geometry optimizations and the subsequent single-point energy calculations can be significantly different. To this end, we have conducted an analysis of the effects of geometry optimization on the values of the computed interaction energies (ΔE_{int}). The performance of three methods, RI-DFT(PBE), RI-DFT+D(PBE), and RI-MP2, is summarized in Table 1. Three more methods, DFT(B3LYP), DFT+D(B3LYP), and M06, were examined, but because the situation is very similar in all aspects to the RI-DFT methods, these results are presented in Table S1 of the Supporting Information. To focus solely on the geometrical effects, all of the displayed interaction energies were computed using RI-MP2/aug-cc-pVDZ.

Table 1. Interaction Energies of the Studied Metal Ions with Model Binding Sites, $\Delta E_{int}(L_n)$, Corresponding to the $\{L_1 + ... + L_n\}^{c-2} + [M(H_2O)_n]^{2+} \rightarrow [ML_n]^c + nH_2O$ Process[a]

metal ion	coord geom	ΔE_{int} optimization method			DFT vs DFT+D		DFT+D vs MP2		DFT vs MP2	
		DFT	DFT+D	MP2	ΔE_{int}	rmsd	ΔE_{int}	rmsd	ΔE_{int}	rmsd
Cd²⁺	$[MX_2]^{+ b}$	−280.2	−280.3	−280.0	0.1	0.022	−0.3	0.026	0.2	0.043
	$[MX_4]^c$	−374.9	−373.7	−375.0	−1.2	0.663	1.3	0.722	−0.2	0.148
	$[MY_4]^{2+ d}$	−17.9	−18.0	−17.8	0.1	0.044	−0.2	0.045	0.1	0.041
Cu²⁺	$[MX_2]^+$	−341.8	−341.9	−339.0	0.1	0.019	−2.9	0.045	2.8	0.035
	$[MX_4]$	−374.3	−374.6	−376.1	0.3	0.077	1.6	0.839	−1.8	0.870
	$[MY_4]^{2+}$	−17.5	−17.6	−18.0	0.1	0.015	0.3	0.998	−0.4	0.997

[a] The $\Delta E_{int}(L_n)$ were calculated using the RI-MP2/aug-cc-pVDZ method and $[M(H_2O)_n]^{2+}$ system as a reference. The equilibrium geometries were obtained using the RI-DFT(PBE)/def-SV(P) method with (DFT+D) and without (DFT) empirical dispersion correction and using the RI-MP2/def-SV(P) method. All of the systems were constrained during optimization to the given coordination geometries (i.e., fixing the L—M—L angles). All the energetical values are in kcal·mol⁻¹. The RMSD values are in Å. [b] $[MX_2]^+$ stands for the $[M^{II}(CH_3S)(H_2O)]^+$ complex. [c] $[MX_4]$ stands for the $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ complex. [d] $[MY_4]^{2+}$ stands for the $[M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ complex.

Although the performance of the DFT methods in calculations of ΔE_{int} values (in comparison with the reference CCSD(T) values; see section 3.2) was notably inferior to the MP2 method, data presented in Table 1 suggest that the effect of geometry optimization on the ΔE_{int} is less pronounced, though not negligible. Six of the 18 inspected pairs show $\Delta\Delta E_{int}$ larger than 1 kcal·mol⁻¹. This corresponds to Cd²⁺, and Cu²⁺ $[M^{II}(CH_3S)-(NH_3)(H_2O)(CH_3COO)]$ square-planar complexes and the already discussed $[Cu^{II}(CH_3S)(H_2O)]^+$ complex. In the case of the former two systems, these differences can be attributed to the fairly small but non-negligible rmsd values. Contrary to the situation in the field of weak noncovalent interactions, these small changes in the geometry in systems containing two negatively charged ligands and a positively charged metal ion have resulted in significant variations in the ΔE_{int} values, which is attributable to the strong Coulomb interactions.

On the other hand, the differences of almost 3 kcal·mol⁻¹ for the $[Cu^{II}(CH_3S)(H_2O)]^+$ complex in conjunction with the minute rmsd values, most likely originate from the intricacies in its electronic structure, which are discussed in section 3.2.

Systems with neutral ligands show much smaller dependence of ΔE_{int} on the molecular geometry. The potential-energy surface of the metal—ligand complex is relatively shallow so that the rmsd value of almost 1 Å ($[Cu^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ system), corresponding to the rotation of one of the ligands, does not have a noticeable effect on the ΔE_{int} (Table 1).

On the other hand, systems with a higher number of ligands and their complexity (charge, polarization, or interligand dispersion forces) can pose a greater challenge and cause stronger dependence on the chosen methods, facing not only the danger of encountering a shift of the local minima on the PES but also the threat of converging to a different minima.

It should be noted that the optimizations presented up to this moment were carried out using constrained L—M—L angles to preserve the desired coordination geometry. The risks mentioned above are expected to be higher for unconstrained optimizations, and starting the geometry optimization with a reasonable initial guess increases in importance. Using the above geometries as starting points and the very same methods, we performed unconstrained optimizations. The same analysis containing the rmsd and E_{int} values can be found in Table S2 in the Supporting Information. The discussion for unconstrained

optimizations is essentially equivalent to the one for constrained optimizations, with the divergence among the methods being visibly pronounced.

Despite the differences, the geometries carry the same fundamental characteristics and the dependence of ΔE_{int} on the method used for geometry optimization is not very large. For this reasons, we consider the use of computationally cheap RI-DFT(PBE) and RI-DFT+D(PBE) for further discussion a reasonable choice.

3.2. Gas-Phase Complexation Energies: CCSD(T) Reference Calculations. In the field of weak intermolecular interactions, CCSD(T) calculations are considered[1] to be the "golden standard". The performance of the CCSD(T) method is less tested and understood in the calculations of the complexation energies of charged metal ions with biomolecular ligands, especially in the case of open-shell transition metal systems, because the request for the single-reference character of the zero-order HF wave function is not automatically fulfilled. The same holds true for the UCCSD(T) schemes applied for systems with an open-shell single-reference character of the wave function (such as many Cu²⁺ complexes). In our opinion, the results must be viewed with a greater degree of caution in comparison with the results obtained for the closed-shell, electronically saturated systems.

Notwithstanding the above limitations, we have carried out CCSD(T) single-point calculations for three smaller systems, linear $[M^{II}(CH_3S)(H_2O)]^+$, tetrahedral $[M^{II}(H_2O)_2(H_2S)-(NH_3)]^{2+}$, and square-planar $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3-COO)]$ (Figure 1a–c) and the corresponding perhydrated complexes: $[M^{II}(H_2O)_2]^{2+}$ and $[M^{II}(H_2O)_4]^{2+}$, using medium to large basis sets (up to aug-cc-pVTZ). The geometry optimizations were conducted at the RI-DFT(PBE)/def-SV(P) level. Four basis sets were used: def2-TZVP, aug-cc-pVDZ, aug-cc-pVTZ, and aug-cc-pVQZ (only for RI-MP2). The calculated values of $\Delta E_{int}(L_n)$, defined by eq 4, are summarized in Table 2.

As can be seen in Table 2, the values of $\Delta E_{int}(L_n)$ vary between −5 and −30 kcal·mol⁻¹ for complexes with neutral ligands up to −400 kcal·mol⁻¹ for square-planar model complexes with two anionic ligands. Several trends can be seen in the computed values (e.g., a generally stronger binding of Cu²⁺ ion, which is in line with the Irving—Williams series; preference of Cd²⁺ for softer ligands, which follows the hard and soft acids and bases

11397

dx.doi.org/10.1021/jp205442p |J. Phys. Chem. A 2011, 115, 11394–11402

Table 2. Interaction Energies of the Studied Metal Ions with Model Binding Sites, $\Delta E_{int}(L_n)$ (for the Reaction $\{L_1 + ... + L_n\}^{c-2} + [M(H_2O)_n]^{2+} \rightarrow [ML_n]^c + nH_2O$) Calculated Using Various ab Initio and DFT Methods and $[M(H_2O)_n]^{2+}$ as References[a]

coord geom[b]	complex	RI-MP2/ aDZ[c]	RI-MP2/ aTZ[d]	RI-MP2/ aQZ[e]	CCSD(T)/ aDZ	CCSD(T)/ aTZ	PBE+D/ TZ[f]	B3LYP+D/ TZ	TPSS+D/ TZ	M06/ TZ	BHLYP/ TZ	M052X/ TZ	MPWB1K/ TZ
LI	$[ZnX_2]^{+g}$	−278.6	−280.1	−280.3	−279.4	−280.5	−292.5	−288.1	−288.3	−290.9	−281.3	−282.2	−282.8
	$[CdX_2]^+$	−280.2	−282.4	−282.8	−279.4	−280.9	−294.9	−290.4	−290.3	−290.6	−283.3	−284.1	−285.7
	$[CuX_2]^+$	−341.8	−342.6	−343.3	−320.3	−320.5	−337.6	−334.8	−333.6	−341.1	−314.2	−316.9	−320.4
	$[FeX_2]^+$	−262.4	−261.5	−265.1	−266.8	−266.2	−293.8	−281.7	−289.7	−282.1	−268.0	−269.2	−272.2
SQ	$[ZnX_4]^h$	−379.9	−380.6	−380.6	−379.8	−380.2	−385.6	−379.1	−384.6	−385.0	−373.9	−380.2	−381.8
	$[CdX_4]$	−374.9	−375.8	−375.7	−374.8	−375.1	−384.3	−377.9	−382.0	−380.2	−370.5	−375.3	−376.2
	$[CuX_4]$	−374.3	−374.6	−374.7	−380.2		−398.9	−387.4	−395.8	−391.5	−372.3	−376.2	−379.9
	$[FeX_4]$	−362.1	−364.6	−364.7			−381.4	−372.8	−378.9	−372.9	−364.4	−366.0	−368.4
TH	$[ZnY_4]^{2+i}$	−16.5	−15.7	−15.6	−15.3	−14.4	−17.1	−13.3	−15.7	−14.6	−9.7	−11.6	−11.6
	$[CdY_4]^{2+}$	−17.9	−17.3	−17.3	−17.2	−16.3	−20.3	−16.6	−18.4	−17.5	−12.6	−14.8	−14.3
	$[CuY_4]^{2+}$	−17.5	−16.5	−16.7	−19.4	−18.0	−29.1	−22.2	−27.1	−25.0	−12.3	−14.0	−15.6
	$[FeY_4]^{2+}$	−7.0	−6.7	−7.4	−6.9	−6.6	−10.8	−6.3	−9.4	−7.3	−2.8	−4.6	−4.6

[a] The equilibrium geometries were obtained using the RI-DFT(PBE)/def-SV(P) method and constraining the systems to the given coordination geometries (i.e., fixing the L-M–L angles). All of the values are in kcal·mol^{-1}. [b] LI, linear; TH, tetrahedral; SQ, square-planar coordination geometry. [c] aug-cc-pVDZ basis set. [d] aug-cc-pVTZ basis set. [e] aug-cc-pVQZ basis set. [f] def2-TZVP basis set. [g] $[MX_2]^+$ stands for the $[M^{II}(CH_3S)(H_2O)]^+$ complex. [h] $[MX_4]$ stands for the $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ complex. [i] $[MY_4]^{2+}$ stands for the $[M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ complex.

principle) that were discussed in previous works[13,28] and are not, therefore, within the primary scope of the current study.

The primary interest is the critical evaluations of selected wave function and DFT methods. First, we observe that the CCSD-(T)/aug-cc-pVTZ values can be considered as reasonably converged. The difference between the CCSD(T)/aug-cc-pVDZ and CCSD(T)/aug-cc-pVTZ values is mostly within 1 kcal·mol^{-1}, whereas the difference between RI-MP2/aug-cc-pVTZ and RI-MP2/aug-cc-pVQZ is mostly within 0.5 kcal·mol^{-1}. The notable exception is the $[Fe^{II}(CH_3S)(H_2O)]^+$ complex, for which $\Delta E(aTZ \rightarrow aQZ) = 3.5$ kcal·mol^{-1}. Therefore, with a certain degree of caution, we can consider the CCSD(T)/aug-cc-pVTZ values as the reference in the following discussions. Admittedly, the degree of uncertainty is higher for the Cu^{2+} and Fe^{2+} species.

Concerning the method performance, it is encouraging to discover that the computationally inexpensive RI-MP2 method is performing quite well in predicting the values of gas-phase complexation energies. At the computational price of a DFT calculation (without the density fitting), one obtains values that are usually within 1.5 kcal·mol^{-1} from the reference values. Again, the notable difference is the linear complex of $[Cu^{II}-(CH_3S)(H_2O)]^+$, for which the MP2 and CCSD(T) values differ by formidable 20 kcal·mol^{-1}. We have carefully analyzed this seemingly simple copper(II)−thiolate system and have found several intricacies in its electronic structure. For example, it takes quite an effort to converge the HF wave function to a (presumably) global minimum. Not having a more thoughtful explanation at hand, we conclude that it is an anticipated high degree of charge transfer and covalency in the Cu(II)−S bond that lies behind the fairly demanding electronic structure of this particular complex, which issues a warning for a straightforward usage of the quantum chemical method for the complexation energies of open-shell metal ions without further critical assessment of the computed values. This observation is corroborated by a similar error found for the Cu(II)−S bond-containing $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ complex, for which we could not obtain a reference CCSD(T)/aug-cc-pVTZ value;

however, the difference between the MP2 and CCSD(T) using the aug-cc-pVDZ basis set (∼6 kcal·mol^{-1}) highlights that a similar problem might have been encountered.

On the other hand, the general agreement of DFT methods with the reference values is less spectacular. Although they perform reasonably well for complexes with uncharged ligands (in line with our previous findings[26] for the significantly more limited set of Zn^{2+} and Ni^{2+} small complexes with neutral ligands), they deviate often by more than 10 kcal·mol^{-1} for systems with anionic ligands (LI, SQ). From the set of eight functionals tested, a certain dependence of the ΔE_{int} difference from reference values on the amount of the exact HF exchange included in the functional can be observed. The best performance is exhibited by MPWB1K (44% of HF exchange) and M052X (56% of HF exchange) with values that are mostly within 2.5 kcal·mol^{-1}. The shifts do not seem to be systematic (large relative standard deviation of errors), but they are relatively small (small mean unsigned error, MUE). BHLYP (50% of HF exchange) also gives very good results, with somewhat larger MUE. The popular B3LYP with only 20% of HF exchange still produces reasonable results, with MUE around 5 kcal·mol^{-1}. The performance is not determined solely by the HF exchange though, e.g., M06 (27%) produced results with larger MUE and STD than B3LYP. Lagging behind at the very tail of our list are TPSS, and PBE. Although the relative standard deviations are among the smallest, the overall standard deviation is hampered by MUE values of ca. 10 kcal·mol^{-1}.

TH complexes, containing no charged ligands, seem to be generally less problematic than the other two types of complexes. The nature of the LI system, discussed above, probably also contributes to the overall error. E_{int} values of systems with open shells are accompanied with larger errors than those of closed shells.

In Table S3 (Supporting Information), we show further data that might be of interest for discussions concerning the performance of standard and widely used methods in the calculations of complexation energies. These include the spin-component scaled MP2 (SCS-MP2)[59] or scaled opposite-spin

Table 3. Free Energies Associated with the Complexation of Metal Ions (Cd^{2+}, Cu^{2+}) in Solution, Calculated as $\Delta G = \Delta E_{int} + \Delta G_{solv} + \Delta E_{ZPE} - \Delta(RT \ln(q_{trans}q_{rot}q_{vib}))$, Using Various Protocols and $[M(H_2O)_6]^{2+}$ as the Reference^a

coord geom[b]	complex	COSMO-RS						SMD	
		MP2/aDZ[c]	MP2/aTZ[d]	B3LYP+D/TZ[e]	B3LYP/TZ	MPWB1K/TZ	M052X/cTZ[f]	M052X/cTZ	M052X/6-31G*[g]
LI	$[CdX_2]^{+h}$	24.3	21.7	16.9	8.3	18.8	33.4	−32.8	−28.4
	$[CuX_2]^+$	−18.9	−19.0	3.4	−4.5	9.0	26.3	−17.7	−17.7
SQ	$[CdX_4]^i$	3.7	1.5	−0.4	−1.6	1.4	7.6	−44.8	−46.0
	$[CuX_4]$	2.5	2.4	−3.3	−3.9	−1.3	6.4	−40.0	−42.5
TH	$[CdY_4]^{2+j}$	5.4	5.1	7.2	2.8	7.8	15.2	−3.4	1.1
	$[CuY_4]^{2+}$	−0.4	1.0	1.7	−0.5	2.9	10.6	−1.1	6.1
OH	$[CdX_6]^k$	14.0	12.8	7.9	14.2	14.9	14.5	−29.8	−34.5
	$[CuX_6]$	10.9	12.4	5.7	13.0	11.5	12.6	−22.4	−25.7
TP	$[CdX_5]^{+l}$	9.7	10.3	15.4	22.7	22.9	22.0	−7.8	−3.3
	$[CuX_5]^+$	−5.8	−3.4	1.9	3.9	4.6	8.9	−8.1	−4.3

^{a} The equilibrium geometries were obtained using the RI-DFT(PBE)/def-SV(P) method. No constraints were imposed during the optimizations, and stability checks were performed. The COSMO-RS calculations were performed using the recommended RI-DFT(B−P)/def-TZVP protocol, for which the parameters of the procedure were optimized. All of the values are in kcal·mol^{-1}. ^{b} LI, linear; TH, tetrahedral; SQ, square-planar coordination; OH, octahedral; TP, trigonal bipyramidal starting geometry. The final geometry, though, can be distorted in many cases, especially toward the square-planar shape for Cu(II) systems. ^{c} aug-cc-pVDZ basis set. ^{d} aug-cc-pVTZ basis set. ^{e} def2-TZVP basis set. ^{f} cc-pVTZ basis set for light atoms, cc-pVTZ-PP for Zn^{2+}, Cd^{2+}, and Cu^{2+}. ^{g} 6-31G* basis set for light atoms, def2-TZVP for metal ions. ^{h} $[MX_2]^+$ stands for the $[M^{II}(CH_3S)(H_2O)]^+$ complex. ^{i} $[MX_4]$ stands for the $[M^{II}(CH_3S)(NH_3)(H_2O)(CH_3COO)]$ complex. ^{j} $[MY_4]^{2+}$ stands for the $[M^{II}(H_2O)_2(H_2S)(NH_3)]^{2+}$ complex. ^{k} $[MX_6]$ stands for the $[M^{II}((H_2O)_3(HS)(CH_3COO)(Im)]$ complex. ^{l} $[MX_5]^+$ stands for the $[M^{II}(H_2O)(H_2S)(CH_3COO)(Im)(PhOH)]^+$ complex.

(SOS-MP2)[60] methods (that are usually considered as an improvement over the MP2 method, at least for smaller molecules and noncovalent interactions) and DFT+D methods (DFT methods with empirical dispersion). Both SCS-MP2 and SOS-MP2 yielded interaction energies in poorer agreement with the reference CCSD(T) values than the standard RI-MP2 method, whereas the empirical dispersion correction has only a minor impact on $\Delta E_{int}(L_n)$. In TH and LI systems, which contain only smaller ligands, the overall contribution of the empirical dispersion to the binding energy does not exceed 1 kcal·mol^{-1}. The remaining SQ system is not only the largest of the three systems but also contains softer ligands and, consequently, has also the largest contribution of dispersion to $\Delta E_{int}(L_n)$, which is in the range 3−3.5 kcal·mol^{-1}. It can be expected that this contribution increases with the growing size of the system and the number of soft ligands. However, this contribution is mostly ligand-dependent, changing little (<1 kcal·mol^{-1}) as we vary the central metal ion, suggesting it is not essential for grasping metal-ion selectivity trends, at least for a given choice of ligands. It is also noteworthy that, at least in regard to our reference, dispersion correction does not present an improvement in all cases.

3.3. From Gas-Phase Complexation Energies to the Free Enthalpies (ΔG) in Solution. Moving from gas-phase complexation energies of the metal ions to the free enthalpies (ΔG) in solution is a highly nontrivial task. Experimental free energies are on the order of a few kcal·mol^{-1}, as can be inferred from the dissociation constants (or stability constants), whereas the gas-phase interaction energies are on the order of several tens of kcal·mol^{-1} for neutral species and hundreds of kcal·mol^{-1} for charged species. Not only for this reason is the accurate calculation of solvation and desolvation free enthalpies, especially for the charged species in water,[54] one of the most imminent challenges in contemporary computational chemistry.

Using the standard thermodynamic cycle (gas-phase complexation of metal ion and ligands and solvation of both free

and complexed species to obtain ΔG for complexation in solution) together with two recent (and presumably two of the most accurate) methods for estimating solvation energies, namely, COSMO-RS and SMD, we attempt to provide the calculated values of free enthalpies related to the metal-ion complexation in solution (Table 3). In our calculations, we apply these methods at the same level of theory at which they were parametrized and which is recommended for their use.

Besides the three "canonical" systems discussed in previous sections, we have extended the set by two other complexes, hexacoordinate $[M^{II}((H_2O)_3(HS)(CH_3COO)(Im)]$ and pentacoordinate $[M^{II}(H_2O)(H_2S)(CH_3COO)(Im)(PhOH)]^+$, where Im stands for the imidazole and PhOH is phenol. To evaluate all of the terms in ΔG values, full geometry optimizations were performed (no constraints).

As already mentioned, the overall binding energy can be split into several contributions, $\Delta G = \Delta E_{int} + \Delta G_{solv} + \Delta E_{ZPE} - \Delta(RT \ln(q_{trans}q_{rot}q_{vib}))$. The latter two terms are less likely to be the source of major errors, as their absolute values are much smaller when compared to the other two terms (cf. Tables S4 and S5, Supporting Information) and are not as sensitive to the choice of method (data not shown). The key two terms are ΔE_{int}, benchmarked and analyzed in two previous sections, and ΔG_{solv}. It is difficult, if not impossible, to evaluate the accuracy of the calculated ΔG_{solv} values, because no experimental data are available for the model systems studied here.

What is encouraging is that the final values are within the range of units (or at most tens) of kcal·mol^{-1}, which is the range of experimental dissociation constants of this class of systems (e.g., peptides with divalent metal ions).[61] The difference between the COSMO-RS and SMD solvation models for the charged species is, however, less encouraging. They both perform reasonably well for neutral molecules,[62,63] whereas the anticipated errors in solvation energies of the ionic species using the SMD model should be "only" a few kcal·mol^{-1}.[54] We may conjecture that it might be both the cumulative error in the assembly of the

11399

dx.doi.org/10.1021/jp205442p |*J. Phys. Chem. A* 2011, 115, 11394–11402

Table 4. Free Energies Associated with the Complexation of Metal Ions (Mn^{2+}, Fe^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}) in Solution, Calculated as $\Delta G = \Delta E_{int} + \Delta G_{solv} + \Delta E_{ZPE} - \Delta(RT \ln(q_{trans}q_{rot}q_{vib}))$, Using Various Methods and $[M(H_2O)_6]^{2+}$ as Reference[a]

complex	exp[b]	MP2/aTZ[c]	M06/TZ[d]	B3LYP/TZ[e]	B3LYP+D/TZ[e]	MPWB1K/TZ	M052X/cTZ(SMD)
$[ZnX_6]^{2+f}$	−3.9	−4.8	−3.2	−2.1	−4.2	−1.3	−11.8
$[CdX_6]^{2+}$	−3.9	−6.2	−5.4	−5.6	−7.1	−4.8	−14.7
$[CuX_6]^{2+}$	−5.1	−22.6	−24.9	−20.4	−22.4 (−10.0)	−22.7	−23.1 (−18.2)
$[FeX_6]^{2+}$	−2.3	−0.3	0.4	1.4	−0.9	1.4	−5.7
$[MnX_6]^{2+}$	−1.0	−0.7	0.5	1.0	−1.1	0.8	−8.2
$[ZnY_6]^{2+g}$	−2.4	−5.7	−1.6	−0.3	−4.1	−0.6	−3.0
$[CdY_6]^{2+}$	−2.5	−8.9	−4.8	−4.2	−7.1	−4.3	−5.0
$[CuY_6]^{2+}$	−3.6	−25.1	−24.5	−16.3	−20.6 (−3.2)	−23.3	−11.8 (−11.5)
$[FeY_6]^{2+}$		1.4	1.0	−2.7	0.5		−0.8
$[MnY_6]^{2+}$	−0.7	−5.5	−2.0	−0.5	−4.0	−1.2	−1.7
$[ZnZ_6]^{+h}$	−1.1	12.0	13.2	14.6	12.0	12.7	−11.3
$[CdZ_6]^{+}$	−1.6	13.0	13.8	14.8	11.7	13.6	−10.1
$[CuZ_6]^{+}$	−2.0	1.4	−1.2	2.8	1.2 (7.2)	0.1	−19.1 (−19.5)
$[FeZ_6]^{+}$	−0.8	13.3	14.2	15.1	12.6	12.9	−10.2
$[MnZ_6]^{+}$	−0.8	13.3	14.7	15.0	11.4	13.1	−11.1

[a] The equilibrium geometries were obtained using the RI-DFT(PBE)/def-SV(P) method along with empirical dispersion correction. No constraints were imposed during the optimizations, and stability checks were performed. COSMO-RS was used for the calculation of the solvation energies. The values in brackets are for tetracoordinated systems; see text for details. All of the values are in kcal·mol⁻¹. [b] Experimental values for ionic strength equal to 2 M ($[MX_6]$ systems), 0.5 M ($[MY_6]$ systems), and 0 M ($[MZ_6]$ systems). [c] RI-MP2 method, aug-cc-pVTZ basis set. [d] M06 functional, def2-TZVP basis set. [e] B3LYP functional, def2-TZVP basis set. [f] $[MX_6]^{2+}$ stands for the $[M^{II}(NH_3)_2(H_2O)_4]^{2+}$ complex. [g] $[MY_6]^{2+}$ stands for the $[M^{II}(Im)(H_2O)_5]^{2+}$ complex. [h] $[MZ_6]^{+}$ stands for the $[M^{II}(CH_3COO)(H_2O)_5]^{+}$ complex.

complex from several ionic species (each calculated with non-negligible error) or greater errors in the solvation energies for the doubly charged complexes (as has been mentioned in the case of Ru(II) hexahydrates for PCM-like models).[64] In our opinion, the COSMO-RS values seem to be more realistic, because they would translate into K_D's of ∼10^{10} to 10^{-10} under normal pressure and temperature. However, with respect to the previous discussion, one has to admit errors of more than 10 kcal·mol⁻¹ in the overall values (at least for the charged ligands).

It is interesting to notice that M052X provides results quite different from MP2/aTZ. This is puzzling, as these methods produced very similar values in section 3.2. The differences between these two cases are the use of unconstrained geometries, which has only a minor impact on E_{int}, and the use of hexahydrated metal ion as reference. It is the latter that is responsible for the discrepancy. On closer inspection, the predicted energy of dissociation of one water molecule from the $[M(H_2O)_n]^{2+}$ complex differs by ∼2 kcal·mol⁻¹ between MP2 and M052X. The error thus cumulates and is most pronounced for linear complexes, where the reference states differ by four molecules (error of ∼9 kcal·mol⁻¹) and gradually decreases through the square-planar and tetrahedral complexes (4−5 kcal·mol⁻¹) and vanish for the octahedral complex. This explanation, however, does not account for the large differences in trigonal bipyramidal complexes.

The same inspection of the MPWB1K functional shows that the dissociation energy of the $[M(H_2O)_n]^{2+}$ complexes is almost identical to that for MP2 and so the two methods still provide very similar results. Again, the trigonal bipyramidal system is not accounted for.

Also, the dispersion corrections seem to be of higher importance than in the previous discussion, because hexahydrated metal ions were used as the reference (as opposed to n-hydrated metal ions in the previous two sections, where n is the coordination

number). The different size of the systems may account for the different amount of the dispersion contribution.

3.4. Comparison of the Calculated and Experimental Stability Constants. Having seen that the protocol described in section 3.3 provides ΔG_{int} values in the expected range, we tried to examine the agreement of these values with experiment. Three simple ligands (imidazole, acetate, and ammonia) were chosen, and the corresponding three systems were theoretically investigated: $[M^{II}(NH_3)_2(H_2O)_4]^{2+}$, $[M^{II}(Im)(H_2O)_5]^{2+}$, and $[M^{II}(CH_3COO)(H_2O)_5]^{+}$. The results are summarized in Table 4.

All three ligands behave in agreement with the Irving−Williams series, exhibiting the highest preference for Cu(II), followed by Zn(II) and Cd(II) and tailed by Fe(II) and Mn(II). The range of the experimental values is quite narrow, from −0.7 kcal·mol⁻¹ for $[Mn^{II}(CH_3COO)(H_2O)_5]^{+}$ to −5.1 kcal·mol⁻¹ for $[Cu^{II}(NH_3)_2(H_2O)_4]^{2+}$.

On the other hand, the predicted values are in the range −25 to +15 kcal/mol. A quantitative prediction of the binding free energy thus seems to be beyond the capability of this model. However, the general trends can be obtained. The free binding energy of Cu(II) stands out among the other metal ions. This is partly connected with the solvation energy of the reference hexahydrated complex. Unlike all of the other metal ions investigated herein, Cu(II) does not prefer octahedral coordination geometry and can break down into a square-planar complex with two water molecules coordinated in the second sphere or acquires a square-pyramidal geometry with one water molecule coordinated in the second sphere. In Figure 2, we have depicted several identified local minima of the Cu^{2+} hexahydrates. Depending on the local minima obtained during the optimization, the solvation energies differ by almost 5 kcal·mol⁻¹. This applies mostly to reference, but the distortion is observable also in other Cu(II) systems. In the calculations of final free enthalpy values, one always considers the lowest minimum, but we wish to
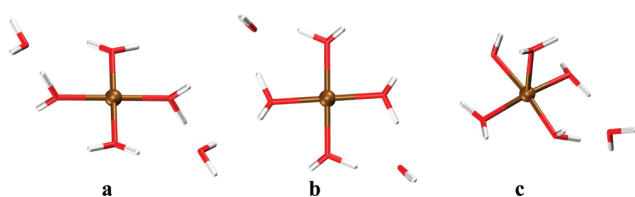
Figure 2. Three geometries identified as local minima of the Jahn—Teller unstable $[Cu(H_2O)_6]^{2+}$ complex using the RI-DFT(PBE) method: (a) gas-phase optimization; (b), (c) optimizations in a polarized continuum (COSMO model).

mention these problems to illustrate that it is not straightforward to identify the global minimum and that it might be difficult to ascertain the one to one structural mapping between the real systems in solvent and the theoretical models.

To tackle this problem, we repeated the calculations for Cu(II) systems (DFT+D(B3LYP)/def2-TZVP and COSMO-RS for the solvation energies), but this time reducing the coordination number to four, thus getting rid of the second coordination sphere. Such an approach is justified because the question of the structure of the first coordination sphere of the Cu(II) hydrate (including coordination number) is a matter of debate that can be, for example, found in ref 38 and references therein. The resultant values, presented in brackets in Table 4, are now in much better agreement with expectations, at least for $[Cu^{II}(NH_3)_2$-$(H_2O)_2]^{2+}$ and $[Cu^{II}(CH_3COO)(H_2O)_3]^+$, where the ΔG values are a few kcal·mol^{-1} lower than for the rest of the metal ions. $[Cu^{II}(Im)(H_2O)_3]^{2+}$ lies in the middle of the range, instead of the expected lower end. The ΔG values for systems with acetate are still way above 0 kcal·mol^{-1}, which can most likely be attributed to its negative charge. The use of tetrahydrated Cu(II) with the COSMO-RS model seems to remedy partially the issues concerning the choice of its reference state. The SMD results are, on the other hand, only slightly affected by this change.

For diammonia complexes, the values given by COSMO-RS for the other four metal ions are within 2 kcal·mol^{-1} from the experimental values for all of the MP2/aug-cc-pTVZ, M06/def2-TZVP, and B3LYP(+D)/def2-TZVP. MPWB1K/def2-TZVP performs slightly worse, with errors up to 4 kcal·mol^{-1}. In comparison with experiment, the relative preference for individual metal ions is well retained. Even with the uncertainty in the Cu(II) reference geometry, it is correctly recognized as the most preferred metal ion. The values given by the SMD model are shifted by 3–11 kcal/mol. The preference for individual metal ions is well preserved, too.

For imidazole complexes, the COSMO-RS differences from the experimental values are up to 6 kcal·mol^{-1} (except Cu(II)). Despite this fact, the relative preference is mostly retained, recognizing Cu(II) as the most preferred metal, followed by Cd(II). The error is larger for MP2/aug-cc-pTVZ and B3LYP-(+D)/def2-TZVP and is around 2 kcal·mol^{-1} for M06/def2-TZVP and MPWB1K/def2-TZVP. SMD performs better in this case, showing error below 3 kcal·mol^{-1} and preserving the order of preference. For both diammonia and imidazole complexes, all three presented methods in conjunction with COSMO-RS perform reasonably well, led by MP2/aug-cc-pTVZ with only a thin margin. SMD, too, performs very well.

For acetate complexes (anionic ligand), the situation is less encouraging. Although Cu(II) is still recognized as the most preferred metal ion, the order of preference is utterly lost for the

other ions. SMD at least predicts overall negative ΔG, but it is accompanied with errors of 8–10 kcal·mol^{-1}. For COSMO-RS, errors of ~15 kcal·mol^{-1} are encountered and, moreover, the formation of these complexes is predicted to be unfavorable. In our opinion, the presence of charged species is problematic for the calculation of solvation energies and is most likely the reason for the failure of both models in this case, which again issues a warning to the straightforward usage of the described protocols for obtaining quantitatively accurate data. One solution to the problem, which is beyond the scope of this study, is the introduction of a few explicit water molecules, coordinating as the second-shell ligands in the complex, and saturating the negative charge of the free ligand in the solution.

4. CONCLUSIONS

Though not providing new concepts or opening new horizons in computational bioinorganic chemistry, we have tried to compile in this work a consistent set of data to illustrate all of the problems encountered in the attempts to assess metal ion complexation and selectivity from the first principles quantitatively. It was convincingly shown that these efforts are far from being trivial or straightforward with most of the computational subtleties arising from the two large canceling terms, gas-phase complexation energies and solvation energies of the species involved.

Using the CCSD(T)/aug-cc-pVTZ values as the reference, we for the most part obtained a reasonable agreement in the calculated complexation energies (defined as the substitution of water molecules in the reference per-hydrated system) for the computationally cheap RI-MP2/aug-cc-pVTZ protocol. The DFT methods performed less satisfactorily, with MPWB1K probably being the functional we may recommend as slightly superior to other functionals tested. Still, extreme caution must be exercised, as shown for complexes with a copper—thiolate bond (with errors up to 20 kcal·mol^{-1}), in interpreting and quantifying the results in a straightforward way.

The effects of the method for geometry optimizations were shown to be less severe, and quite satisfactory results were obtained using computationally cheap RI-DFT(PBE) optimizations.

An accurate description of the solvation seems to be the main stumbling block in our efforts to predict experimental binding constants quantitatively. From the data obtained for our model complexes, we may mildly advocate in favor of using the COSMO-RS method. Still, the differences between the two standard solvation methods studied are alarming, as well as the differences of ~15 kcal·mol^{-1} in the comparison of the calculated and experimental data for the model complex with anionic ligands. Finally, for Jahn—Teller unstable octahedral complexes, it has been shown that the selection of the correct geometry for the reference state is not an easy task.

Despite all the problems mentioned above, we have witnessed considerable progress in theoretical bioinorganic chemistry that fills with optimism that, in near future, we will converge to the same level of computational accuracy as has been achieved in the area of weak noncovalent interactions.

■ ASSOCIATED CONTENT

ⓢ Supporting Information. Equilibrium geometries of all the molecules studied and four tables of interaction energies,

11401

dx.doi.org/10.1021/jp205442p |*J. Phys. Chem. A* 2011, 115, 11394–11402

electronic energy contributions to ΔG_{int}, and solvation energy, ZPE, and $RT \ln Q$ contributions to ΔG_{int} (S1–S5). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*Tel.: +420-220-183-263. Fax: +420-220-183-578. E-mail: rulisek@ uochb.cas.cz.

■ ACKNOWLEDGMENT

■ REFERENCES

(1) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. *Chem. Rev.* **2010**, *110*, 5023–5063.

(2) Pitoňák, M; Řezáč, J.; Hobza., P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9611–9614.

(3) Gráfová, L.; Pitoňák, M; Řezáč, J.; Hobza., P. P. *J. Chem. Theory Comput.* **2010**, *6*, 2365–2376.

(4) Morgado, C. A.; Jurečka, P.; Svozil, D.; Hobza, P.; Šponer, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3522–3534.

(5) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.

(6) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. *J. Comput. Chem.* **2007**, *28*, 555–569.

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

(8) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427–2438.

(9) Fanfrlík, J; Bronowska, A. K.; Řezáč, J.; Přenosil, O.; Konvalinka, J; Hobza, P. *J. Phys. Chem. B* **2010**, *114*, 12666–12678.

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

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

(12) Glusker, J. P. *Adv. Protein Chem.* **1991**, *42*, 1–76.

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

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

(15) Williams, R. J. P. *Biometals* **2007**, *20*, 107–112.

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

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

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

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

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

(21) Senn, H. M.; Thiel, W. *Angew. Chem.-Int. Ed.* **2009**, *48*, 1198–1229.

(22) Ryde, U. *Curr. Opin. Chem. Biol.* **2003**, *7*, 136–142.

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

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

(25) Yin, H.; Slusky, J. S.; Berger, B. W.; Walter, R. S.; Vilaire, G.; Litvinov, R. I.; Lear, J. D.; Caputo, G. A.; Bennett, J. S.; DeGrado, W. F. *Science* **2007**, *315*, 1817–1822.

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

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

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

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

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

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

(32) 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.

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

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

(35) Dudev, T.; Lim, C. *J. Am. Chem. Soc.* **2006**, *128*, 1553–1561.

(36) Dudev, T.; Lim, C. *J. Phys. Chem. B* **2009**, *113*, 11754–11764.

(37) Wladkowski, B. D.; Krauss, M.; Marti, M. A.; Stevens, W. J. *J. Phys. Chem.* **1995**, *99*, 6273–6276.

(38) Rios-Font, R.; Sodupe, M.; Rodríguez-Santiago, L.; Taylor, P. R. *J. Phys. Chem. A* **2010**, *114*, 10857–10863.

(39) Jeanvoine, Y.; Spezia, R. *J. Mol. Struct. (THEOCHEM)* **2010**, *954*, 7–15.

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

(41) Eichkorn, K.; Weigen, F.; Treutler, O.; Ahlrichs, R. *Theor. Chim. Acta* **1997**, *97*, 119–124.

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

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

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

(45) (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.

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

(47) Staroverov, V. N.; Scuseria, G. E.; Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2003**, *119*, 12129–12137.

(48) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215–241.

(49) Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364–382.

(50) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.

(51) Woon, D. E.; Dunning, T. H., Jr. *J. Chem. Phys.* **1993**, *98*, 1358–1371.

(52) Kendall, R. A.; Dunning, T. H., Jr.; Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796–6806.

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

(54) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224–2235.

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

(56) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, *113*, 6378–6396.

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

(58) Åkesson, R; Pettersson, L. G. M.; Sandström, M.; Wahlgren, U. *J. Am. Chem. Soc.* **1994**, *116*, 8691–8704.

(59) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095–9102.

(60) Jung, Y.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793–9802.

(61) Petros, A. K.; Reddi, A. R.; Kennedy, M. L.; Hyslop, A. G.; Gibney, B. R. *Inorg. Chem.* **2006**, *45*, 9941–9958.

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

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

(64) Srnec, M.; Chalupský, J.; Fojta, M.; Zendlová, L.; Havran, L.; Hocek, M.; Kývala, M.; Rulíšek, L. *J. Am. Chem. Soc.* **2008**, *130*, 10947–10954.

11402

dx.doi.org/10.1021/jp205442p |*J. Phys. Chem. A* 2011, 115, 11394–11402

Correction to "Interaction of Metal Ions with Biomolecular Ligands: How Accurate Are Calculated Free Energies Associated with Metal Ion Complexation?"

Ondrej Gutten, Ivana Bešševá, and Lubomír Rulíšek*

Scrutinizing the values published in Table 4, we realized a mistake in our script that was converting stability constants obtained

Table 4. Free Energies Associated with the Complexation of Metal Ions (Mn^{2+}, Fe^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}) in Solution, Calculated as $\Delta G = \Delta E_{int} + \Delta G_{solv} + \Delta E_{ZPE} - \Delta(RT \ln(q_{trans}q_{rot}q_{vib}))$, Using Various Methods and $[M(H_2O)_6]^{2+}$ as Reference[a]

complex	exp[b]	MP2/aTZ[c]	M06/TZ[d]	B3LYP/TZ[e]	B3LYP+D/TZ[e]	MPWB1K/TZ	M052X/cTZ(SMD)
$[ZnX_6]^{2+,f}$	−6.7	−4.8	−3.2	−2.1	−4.2	−1.3	−11.8
$[Cd\,X_6]^{2+}$	−6.7	−6.2	−5.4	−5.6	−7.1	−4.8	−14.7
$[CuX_6]^{2+}$	−10.6	−22.6	−24.9	−20.4	−22.4 (−10.0)	−22.7	−23.1(−18.2)
$[FeX_6]^{2+}$	−3.4	−0.3	0.4	1.4	−0.9	1.4	−5.7
$[MnX_6]^{2+}$	−2.0	−0.7	0.5	1.0	−1.1	0.8	−8.2
$[ZnY_6]^{2+\,g}$	−3.6	−5.7	−1.6	−0.3	−4.1	−0.6	−3.0
$[Cd\,Y_6]^{2+}$	−3.7	−8.9	−4.8	−4.2	−7.1	−4.3	−5.0
$[CuY_6]^{2+}$	−5.7	−25.1	−24.5	−16.3	−20.6 (−3.2)	−23.3	−11.8(−11.5)
$[FeY_6]^{2+}$			1.4	1.0	−2.7	0.5	−0.8
$[MnY_6]^{2+}$	−1.8	−5.5	−2.0	−0.5	−4.0	−1.2	−1.7
$[ZnZ_6]^{+\,h}$	−2.2	12.0	13.2	14.6	12.0	12.7	−11.3
$[CdZ_6]^{+}$	−2.6	13.0	13.8	14.8	11.7	13.6	−10.1
$[CuZ_6]^{+}$	−3.0	1.4	−1.2	2.8	1.2 (7.2)	0.1	−19.1(−19.5)
$[FeZ_6]^{+}$	−1.9	13.3	14.2	15.1	12.6	12.9	−10.2
$[MnZ_6]^{+}$	−1.9	13.3	14.7	15.0	11.4	13.1	−11.1

[a]The equilibrium geometries were obtained using the RI-DFT(PBE)/def-SV(P) method along with empirical dispersion correction. No constraints were imposed during the optimizations, and stability checks were performed. COSMO-RS was used for the calculation of the solvation energies. The values in parentheses are for tetracoordinated systems; see the text for details. All of the values are in kcal mol^{-1}. [b]Experimental values for ionic strength equal to 2 M ($[MX_6]$ systems), 0.5 M ($[MY_6]$ systems), and 0 M ($[MZ_6]$ systems). [c]RI-MP2 method, aug-cc-pVTZ basis set. [d]M06 functional, def2-TZVP basis set. [e]B3LYP functional, def2-TZVP basis set [f]$[MX_6]^{2+}$ stands for the $[M^{II}(NH_3)_2(H_2O)_4]^{2+}$ complex. [g]$[MY_6]^{2+}$ stands for the $[M^{II}(Im)(H_2O)_5]^{2+}$ complex. [h]$[MZ_6]^{+}$ stands for the $[M^{II}(CH_3COO)(H_2O)_5]^{+}$ complex.

from the literature into free energy changes associated with corresponding complexation reactions. We remedy our mistake and the correct values (column 2) listed in the new Table 4 presented below. In most cases, the values are ∼1 kcal mol^{-1} more negative, with few exceptions that amount to 2−6 kcal mol^{-1} (more negative). The correct values are generally in slightly better agreement with the calculated values. The corrections leave our observations and conclusions mostly unchanged.

Predicting the Stability Constants of Metal-Ion Complexes from First Principles
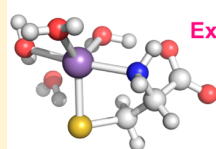
Ondrej Gutten and Lubomír Rulíšek*

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
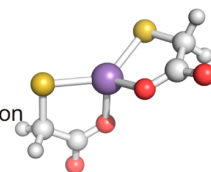
Supporting Information

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 (β). 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 inter-actions 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 β). 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.

The abstract graphic shows:
$$[M(H_2O)_6]^{2+} + \{n\ L\}^c \rightarrow [ML_n(H_2O)_m]^{c+2} + (6-m)\ H_2O$$
Experiment: Stability Constant ($\log \beta_{exp}$) (ΔG_{exp})
Theory: RI-BP86 + COSMO-RS solvation (ΔG_{theory}, $\log \beta_{exp}$)
$$| \Delta G_{theory} - \Delta G_{exp} | = ?$$

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 bio-chemical 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 (β) 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 ΔG 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 com-putational 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.

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 (ΔG) 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 ΔG ($-RT \ln \beta$) 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 $\{L_i\} \equiv L$ (introduced in Figure 1) is defined as

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

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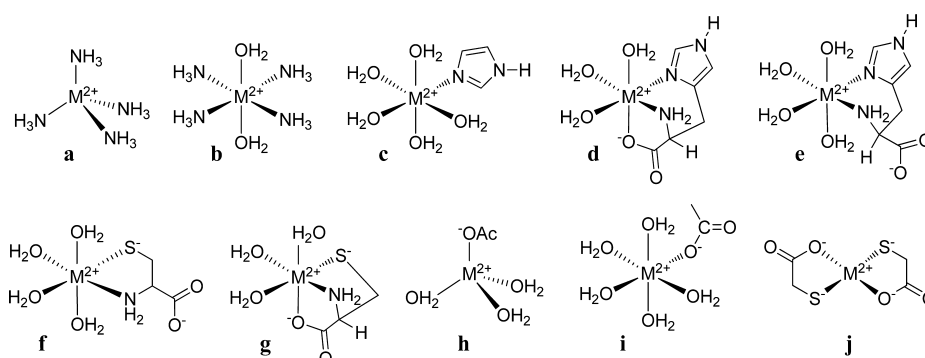
Figure 1. Model complexes: (a) $[M(NH_3)_4]^{2+}$; (b) $[M(NH_3)_4(H_2O)_2]^{2+}$; (c) $[M(Imi)(H_2O)_5]^{2+}$; (d) $[M(His)(H_2O)_3]^+$; (e) $[M(His)(H_2O)_4]^+$; (f) $[M(Cys)(H_2O)_4]$; (g) $[M(Cys)(H_2O)_3]$; (h) $[M(CH_3COO)(H_2O)_3]^+$; (i) $[M(CH_3COO)(H_2O)_5]^+$; (j) $[M(SCH_2COO)_2]^{2-}$.

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:

$$G_{M,L}^{calc} = E_{el} + G_{solv} \tag{1b}$$

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

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

$$\Delta G_{M,L}^{exp} = -RT \ln K_{M,L} \approx -RT \ln \beta_{M,L} = -RT \ln \frac{[ML_n]_\theta}{[M]_\theta [L]_\theta^{\,n}} \tag{2}$$

where K stands for the thermodynamical stability constant, β for the experimental (concentration) stability constant, and $[X]_\theta$ 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, $\Delta G_{M,L}^{calc}$ was calculated as the difference of the Gibbs free energies, $G_{M,L}^{calc}$, as defined in eq 1, of the products and reactants in the general complexation reaction:

$$[M(H_2O)_6]^{2+} + \{nL\}^c \rightarrow [ML_n(H_2O)_m]^{c+2} + (6-m)H_2O \tag{3}$$
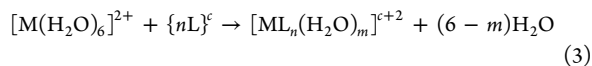
where c is the total formal charge of the ligands. The numbers of ligands, n, and H_2O molecules, m, are specified in the corresponding tables.

2.2. Model Systems. The set of model systems comprised six complexes ($[M(NH_3)_4]^{2+}$, $[M(Imi)]^{2+}$, $[M(His)]^+$, $[M(Cys)]$, $[M(CH_3COO)]^+$, and $[M(SCH_2COO)_2]^{2-}$). 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 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 $\log(\beta_L^{max}/\beta_L^{min}) = 1$–35, where β_L^{max} and β_L^{min} are the maximum (mostly in the Hg^{2+} complex) and minimum (mostly 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, $\Delta G_{M,L}^{exp}$, are derived from the experimental values of $\log \beta_{M,L}$ (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 Ni^{2+} and 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, $|\Delta G_L^{exp,max} - \Delta G_L^{exp,min}|_{(Mn^{2+}\leftrightarrow Cd^{2+})}$, varies significantly among the studied systems (from 1.1 kcal·mol^{-1} for the $[M(CH_3COO)]^+$ system to 15 kcal·mol^{-1} for $[M(NH_3)_4]^{2+}$). These magnitudes are even more pronounced if we include 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, $\Delta G_{M,L}^{calc}$, calculated according to eqs 1b and 3 using RI-BP86 for gas-phase electronic energies (the corresponding values of $\Delta G_{M,L}^{calc}$ obtained using RI-MP2 for the gas-phase electronic energies and $\Delta G_{M,L}^{calc,\mu}$ 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), $\Delta G_{M,L}^{exp}$, of the Studied Metal Ions As Calculated from Experimental Stability Constants[a]

complex	I^b	$\Delta G_{M,L}^{exp}$ [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, $\Delta G_{M,L}^{calc}$ (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$_{Imi}$) 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 $\Delta G_{M,L}^{calc}$ 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 $[M(H_2O)_3(Cys)]\cdot15H_2O$ 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 $\Delta G_{M,L}^{calc}$ (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 $G_{M,L}^{calc}$ (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
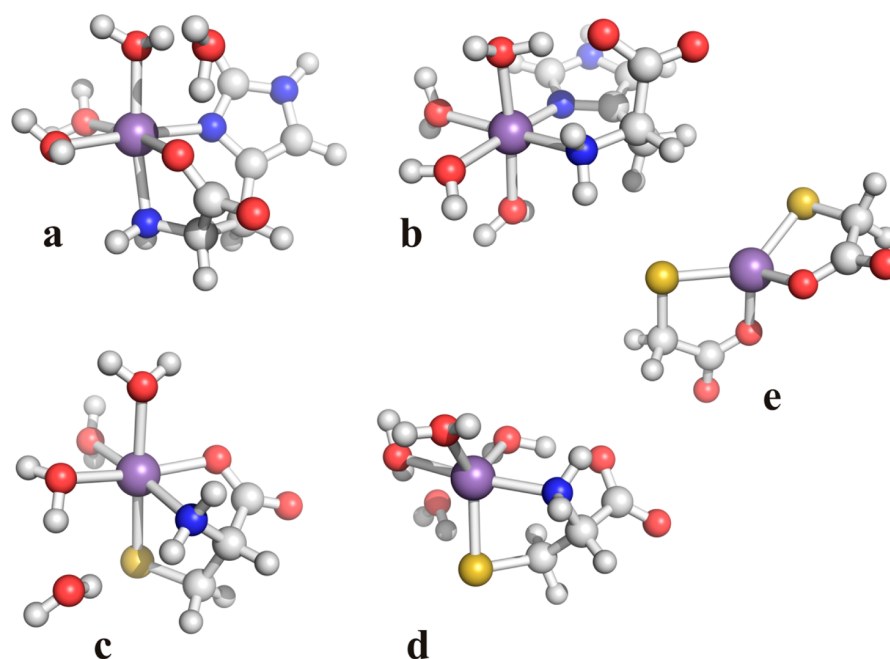
Figure 2. Representative equilibrium geometries for selected complexes studied in this work: (a) $[Mn(His)(H_2O)_4]^+$, (N,N,O) binding mode, one H_2O in the second coordination sphere; (b) $[Mn(His)(H_2O)_4]^+$, (N,N) binding mode; (c) $[Mn(Cys)(H_2O)_4]$, (N,S,O) binding mode, one H_2O in the second 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

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

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:

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

"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

10351

dx.doi.org/10.1021/ic401037x | *Inorg. Chem.* 2013, 52, 10347−10355

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

acid−conjugate base	ΔG^{exp}	ΔG^{calc}	$\Delta\Delta G^{exp/calc}$	complexation	LSS$_L$[a]
NH$_3$ → NH$_4^+$	−12.6	−17.3	4.7	[M(H$_2$O)$_6$]$^{2+}$ + 4NH$_3$ → [M(NH$_3$)$_4$(H$_2$O)$_2$]$^{2+}$ + 2H$_2$O	2.8
Imi → ImiH$^+$	−9.5	−12.6	3.1	[M(H$_2$O)$_6$]$^{2+}$ + Imi → [M(Imi)(H$_2$O)$_5$]$^{2+}$ + H$_2$O	0.3
His$^-$ → HisH$_2^{2+}$	−23.0	27.1	4.1	[M(H$_2$O)$_6$]$^{2+}$ + His$^-$ → [M(His)(H$_2$O)$_3$]$^{2+}$ + 3H$_2$O	3.0
Cys^{2-} → Cys$^+$	−27.8	−23.2	−4.6	[M(H$_2$O)$_6$]$^{2+}$ + Cys^{2-} → [M(Cys)(H$_2$O)$_3$] + 3H$_2$O	−8.0
CH$_3$COO$^-$ → CH$_3$COOH	−6.5	−4.6	−1.9	[M(H$_2$O)$_6$]$^{2+}$ + Ac$^-$ → [M(Ac)$_4$(H$_2$O)$_5$]$^+$ + H$_2$O	−8.6
SCH$_2$COO^{2-} → HSCH$_2$COOH	−19.3	−9.7	−9.6	[M(H$_2$O)$_6$]$^{2+}$ + 2tg^{2-} → [M(tg)$_2$]$^{2-}$ + 6H$_2$O[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$_2$COO)$^{2-}$.

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$_2$O 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 pK_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 pK_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 (pK_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:

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

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**:

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

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$_3$COO)]$^+$) 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·mol^{-1})^a

system	$MSS_{M,L}$								RMS_L	LSS_L
	Mn	Fe	Co	Ni	Cu	Zn	Cd	Hg		
$[M(NH_3)_4]^{2+}$	−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:

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

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

10353

dx.doi.org/10.1021/ic401037x | *Inorg. Chem.* 2013, 52, 10347−10355

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

complex	$\Delta G_{M,L}^{C}$								RMS_L^C	RMS_L^{orig}
	Mn	Fe	Co	Ni	Cu	Zn	Cd	Hg		
$[M(NH_3)_4]^{2+}$	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

■ 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.

Cyclam Derivatives with a Bis(phosphinate) or a Phosphinato–Phosphonate Pendant Arm: Ligands for Fast and Efficient Copper(II) Complexation for Nuclear Medical Applications

Tomáš David,[†] Vojtěch Kubíček,[†] Ondrej Gutten,[‡] Přemysl Lubal,[§,‖] Jan Kotek,[†] Hans-Jürgen Pietzsch,[⊥] Lubomír Rulíšek,[‡] and Petr Hermann*,[†]

[†]Department of Inorganic Chemistry, Faculty of Science, Charles University in Prague, Hlavova 2030, 12840 Prague, Czech Republic

[‡]Institute of Organic Chemistry and Biochemistry AS CR, Flemingovo náměstí 2, 16610 Prague, Czech Republic

[§]Department of Chemistry, Masaryk University, Kotlářská 2, 61137 Brno, Czech Republic

[‖]Central European Institute of Technology (CEITEC), Masaryk University, Kamenice 5, 62500 Brno, Czech Republic

[⊥]Institute of Radiopharmaceutical Cancer Research, Helmholtz-Zentrum Dresden-Rossendorf, Bautzner Landstrasse 400, 01328 Dresden, Germany

ⓢ *Supporting Information*

ABSTRACT: Cyclam derivatives bearing one geminal bis-(phosphinic acid), $-CH_2PO_2HCH_2PO_2H_2$ (H_2L^1), or phosphinic–phosphonic acid, $-CH_2PO_2HCH_2PO_3H_2$ (H_3L^2), pendant arm were synthesized and studied as potential copper(II) chelators for nuclear medical applications. The ligands showed good selectivity for copper(II) over zinc(II) and nickel(II) ions (log K_{CuL} = 25.8 and 27.7 for H_2L^1 and H_3L^2, respectively). Kinetic study revealed an unusual three-step complex formation mechanism. The initial equilibrium step leads to *out-of-cage* complexes with Cu^{2+} bound by the phosphorus-containing pendant arm. These species quickly rearrange to an *in-cage* complex with cyclam conformation **II**, which isomerizes to another *in-cage* complex with cyclam conformation **I**. The first *in-cage* complex is quantitatively formed in seconds (pH ≈5, 25 °C, Cu:L = 1:1, c_M ≈ 1 mM). At pH >12, **I** isomers undergo nitrogen atom inversion, leading to **III** isomers; the structure of the **III**-$[Cu(HL^2)]$ complex in the solid state was confirmed by X-ray diffraction analysis. In an alkaline solution, interconversion of the **I** and **III** isomers is mutual, leading to the same equilibrium isomeric mixture; such behavior has been observed here for the first time for copper(II) complexes of cyclam derivatives. Quantum-chemical calculations showed small energetic differences between the isomeric complexes of H_3L^2 compared with analogous data for isomeric complexes of cyclam derivatives with one or two methylphosphonic acid pendant arm(s). Acid-assisted dissociation proved the kinetic inertness of the complexes. Preliminary radiolabeling of H_2L^1 and H_3L^2 with ^{64}Cu was fast and efficient, even at room temperature, giving specific activities of around 70 GBq of ^{64}Cu per 1 μmol of the ligand (pH 6.2, 10 min, ca. 90 equiv of the ligand). These specific activities were much higher than those of $H_3\mathbf{nota}$ and $H_4\mathbf{dota}$ complexes prepared under identical conditions. The rare combination of simple ligand synthesis, very fast copper(II) complex formation, high thermodynamic stability, kinetic inertness, efficient radiolabeling, and expected low bone tissue affinity makes such ligands suitably predisposed to serve as chelators of copper radioisotopes in nuclear medicine.
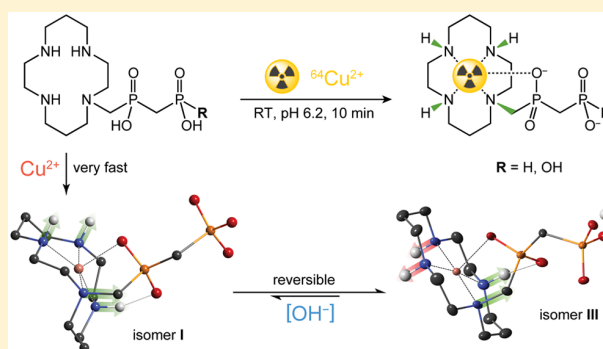
■ INTRODUCTION

Polyazamacrocycles with coordinating pendant arms can encapsulate a wide range of metal ions in their macrocyclic cavity, forming thermodynamically very stable complexes that often show high kinetic inertness.[1] Such properties are highly desirable for utilization in medicine and molecular biology. Therefore, these ligands and their complexes have long been investigated to optimize their properties for particular applications, for example, as magnetic resonance imaging

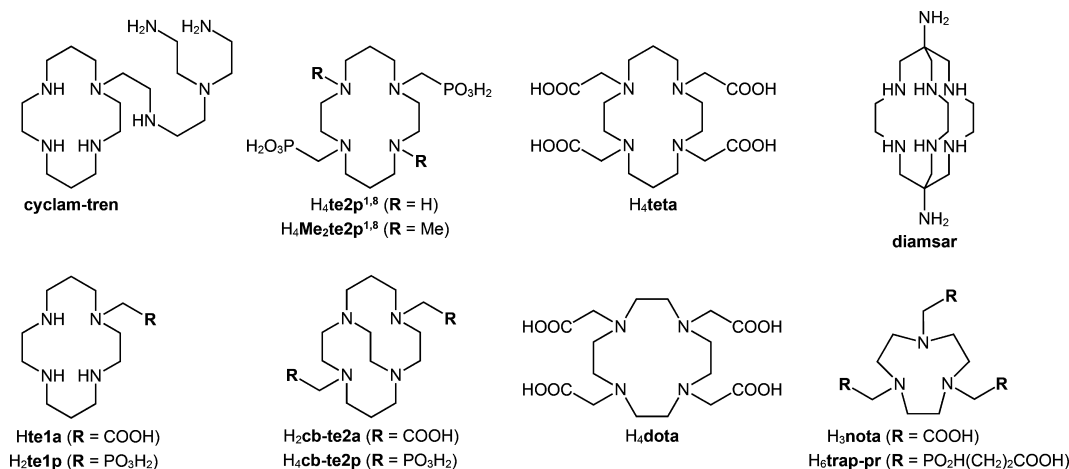contrast agents,[2,3] carriers of metal radionuclides for diagnosis and targeted therapy,[4] or luminescent probes.[2,5]

Among the methods used in modern medicine, positron-emission tomography (PET) is a powerful diagnostic tool. The method relies on the application of proton-rich isotopes that undergo β^+ decay. Collision of the emitted positron with an electron from the surrounding tissue results in a pair of

Chart 1. Structures of Various Macrocycles Discussed in the Text

cyclam-tren

H_4**te2p**[1,8] (R = H)
H_4**Me$_2$te2p**[1,8] (R = Me)

H_4**teta**

diamsar

H**te1a** (R = COOH)
H_2**te1p** (R = PO$_3$H$_2$)

H_2**cb-te2a** (R = COOH)
H_4**cb-te2p** (R = PO$_3$H$_2$)

H_4**dota**

H_3**nota** (R = COOH)
H_6**trap-pr** (R = PO$_2$H(CH$_2$)$_2$COOH)

collinear γ photons, which allow exact determination of the annihilation site(s) and, thus, show radioisotope distribution in the body. PET requires radioisotopes with suitable half-life, low energy of the emitted positrons, and good availability. Various nonmetallic (e.g., ^{18}F, ^{11}C, and ^{15}O) and metallic (e.g., ^{68}Ga, [44]Sc, and [89]Zr) radioisotopes have been utilized.[1,4,6−9] Among metallic positron emitters, [64]Cu (61% β^+) is of interest because of its long half-life ($\tau_{1/2}$ = 12.8 h) and low positron energy (E_{av} = 0.65 MeV), resulting in high image resolution. Besides, other copper radioisotopes are also interesting positron-emitting [60]Cu ($\tau_{1/2}$ = 23.7 min; 100% β^+) and [61]Cu ($\tau_{1/2}$ = 3.3 h; 100% β^+) and the β^- emitter [67]Cu ($\tau_{1/2}$ = 61.8 h; 100% β^-), which may be used for internal radionuclide therapy.

Metal radioisotopes cannot usually be applied in the form of the free ion because of nonspecific deposition in tissues. To achieve the desired biodistribution, the metal ion must be bound in a thermodynamically stable and kinetically inert complex, and the complex is commonly conjugated to a targeting vector. Over the years, several classes of chelators have been suggested for the complexation of copper radioisotopes.[1,8,10,11] Initially, the chelators were based on well-known acyclic or macrocyclic ligands, such as H$_5$**dtpa**, cyclam, H$_4$**teta**, and H$_4$**dota**; H$_4$**dota** monoamides (Chart 1) are the most commonly used chelators for radioactive copper. However, the properties of complexes of these early ligands are not optimal for copper(II), mainly from the viewpoints of the rate of complexation (too slow for most macrocyclic ligands), kinetic inertness (too low for complexes of acyclic ligands or those with too many donor atoms), and redox vulnerability toward monovalent copper [complexes of copper(I) are easily decomposed]. Thus, other ligands have been investigated. Derivatives of cross-bridged cyclam are the largest family of such ligands (Chart 1). In these ligands, the central ion is encapsulated in a well-preorganized ligand cage, which leads to kinetically very inert complexes, but this is coupled, except in several very recent examples involving phosphonate pendant arms,[12] with very slow copper(II) complexation. Other successful chelators are based on the hexaazasarcophagine skeleton (e.g., diamsar in Chart 1).[13] More recently, H$_3$**nota** (Chart 1) derivatives have become very popular because their complexes are rapidly formed under mild conditions and they are stable in vivo.[14] However, among these ligands, the cyclam derivatives offer the best selectivity for

copper(II) over zinc(II) and nickel(II), which are common metallic impurities in no-carrier-added (NCA) radiocopper solutions. Thus, new cyclam derivatives with one or two coordinating pendant arms [to fulfill the copper(II) requirement for coordination numbers of 5 or 6] have recently been suggested for radiocopper complexation.[15,16]

Owing to the rather short half-lives of metallic PET radioisotopes, the major limitation of macrocyclic chelators is their slow complexation, and ligands with improved labeling efficiency are highly desired. Some time ago, we showed that cyclam derivatives with methylphosphonic acid pendant arms, H$_4$**te2p**[1,8] or H$_2$**te1p** (Chart 1), show fast copper(II) complexation.[17−19] On the basis of the generally accepted two-step complex formation with pendant-armed macrocyclic ligands, this fast complexation could be explained in terms of the interaction between the metal ion and pendant phosphonate group(s) outside of the macrocycle cage. Thus, these groups assist in the transfer of the metal ion from the bulk solution into the ligand cavity. In addition, these cyclam-based ligands are very selective for copper(II) ion, and the resulting complexes are thermodynamically stable and kinetically inert.[17−22] A pendant arm moiety with even better complexation ability can further improve the complexation rates. Recently, we confirmed this assumption for **trap** derivatives (e.g., H$_6$**trap-pr** in Chart 1), a family of H$_3$**nota** phosphinic acid derivatives, which exhibit accelerated ^{68}Ga labeling because of a stronger *out-of-cage* metal ion−ligand interaction between gallium(III) and their pendant arms.[23] Moreover, some conjugates based on **trap** derivatives have recently entered into clinical utilization for ^{68}Ga-PET imaging.[24]

On the basis of the collected data, we propose herein a new type of pendant arm to improve the radiometal labeling efficiency, that is, geminal phosphorus acid groups, specifically methylenebis(phosphinic acid) or methylene(phosphinic−phosphonic acid) moieties. These groups are very acidic and, therefore, they are only weakly metal-binding chelating groups.[25] At the same time, this property implies that they should be able to bind metal ions even in acidic solutions, yet they should not compete with much better ligands, such as macrocycles. Such a chelating pendant arm can capture very low concentrations of metal radioisotopes in solution and, if attached to a macrocycle, can assist in acceleration of the overall metal complexation because of the increased stability of an *out-of-cage* complex intermediate. Fast and efficient complexation

under mild conditions is a very desirable property for ligands intended to be used for the binding of short-lived metal radioisotopes. Cyclam derivatives H_2L^1 and H_3L^2 bearing these pendant arms (Chart 2) were envisaged as being suitable for

Chart 2. Structures of the Title Ligands, H_2L^1 and H_3L^2, Showing the Ring Nitrogen Atom Numbering
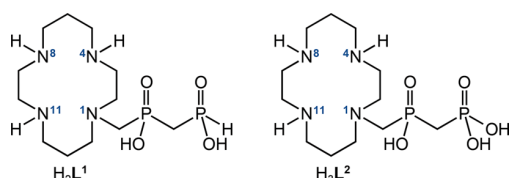
divalent copper. In this study, we present a comprehensive investigation of these complexes, which has included the evaluation of an unusual isomerism (for a definition of isomers of metal ion−cyclam complexes, see Figure S1) and their thermodynamic and kinetic properties, coupled with theoretical calculations. The assumption of efficient binding of a copper radioisotope was confirmed in preliminary radiochemical experiments.

■ RESULTS

Synthesis of the Ligands and Their Complexes. A direct Mannich-type reaction (Scheme 1) of cyclam with $CH_2(PO_2H_2)_2$ afforded ligand H_2L^1 in 70−75% conversion (based on formaldehyde, monitored by ^{31}P NMR). Using excesses of cyclam and methylenebis(phosphinic acid), with formaldehyde as the limiting component, the reaction was very clean because no formation of multiply substituted cyclic byproducts, hydroxomethylphosphinic acid, or N-methylated byproducts was observed. Simple purification on ion exchangers enabled the recovery of cyclam as well as methylenebis(phosphinic acid). Mild oxidation of the P−H bond in H_2L^1 with aqueous $HgCl_2$ under acidic conditions with subsequent removal of mercury(II) ions with H_2S led to H_3L^2.[26]

The Cu^{II}-H_2L^1 and Cu^{II}-H_3L^2 complexes were isolated in two isomeric forms. Blue complexes having cyclam ring conformation **I** were prepared from the zwitterionic forms of the ligands at room temperature. They were converted to violet isomers with cyclam ring conformation **III** in an alkaline solution. The cyclam ring conformations **I** and **III** are shown in Figure S1, and this notation is used hereinafter for assignment of these isomers. The isomerism of the complexes is discussed in detail later. The isomerization reaction always led to an equilibrium mixture of the blue and violet complexes (approximate equilibrium **I/III** isomer molar ratios were 20:80 and 30:70 for the Cu^{II}-H_2L^1 and -H_3L^2 complexes, respectively). Because their mutual interconversion is very slow

at lower pH, their mixtures were separated by chromatography on SiO_2. Fractions were checked by thin-layer chromatography (TLC), the most sensitive technique to distinguish the isomers (Figure S2). Single crystals of the violet complex of composition $[Cu(HL^2)]·5H_2O$ suitable for X-ray diffraction study were isolated by slow evaporation of the volatiles from an aqueous solution (for experimental details, see the Supporting Information, SI). Electronic spectral data of the isomeric complexes are summarized in the SI (Table S1 and Figure S3).

Thermodynamic Studies. The acid−base properties of the ligands and thermodynamic stabilities of their complexes were studied by potentiometry (for details, see the SI). Protonation constants of the title ligands and their comparison with those of similar chelators are given in Tables S2 and S3, and the corresponding distribution diagrams are shown in Figure S4. To help with the interpretation of complex formation kinetic data, $^{13}C\{^1H\}$ and $^{31}P\{^1H\}$ NMR titrations (Figure S5) were carried out with H_3L^2 to determine the sites of consecutive protonation. Because it is known that the bis(phosphonate) group can strongly complex Na^+ or K^+ ions,[27] CsOH was used as a base to minimize such interactions.

Complexation of copper(II) and zinc(II) ions was fast, even in acidic solutions, and the stability constants of the complexes were determined by direct titration (an in-cell method). For nickel(II) systems, complexation was too slow for direct titration and, thus, an out-of-cell equilibrium method had to be used. The results are presented in Tables 1 and S2, and the corresponding distribution diagrams are shown in Figures 1 and S6 and S7. About 20% of free copper(II) ion was present at the beginning of titrations at pH 1.5, and it was fully complexed below pH 2.5.

Formation Kinetics of Copper(II) Complexes. The rate of complexation is one of the most important parameters to be evaluated if ligands are intended to be utilized for radiometal binding. Thus, complexation kinetics was investigated under pseudo-first-order conditions ($c_L = 0.1$ mM; $c_M = 1.0−5.0$ mM) in the range of pH 2−7 ($t = 25 \pm 0.1$ °C; $I = 0.1$ M KCl). Stopped-flow measurements showed an unexpected mechanism because the overall process had to be divided into a preequilibrium step and two kinetically distinct reactions (Figures 2 and 3). Biexponential fitting (yielding rate constants for two kinetic steps, $^fk_{obs}$ and $^{isl}k_{obs}$; see below) had to be used because the monoexponential treatment led to unsatisfactory results (Figures S8 and S9). This observation could be explained in terms of a three-step complexation mechanism (Scheme 2). Initially, copper(II) interacts with H_2L^1 or H_3L^2 in a very fast preequilibrium step to form intermediate $1^\#$; this rearranges into intermediate $2^\#$, which then isomerizes to the final blue **I** complex.

To fit the experimental data and to extract kinetic parameters for the first formation step, rate laws were derived, taking into account the above general mechanism. A full description of the

Scheme 1. Synthesis of H_2L^1 and H_3L^{2a}
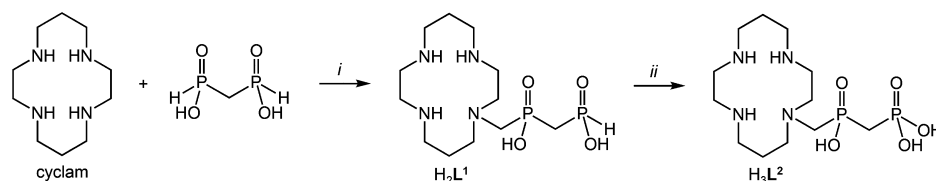
[a](i) paraformaldehyde, concentrated aqueous HCl, 80 °C; (ii) (1) $HgCl_2$, H_2O; (2) H_2S(g).

Table 1. Stability Constants (log K_{ML}) or Stepwise Protonation Constants (log K_n) of Complexes of the Title Ligands [25 °C, I = 0.1 M (Me$_4$N)Cl]

	copper(II)		zinc(II)		nickel(II)	
equilibrium[a]	H$_2$L^1	H$_3$L^2	H$_2$L^1	H$_3$L^2	H$_2$L^1	H$_3$L^2
M + L ⇌ [M(L)]	25.83	27.66	18.12	19.84	21.94	24.01
[M(L)] + H ⇌ [M(HL)]		6.97	3.74	7.17	2.04	6.61
[M(HL)] + H ⇌ [M(H$_2$L)]		1.66		3.68		
[M(L)(OH)] + H ⇌ [M(L)] + H$_2$O	12.26	12.63				

[a]Charges are omitted for simplicity.
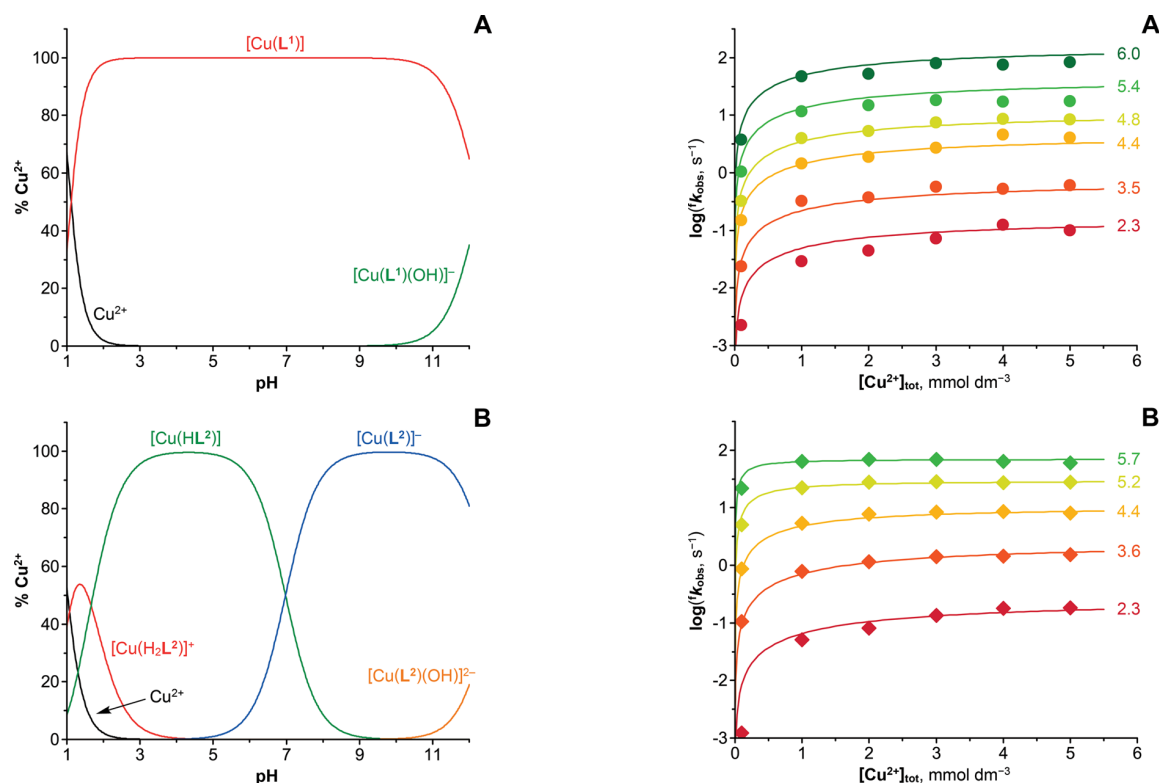
Figure 1. Distribution diagrams of the Cu^{II}-H$_2$L^1 (A) and Cu^{II}-H$_3$L^2 (B) systems (c_L = c_{Cu} = 4 mM).

Figure 2. Formation of the intermediate 2$^{\#}$: dependence of the pseudo-first-order formation kinetic constants $^f k_{obs}$ of H$_2$L^1 (A) and H$_3$L^2 (B) on the concentration of copper(II) ions measured at various pH values (colored numbers on the right). The solid lines represent fits according to eq 1 (H$_2$L^1) or eq 2 (H$_3$L^2). The experimental points obtained for [Cu^{2+}] = 0.1 mM are shown to illustrate the saturation nature of the curves; however, they were not included in the fitting because these experiments do not meet the pseudo-first-order conditions.

procedure is given in the SI. For the 1$^{\#}$-(Cu-H$_3$L^2) species, a deprotonation equilibrium involving a proton bound to the distant phosphonate group (characterized by dissociation constant $^{\#}K_{aP}$) is operative in the investigated pH range.[27] Observed reaction rates for the first step were accelerated even in the pH range in which diprotonated or triprotonated (for H$_2$L^1 and H$_3$L^2, respectively) ligand species are exclusively present (Figure S4). Therefore, less protonated intermediate 1$^{\#}$ species, 1$^{\#}$-(Cu-HL) and 1$^{\#}$-(Cu-H$_2$L), had to be considered, which play an important role in the reaction step because log($^f k_{obs}$) increases linearly with the pH (Figure S10). This is in accordance with kinetic/mechanistic investigations of a number of other metal ion–macrocyclic ligand systems, in which these species proved to be kinetically the most important.[28] Thus, for the step leading to intermediate 2$^{\#}$, eqs 1 and 2 (see the SI) were used for the final treatment of the kinetic data.

$$^f k_{obs} = \frac{^{\#}K_{MH_2L}[Cu^{2+}]_{tot}(^f k_1 + {}^f k_2[H^+]/K_{a2})}{(1 + {}^{\#}K_{MH_2L}[Cu^{2+}]_{tot})[H^+]/K_{a2}} \quad (1)$$

$$^f k_{obs} = \frac{^{\#}K_{MH_2L}[Cu^{2+}]_{tot}(^f k_1 K_{a2}/[H^+] + {}^f k_2 + {}^f k_3[H^+]/{}^{\#}K_{aP})}{1 + [H^+]/K_{a3} + {}^{\#}K_{MH_2L}[Cu^{2+}]_{tot}(1 + [H^+]/{}^{\#}K_{aP})} \quad (2)$$

Here, $^f k_{obs}$ is the pseudo-first-order rate constant, $^f k_1$, $^f k_2$ and $^f k_3$ correspond to the reactivities of the differently protonated intermediate 1$^{\#}$ species, K_{a2} and K_{a3} are the dissociation constants of the H$_2$L and H$_3$L ligand species (Table S3), $^{\#}K_{MH_2L}$ is the equilibrium constant for the formation of the 1$^{\#}$-(Cu-H$_2$L) species from copper(II) and H$_2$L ligand species, and $^{\#}K_{aP}$ is the dissociation constant of the intermediate 1$^{\#}$-(Cu-H$_3$L^2) species.
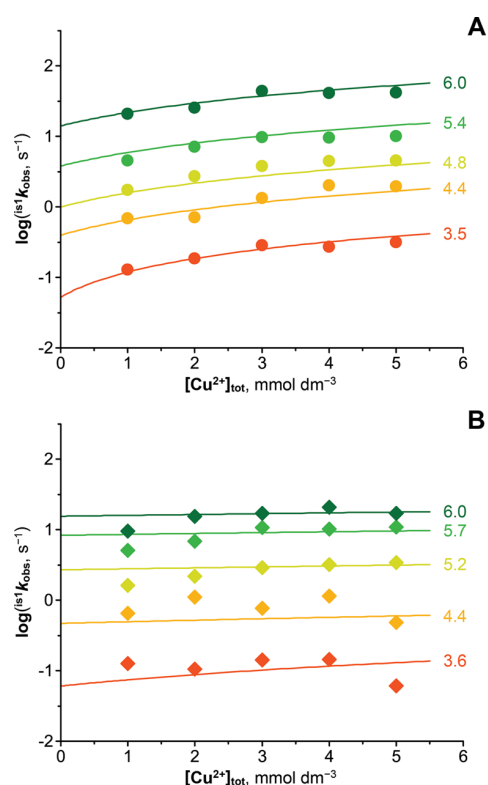
Figure 3. Formation of the I isomer from the intermediate $2^{\#}$: dependence of the pseudo-first-order formation kinetic constants $^{is1}k_{obs}$ of H_2L^1 (A) and H_3L^2 (B) on the concentration of copper(II) ions measured at various pH values (colored numbers on the right). The solid lines represent fits according to eq 3.

The second kinetically important step can be viewed as an isomerization process. The isomerization was slower than the formation of intermediate $2^{\#}$ and should involve a change of the chirality at a coordinated nitrogen atom(s). Intuitively, such isomerization is expected to proceed without a change of the

protonation state of the species and, thus, should not depend on the pH. However, the reaction became faster with increasing pH (Figures 3 and S11). Because the same dependence was observed for complexes of both ligands, it should not be connected with possible deprotonation of the distant phosphonate group in the intermediate $1^{\#}$-$[Cu(HL^2)]$ complex (Scheme 2). Instead, it is more likely to be a catalytic effect of hydroxide ions, and such a mechanism has been suggested for complexation of a number of macrocyclic ligands.[29] Furthermore, the rate of this step also increases with increasing copper(II) concentration. This might be rationalized in terms of the formation of dinuclear species in which one metal ion is coordinated in the macrocyclic cavity, whereas the other is coordinated by the pendant bis(phosphinate) or phosphonate/phosphinate group. The temporary formation of such a weak copper(II) complex would then accelerate the nitrogen atom inversion. Taking into account all of the pathways, eq 3 was derived (see the SI) and used to fit the experimental data describing the final isomerization step

$$^{is1}k_{obs} = {}^{is1}k_0 + {}^{is1}k_{OH}[OH^-] + {}^{is1}k_{Cu}[Cu^{2+}]_{tot}$$
$$+ {}^{is1}k_{Cu,OH}[OH^-][Cu^{2+}]_{tot} \tag{3}$$

where $^{is1}k_0$ is the rate constant describing the unassisted isomerization, $^{is1}k_{OH}$ is the rate constant describing the hydroxide-assisted reaction, $^{is1}k_{Cu}$ is the rate constant describing the metal-assisted isomerization, and $^{is1}k_{Cu,OH}$ is the rate constant describing the simultaneous (metal and hydroxide ions) assistance. However, from the fitting results, it was found that the spontaneous isomerization plays a negligible role ($^{is1}k_0 \approx 0$). The parameters obtained by fitting of the $^{f}k_{obs}$ and $^{is1}k_{obs}$ values utilizing eqs 1−3 are presented in Table 2.

Isomerization Kinetics in an Alkaline Solution. In alkaline solutions (pH >12), even at room temperature, the blue **I** isomers were converted to the violet **III** isomers. However, this conversion always led to an equilibrium mixture of **I** and **III** species. Spectral changes with time (Figure S12) showed that the same equilibrium mixture was reached irrespective of whether the **I** or **III** isomer was used as the

Scheme 2. Formation of the I Isomers of $[Cu(L^1)]$ (A) and $[Cu(L^2)]^-$ (B) Complexes[a]
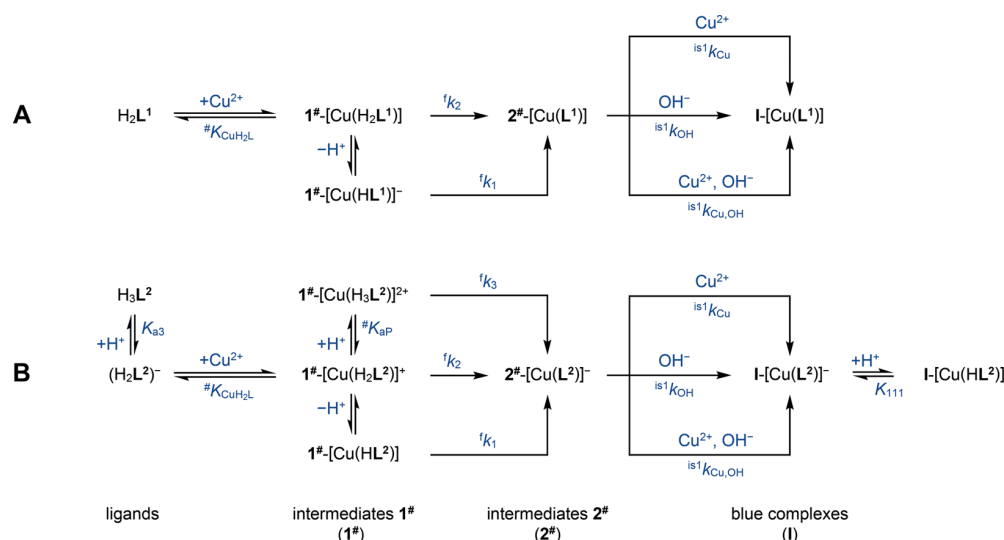
[a]Additional protonation equilibria in part B (constants K_{a3}, $^{\#}K_{aP}$, and K_{111}) involve a proton located on the distant phosphonate group.

Table 2. Kinetic Parameters Describing the Stepwise Formation of the I Complexes

	formation of intermediate 2#			formation of the I complexes		
constant	H_2L^1	H_3L^2	constant	H_2L^1	H_3L^2
$^fk_1$ [s^{-1} mol^{-1} dm^3]	$1.0(1) \times 10^7$	$5.7(3) \times 10^7$	$^{is1}k_{OH}$ [s^{-1} mol^{-1} dm^3]	$1.5(1) \times 10^9$	$1.7(1) \times 10^9$
$^fk_2$ [s^{-1} mol^{-1} dm^3]	0.13(2)	14(3)	$^{is1}k_{Cu}$ [s^{-1} mol^{-1} dm^3]	38(7)	12(5)
$^fk_3$ [s^{-1} mol^{-1} dm^3]		0.11(2)	$^{is1}k_{Cu,OH}$ [s^{-1} mol^{-2} dm^6]	$8.3(9) \times 10^{11}$	$4.9(1) \times 10^{10}$
$^\#K_{CuH_2L}$ [mol^{-1} dm^3]	$4(1) \times 10^2$	$2.4(6) \times 10^6$			
log($^\#K_{CuH_2L}$)	2.6	6.4			
$^\#K_{aP}$ [mol dm^{-3}]		$6(2) \times 10^{-5}$			
$p(^\#K_{aP})$		4.2			

starting material. The isomerization was investigated in detail in the [OH$^-$] range of 10−100 mM at constant ionic strength $I = 0.5$ M K(OH,Cl) and temperature ($t = 25$ °C). In the studied pH range, the equilibrium **I/III** isomer molar ratios proved to be independent of the hydroxide concentration and were approximately 0.2:0.8 and 0.3:0.7 for the [Cu(**L**^{1})] and [Cu(**L**^{2})]$^-$ complexes, respectively. The measurements were performed in both directions, that is, starting from the pure **I** isomer or the pure **III** isomer. The isomerization processes were characterized by rate constants $^{is2}k_{obs1}$ and $^{is2}k_{obs-1}$ for the **I → III** process and the reverse reaction, respectively. The abundance of one component in the system could be expressed as shown in eq 4, where c^0 is the starting concentration of a complex.[30]

$$\frac{c}{c^0} = \left[\frac{1}{^{is2}k_{obs1} + {}^{is2}k_{obs-1}}\right]\left[{}^{is2}k_{obs-1} + {}^{is2}k_{obs1}e^{-({}^{is2}k_{obs1} + {}^{is2}k_{obs-1})t}\right] \tag{4}$$

Equation 4 can be rewritten as the general exponential expression $c = a + be^{(-kt)}$, where the constant k is expressed as $k = {}^{is2}k_{obs1} + {}^{is2}k_{obs-1}$. Fitting the experimental data to the general expression yielded a constant k, which was used to estimate the rate constants $^{is2}k_{obs1}$ and $^{is2}k_{obs-1}$. Hence, constants describing the kinetics of the forward and reverse reactions leading to the same equilibrium mixture are involved in the expression and, therefore, both constants could be determined from a single experiment. The two sets of experiments (i.e., reactions started from **I** or **III** isomers) gave fully consistent results (Figure 4), which proved the validity of the chosen model.

The $^{is2}k_{obs}$ values for complexes of both ligands showed nonlinear dependences on the concentration of hydroxide ions. On the basis of the potentiometric results, the studied complexes formed monohydroxido species (Figure 1) in the given pH range and, therefore, the corresponding $^{OH}K_{I,1}$ constant could be estimated from the potentiometric data. However, analysis of the kinetic data showed that isomerization was also assisted by a higher number of hydroxide anions. The proposed isomerization mechanism is depicted in Scheme 3, and the process can be described by eq 5:

$$^{is2}k_{obs1} = {}^{is2}k_1[\text{CuL(OH)}] + {}^{is2}k_2[\text{CuL(OH)}][\text{OH}^-] + {}^{is2}k_3[\text{CuL(OH)}][\text{OH}^-]^2 \tag{5}$$

where $^{is2}k_1 = k_1$, $^{is2}k_2 = k_2{}^{OH}K_{I,2}$, and $^{is2}k_3 = k_3{}^{OH}K_{I,2}{}^{OH}K_{I,3}$ for the **I → III** process (Scheme 3) and an analogous equation could be applied to the reverse process (for $^{is2}k_{obs-1}$). The results are compiled in Table 3. Fitting of the data obtained for the [Cu(**L**^{2})]$^-$ complex gave only constants $^{is2}k_2$ and $^{is2}k_3$. For the [Cu(**L**^{1})] complex, all three constants were found.
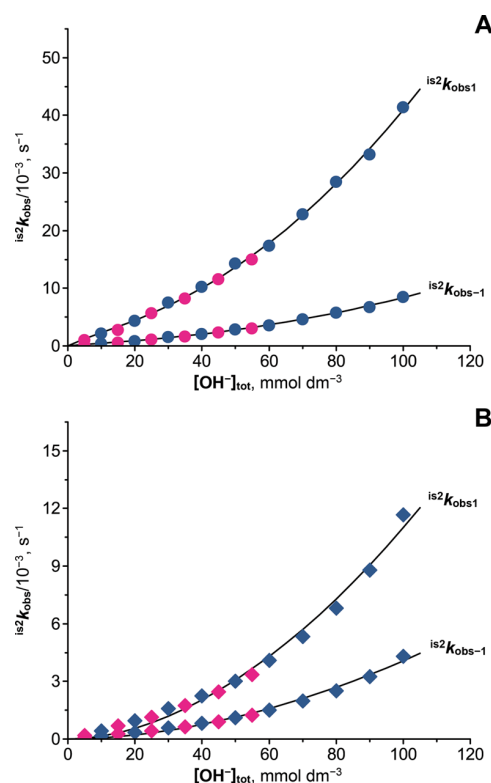
Figure 4. Mutual isomerization of [Cu(**L**^{1})] (A) and [Cu(**L**^{2})]$^-$ (B) complexes: reaction started from the **I** isomers (blue symbols) or the **III** isomers (violet symbols). The solid lines represent fits according to eq 5.

Scheme 3. Mechanisms of the I → III and III → I Isomerization Processes[a]
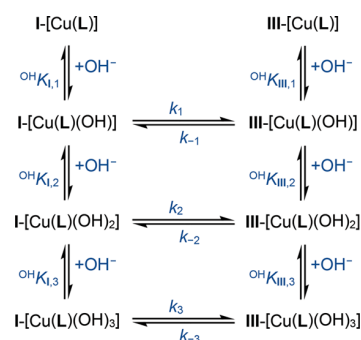
[a]The charges of the complex species are omitted for clarity.

Table 3. Rate Constants of the I → III and III → I Isomerization Processes

constant	$[Cu(L^1)]$	$[Cu(L^2)]^-$
I → III		
$^{is2}k_1$ [s^{-1} mol^{-2} dm^3]	0.98 ± 0.18	
$^{is2}k_2$ [s^{-1} mol^{-2} dm^6]	22 ± 7	14 ± 2
$^{is2}k_3$ [s^{-1} mol^{-3} dm^9]	645 ± 56	172 ± 22
III → I		
$^{is2}k_{-1}$ [s^{-1} mol^{-1} dm^3]	0.20 ± 0.04	
$^{is2}k_{-2}$ [s^{-1} mol^{-2} dm^6]	4.6 ± 1.3	5.3 ± 0.6
$^{is2}k_{-3}$ [s^{-1} mol^{-3} dm^9]	132 ± 11	63 ± 8

However, isomerization of the $[Cu(L^1)(OH)]^-$ species was very slow, and its contribution to the overall isomerization rate was negligible.

Dissociation Kinetics. The kinetics of acid-assisted dissociation of complexes I-$[Cu(L^1)]$ and I-$[Cu(L^2)]^-$ was investigated in 0.1–5.0 M $HClO_4$ at I = 5 M $(H,Na)ClO_4$. However, in the examined pH range, complex I-$[Cu(L^1)]$ was found (TLC analysis) to undergo oxidation, leading to a progressively increasing abundance of I-$[Cu(L^2)]^-$ in the course of the reaction, which also underwent dissociation. Neither the oxidation process nor the extent of decomplexation of the I-$[Cu(L^1)]$ complex could be readily quantified because of the very similar UV/vis spectra of the relevant species. However, the overall rate of I-$[Cu(L^1)]$ complex dissociation (including some contributions from the oxidation−dissociation pathway) was very similar to that found for the pure I-$[Cu(L^2)]^-$ complex (Figure S13). This indicates a similar decomplexation resistance and mechanism for both complexes, suggesting no significant alteration of the overall kinetic inertness due to the presence of distant P−H or P−OH bonds.

Dissociation of the I-$[Cu(L^2)]^-$ complex was studied over the temperature range of 60–90 °C.[17−19,31] The saturation shape of the curves (Figure 5) indicates that the complex was
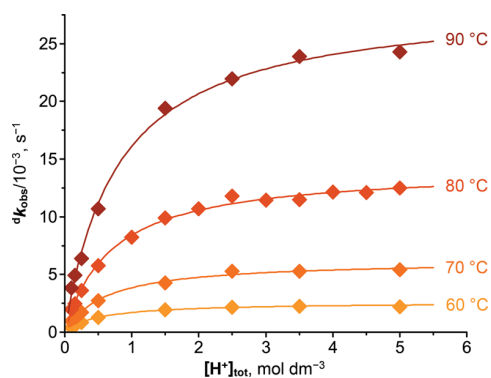
Figure 5. Acid-assisted dissociation of the I-$[Cu(L^2)]^-$ complex [I = 5 M $(H,Na)ClO_4$].

Scheme 4. Acid-Assisted Decomplexation of the I-$[Cu(L^2)]^-$ Complex (Here, Probably $n > 1$)
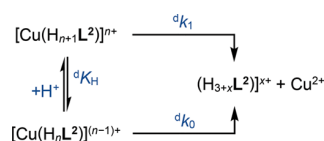
present in two forms differing in the number of bound protons (Scheme 4). However, the number of protons in the studied pH range is difficult to determine. Thus, the data were treated according to eq 6:

$$^dk_{obs} = (^dk_0 + {}^dk_1{}^dK_H[H^+])/(1 + {}^dK_H[H^+]) \quad (6)$$

where $^dk_0$ is a constant corresponding to dissociation of the less protonated complex, $^dk_1$ is a constant corresponding to dissociation of the more protonated complex, and ^dK_H is the corresponding protonation constant. Fitting of the results yielded negligible values of $^dk_0$, indicating that dissociation of the more protonated species was the dominant process. The obtained values of $^dk_1$ and ^dK_H are summarized in Table S5, and analysis of the temperature-dependent data provided activation parameters (Table 4).

To directly compare the kinetic inertness of both the I- and III-$[Cu(L^2)]^-$ complexes with those of related complexes, the dissociation was studied at 90 °C in 5.0 M aqueous $HClO_4$ or 5.0 M aqueous HCl because it is known that HCl significantly accelerates the decomplexation process.[17,32] Decomplexation half-lives of around 29 and 7 s for I-$[Cu(L^2)]^-$ and of around 63 and 4 min for III-$[Cu(L^2)]^-$ were obtained in $HClO_4$ and HCl, respectively. Values for the III-$[Cu(L^2)]^-$ isomer represented a lower limit because dissociation of the III isomer proceeded simultaneously with isomerization to isomer I, which subsequently dissociated at a different (faster) rate. The isomerization was confirmed by a shift in the absorption band maximum and by TLC.

Quantum-Chemical Calculations. Theoretical calculations were performed on copper(II) complexes of H_3L^2 and related cyclam methylenephosphonic acid derivatives, H_2**te1p** and H_4**te2p**[1,8] (Chart 1). Because the III → I isomerization was observed to proceed mainly in alkaline solutions, the calculations were carried out on the fully deprotonated forms of the complexes. Several conformations were examined for each of these isomers, including both penta- and, in a few cases, hexacoordinated structures. In each system, up to eight isomers are possible, as shown in Chart 3. No metal-bound water molecules were explicitly included in the models because binding of a water molecule at an axial position of copper(II) complexes is generally weak and, in some isomers, even sterically inaccessible. The isomers themselves differ in the chirality at the coordinated nitrogen atoms, and because of the asymmetry of structures formally belonging to the II isomer, there are four possible isomers differing in the orientation of an N-substituent with regard to the substituent orientation at the other three amine groups.

For calculation of the gas-phase energies, pure generalized gradient approximation (GGA) and meta-GGA functionals (BP86 and TPSS) were preferred over methods including Hartree−Fock exchange based on the values of D1(MP2) diagnostics,[33] which indicated potential multiconfigurational character of the wave function (data not shown). The solvent (water) was taken into account using one of the best currently available solvation models, the COSMO-RS method. Of the two functionals, the TPSS values are presented owing to a better performance for metal-ion complexes,[34] although the BP86 results were qualitatively identical, lying within 1−2 kcal mol^{-1} of those by the former method; the results are compiled in Tables 5 and S6−S8 and full sets of isomers for the H_2**te1p** and H_4**te2p**[1,8] complexes are shown in Figures S14 and S15, respectively. Calculated structures of the discussed isomers of complexes of all three ligands are shown in Figures S16−S18.

Table 4. Activation Parameters for Acid-Assisted Dissociation of the I-[Cu(L^2)] and Related Complexes

parameter	I-[Cu(L^2)]$^-$	I-[Cu(te2p^{1,8})]$^{2-}$ [17]	I-[Cu(Me$_2$te2p^{1,8})]$^{2-}$ [18]	I-[Cu(te2p^{1,8ABn})]^{32}
E_A [kJ mol^{-1}]^a	81 ± 1	72.0, 85^d	60	78
$\Delta H^{\#}$ [kJ mol^{-1}]^b	78 ± 1	69.5, 82^d	57	75
$\Delta S^{\#}$ [J K^{-1} mol^{-1}]^b	−61 ± 2	−71, −52^d	−95	−52
ΔH [kJ mol^{-1}]^c	−13.6 ± 0.7	−8.3	−17	−8
ΔS [J K^{-1} mol^{-1}]^c	−36 ± 2	−22.7	−30	−39

^{a}Arrhenius model: $\ln(^dk) = -(E_A/RT) + \ln A$. ^{b}Eyring model: $\ln(^dk/T) = -(\Delta H^{\#}/RT) + \Delta S^{\#}/R + \ln(k_B/h)$. $^c\ln K = -(\Delta H/RT) + \Delta S/R$. ^{d}Data for two dissociation pathways.

Chart 3. Bosnich's Nomenclature for Isomers of Cu^{II}-H$_2$L^1 (R = H) and Cu^{II}-H$_2$L^2 (R = OH or O$^-$) Complexes^a

isomer I

isomer ^{1}II

isomer ^{4}II

isomer ^{8}II

isomer ^{11}II

isomer III
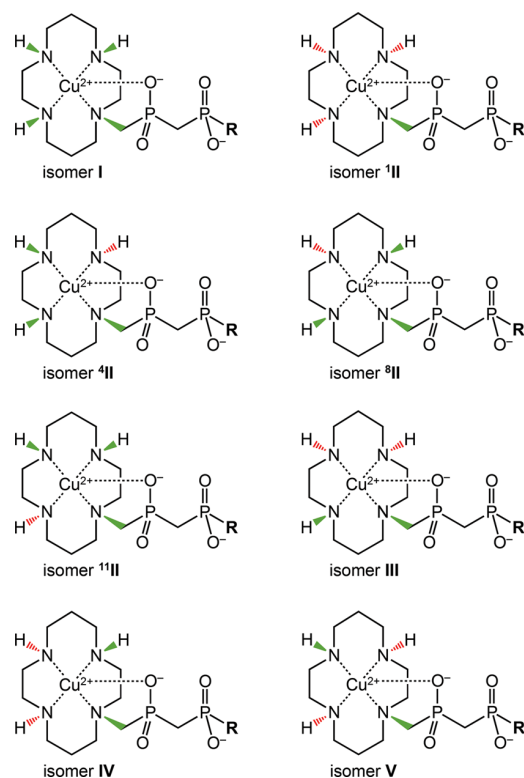
isomer IV

isomer V

^{a}Nitrogen atom substituents point either above (green) or below (red) the macrocyclic plane. For the cyclam II conformation, more isomers are possible, and they are distinguished by an upper index specifying the nitrogen atom at which the substituent is directed to the opposite side to those at the other three.

Table 5. Relative Free Energies (in kcal mol^{-1}; RI-TPSS/ def2-TZVP + COSMO-RS) for Isomers of Copper(II) Complexes of H$_3$L^2, H$_2$te1p, and H$_4$te2p^{1,8}

isomer	[Cu(L^2)]$^-$	[Cu(te1p)]	[Cu(te2p^{1,8})]$^{2-}$
I	1.4	2.1	5.4
^{1}II	11.8	10.1	8.0
^{4}II	2.8	5.3	11.2
^{8}II	8.4	7.3	11.9
^{11}II	5.4	6.1	9.8
III	0.0	0.0	0.0
IV	10.4	9.6	10.5
V	13.0	11.4	15.4

parameters are compared with the calculated values in Table S9. For complexes of all three ligands, the III isomers were identified as the most stable (i.e., having the lowest free energy), followed by the I isomers. The smallest difference between these two isomers was found for the complex of H$_3$L^2, and the largest difference was found for the complex of H$_4$te2p^{1,8}, for which no III → I isomerization has been observed in solution. These are followed by several species with cyclam conformation II, supporting the assignment of the kinetic *in-cage* intermediate during Cu^{II}-H$_3$L^2 complex formation as the II isomer (see below). Conformers IV and V are energetically too high.

Radiochemical Experiments. Preliminary complexation experiments with NCA ^{64}Cu [molar copper(II) concentration of less than nM] were carried out at 25 and 70 °C, at a pH commonly used for labeling experiments (pH 6.2) and with various molar ligand excesses over the molar amount of ^{64}Cu (10 MBq corresponds to approximately 1 pmol of ^{64}Cu); for further experimental details, see the SI. The results clearly showed (Figure 6) that the labeling of both ligands was similar and that higher temperature led to better efficiency, that is, to a higher specific activity because a lower ligand excess had to be used for full labeling (Table S10). To compare the usefulness of our new ligands with that of established ligands, labeling experiments were also carried out with H$_3$nota and H$_4$dota under identical conditions and with the same batches of ^{64}Cu (Figures 6 and S19 and Table S11).

■ DISCUSSION

Recently, we have suggested that weakly complexing groups in pendant arms may accelerate the incorporation of a metal ion into a macrocyclic cavity.[23a] To examine this idea for cyclam and divalent copper, ligands H$_2$L^1 and H$_3$L^2 bearing a geminal bis(phosphorus acid) moiety were prepared by a simple and scalable synthesis. Monosubstitution of the cyclam could be efficiently controlled by the amount of formaldehyde used in the synthesis. The ligand H$_2$L^1 is fully stable in the solid state as well as in aqueous solution at any pH, unlike some other geminal bis(H-phosphinic acid)s, in which the P−C bond is slowly hydrolyzed.[25b,35] Oxidation of the P−H bond was tested with H$_2$O$_2$ or halogens but always led to mixtures caused by cleavage of the N−C−P moiety that were difficult to purify. However, no cleavage of the N−C−P moiety was observed during mild oxidation with HgCl$_2$. Both ligands were isolated in the zwitterionic form.

Two isomeric copper(II) complexes of each ligand were isolated: blue complexes were formed at room temperature and at pH <7, and violet species were obtained from reaction mixtures equilibrated at pH >12. Blue-to-violet isomerization is commonly observed in copper(II) complexes of cyclam-based ligands.[36,37] The blue species assigned to I isomers (Figure S1) are formed as low-temperature kinetic species, which are
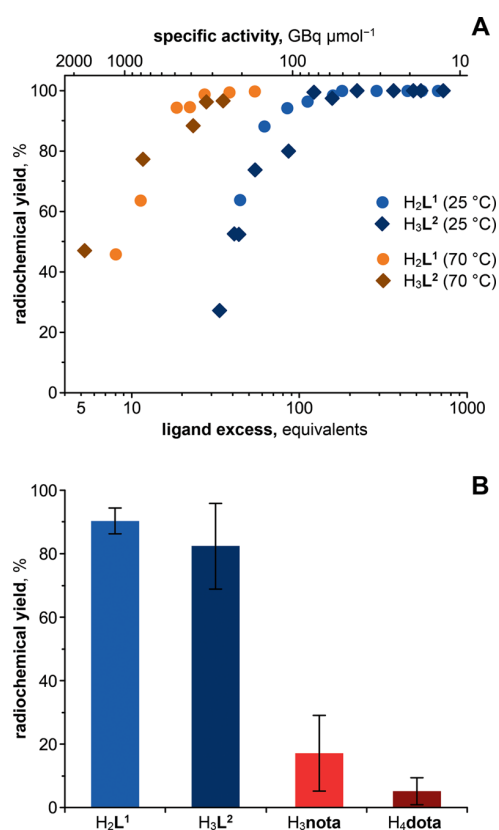
Figure 6. Dependence of the radiolabeling efficiencies of H_2L^1 and H_3L^2 with NCA ^{64}Cu on the ligand excess (A; 0.5 M MES buffer, pH 6.2, labeling time of 10 min; six different batches of ^{64}Cu were used) and comparison of the labeling with H_2L^1, H_3L^2, H_3**nota**, and H_4**dota** (B; 0.5 M MES buffer, pH 6.2, 25 °C, ca. 90 equiv of ligands, labeling time of 10 min; the data are averaged over three experiments, each performed with a freshly prepared batch of 9−11 MBq ^{64}Cu).

converted into violet **III** isomers (Figure S1) at high temperature.[36,37] These isomers differ in the relative orientations of the substituents on their ring nitrogen atoms and, thus, in their chelate ring conformations (Chart 3). Such isomerization has also been observed for copper(II) complexes of phosphonic/phosphinic acid cyclam derivatives.[17] However, unexpected behavior was observed here because **I** isomers of the title ligand complexes were transformed into the **III** isomer complexes even at room temperature, although the reaction only proceeded in aqueous solutions at pH >12 and always led to **I/III** isomer mixtures. Surprisingly, the **III** isomer complexes isomerized in alkaline solutions to the same isomeric mixture as that obtained starting from the **I** isomers. Partial **III**-to-**I** conversion was also observed in very acidic solutions at elevated temperatures during decomplexation studies. To the best of our knowledge, the **III**-to-**I** conversion has only been observed for copper(II) complexes of *C*-hexamethylated cyclam derivatives.[38] At neutral pH, no isomerization of any stock solution of a **I** or **III** isomer stored at or below room temperature was observed over a time span of months. Some isomerization could only be detected (TLC) after heating a neutral aqueous solution under reflux for several days.

The structure of the violet **III**-$[Cu(HL^2)]$ complex was confirmed by X-ray diffraction study. The central copper(II) ion is pentacoordinated by the four macrocycle amino groups

in the equatorial plane and one oxygen atom of the phosphinate pendant arm in an apical position (Figure 7). The metal ion
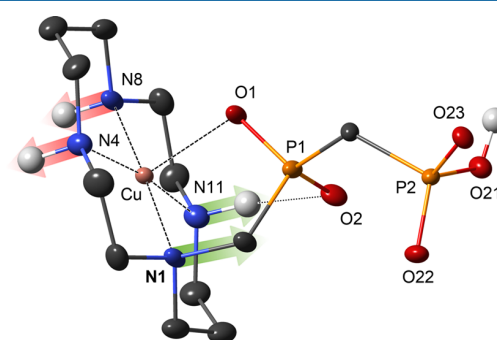

Figure 7. Molecular structure of $[Cu(HL^2)]$ found in the crystal structure of $[Cu(HL^2)]\cdot5H_2O$. For simplicity, hydrogen atoms (light gray) are shown only if bound to nitrogen or oxygen atoms.

resides only slightly above the N_4 base (0.10 Å) toward the phosphinate oxygen atom. The equatorial coordination distances $d(Cu-N)$ are significantly shorter (2.00−2.10 Å) than the axial one $[d(Cu-O1) = 2.24$ Å]. Below the N_4 base, an oxygen atom of a neighboring complex molecule is located in a pseudocoordinating distant position $[d(Cu\cdots O) = 2.99$ Å]. Thus, the structure could be viewed as an extremely elongated tetragonal bipyramid (Figure S20) with significant Jahn−Teller elongation of the axial P−O phosphinate chelate bond. The mutual orientation of the nitrogen atom substituents clearly points to conformation **III** of the cyclam chelate rings. The noncoordinated oxygen atom of the phosphinate group is involved in an intramolecular medium-to-strong hydrogen bond $[d(N11\cdots O2) = 3.04$ Å; $\angle(N11-H\cdots O2) = 148°]$. The whole structure is stabilized by a 3D network of hydrogen bonds. Selected geometric parameters of the copper(II) coordination sphere are listed in Table S12. The equatorial Cu−N distances in **III**-$[Cu(HL^2)]$ are very similar to those in copper(II) complexes of H_4**te2p**[1,8] and H_2**te1p** having cyclam ring conformation **III** (ca. 2.01 Å for N_{sec} and ca. 2.09 Å for N_{tert})[17,20] as well as those in complexes of H_4**teta** (2.06 Å)[39] or *C*-hexamethylated cyclam with one *N*-acetate pendant (ca. 2.04 Å for N_{sec} and 2.09 Å for N_{tert}).[40] Similarly, the Cu−O_{axial} bond (2.24 Å) is comparable to the analogous bonds in the **III**-$[Cu(Hte1p)]^+$ (2.28 Å) and H_4**teta** (2.27 Å) complexes[20,39] but longer than the Cu−O_{axial} distance in the complex of the methylated cyclam monoacetate (2.16 Å).[40]

The absorption spectra of the isomeric complexes correspond well to those published for complexes of other ligands with confirmed cyclam ring conformations.[17,41] The d−d absorption maxima of the blue isomers (ca. 590 nm) match those of the **I** isomers of the copper(II) complexes of H_4**te2P**[1,8] (596 nm)[17] and a mono(methylpyridine) cyclam derivative (604 nm),[41] in which the pendant arm is coordinated in the axial position. Therefore, the blue isomers should exhibit the **I** conformation of the cyclam ring with axial coordination of the pendant phosphinate group. The violet isomers should have the **III** conformation of the cyclam ring. Their d−d absorption maxima (ca. 535 nm) point to an equatorial ligand field intermediate between those for cyclam and H**te1a** copper(II) complexes (510 and 547 nm, respectively).[42] This can be attributed to the weaker axial interaction of the phosphinate group compared to that of the acetate group.

Having established that the **I** isomers of the copper(II) complexes are exclusively formed at room temperature and at pH <12, the thermodynamic properties of the ligands and complexes were studied. The protonation constants of the ligands are typical of those for polyazamacrocycles (Table S3). Ligand protonation sites were assigned on the basis of changes in the NMR chemical shifts of H_3L^2 (Figure S5). In tetraazamacrocycles, the first two protons are bound to ring amine groups, as supported by NMR shift changes in the alkaline region. However, the shapes of the NMR titration curves also indicate that the distant phosphonate/phosphinate groups alter the overall electronic properties of the ligands and/or are involved in the intramolecular hydrogen-bonding network. This leads to rather high basicity of the ring nitrogen atoms ($\log K_1 + \log K_2 = 22.93$ and 24.56 for H_2L^1 and H_3L^2, respectively). The amine basicity is significantly higher than that of cyclam itself. For H_2L^1, the ring amine basicity is comparable to that of Hte1a (23.05),[42] and for H_3L^2, it is even higher than that of the monophosphonate analogue, H_2te1p (24.25).[20] This is somewhat surprising because the presence of a methylphosphinate pendant arm on a macrocyclic amine typically results in a significant decrease in the ring amine basicity because of the presence of the electron-withdrawing phosphinate moiety.[43] For H_3L^2, the third proton ($\log K_a = 7.12$) is attached to the terminal phosphonate group because the $\log K_a$ value corresponds to those of other similar ligands[17−20,22] and the NMR chemical shift changes clearly relate only to the atoms in the pendant arm. Further protonations ($\log K_a = 1.6−2.9$) probably take place on the rest of the ring amino groups and/or, for H_3L^2, on the phosphonate group. Protonation of the phosphinate group(s) takes place at even lower pH. Therefore, the groups can interact with metal ions even at very low pH. An extended discussion on the NMR titration data is given in the SI.

Both ligands give very stable copper(II) complexes (Table 1), comparable in stability to complexes of related cyclam derivatives ($\log K_{CuL}$: cyclam, 27.2; H_2te1p, 27.3; H_4te2p[1,8], 25.4).[17,20,44] Ligand selectivity toward copper(II) over zinc(II) and nickel(II) ions (by ca. 7 and 4 orders of magnitude, respectively) is excellent and similar to those of other cyclam derivatives. Under equilibrium conditions, the protonation constants of the H_2L^1/H_3L^2 complexes might be connected to either protonation of the ring amine group or the distant phosphinate/phosphonate.[17,45] The terminal phosphonate group of H_3L^2 is protonated at neutral pH and is probably not coordinated because the corresponding $\log K_1$ values (6.6−7.2) of the complexes are in the typical range for free phosphonate groups. The axial bond to the phosphinate oxygen atom in the copper(II) complexes is rather weak and, therefore, it can be replaced by OH^- to form hydroxido complexes in strongly alkaline solutions.

The efficiency of radiometal complexation is probably the most important parameter to be evaluated for ligands considered as radiometal chelators. Ligand structural parameters governing metal-ion incorporation can be estimated from kinetic data obtained under defined conditions, which also allow determination of the mechanism of complexation. The mechanism of complexation of H_2L^1 and H_3L^2 with copper(II) (Scheme 2) differs from those reported for ligands with simple pendant arms.[17−19,32] The data are similar to those for stepwise complexation of copper(II) by cyclam bearing a "tren" pendant arm (cyclam-tren; Chart 1). Complexation starts with the formation of a rather stable Cu^{II}-tren out-of-cage (ooc) complex,

in which only the tren moiety is coordinated.[46] Here, intermediate **1**[#] formed in the preequilibrium step should be weakly chelated only through the oxygen atoms of the geminal bis(phosphinate) or phosphino−phosphonate group in an *out-of-cage* fashion (Scheme 5), analogous to coordination modes

Scheme 5. Formation and Isomerization of Copper(II) Complexes of H_2L^1 and H_3L^{2a}
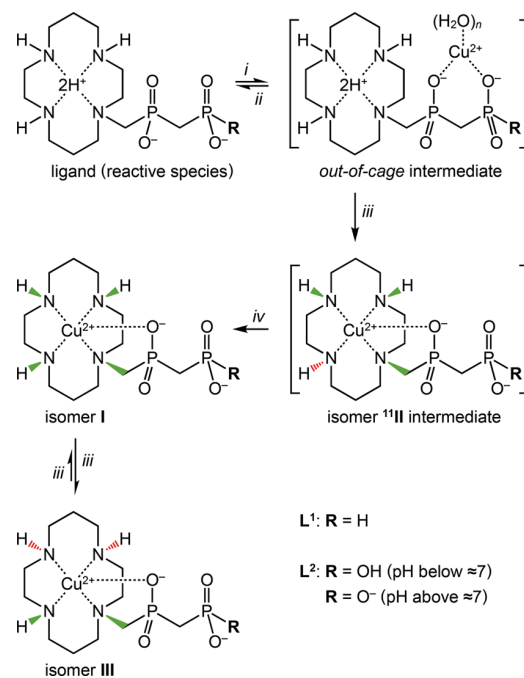
^{a}(i) $+Cu^{2+}$, (ii) $-Cu^{2+}$, (iii) OH^-, (iv) Cu^{2+}, OH^-, or both Cu^{2+} and OH^-.

recently observed for bis(phosphinate)s[25] and for the more frequently investigated geminal bis(phosphonate)s.[27,47−49] The stability constants of these *out-of-cage* intermediates (log $^{\#}K_{CuH_2L} = 2.6$ and 6.4 for H_2L^1 and H_3L^2, respectively) are in the expected range for chelation of copper(II) with geminal bis(phosphinate/phosphonate)s.[25]

Higher stability of the ooc-(Cu-H_3L^2) complex is imparted by the higher charge and basicity of the phosphonate group compared with those of the phosphinate group. The dissociation constant of the triprotonated ooc-(Cu-H_3L^2) species $[p(^{\#}K_{aP}) = 4.2]$ corresponds well with those of coordinated phosphonate groups.[27] In the investigated pH range, the cyclam ring should be predominantly diprotonated and, thus, the *out-of-cage* intermediates might have the structure depicted in Scheme 5.

The next step in the mechanism is transfer of the metal ion into the macrocyclic cavity, which proceeds in an unusual stepwise manner (Scheme 5) not previously seen for similar systems. An intermediate complex is formed very quickly, which, somewhat more slowly, rearranges to the final blue **I** complex. The intermediate complex **2**[#] shows spectral characteristics (Figure S8) corresponding to coordination of the four ring amine groups in the equatorial plane and, therefore, it should be one of the five major isomers of cyclam complexes (Chart 3 and Figure S1). If constrained cyclams (e.g., cross-bridged ligands in Chart 1) are not considered, the

IV and **V** isomers of copper(II) complexes of cyclam derivatives are very rare and have been observed only for highly *N*- or *C*-substituted cyclams.[50,51] Most of the copper(II) complexes have conformations **I** or **III**. Because the final product formed at pH <7 has cyclam ring conformation **I** and isomerizes to the **III** complex only at high pH, the blue intermediate **2**[#] probably has conformation **II**. Indeed, this conformation has previously been observed in copper(II) complexes of some substituted cyclams[39,52] as well as for cyclam itself.[53] Apart from the limited spectral characteristics, no more structural information could be obtained. Therefore, quantum-chemical calculations on complexes of the title and similar ligands were undertaken. For H_3L^2, H_2te1p, and $H_4te2p^{1,8}$, there are four possible arrangements of the nitrogen substituents in pentacoordinated complexes complying with the general cyclam ring conformation **II**. Among them, the [4]**II** isomer has the lowest relative energy (Chart 3, Figure S16, and Table 5). However, such a structure has similar probabilities of rearranging to either the **I** or **III** isomers, whereas only isomerization to the **I** complex was observed. The [11]**II** isomer has somewhat higher relative energy, but it features intramolecular hydrogen bonds to two N—H groups (Figure 8) that might contribute to the structural
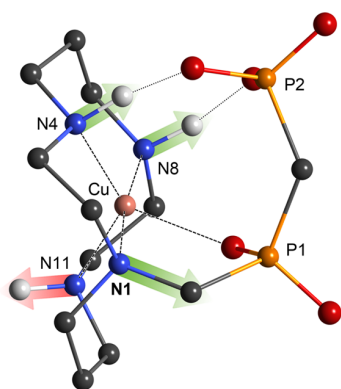

Figure 8. Calculated structure of the [11]**II**-$[Cu(L^2)]^-$ complex. Carbon-bound hydrogen atoms are omitted for simplicity.

stabilization. Inversion at nitrogen atom N11 leads only to the observed **I** isomer. Therefore, we suggest that the intermediate **2**[#] could be the [11]**II** isomer shown in Scheme 5 and Figure 8. A complex with cyclam ring conformation **II** was not observed as an intermediate during formation of the **I** isomer in the Cu^{II}-$H_4te2p^{1,8}$ system,[17] whereby the relative energy of the **II** isomer is much higher than that of the **I** isomer.

The step leading to the final complex, [11]**II** → **I** isomerization, is clearly hydroxide-catalyzed because the major contributions to the overall reaction rate of the isomerization reaction are associated with hydroxide-assisted pathways. The relevant kinetic parameters are very similar for complexes of both ligands (Table 2). For instance, at pH 4.8, this process is only slightly faster for the Cu^{II}-H_3L^2 complex (99% conversion, 44 and 12 s for H_2L^1 and H_3L^2, respectively, for molar ratio L:Cu = 1:1). This might be attributed to the better coordination ability of the phosphonate group, increasing the stability of its *out-of-cage* intermediate ([#]K_{CuH_2L}; Table 2) and/or a much higher value of [f]k_2, reflecting the better ability of the distant phosphonate group to transfer a proton from the macrocyclic cavity to the bulk solvent compared to the ability of the distant phosphinate group to do so. The anion of a cyclam derivative

with two phosphonate groups, also diprotonated on the ring amine nitrogen atoms, $(H_2te2p^{1,8})^{2-}$,[17] is even about 10^4 times more reactive than $(H_2L^2)^-$, pointing to even more effective proton transfer to the bulk solvent due to two phosphonate groups located close to the cycle. The reactivity of the diprotonated species of the bis(methylphosphinate) ligand,[32] $H_2te2p^{1,8ABn}$ (Chart 1), is comparable to that of the $(H_2L^2)^-$ anion, but it is about 600 times higher than that of the H_2L^1 species, further pointing to a significant role of the phosphonate group in the complexation mechanism of H_3L^2.

A qualitative comparison of the reaction rates (Figure S21) indicates that H_2L^1 and H_3L^2 form *in-cage* complexes (**II** isomers) much more rapidly than any other ligand and that H_3L^2 reacts about an order of magnitude faster than H_2L^1 at any pH. If formation of the **II** complexes is taken into account, full complexation (99%) is completed in around 14 and 2 s for H_2L^1 and H_3L^2, respectively (molar ratio Cu:L = 1:1; c_L = 0.1 mM; pH 4.8; 25 °C). Analogous kinetic data have only been published for **cyclam-tren** (Chart 1), a ligand with a strongly coordinating pendant arm.[46] **Cyclam-tren** binds copper(II) in its ligand cavity much more slowly (ca. 300 s; Cu:L = 1:1, c_L = 1 mM, pH 4.96, 25 °C) than H_2L^1 or H_3L^2. This observation suggests that the coordinating ability of the pendant arms should be appropriately tuned to balance the mutual stability of *out-of-cage* and *in-cage* complexes.

Surprisingly, the **III** isomers of the H_2L^1 and H_3L^2 complexes can be reverted to the **I** isomers, at variance with the common expectation that the **III** isomers are the final products. To the best of our knowledge, a **III** → **I** rearrangement in copper(II) complexes of cyclam-based ligands has only been observed for *C*-polymethylated cyclams.[38,54] This must be due to subtle differences in the structures of the isomeric complexes, reflected in their relative energies. Calculations confirmed the generally accepted opinion that **III** isomers have the lowest relative energy (Table 5). The energetic difference between the **I** and **III** isomers decreases in the order $H_4te2p^{1,8}$ > H_2te1p > H_3L^2, and the very small energetic difference between the isomers of the Cu^{II}-H_3L^2 complex points to the possibility of their mutual interconversion.

A small degree of isomerization was observed when neutral solutions of the complexes were heated under reflux for several days, and more extensive isomerization was observed in the course of acid-assisted dissociation experiments with the **III**-$[Cu(L^2)]$ complex. The isomerization proceeded smoothly at pH >12 under hydroxide anion catalysis, however, as a complicated process with no simple interpretation. In hydroxide-catalyzed reactions, the hydroxide anion may play several roles: (i) it may replace the coordinated phosphinate pendant arm, (ii) it may coordinate in the other axial position, and/or (iii) it may deprotonate coordinated secondary amine group(s). Monohydroxido species are not especially kinetically active (Table 3). Two other hydroxide anions may assist deprotonation and pyramidization of a coordinated secondary amino group (a classical CB mechanism), leading to inversion at the nitrogen atom. Such a mechanism involving amine inversion has been suggested for isomerization of copper(II) complexes of *C*-polymethylated cyclam derivatives in highly alkaline media.[54] The isomerization is somewhat slower for the $[Cu(L^2)]^-$ complex because of the higher negative charge of the phosphonate group, which prevents access by hydroxide anions. Such mutual interconversion of isomers has also been observed for zinc(II) and nickel(II) complexes of cyclams *N*-monosubstituted with a noncoordinating pendant arm.[55]

The kinetic inertness of complexes is commonly checked through decomplexation in highly acidic solutions, that is, through proton-assisted dissociation. Here, such measurements could only be rigorously carried out for $\mathbf{I}$-$[\mathrm{Cu}(\mathbf{L^2})]^-$. Under the conditions used for the measurements, both isomers of $[\mathrm{Cu}(\mathbf{L^1})]$ underwent oxidation of the P−H bond by atmospheric oxygen. The activation parameters for dissociation of complexes of the title and related phosphonate ligands[17,18,32] are similar (Table 4), indicating analogous dissociation mechanisms. Decomplexation of $\mathbf{I}$-$[\mathrm{Cu}(\mathbf{L^2})]^-$ starts with protonation of the coordinated phosphinate group, followed by transfer of the proton to a ring amine group with concomitant decomposition of the protonated intermediate. Such a scenario is supported by the absence of the $^d k_0$ dissociation pathway, corresponding to a $\mathbf{I}$-$[\mathrm{Cu}(\mathrm{H}\mathbf{L^2})]$ species. It is known that $\mathbf{III}$ isomers of copper(II) complexes of cyclam-based ligands are very kinetically inert (half-life on the order of weeks or months, even at high temperatures).[17,32] Thus, the $\mathbf{III}$-$[\mathrm{Cu}(\mathbf{L^2})]^-$ complex was expected to exhibit inertness similar to those of other $\mathbf{III}$ complexes. Because of its partial conversion to the $\mathbf{I}$ isomer during kinetic experiments, detailed kinetic studies were not performed. Therefore, only qualitative experiments[56] were carried out in the presence of two anions because it is known that the presence of chloride anions significantly enhances such decomplexation processes. It was confirmed that the $\mathbf{III}$-$[\mathrm{Cu}(\mathbf{L^2})]^-$ complex was more kinetically inert than its $\mathbf{I}$ isomer. However, $\mathbf{III}$ complexes of both 1,8-disubstituted cyclam derivatives[17,32] are significantly more kinetically inert than the $\mathbf{III}$-$[\mathrm{Cu}(\mathbf{L^2})]^-$ complex, probably because of the octahedral coordination sphere completely encapsulating the central metal ion. Nevertheless, the kinetic inertness of the title complexes is fully sufficient for possible in vivo applications.

The limiting factor for the application of geminal bis-(phosphonate) complexes in molecular imaging is their binding to calcified tissues. To evaluate the possible bone affinity of the studied complexes, an aqueous suspension of hydroxyapatite was used as a model of bone tissue (for details, see the SI). All of the copper(II) complexes showed negligible adsorption under conditions that led to complete adsorption of lanthanide-(III) macrocyclic complexes of similar size bearing a geminal bis(phosphonate) group (Figure S22).[57] This indicates that both ligands/complexes with bis(phosphinate) or phosphinato-to−phosphonate moieties should show negligible sorption on bones, similar to other bis(phosphinate)-containing compounds;[35] therefore, compounds with these groups are suitable for biomedical applications in which bone sorption is undesirable.

To confirm the promising data for copper(II) complexation obtained with nonradioactive copper at millimolar concentrations and to shift the ligands more closely toward real-life applications, the behavior of the title ligands was tested with NCA $^{64}\mathrm{Cu}$ (for details, see the SI). Labeling was carried out at two temperatures, one suitable for the labeling of oligopeptides (70 °C) and the other for the labeling of antibody conjugates (room temperature). To achieve a high specific activity of the labeled molecule, ligand excess over the molar amount of the metallic radioisotope was kept as low as possible, whereupon very striking results were obtained. Even at room temperature, labeling was very fast, leading to a high specific activity of around 70 GBq of $^{64}\mathrm{Cu}$ per 1 μmol of ligand, and a much higher specific activity of around 400 GBq per 1 μmol of ligand was obtained at high temperature (70 °C). However, a comparison of the radiolabeling efficiencies between ligands reported in the literature is often difficult because of the (even slightly) different conditions used. Two "standard" macrocyclic ligands, $\mathrm{H_3}\mathbf{nota}$ and $\mathrm{H_4}\mathbf{dota}$, were chosen for comparison in these very preliminary experiments because their derivatives are the most commonly used copper radioisotope chelators. To obtain directly comparable data, all four ligands were labeled with the same batches of freshly prepared NCA $^{64}\mathrm{Cu}$, following an identical protocol. To distinguish between the ligands, conditions for such a comparison were chosen on the basis of the results obtained for $\mathrm{H_2}\mathbf{L^1}$ and $\mathrm{H_3}\mathbf{L^2}$ (Figure 6A). The comparison (Figure 6B) clearly shows a better radiolabeling efficiency (significantly higher radiochemical yields for the same ligand-to-metal ratio) of the title ligands in comparison with $\mathrm{H_3}\mathbf{nota}$ and $\mathrm{H_4}\mathbf{dota}$. These results point to the advantages of phosphorus acid-containing pendant arms in ligands intended for radiochemical applications and are also consistent with data published on the labeling of cross-bridged cyclam phosphonate derivatives with $^{64}\mathrm{Cu}$.[12]

■ CONCLUSIONS

The studied cyclam derivatives bearing one methylenebis-(phosphinic acid) or methylene(phosphinic−phosphonic acid) pendant arm have been prepared by a simple and scalable synthesis. They show high amine basicity, high stability of complexes with divalent transition-metal ions, and high selectivity for copper(II) ions. The ligands show significantly increased rates of copper(II) complexation compared with other cyclam derivatives. Thus, these ligands are among the best macrocycles suggested for copper(II) binding. The fast complex formation can be attributed to the weakly coordinating bis(phosphinate) or phosphinato−phosphonate pendant arms, forming an *out-of-cage* complex. The complex formation further proceeds through rarely observed *in-cage* intermediates with cyclam ring conformation $\mathbf{II}$ to give the final $\mathbf{I}$ isomer of the complexes. For the first time, "reverse" conversion of copper(II) cyclam complexes in conformation $\mathbf{III}$ to complexes in conformation $\mathbf{I}$ has been observed. This is facilitated by the ability of the pendant phosphonate−phosphinate anions to form intramolecular hydrogen bonds. This has also been supported by detailed theoretical calculations that included a thorough conformational search and the use of a state-of-the-art solvation model (COSMO-RS).

The presented data suggest a high application potential of the studied ligands. The copper(II) complexes are formed very quickly and are kinetically inert. They show no measurable hydroxyapatite sorption and, thus, negligible bone uptake in vivo could be expected. The ligands exhibit a high selectivity for divalent copper over other divalent metal ions, implying that high specific activity should be obtained during radiolabeling, as was confirmed by preliminary labeling experiments. In line with our previous data,[19,23] we have shown here that the concept of weakly binding pendant arms is helpful in the design of new ligands for radiochemical applications. Metal radioisotope incorporation is accelerated through the formation of weak *out-of-cage* complexes, whereby the coordinating pendant arms are essentially "fishing out" dilute radiometals in an "open-chain ligand" fashion, increasing the effective metal-ion concentration close to the macrocyclic cavity. Thus, radiolabeling of $\mathrm{H_2}\mathbf{L^1}$ and $\mathrm{H_3}\mathbf{L^2}$ led to much higher specific activities compared to those with $\mathrm{H_3}\mathbf{nota}$ and $\mathrm{H_4}\mathbf{dota}$. Future designs of ligands for efficient complexation of various metal radioisotopes might be based on this phenomenon.

■ EXPERIMENTAL SECTION

General Procedures. H_3**nota** and H_4**dota** were available from previous studies. The commercially available (Fluka, Aldrich, Chematech, and Strem) chemicals had synthetic purity and were used as received. Paraformaldehyde was filtered from an aged formaldehyde aqueous solution and dried over P_2O_5 in a vacuum desiccator. Methylenebis(phosphinic acid) was quantitatively prepared by careful hydrolysis of commercial $Cl_2PCH_2PCl_2$ (Strem) according to the published procedure.[58] Commercial chemicals (Merck, Fluka, and Aldrich) for physicochemical measurements had at least analytical purity. 1D and 2D NMR experiments were run on Bruker Avance 500 or Bruker Avance III 600 with cold probes (^1H and $^{13}C\{^1H\}$) or on Varian S300 (^{31}P and $^{31}P\{^1H\}$) spectrometers (chemical shifts in ppm and coupling constants in Hz; for an atom-labeling scheme, see Figure S5). Electrospray ionization mass spectrometry (ESI-MS) spectra were recorded on a Bruker Esquire 3000 spectrometer with ion-trap detection in negative or positive mode. Merck aluminum foils with silica gel 60 F_{254} were used for TLC. For analysis of the copper(II) complexes, 2D TLC techniques were used to check the isomerization stability in the mobile phase used: analyses were performed on square-shaped TLC plate; after the first run, the mobile phase was carefully evaporated from the plate by a flow of air, and the analysis was repeated on the same plate rotated by 90° (in the same mobile phase). In addition, in this setup, the presence of diagonal spots means that the compounds are stable in the examined mobile phase. For UV/vis characterization of the complexes, experiments were carried out on a Specord 50Plus spectrophotometer (210−1100 nm, Analytic Jena, Germany) and the absorption peaks are presented in the format λ_{max} (ε) [nm (mol^{-1} dm^3 cm^{-1})]. The exact copper(II) complex concentrations of stock solutions were determined by atomic absorption spectrometry (AAS) on an AAS3 spectrometer (Carl Zeiss, Germany); content obtained in mg$_{Cu}$ dm^{-3}. Throughout the paper, pH means $-\log[H^+]$ and $pD = pH + 0.4$. Elemental analyses were performed at Palacky University (Olomouc, Czech Republic).

^{64}Cu was produced in a Cyclone 18/9 cyclotron (Helmholtz-Zentrum Dresden-Rossendorf) by a $^{64}Ni(p,n)^{64}Cu$ nuclear reaction, giving specific activities of 150−250 GBq mM^{-1} Cu diluted in aqueous HCl (10 mM).[59] The radiolabeling of the chelators was monitored by radio-TLC (SiO$_2$ 60 W F$_{254}$ S, Merck) and evaluated using a radioactivity thin-layer analyzer (Rita Star, Raytest). The radioactivity was counted with an ISOMED 2010 (Nuklear-Medizintechnik Dresden GmbH).

Synthesis of [[(1,4,8,11-Tetraazacyclotetradecan-1-yl)methyl]-{[(hydroxy)(hydro)phosphoryl]methyl}phosphinic Acid (H$_2$L^1). To an ice-bath-cooled flask charged with cyclam (11.2 g, 55.7 mmol, 4.0 equiv) and methylenebis(phosphinic acid) (4.00 g, 27.8 mmol, 2.0 equiv) was slowly added cold concentrated aqueous HCl (650 mL). After the end of the exothermic reaction, paraformaldehyde (422 mg, 14.1 mmol, 1.0 equiv) was added in one portion, and the flask was quickly closed with a stopper. The resulting suspension was then stirred at 80 °C overnight. After cooling to room temperature, the reaction mixture was concentrated to dryness and residual volatiles were removed by repeated coevaporation with water. The crude product was purified on a strong anion-exchange resin (Dowex 1, ≈250 mL, OH$^-$ form, elution H$_2$O → 10% aqueous AcOH). Unreacted cyclam was eluted with water, and the crude product was eluted with aqueous AcOH. The acetic acid fraction was concentrated to dryness, residual volatiles were removed by repeated coevaporation with water, and the residue was purified on a strong cation-exchange resin (Dowex 50, ≈250 mL, H$^+$ form, elution H$_2$O → 10% aqueous pyridine). Unreacted methylenebis(phosphinic acid) was eluted with water, and the product was eluted with aqueous pyridine. The pyridine fraction was concentrated to dryness, and residual volatiles were coevaporated several times with water. The resulting colorless residue was dried for several days on a vacuum pump at 80 °C, and the solid was further dried in a vacuum desiccator over P$_2$O$_5$. The product was obtained as a fine white hygroscopic powder (3.95 g, 73%). Elem anal. Calcd for C$_{12}$H$_{30}$N$_4$O$_4$P$_2$·1.5H$_2$O (M_r = 383.4): C, 37.6; H, 8.7; N, 14.6. Found: C, 37.9; H, 8.8; N, 14.3. NMR (D$_2$O; pD ≈ 5.0, 97 °C):

^1H δ 1.91 (H6, m, 2H), 2.01 (H13, m, 2H), 2.03 (P−CH$_2$−P, t, 2H, $^2J_{HP}$ = 17 Hz), 2.79 (H14, m, 2H), 2.88 (N−CH$_2$−P, d, 2H, $^2J_{HP}$ = 6 Hz), 2.92 (H2, m, 2H), 2.93 (H9, m, 2H), 2.94 (H7, m, 2H), 3.13 (H10, m, 2H), 3.16 (H12, m, 2H), 3.23 (H3, m, 2H), 3.26 (H5, m, 2H); $^{13}C\{^1H\}$ δ 23.7 (C13, s), 25.7 (C6, s), 36.3 (P−CH$_2$−P, t, $^1J_{CP}$ = 117 and 77 Hz), 44.9 (C9, s), 46.5 (C3, s), 47.2 (C10, s), 47.5 (C12, s), 49.4 (C7, s), 49.7 (C5, s), 54.8 (N−CH$_2$−P, d, $^1J_{CP}$ = 108 Hz), 55.2 (C2, s), 57.4 (C14, d, $^3J_{CP}$ = 12 Hz); ^{31}P δ 19.1 (PH, dtd, 1P, $^1J_{PH}$ = 530 Hz, $^2J_{PH}$ = 17 Hz, $^2J_{PP}$ = 5 Hz), 30.5 (P−CH$_2$−N, m, 1P). ESI-MS (negative ion) m/z 355.6 ([M − H]$^-$); (positive mode) m/z 357.9 ([M + H]$^+$), 395.7 ([M + K]$^+$). TLC (SiO$_2$, 1:1 EtOH/concentrated aqueous NH$_4$OH): R_f = 0.7.

Synthesis of ({[(1,4,8,11-Tetraazacyclotetradecan-1-yl)methyl]-(hydroxy)phosphoryl}methyl)phosphonic Acid (H$_3$L^2). A hot solution of HgCl$_2$ (1.63 g, 6.00 mmol, 1.6 equiv) in water (30 mL) was added to a solution of H$_2$L^1·1.5H$_2$O (1.47 g, 3.83 mmol, 1.0 equiv) in 1% aqueous HCl (30 mL), and the mixture was stirred at 60 °C overnight. After cooling to room temperature, precipitated Hg$_2$Cl$_2$ was carefully removed by filtration and washed with water. The filtrate was saturated with H$_2$S, and the precipitated HgS was filtered off and washed with water. The colorless filtrate was concentrated to dryness, and the residue was purified on a strong anion-exchange resin (Dowex 1, ≈100 mL, OH$^-$ form, elution H$_2$O → 10% aqueous AcOH). The acetic acid fraction was concentrated to dryness, residual volatiles were removed by repeated coevaporation with water, and the residue was purified on a strong cation-exchange resin (Dowex 50, ≈100 mL, H$^+$ form, elution H$_2$O → 10% aqueous pyridine). The pyridine fraction was concentrated to dryness, and residual volatiles were removed by repeated coevaporation with water. The resulting colorless residue was dried for several days on a vacuum pump at 80 °C, and the solid was further dried in a vacuum desiccator over P$_2$O$_5$. The product was obtained as a fine white hygroscopic powder (1.34 g, 90%). Elem anal. Calcd for C$_{12}$H$_{30}$N$_4$O$_5$P$_2$·H$_2$O (M_r = 390.4): C, 37.0; H, 8.3; N, 14.2. Found: C, 37.3; H, 8.6; N, 13.9. NMR (D$_2$O + CsOD; pD ≈ 8.0, 97 °C): ^1H δ 1.90 (H6, m, 2H), 1.98 (P−CH$_2$−P, t, 2H, $^2J_{HP}$ = 18 Hz), 2.05 (H13, m, 2H), 2.74 (H14, m, 2H), 2.90 (H2, m, 2H), 2.91 (H9, m, 2H), 2.92 (H7, m, 2H), 3.01 (N−CH$_2$−P, d, 2H, $^2J_{HP}$ = 7 Hz), 3.06 (H12, m, 2H), 3.12 (H10, m, 2H), 3.21 (H3, m, 2H), 3.24 (H5, m, 2H); $^{13}C\{^1H\}$ δ 23.6 (C13, s), 25.8 (C6, s), 34.0 (P−CH$_2$−P, dd, $^1J_{CP}$ = 117 and 80 Hz), 44.7 (C9, s), 46.6 (C3, s), 46.8 (C10, s), 47.1 (C12, s), 49.6 (C7, s), 49.9 (C5, s), 53.5 (N−CH$_2$−P, d, $^1J_{CP}$ = 104 Hz), 55.4 (C2, s), 57.0 (C14, d, $^3J_{CP}$ = 13 Hz); $^{31}P\{^1H\}$ δ 13.0 (HO−P−OH, d, 1P, $^2J_{PP}$ = 4 Hz), 33.9 (N−CH$_2$−P, d, 1P, $^2J_{PP}$ = 4). ESI-MS: (positive ion) m/z 373.3 ([M + H]$^+$), 395.2 ([M + Na]$^+$), 411.2 ([M + K]$^+$). TLC (SiO$_2$, 1:1 EtOH/concentrated aqueous NH$_4$OH): R_f = 0.4.

Synthesis of I-[Cu(L^1)]. To a solution of ligand H$_2$L^1·1.5 H$_2$O (138 mg, 360 μmol, 1.0 equiv) in water (4 mL) was added aqueous pyridine (10%, 3 mL). Solid Cu(OAc)$_2$·H$_2$O (104 mg, 521 μmol, 1.4 equiv) was added, the mixture was carefully concentrated to dryness in vacuo (the bath temperature was not raised above 35 °C), and residual volatiles were removed by repeated coevaporation with water. The residue was purified on a weak cation-exchange resin (Amberlite CG50, H$^+$ form, ≈20 mL, elution with H$_2$O). Fractions with pure product were combined and carefully concentrated to dryness in vacuo (the bath temperature was not raised above 35 °C). The dark-blue oily residue was dissolved in water (100 mL) and subsequently lyophilized. The resulting solid product was further dried in a vacuum desiccator over P$_2$O$_5$. The product was obtained as a dark-blue hygroscopic powder (108 mg, 64%). Elem anal. Calcd for C$_{12}$H$_{28}$N$_4$O$_4$P$_2$Cu·3H$_2$O (M_r = 471.9): C, 30.5; H, 7.3; N, 11.9. Found: C, 30.8; H, 7.0; N, 11.8. 2D TLC (SiO$_2$, 7:3:3 iPrOH/concentrated aqueous NH$_4$OH/H$_2$O): R_f = 0.6 (blue spot). ESI-MS: (negative ion) m/z 415.6 ([M − H]$^-$), 451.6 ([M + Cl]$^-$); (positive ion) m/z 417.8 ([M + H]$^+$), 439.8 ([M + Na]$^+$), 455.7 ([M + K]$^+$). UV/vis (H$_2$O, pH 7.4): 270 (5.6 × 10^3), 590 (1.5 × 10^2).

Synthesis of I-[Cu(HL^2)]. To a solution of ligand H$_3$L^2·H$_2$O (216 mg, 553 μmol, 1.0 equiv) in water (6 mL) was added aqueous pyridine (10%, 4 mL). Solid Cu(OAc)$_2$·H$_2$O (159 mg, 796 μmol, 1.4 equiv) was added, the mixture was carefully concentrated to dryness in vacuo (the bath temperature was not raised above 35 °C), and residual

volatiles were removed by repeated coevaporation with water. The residue was purified on a weak cation-exchange resin (Amberlite CG50, H$^+$ form, ≈20 mL, elution with H$_2$O). Fractions with pure product were combined and carefully concentrated to dryness in vacuo (the bath temperature was not raised above 35 °C). The dark-blue oily residue was dried on a vacuum pump for 1 day at 35 °C, and the resulting solid was further dried in a vacuum desiccator over P$_2$O$_5$. The product was obtained as a dark-blue hygroscopic powder (195 mg, 78%). Elem anal. Calcd for C$_{12}$H$_{28}$N$_4$O$_5$P$_2$Cu·2H$_2$O (M_r = 469.9): C, 30.7; H, 6.9; N, 11.9. Found: C, 30.5; H, 6.6; N, 11.9. 2D TLC (SiO$_2$, 7:3:3 *i*-PrOH/concentrated aqueous NH$_4$OH/H$_2$O): R_f = 0.3 (blue spot). ESI-MS: (negative ion) *m/z* 431.6 ([M − H]$^-$), 467.6 ([M + Cl]$^-$); (positive ion) *m/z* 433.8 ([M + H]$^+$), 455.8 ([M + Na]$^+$), 471.7 ([M + K]$^+$). UV/vis (H$_2$O, pH 7.4): 270 (5.6 × 10^3), 590 (1.4 × 10^2).

Synthesis of III-[Cu(L^1)]. Aqueous CuCl$_2$ (101.8 mM, 6.3 mL, 640 μmol, 1.0 equiv) was added to ligand H$_2$L^1·1.5 H$_2$O (246 mg, 642 μmol, 1.0 equiv). Then, aqueous NaOH (2 M, approximately 15 mL, to reach pH ≈12−13) was added, and the mixture was stirred for 1 h at room temperature. The resulting solution containing a mixture of I-[Cu(L^1)] and III-[Cu(L^1)] (monitored by TLC) was concentrated to dryness in vacuo, and the residue was purified by column chromatography [SiO$_2$, ≈250 g, 7:3:3 *i*PrOH/concentrated aqueous NH$_4$OH/H$_2$O; R_f (III-[Cu(L^1)]) = 0.5]. Fractions with pure product were combined and carefully concentrated in vacuo to ≈25 mL (the bath temperature was not raised above 35 °C), and the resulting solution was immediately purified on a weak cation-exchange resin (Amberlite CG50, H$^+$ form, ≈100 mL, elution with H$_2$O). Fractions with pure product were combined, carefully concentrated in vacuo to ≈50 mL (the bath temperature was not raised above 35 °C), and subsequently lyophilized. The resulting solid was further dried in a vacuum desiccator over P$_2$O$_5$. The product was obtained as a dark-violet hygroscopic powder (106 mg, 35%). Elem anal. Calcd for C$_{12}$H$_{28}$N$_4$O$_4$P$_2$Cu·0.5NH$_3$·2.5H$_2$O (M_r = 471.4): C, 30.6; H, 7.4; N, 13.4. Found: C, 30.2; H, 7.1; N, 13.1. 2D TLC (SiO$_2$, 7:3:3 *i*PrOH/concentrated aqueous NH$_4$OH/H$_2$O): R_f = 0.5 (violet spot). ESI-MS: (negative ion) *m/z* 415.6 ([M − H]$^-$), 451.6 ([M + Cl]$^-$); (positive ion) *m/z* 417.8 ([M + H]$^+$), 439.8 ([M + Na]$^+$), 455.7 ([M + K]$^+$). UV/vis (H$_2$O, pH 7.4): 265 (6.3 × 10^3), 535 (1.2 × 10^2).

Synthesis of III-[Cu(HL^2)]. Aqueous CuCl$_2$ (101.8 mM, 5.7 mL, 582 μmol, 1.0 equiv) was added to ligand H$_3$L^2·H$_2$O (249 mg, 638 μmol, 1.1 equiv). Then, aqueous NaOH (2 M, approximately 15 mL, to reach pH ≈12−13) was added, and the mixture was stirred for 1 h at room temperature. The resulting solution containing a mixture of I-[Cu(L^2)]$^-$ and III-[Cu(L^2)]$^-$ (monitored by TLC) was concentrated to dryness in vacuo, and the residue was purified by column chromatography [SiO$_2$, ≈200 g, 7:3:3 *i*PrOH/concentrated aqueous NH$_4$OH/H$_2$O; R_f (III-[Cu(L^2)]$^-$) = 0.2]. The fraction containing pure product was carefully concentrated in vacuo to ≈25 mL (the bath temperature was not raised above 35 °C), and the resulting solution was immediately purified on a strong cationic exchanger (Dowex 50, ≈50 mL, Hpyridine$^+$ form, elution with H$_2$O). Fractions with pure product were combined and carefully concentrated in a vacuum to ≈25 mL (the bath temperature was not raised above 35 °C), and the resulting solution was immediately purified on a weak cation-exchange resin (Amberlite CG50, H$^+$ form, ≈100 mL, elution with H$_2$O). Fractions with product were combined, carefully concentrated in a vacuum to ≈50 mL (the bath temperature was not raised above 35 °C), and subsequently lyophilized. The resulting solid was further dried in a vacuum desiccator over P$_2$O$_5$. The product was obtained as a dark-violet hygroscopic powder (126 mg, 43%). Elem anal. Calcd for C$_{12}$H$_{28}$N$_4$O$_5$P$_2$Cu·3.5H$_2$O (M_r = 496.9): C, 29.0; H, 7.1; N, 11.3. Found: C, 28.6; H, 6.7; N, 11.1. 2D TLC (SiO$_2$, 7:3:3 *i*PrOH/concentrated aqueous NH$_4$OH/H$_2$O): R_f = 0.2 (violet spot). ESI-MS: (negative ion) *m/z* 431.6 ([M − H]$^-$), 467.6 ([M + Cl]$^-$); (positive ion) *m/z* 433.8 ([M + H]$^+$), 455.8 ([M + Na]$^+$), 471.7 ([M + K]$^+$). UV/vis (H$_2$O, pH 7.4): 265 (6.1 × 10^3), 535 (1.2 × 10^2).

Formation Kinetics. Experiments were performed on a Bio Sequential SX-20 stopped-flow spectrophotometer (Applied Photophysics) equipped with a 150-W xenon lamp and a diode-array

accessory detector. The measurements were performed at 25 ± 0.1 °C, I = 0.1 M KCl, and pH 2.2−6.4. 50-fold molar excesses of the appropriate buffers were used [chloroacetic acid, acetic acid, or 2-(*N*-morpholino)ethanesulfonic acid (MES)] to buffer stock solutions of the ligands. Aqueous CuCl$_2$ was used as a source of copper(II) ions. The spectra were recorded both in the full wavelength range of 200−800 nm and at a single wavelength. Experiments were run at c_L = 0.1 mM and at c_{Cu} = 0.1−5 mM. Changes in the intensity of the ligand-to-metal charge-transfer or d−d bands of the complexes at λ = 300 or 590 nm, respectively, with time were used for the fitting. Data were fitted by *Pro-KII* software (Applied Photophysics). The results presented in Tables 2 and 3 and Figures 3 and 4 originated from simultaneous fitting of all formation kinetic data.

Dissociation Kinetics of I- and III-[Cu(HL^2)]. Stock solutions were prepared from the isolated complexes and stored in a refrigerator. Experiments were run on an HP 8453A diode-array spectrophotometer (Agilent) equipped with a thermoregulator. The measurements were carried out at I = 5.0 M (H,Na)ClO$_4$ or (H,Na)Cl and in the temperature range of (60−90) ± 0.1 °C. For measurements in the UV region, experiments were run at $c_{complex}$ = 0.1 mM, and the spectra were recorded in the range of 250−350 nm. For measurements in the visible region, experiments were run at $c_{complex}$ = 1.0 mM and over the wavelength range of 300−1000 nm. The following wavelengths were used for fitting: **I** isomers, λ = 270 or 590 nm; **III** isomer, λ = 270 or 535 nm.

Isomerization Kinetics. Experiments were run on a Varian Cary 50 UV/vis spectrophotometer (Agilent) equipped with a thermoregulator. The measurements were performed at 25 ± 0.1 °C with c_{CuL} = 5.0 mM and at [OH$^-$] in the range of 20−100 mM. If not otherwise stated, the experiments were performed at ionic strength I = 0.5 M K(OH,Cl); several points were measured in I = 0.5 M K(OH,NO$_3$) to evaluate the influence of anions, and the same results as those for chloride were obtained. The spectra were recorded in the wavelength range of 500−700 nm. For determination of the isomerization constants, changes in the intensity of the d−d bands at 650 nm were used for the fitting.

Computational Details. All calculations reported herein were performed using the *TURBOMOLE 6.5* program.[60] Whenever applicable, calculations were expedited by expanding the Coulomb integrals in an auxiliary basis set, using the resolution-of-identity (RI) approximation (density fitting). Grimme's D3 dispersion was applied to all calculations.[61] The geometry optimizations were carried out at the BP86[62] level of theory in conjunction with the def-TZVP[63] basis set and implicit aqueous solvent using the COSMO[64] method [r(Cu^{2+}) = 2.0 Å; ε = 80.0]. Free energies of interaction with the solvent (water) were calculated using the COSMO-RS method[65] (conductor-like screening model for realistic solvation), as implemented in the *COSMOtherm* program,[66] using the "BP_TZVP_C30_1201.ctd" parametrization file. The input files were provided by single-point COSMO calculations at the corresponding RI-BP86/def-TZVP level with dielectric constant parameter ε = ∞ (ideal conductor) or ε = 1 (vacuum). Gas-phase energies were calculated at the RI-TPSS[67]/def2-TZVP[68] and RI-BP86/def-TZVP levels of theory (Tables S6−S8). Zero-point energies as well as translational, rotational, and vibrational partition functions were obtained from harmonic and ideal-gas approximations based on frequency analysis performed in the solvent. The following definition of the Gibbs free energy was used: $G = E_{el} + ZPE − RT \ln(q_{rot}q_{vib}q_{trans}) + pV + \Delta G_{solv}$, where E_{el} is the gas-phase electronic energy, ZPE is the zero-point vibrational energy, $RT \ln q$ is the thermal correction to partition functions and entropy terms, and ΔG_{solv} is the solvation free energy.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.inorgchem.5b01791.

Additional experimental details, figures of the definition of isomers of cyclam complexes, solid-state packing of the Cu^{II}-H$_3$L^2 complex, electronic spectra of complexes, distribution diagrams, NMR titration, kinetic measurements, HAP sorption experiments, calculated structures of isomers of phosphonic acid copper(II) complexes, and ^{64}Cu labeling experiments and tables of electronic spectra of complexes, X-ray data on of the Cu^{II}-H$_3$L^2 complex, potentiometric results, calculated free energies of isomers of phosphonic acid copper(II) complexes, comparison of calculated and measured structural parameters of isomers of phosphonic acid copper(II) complexes, radiochemical data, and derivation for the mathematical treatment of kinetic data (PDF)

X-ray crystallographic data in CIF format (CIF)

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: petrh@natur.cuni.cz.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Wadas, T. J.; Wong, E. H.; Weisman, G. R.; Anderson, C. J. *Chem. Rev.* **2010**, *110*, 2858−2902.

(2) Heffern, M. C.; Matosziuk, L. M.; Meade, T. J. *Chem. Rev.* **2014**, *114*, 4496−4539.

(3) (a) *The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging*, 2nd ed.; Merbach, A., Helm, L., Tóth, É., Eds.; Wiley: Hoboken, NJ, 2013. (b) Hermann, P.; Kotek, J.; Kubíček, V.; Lukeš, I. *Dalton Trans.* **2008**, 3027−3047.

(4) (a) Dash, A.; Chakraborty, S.; Pillai, M. R. A.; Knapp, F. F., Jr. *Cancer Biother. Radiopharm.* **2015**, *30*, 47−71. (b) Cutler, C. S.; Hennkens, H. M.; Sisay, N.; Huclier-Markai, S.; Jurisson, S. S. *Chem. Rev.* **2013**, *113*, 858−883.

(5) (a) Bünzli, J.-C. G.; Eliseeva, S. V. *Chem. Sci.* **2013**, *4*, 1939−1949. (b) Butler, S. J.; Parker, D. *Chem. Soc. Rev.* **2013**, *42*, 1652−1666.

(6) Mewis, R. E.; Archibald, S. J. *Coord. Chem. Rev.* **2010**, *254*, 1686−1712.

(7) Ramogida, C. F.; Orvig, C. *Chem. Commun.* **2013**, *49*, 4720−4739.

(8) Price, E. W.; Orvig, C. *Chem. Soc. Rev.* **2014**, *43*, 260−290.

(9) Zeglis, B. M.; Houghton, J. L.; Evans, M. J.; Viola-Villegas, N.; Lewis, J. S. *Inorg. Chem.* **2014**, *53*, 1880−1899.

(10) Wadas, T. J.; Wong, E. H.; Weisman, G. R.; Anderson, C. J. *Curr. Pharm. Des.* **2007**, *13*, 3−16.

(11) Ma, M. T.; Donnelly, P. S. *Curr. Top. Med. Chem.* **2011**, *11*, 500−520.

(12) (a) Stigers, D. J.; Ferdani, R.; Weisman, G. R.; Wong, E. H.; Anderson, C. J.; Golen, J. A.; Moore, C.; Rheingold, A. L. *Dalton Trans.* **2010**, *39*, 1699−1701. (b) Ferdani, R.; Stigers, D. J.; Fiamengo, A. L.; Wei, L.; Li, B. T. Y.; Golen, J. A.; Rheingold, A. L.; Weisman, G. R.; Wong, E. H.; Anderson, C. J. *Dalton Trans.* **2012**, *41*, 1938−1950. (c) Guo, Y.; Ferdani, R.; Anderson, C. J. *Bioconjugate Chem.* **2012**, *23*, 1470−1477. (d) Jiang, M.; Ferdani, R.; Shokeen, M.; Anderson, C. J. *Nucl. Med. Biol.* **2013**, *40*, 245−251. (e) Zeng, D.; Ouyang, Q.; Cai, Z.; Xie, X.-Q.; Anderson, C. J. *Chem. Commun.* **2014**, *50*, 43−45. (f) Cai, Z.; Ouyang, Q.; Zeng, D.; Nguyen, K. N.; Modi, J.; Wang, L.; White, A. G.; Rogers, B. E.; Xie, X.-Q.; Anderson, C. J. *J. Med. Chem.* **2014**, *57*, 6019−6029. (g) Cai, Z.; Li, B. T. Y.; Wong, E. H.; Weisman, G. R.; Anderson, C. J. *Dalton Trans.* **2015**, *44*, 3945−3948.

(13) (a) Di Bartolo, N. M.; Sargeson, A. M.; Donlevy, T. M.; Smith, S. V. *J. Chem. Soc., Dalton Trans.* **2001**, 2303−2309. (b) Ma, M. T.; Cooper, M. S.; Paul, R. L.; Shaw, K. P.; Karas, J. A.; Scanlon, D.; White, J. M.; Blower, P. J.; Donnelly, P. S. *Inorg. Chem.* **2011**, *50*, 6701−6710. (c) Alt, K.; Paterson, B. M.; Ardipradja, K.; Schieber, C.; Buncic, G.; Lim, B.; Poniger, S. S.; Jakoby, B.; Wang, X.; O'Keefe, G. J.; Tochon-Danguy, H. J.; Scott, A. M.; Ackermann, U.; Peter, K.; Donnelly, P. S.; Hagemeyer, C. E. *Mol. Pharmaceutics* **2014**, *11*, 2855−2863.

(14) (a) Prasanphanich, A. F.; Nanda, P. K.; Rold, T. L.; Ma, L.; Lewis, M. R.; Garrison, J. C.; Hoffman, T. J.; Sieckman, G. L.; Figueroa, S. D.; Smith, C. J. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 12462−12467. (b) Cooper, M. S.; Ma, M. T.; Sunassee, K.; Shaw, K. P.; Williams, J. D.; Paul, R. L.; Donnelly, P. S.; Blower, P. J. *Bioconjugate Chem.* **2012**, *23*, 1029−1039. (c) Zhang, Y.; Hong, H.; Engle, J. W.; Bean, J.; Yang, Y.; Leigh, B. R.; Barnhart, T. E.; Cai, W. *PLoS One* **2011**, *6*, e28005.

(15) (a) Pandya, D. N.; Kim, J. Y.; Park, J. C.; Lee, H.; Phapale, P. B.; Kwak, W.; Choi, T. H.; Cheon, G. J.; Yoon, Y.-R.; Yoo, J. *Chem. Commun.* **2010**, *46*, 3517−3519. (b) Pandya, D. N.; Kim, J. Y.; Kwak, W.; Park, J. C.; Gawande, M. B.; An, G. I.; Ryu, E. K.; Yoo, J. *Nucl. Med. Mol. Imaging* **2010**, *44*, 185−192. (c) Pandya, D. N.; Bhatt, N.; Dale, A. V.; Kim, J. Y.; Lee, H.; Ha, Y. S.; Lee, J.-E.; An, G. I.; Yoo, J. *Bioconjugate Chem.* **2013**, *24*, 1356−1366. (d) Dale, A. V.; Pandya, D. N.; Kim, J. Y.; Lee, H.; Ha, Y. S.; Bhatt, N.; Kim, J.; Seo, J. J.; Lee, W.; Kim, S. H.; Yoon, Y.-R; An, G. I.; Yoo, J. *ACS Med. Chem. Lett.* **2013**, *4*, 927−931.

(16) Lima, L. M. P.; Esteban-Gómez, D.; Delgado, R.; Platas-Iglesias, C.; Tripier, R. *Inorg. Chem.* **2012**, *51*, 6916−6927.

(17) Kotek, J.; Lubal, P.; Hermann, P.; Císařová, I.; Lukeš, I.; Godula, T.; Svobodová, I.; Táborský, P.; Havel, J. *Chem. - Eur. J.* **2003**, *9*, 233−248.

(18) Svobodová, I.; Havlíčková, J.; Plutnar, J.; Lubal, P.; Kotek, J.; Hermann, P. *Eur. J. Inorg. Chem.* **2009**, *24*, 3577−3592.

(19) Paúrová, M.; Havlíčková, J.; Pospíšilová, A.; Vetrík, M.; Císařová, I.; Stephan, H.; Pietzsch, H.-J.; Hrubý, M.; Hermann, P.; Kotek, J. *Chem. - Eur. J.* **2015**, *21*, 4671−4687.

(20) Füzerová, S.; Kotek, J.; Císařová, I.; Hermann, P.; Binnemans, K.; Lukeš, I. *Dalton Trans.* **2005**, 2908−2915.

(21) Svobodová, I.; Lubal, P.; Plutnar, J.; Kotek, J.; Havlíčková, J.; Hermann, P.; Lukeš, I. *Dalton Trans.* **2006**, 5184−5197.

(22) Havlíčková, J.; Medová, H.; Vitha, T.; Kotek, J.; Císařová, I.; Hermann, P. *Dalton Trans.* **2008**, 5378−5386.

(23) (a) Notni, J.; Hermann, P.; Havlíčková, J.; Kotek, J.; Kubíček, V.; Plutnar, J.; Loktionova, N. S.; Riss, P. J.; Rösch, F.; Lukeš, I. *Chem. - Eur. J.* **2010**, *16*, 7174−7185. (b) Notni, J.; Šimeček, J.; Hermann, P.; Pohle, K.; Wester, H.-J. *Chem. - Eur. J.* **2011**, *17*, 14718−14722. (c) Šimeček, J.; Schulz, M.; Notni, J.; Plutnar, J.; Kubíček, V.; Havlíčková, J.; Hermann, P. *Inorg. Chem.* **2012**, *51*, 577−590. (d) Šimeček, J.; Hermann, P.; Wester, H.-J.; Notni, J. *ChemMedChem* **2013**, *8*, 95−103. (e) Máté, G.; Šimeček, J.; Pniok, M.; Kertész, I.; Notni, J.; Wester, H.-J.; Galuska, L.; Hermann, P. *Molecules* **2015**, *20*, 13112−13126.

(24) Notni, J.; Šimeček, J.; Wester, H.-J. *ChemMedChem* **2014**, *9*, 1107−1115; *ChemMedChem* **2014**, *9*, 2614.

(25) (a) David, T.; Procházková, S.; Havlíčková, J.; Kotek, J.; Kubíček, V.; Hermann, P.; Lukeš, I. *Dalton Trans.* **2013**, *42*, 2414−2422. (b) David, T.; Procházková, S.; Kotek, J.; Kubíček, V.; Hermann, P.; Lukeš, I. *Eur. J. Inorg. Chem.* **2014**, *2014*, 4357−4368.

(26) Motekaitis, R. J.; Murase, I.; Martell, A. E. *Inorg. Nucl. Chem. Lett.* **1971**, *7*, 1103−1107.

(27) Kubíček, V.; Kotek, J.; Hermann, P.; Lukeš, I. *Eur. J. Inorg. Chem.* **2007**, *2007*, 333−344.

(28) (a) Wu, S.-L.; Horrocks, W. DeW. *Inorg. Chem.* **1995**, *34*, 3724−3732. (b) Moreau, J.; Guillon, E.; Pierrard, J.-C.; Rimbault, J.; Port, M.; Aplincourt, M. *Chem. - Eur. J.* **2004**, *10*, 5218−5232.

(29) Brücher, E.; Tircsó, G.; Baranyai, Z.; Kovács, Z.; Sherry, A. D. In *The Chemistry of Contrast Agents in Medical Magnetic Resonance Imaging*, 2nd ed.; Merbach, A., Helm, L., Tóth, É., Eds.; Wiley: Chichester, U.K., 2013; pp 157−208.

(30) Atkins, P. *Physical Chemistry*, 7th ed.; Oxford University Press: Oxford, U.K., 2002; pp 876−877.

(31) Lubal, P.; Kývala, M.; Hermann, P.; Holubová, J.; Rohovec, J.; Havel, J.; Lukeš, I. *Polyhedron* **2001**, *20*, 47−55.

(32) Blahut, J.; Císařová, I.; Hermann, P.; Kotek, J. Submitted.

(33) (a) Janssen, C. L.; Nielsen, I. M. B. *Chem. Phys. Lett.* **1998**, *290*, 423−430. (b) Leininger, M. L.; Nielsen, I. M. B.; Crawford, T. D.; Janssen, C. L. *Chem. Phys. Lett.* **2000**, *328*, 431−436.

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

(35) David, T.; Křečková, P.; Kotek, J.; Kubíček, V.; Lukeš, I. *Heteroat. Chem.* **2012**, *23*, 195−201.

(36) Bosnich, B.; Poon, C. K.; Tobe, M. *Inorg. Chem.* **1965**, *4*, 1102−1108.

(37) Meyer, M.; Dahaoui-Gindrey, V.; Lecomte, C.; Guilard, R. *Coord. Chem. Rev.* **1998**, *178−180*, 1313−1405.

(38) Liang, B.-F.; Chung, C.-S. *J. Chem. Soc., Dalton Trans.* **1980**, 1349−1351.

(39) Silversides, J. D.; Allan, C. C.; Archibald, S. J. *Dalton Trans.* **2007**, 971−978.

(40) Panneerselvam, K.; Lu, T.-H.; Chi, T.-Y.; Chung, C.-S.; Chen, Y.-J.; Kwan, K.-S. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **1998**, *54*, 25−27.

(41) El Ghachtouli, S.; Cadiou, C.; Déchamps-Olivier, I.; Chuburu, F.; Aplincourt, M.; Roisnel, T. *Eur. J. Inorg. Chem.* **2006**, *2006*, 3472−3481.

(42) Studer, M.; Kaden, T. A. *Helv. Chim. Acta* **1986**, *69*, 2081−2086.

(43) Lukeš, I.; Kotek, J.; Vojtíšek, P.; Hermann, P. *Coord. Chem. Rev.* **2001**, *216-217*, 287−312.

(44) Hancock, R. D.; Motekaitis, R. J.; Mashishi, J.; Cukrowski, I.; Reibenspies, J. H.; Martell, A. E. *J. Chem. Soc., Perkin Trans. 2* **1996**, 1925−1929.

(45) (a) Kotková, Z.; Pereira, G. A.; Djanashvili, K.; Kotek, J.; Rudovský, J.; Hermann, P.; Vander Elst, L.; Muller, R. N.; Geraldes, C. F. G. C.; Lukeš, I.; Peters, J. A. *Eur. J. Inorg. Chem.* **2009**, *2009*, 119−136. (b) Kubíček, V.; Havlíčková, J.; Kotek, J.; Tircsó, G.; Hermann, P.; Tóth, É.; Lukeš, I. *Inorg. Chem.* **2010**, *49*, 10960−10969.

(46) Siegfried, L.; Honecker, M.; Schlageter, A.; Kaden, T. A. *Dalton Trans.* **2003**, 3939−3948.

(47) Silvestre, J.-P.; Dao, N. Q.; Leroux, Y. *Heteroat. Chem.* **2001**, *12*, 73−89.

(48) Matczak-Jon, E.; Videnova-Adrabinska, V. *Coord. Chem. Rev.* **2005**, *249*, 2458−2488.

(49) (a) Wozniak, M.; Tridot, G.; Nicole, J. *Bull. Soc. Chim. France* **1972**, 1145−1152. (b) Afonin, E. G.; Aleksandrov, G. G.; Sergienko, V. S. *Russ. J. Coord. Chem.* **1997**, *23*, 794−800. (c) Sergienko, V. S.; Aleksandrov, G. G.; Afonin, E. G. *Russ. J. Coord. Chem.* **1998**, *24*, 429−434. (d) Afonin, E. G.; Aleksandrov, G. G.; Sergienko, V. S. *Russ. J. Coord. Chem.* **1999**, *25*, 863−869.

(50) Bakaj, M.; Zimmer, M. *J. Mol. Struct.* **1999**, *508*, 59−72.

(51) Liang, X.; Sadler, P. J. *Chem. Soc. Rev.* **2004**, *33*, 246−266.

(52) (a) Clay, R.; Murray-Rust, J.; Murray-Rust, P. *J. Chem. Soc., Dalton Trans.* **1979**, 1135−1139. (b) Bernhardt, P. V.; Sharpe, P. C. *Chem. Commun.* **1996**, 1267−1268.

(53) Hart, S. L.; Haines, R. I.; Decken, A.; Wagner, B. D. *Inorg. Chim. Acta* **2009**, *362*, 4145−4151.

(54) (a) Liang, B.-F.; Margerum, D. W.; Chung, C.-S. *Inorg. Chem.* **1979**, *18*, 2001−2007. (b) Liang, B.-F.; Chung, C.-S. *Inorg. Chem.* **1980**, *19*, 1867−1871. (c) Lee, C.-S.; Chung, C.-S. *Inorg. Chem.* **1984**, *23*, 639−644.

(55) (a) Connolly, P. J.; Billo, E. J. *Inorg. Chem.* **1987**, *26*, 3224−3227. (b) Liang, X.; Weishäupl, M.; Parkinson, J. A.; Parsons, S.; McGregor, P. A.; Sadler, P. J. *Chem. - Eur. J.* **2003**, *9*, 4709−4717. (c) Hunter, T. M.; McNae, I. W.; Simpson, D. P.; Smith, A. M.; Moggach, S.; White, F.; Walkinshaw, M. D.; Parsons, S.; Sadler, P. J. *Chem. - Eur. J.* **2007**, *13*, 40−50.

(56) Woodin, K. S.; Heroux, K. J.; Boswell, C. A.; Wong, E. H.; Weisman, G. R.; Niu, W. J.; Tomellini, S. A.; Anderson, C. J.; Zakharov, L. N.; Rheingold, A. L. *Eur. J. Inorg. Chem.* **2005**, *2005*, 4829−4833.

(57) Vitha, T.; Kubíček, V.; Hermann, P.; Kolar, Z. I.; Wolterbeek, H. T.; Peters, J. A.; Lukeš, I. *Langmuir* **2008**, *24*, 1952−1958.

(58) King, C.; Roundhill, D. M.; Fronczek, F. R. *Inorg. Chem.* **1986**, *25*, 1290−1292.

(59) Thieme, S.; Walther, M.; Pietzsch, H.-J.; Henniger, J.; Preusche, S.; Mading, P.; Steinbach, J. *Appl. Radiat. Isot.* **2012**, *70*, 602−608.

(60) *TURBOMOLE 6.5*; University of Karlsruhe and Forschungszentrum Karlsruhe GmbH: Karlsruhe, Germany2013; available from http://www.turbomole.com/.

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

(62) (a) Becke, A. D. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098−3100. (b) Perdew, J. P. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1986**, *33*, 8822−8824.

(63) (a) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829−5835. (b) Eichkorn, K.; Weigend, F.; Treutler, O.; Ahlrichs, R. *Theor. Chem. Acc.* **1997**, *97*, 119−124.

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

(65) (a) Klamt, A. *J. Phys. Chem.* **1995**, *99*, 2224−2235. (b) Klamt, A.; Jonas, V.; Buerger, T.; Lohrenz, J. C. W. *J. Phys. Chem. A* **1998**, *102*, 5074−5085.

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

(67) (a) Perdew, J. P.; Wang, Y. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1992**, *45*, 13244. (b) Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

(68) (a) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305. (b) Weigend, F. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1057−1065.

PAPER

CrossMark
click for updates

How simple is too simple? Computational perspective on importance of second-shell environment for metal-ion selectivity†

Ondrej Gutten and Lubomír Rulíšek*

The metal-ion selectivity in biomolecules represents one of the most important phenomena in bioinorganic chemistry. The open question to what extent is the selectivity in the complex bioinorganic structures such as metallopeptides determined by the first-shell ligands of the metal ion is answered herein using six model peptides complexed with the set of divalent metal ions (Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, and Hg^{2+}) and their various first-shell representations. By calculating the differences among the free energies of complexation of metal ions in these peptides and their model (truncated) systems it is quantitatively shown that the definition of the first shell is paramount to this discussion and revolves around the chemical nature of the binding site. Despite the vast conceivable diversity of peptidic structures, that suggest certain fluidity of this definition, major contributing factors are identified and assessed based on their importance for capturing metal-ion selectivity. These factors include soft/hard character of ligands and various non-covalent interactions in the vicinity of the binding site. The relative importance of these factors is considered and specific suggestions for effective construction of the models are made. The relationship of first-shell models and their corresponding parent peptides is discussed thoroughly, both with respect to their chemical similarity and potential disparity introduced by generally "non-alignable" conformational flexibility of the two systems. It is concluded that, in special cases, this disparity can be negligible and that heeding the chemical factors contributing to selectivity during construction of the model can successfully result in models that retain the affinity profile for various metal ions with high fidelity.

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. E-mail: rulisek@uochb.cas.cz; Fax: +420-220-183-578; Tel: +420-220-183-263

1. Introduction

Computational modeling represents an indispensable tool in discovering fundamental physico-chemical principles underlying chemical and biochemical processes.[1] One of the important biological phenomena is an uptake and binding of metal ions in biomolecules.[2] Since various metal ions play various roles in biological machinery, Nature has to fine-tune the selectivity of metal-binding sites present in proteins and RNA/DNA for the specific metal ion.[3] Therefore, deciphering the mechanisms and factors behind the metal ion selectivity[4–6] is a highly desirable task which may ultimately lead to answering the fundamental question 'Why Nature selected specific metal ions for performing specific tasks?'[7]

Most of the experimental and computational findings in the area of metal-ion selectivity have been very recently reviewed by Dudev and Lim.[6] As highlighted in their review, there are two 'external' factors (*i.e.*, independent of the constitution of particular metal-binding site) one always needs to take into account in any considerations of the metal-ion selectivity: (i) average concentrations of metal ions in intracellular and extracellular fluids,[8] and (ii) the inherent 'binding properties' of metal ions. Concerning the former, it can be reminded that the concentrations of unbound metal ions in cytosol range from millimolar (Na$^+$, K$^+$, Mg^{2+}) through micro- (Mn^{2+}, Fe^{2+}, Ca^{2+}) nano- (Co^{2+}, Ni^{2+}) to femto- (Zn^{2+}) and attomolar (Cu$^+$/Cu^{2+}).[9] The latter external factor is mostly exemplified by the Irving–Williams series[10] which qualitatively ranks the stabilities of complexes formed by divalent metal ions. This series has its physico-chemical origin in the second ionization enthalpies of the metals[11] and predicts the following order:[12]

$$Mg^{2+} < Mn^{2+} < Fe^{2+} < Co^{2+} < Ni^{2+} < Cu^{2+} > Zn^{2+}.$$
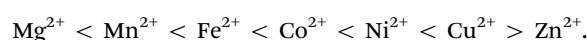
Thus, copper(II) in general forms the most stable complexes with a 'generic' set of ligands, followed by Zn^{2+}/Ni^{2+} whereas complexes of Mg^{2+} are expected to have the lowest stability

This journal is © the Owner Societies 2015

Phys. Chem. Chem. Phys., 2015, **17**, 14393–14404 | **14393**

constants (β). In tuning the metal-ion selectivity, one has to consider the Irving–Williams series as a 'baseline' with respect to which the modifications in the computed or measured metal-ion stabilities are evaluated.

Last, but not least authors of the title review[6] clearly summarize most of the areas where metal-ion selectivity is of key importance, notably in function of sodium/potassium/calcium ion channels[13–15] and in structure–function of almost all metalloproteins (both regulatory and enzymatic).[16–18]

Over the last two decades, many studies were reported that addressed the problem of metal-ion selectivity from a computational and quantum chemical perspective.[19–28] These most often involved quantum chemical calculations of the small model complexes, both *in vacuo* and polarized dielectric continuum (to address the effects of the environment, such as solution or protein). Quite often, the results were qualitatively correlated with the experimentally determined stability constants[21] or phenomenological information obtained from the abundance of metal ions in the sites of metalloproteins.[29–31] In a recent study,[32] an attempt was made to conceive a robust and accurate computational protocol that would yield stability constants (β) of selected metal ions (Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, and Hg^{2+}) in small model complexes. The protocol was benchmarked on the series of complexes with the known experimental values of β. It was concluded that current computational approaches are likely to suffer from both metal-dependent and ligand-dependent systematic shifts and the straightforward 'ab initio' predictions of the 'absolute' values of these thermodynamic properties are likely still beyond their grasp. At the same time, it was demonstrated that a relatively easy procedure can be followed that partially accounts for these systematic shifts and the metal-ion selectivity for a particular model site can be in many cases predicted to 1–2 kcal mol^{-1} accuracy.[32]

Thus, under an optimistic assumption that computational issues pertinent to the binding of metal ions in small complexes (*e.g.* metal-binding sites represented by the first-shell ligands) were at least partially solved, a question emanates to what extent are the second-shell effects (which may also include water mediated binding of certain ions to metal binding sites or ion channels) important for the metal-ion selectivity of the site. Some of these issues were addressed in computational studies of ion channels[33,34] where second-shell residues are expected to have stabilizing effect on certain type of coordination geometries which, in turn, favor the binding of specific metal ion (and not of its counterpart). In an analogous way, the computations together with statistical survey in Protein Data Bank were used to correlate structure and composition of the outer coordination sphere of metal sites in metalloproteins with those of the inner sphere.[35–37] Despite these achievements, the comprehensive study that may yield the robust computational protocol to treat these effects rigorously and quantitatively, is lacking.

This immediately leads to the central question addressed in this work: "How much of the metal-ion selectivity is captured by the first-shell ligand residues of a particular site in metalloprotein?" In answering such a question, one needs to quantitatively assess the stability constants of metal ions in peptidic scaffolds from the *first principles*. To accomplish such a computational task involves exploring the limits of contemporary quantum chemistry and computational modeling, notably accurate calculations of solvation energies of charged systems.

While the ultimate (and perhaps still distant) goal of this and related studies is the quantitative prediction of stability constants of metal ions in the peptide structures with an accuracy of 1–2 pK (log β) units, we expect that several important questions can (and will) be raised and answered on the way towards its accomplishment. As mentioned above, these include (i) the suitability of the first-shell-only representation (that is of the model complex $[M(Y_i)_n]^{c+}$ where Y_i are truncated metal binding amino acids) of the whole metal-binding peptide in calculations of complexation energies and stability constants with special attention paid to their relative values (which, in turn, determine the metal-ion selectivity) and (ii) justifiability of various constraints during the geometry optimization to preserve the original coordination environment in the whole [metal + peptide] complex. The positive answer to the first question would not only provide us with an exciting opportunity to quickly scan various metal sites (both catalytic and functional) for their inherent metal-binding properties and shed further insight into the role of metal ions in biomolecules, but also open new avenues in bottom-up approach for design of novel specific metal-binding sites.

Six model systems studied in this work, schematically depicted in Fig. 1, and described in more details in Section 2.2., are considered. They were carefully selected to represent the considerable part of the 'spectrum' of the experimentally observed binding modes. Two of them represent *in silico* designed metal-binders that were tested experimentally both in gas-phase and solution.[39] Unfortunately, no experimental structural information exists on the nature of the binding of metal ions by these peptides. The other four systems represent the continuous metal-binding amino acid sequences in the cores of selected metalloproteins for which the crystal structure was available (*cf.*, Section 2.2.).

The series of metal ions with the potential of metal-binding by the peptides includes Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, and Hg^{2+} ions as they represent the most common divalent ions and coincides with the selection in our previous work.[32]

2. Computational details and methodological issues

2.1. Quantum chemical methods

All calculations reported in this work were performed using the TURBOMOLE 6.5 program. The quantum chemical calculations were performed using the density functional theory (DFT). The geometry optimizations were carried out at the DFT level, employing the density-fitted (*vide infra*) BP86 functional (RI-BP86)[40,41] in combination with def-TZVP basis set.[42,43] In case of peptide systems (*vide infra*) only the SV basis set[44] was used for atoms other than the metal ion (for which def-TZVP was used) in geometry optimizations whereas the single point calculations of peptide systems were also carried out using the def-TZVP basis set on all atoms. The effect of solvent (water)
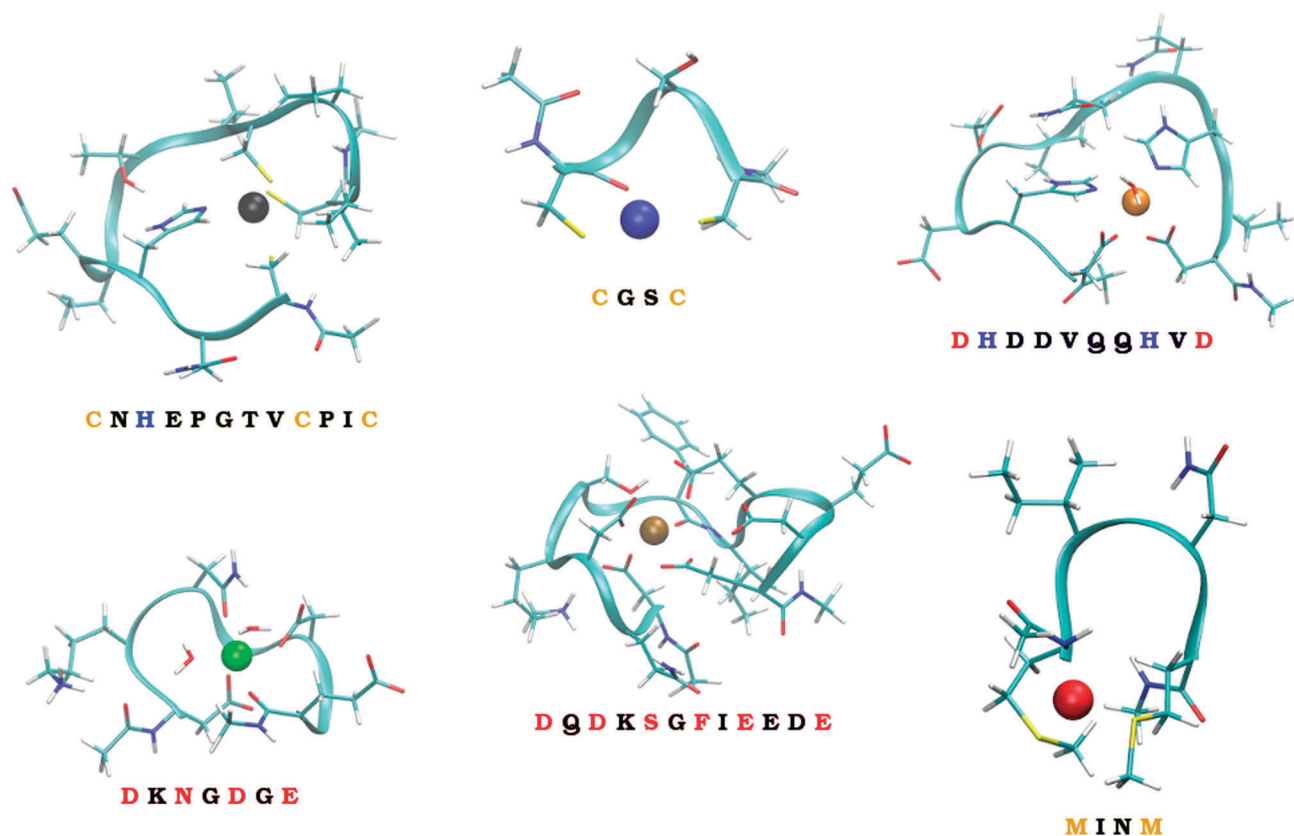
14394 | *Phys. Chem. Chem. Phys.*, 2015, **17**, 14393–14404

This journal is © the Owner Societies 2015

Fig. 1 The six studied peptides. The peptides are abbreviated by the amino–acid residues participating in binding of a metal ion. The larger peptides (**CHCC**, **DHHD**, **DNDO**, **DDSOEE**) are based on structures found in MESPEUS database.[38] The smaller two peptides (**CC**, **MM**) come from previous research in our laboratory.[39] Vid Section 2.2. for details.

environment was modeled by COSMO method (the atomic radii of 2.0 Å for Mn, Fe, Co, Ni and Zn, 2.2 Å for Cd, 2.4 Å for Hg and the defaults for the rest, and the dielectric constant parameter $\varepsilon = 80.0$). The single-point energies were calculated using the same protocol, with dielectric constant of $\varepsilon = \infty$ (ideal conductor) or $\varepsilon = 1$ (vacuum). All calculations were expedited by expanding the Coulomb integrals in an auxiliary basis set, using the resolution-of-identity (RI) approximation (density fitting).[45] Grimme's D3 dispersion has been applied.[46]

All metal-ions (Mn, Fe, Co, Ni, Cu, Zn, Cd, Hg) were considered in their +2 oxidation state. For open-shell metal ions in the series (Mn–Cu), high-spin states were considered, in line with our previous work.[26,32] These are assumed to be ground electronic states for all of the studied complexes containing mostly low-field ligands. Stuttgart/Dresden pseudopotentials for Cd and Hg ions were applied.[47] We consider it as a plausible approximation to relativistic effects, since non-scalar relativistic effects (such as spin–orbit coupling) are estimated to play negligible role for the reaction energetics. The use of ECPs is also in line with our previous benchmark studies.[26,32]

Free energies of interaction with solvent of all studied species were calculated using the COSMO-RS method[48,49] (conductor-like screening model for realistic solvation) as implemented in the COSMO*therm* program,[50] using the ''BP_TZVP_C30_1201.ctd'' parametrization file.

2.2. Studied systems

To quantify how much of the selectivity is lost upon reducing a system to its model we have chosen a number of peptides, ranging in size and coordination shell of the metal ion. Out of the total of 6 peptides, four were inspired by PDB structures, while the remaining two were obtained from previous theoretical studies.

The selected PDB structures were required to possess a binding site for a metal ion with all of the binding partners within a relatively short sequence of amino acids. The structures were identified using the information contained in the MESPEUS database.[38] The whole metalloprotein (PDB) structure was reduced (truncated) to the minimalistic metal-binding continuous sequence and metal-bound water molecules; terminated with acetyl group at the N-terminus and *N*-methyl at the C-terminus. Namely, the four peptide sequences considered were **CNHEPGTVCPIC** (PDB code: 1G71; referred to as **CHCC** according to the metal-binding residues), **DQDKSGFIEEDE** (PDB code: 2PAL, referred to as **DDSOEE**, O standing for backbone carbonyl), **DKNGDGE** (PDB code: 1IGV; includes 2 water molecules; referred to as **DNDO**), **DH**DDVQQHVD (PDB code: 1B9M; includes 1 water molecule; referred to as **DHHD**).

The remaining two peptides were **CGSC** (referred to as **CC**) and **M**INM (referred to as **MM**). The binding modes of these two peptides are merely putative. Nevertheless, these two peptides were previously synthesized and binding of ions in the gas phase

experimentally determined by mass spectrometry.[39] Unfortunately, neither the structural information, nor the solution thermo-dynamics was obtained for these two systems.

Metal-binding cysteine side chains of **CHCC** and **CC** peptides (which were the only Cys residues on both peptides), as well as all aspartate and glutamate side-chains of **CHCC**, **DHHD**, **DNDO** and **DDSOEE** peptides, were considered in deprotonated form. The studied systems are depicted in Fig. 1. Each system was optimized with each of the eight metal ions, resulting in 6 (peptides) × 8 (metal ions) = 48 structures.
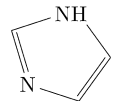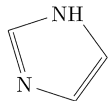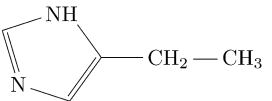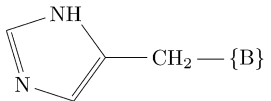
2.3. First-shell models of peptides

A total of four different first-shell models were constructed. In all cases, the truncated fragment was terminated with hydrogen atom(s). The simplest model, referred to as *TINY*, consists of ligands represented by the smallest possible functional fragment,

ranging from (HS$^-$) as a model for deprotonated cysteine side-chain, to imidazole as a model for histidine side-chain. The second model, referred to as *SMALL*, differs from the *TINY* by an addition of a methyl residue (along the truncated side chain). The models denoted as *ALPHA* were then truncated at the C$_\alpha$ atoms of the amino acid (*i.e.* side chains were capped by the $-$C$_\alpha$H$_3$ group). Finally, the *FULL_AA* model contains the full amino acid residue with both of the peptide bonds included (capped by H atoms on both N- and C-'terminus'). The models are presented in Scheme 1.

In general, only the substituted hydrogens were optimized in these first-shell models, whereas the remaining atoms were kept frozen in the Cartesian coordinates of their parent system.

The models, by definition, lack the interaction with distant groups and their conformational freedom is unrelated to that of the parent system (see Section 2.4.2). These features are essential

Scheme 1   Ligands and their representation in individual models. {B} signifies attachment of the peptide backbone, *i.e.*
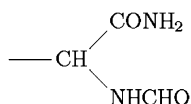
for the models to serve their primary purpose – a comprehensive tool for isolating and studying metal-specific interactions of the binding site and the metal ion, and identifying their features and behavior in various settings. The models can be utilized in other ways too, although one has to keep in mind the limitations these features imply.

Firstly, the models can be used in the 'top-down' approach, where they can serve as small and cheap models for the selectivity of a specific conformation of the parent system. If more than a single conformation of the binding site is relevant, a single calculation of the parent system for each conformation can be complemented with the model systems for each of the metal ions. This is cheaper than full scale calculation for each metal ion in each conformation.

Alternatively, in the 'bottom-up' approach, the models can be used as templates for constructing (e.g. peptidic) scaffolds that can support the binding geometry. Since the model is constructed in a way that it contains almost all selective interactions, the resulting system (e.g. polypeptide chain with the metal-binding site) may retain the selectivity properties identical to those of the model. Rigid scaffolds can also eliminate the need for conformational sampling.

2.4. Studied physico-chemical quantities and the computational setup

2.4.1 Gibbs free energy.
The aim of this study is to gain insight into factors contributing to diverse affinity of peptidic systems for various divalent metal ions. This leads to adapting the definition of Gibbs free energy (sometimes denoted free enthalpy) of a system with metal ion M and set of ligands $\{L_i\} \equiv \mathbf{L}$ as:

$$G_{M,L} = E_{el} + G_{IS} \tag{1}$$

where E_{el} stands for single-point gas-phase electronic energy of solvent-optimized structure and G_{IS} stands for free energy of interaction with solvent (i.e. solvation free energy without the free energy of structure relaxation).

The definition deliberately ignores zero-point vibrational energy and thermal corrections to vibrational, translational and rotational partition function. As discussed thoroughly in previous work,[32] these two terms present a non-trivial technical challenge – stemming from finding minima for large (and ideally also solvated) systems, and differences between solvent and gas-phase structures. As these are not expected to be heavily metal-dependent, we preferred to use eqn (1) as the practical (albeit not theoretically pure) approach and plausible approximation to the free energy of the complex. In our previous work, this simple approach somewhat surprisingly yielded better relative stability constants (differences in experimental measure of binding/complexation Gibbs free energies) for a diverse set of model complexes (8 metal ions + 6 sets of ligands) when compared to experimental data.[32]

In discussing metal-ion selectivity, it is usually advantageous to consider relative values of quantities, rather than their absolute values. These can be brought into spotlight by shifting the quantity equally for all metal ions – an operation that does not change the differences among the values of a quantity.

Where applicable, we use quantities shifted in this fashion and indicate the fact by "REL" superscript:

$$X_{M,L}^{REL} = X_{M,L} - \max_N\{X_{N,L}\} \tag{2}$$

i.e. $X_{M,L}^{REL}$ is zero for the maximum element of the $\{X_{M,L}\}$ set and negative for the rest.

2.4.2 Conformational entropy.
Another important issue to be decided is the treatment of conformational sampling. In our opinion, there are two main reasons for *not* employing any conformational sampling in our study. Firstly, while the protocol has been shown to be quite robust for calculation of relative free energies of binding, it shows significant systematic errors when applied to calculation of absolute values of the quantity. Secondly, the quest lies in identifying parts of the peptides that contribute to selectivity. The utility of the models presented in Section 2.3. is to provide relationship between metal-ion selectivity and structure of the system, not to contain the information about the conformational freedom of their parent system, which is not possible for this kind of model even in principle. In other words, the free energy of binding is a Boltzmann-weighted average of ensemble of peptide structures; the models can retain a major part of selectivity of individual members of the ensemble, but are principally unable to provide the Boltzmann weights. Thus, even an ideal model would possess the same metal-ion selectivity only in case where these weights are identical (e.g. a perfectly rigid parent peptide); i.e. would reproduce the metal-ion selectivity only if the Boltzmann weights for individual conformations are provided.

This realization is in no conflict with the intended purpose of the study – quite the contrary, it allows isolating factors which can influence selectivity but are not part of the model. Consider two cases of a peptide binding a different metal-ion in each case. These two systems differ not only in a metal ion bound, but also in geometry of the peptide/ligands. Comparing such two systems can be viewed as a 2-step transition process: (I) changing a metal ion and (II) relaxing a structure (Fig. 2).

The most important difference between a peptide and its model is the presence of the 2nd shell (i.e. the part of a peptide not included in the model) in the former system. Thus, the overall difference in selectivity of a peptide and its model will be determined by the interaction of the 2nd shell with a metal ion (and the 1st shell) during both of the substeps – giving rise to two contributions: a metal-induced selectivity (step I in Fig. 2) and a structure-induced selectivity (step II in Fig. 2). It is instructive to examine these two contributions separately.

Metal-induced selectivity. The interaction of an ion with 2nd shell can be significant. However, minimizing 2nd-shell metal-induced selectivity (step I in Fig. 2) requires merely absence of metal-specific interactions with the 2nd shell, which can be achieved by proper design of a model. This contribution can be studied by examining unrelaxed systems with an exchanged metal-ion ("hybrid" in Fig. 2) and is the main focus of Section 3.2. The intermediate "hybrid" system is not a minimum on a potential energy surface, since the system with an exchanged metal ion is deliberately *not* optimized. This setting is pivotal
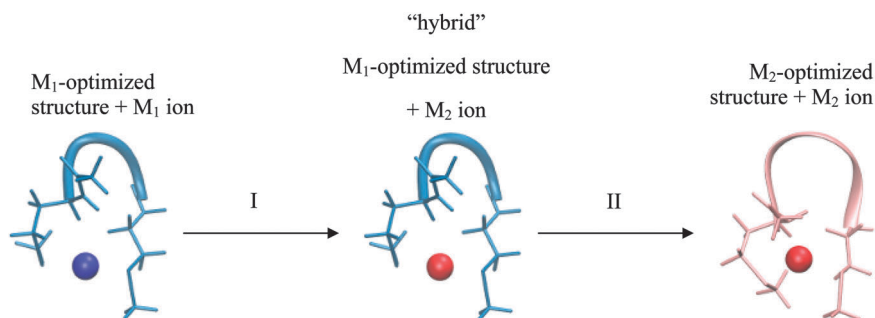
"hybrid"

M$_1$-optimized   M$_1$-optimized structure   M$_2$-optimized
structure + M$_1$ ion   + M$_2$ ion   structure + M$_2$ ion

I   II

Fig. 2 Contributions to selectivity. Free energy difference between metal-binding systems can be decomposed into two steps: (I) changing a metal ion and (II) changing the geometry. These can be examined separately, as free energy is a state function.

for eliminating the structural effects, which allows for in-depth analysis of selectivity relationships. Quantities pertaining to such systems are indicated with "hybrid" superscript.

Structure-induced selectivity. Magnitude of the 2nd-shell structure-induced selectivity (step II in Fig. 2) is more dependent on the actual peptide rather than on the nature of the model, and is strictly zero only for perfectly rigid peptides. Thus, rather than asking what is the impact of different free energies of relaxation on selectivity in case of (arbitrary) conformations of the studied systems, we seek to gain insight into what is the bottom limit of this influence. The issue is discussed in Section 3.3.

Hybrid systems. Each of the 48 structures (see Section 2.2.) was used for construction and calculation of "hybrid" systems. This resulted in 48 × 8 (each ion is substituted by each of the 7 remaining ones) × 5 (representations, *i.e.* original peptide, *TINY*, *SMALL*, *ALPHA*, and *FULL_AA* models) = 1920 systems for the discussion of metal-induced selectivity (see Section 3.2.).

3. Results and discussion

3.1. Identifying relevant and descriptive physico-chemical quantities

The Gibbs free energy of a system defined in eqn (1) is the primary thermodynamic quantity obtained from the calculations, however, it is not a convenient one to discuss retention of selectivity properties of the model systems.

Rather than using *e.g.* hexahydrated metal ion as reference and examining the parent system and the model separately *i.e.* $(G_{M,L,peptide} - G_{M,(H_2O)_6})^{REL}$ and $(G_{M,L,model} - G_{M,(H_2O)_6})^{REL}$, it is beneficial to consider the difference of these two, *i.e.*

$$G_{M,L,model}^{2nd} = (G_{M,L,peptide} - G_{M,L,model})^{REL} \qquad (3)$$

Using the parent system as a reference state, this quantity directly describes the interaction of a metal ion with the 2nd shell. In an ideal case, the model would capture all of the selectivity of a peptide, resulting in this quantity being invariantly zero for all metal ions (M) and systems (L).

In reality, this will not be the case, and the variation of this quantity, examined as average absolute deviation and maximum absolute deviation (eqn (4)), will determine the faithfulness of a model. As will turn out later it might be advantageous that the average is taken over all metal ions (M) but not over different systems (L), as the nature of the systems and even specific structural details will prove to be relevant for the discussion.

$$AAD_{L,model} = \overline{\left| G_{M,L,model}^{2nd} - \overline{G_{M,L,model}^{2nd}} \right|} \qquad (4)$$

$$MAD_{L,model} = \max\left\{ \left| G_{M,L,model}^{2nd} - \overline{G_{M,L,model}^{2nd}} \right| \right\}_M.$$

Using quantities defined in eqn (3) and (4) and shown in Table 1, it can now be immediately seen that the model works relatively well for **CC**, **MM** and **CHCC** systems (incidentally, these have the largest range of $(G_{M,L} - G_{M,(H_2O)_6})^{REL}$, *i.e.* the most selective, *cf.* Table 1). On the other hand, the average absolute deviation is ~3 kcal mol^{-1} (maximum absolute deviation >6 kcal mol^{-1}) for **DNDO** and **DDSOEE** systems, which is even more significant in light of their low selectivity (range of $(G_{M,L} - G_{M,(H_2O)_6})^{REL}$ ~ 15 kcal mol^{-1}).

$G_{M,L}^{2nd}$ describes the information about the metal selectivity retained by a model much more clearly than the original quantity, $(G_{M,L} - G_{M,(H_2O)_6})^{REL}$. The set of eight values (one for each metal ion) can be comprehensively reduced to $AAD_{L,model}$ and $MAD_{L,model}$, quantities that will be the cornerstone of the ensuing discussion.

3.2. Selectivity factors studied on "hybrid" systems

The results obtained in the previous section are not sufficient to fully comprehend the role of 1st-shell ligands, as the data compare systems with different metal ions and different structures (*i.e.* include both steps I and II from Fig. 2). The structural changes are not unimportant, but it is impossible to undertake full conformational sampling at the given methodological level. Therefore, we choose to separate the structural contribution to selectivity, which we can study only to limited extent, from the chemical effect of exchanging a metal ion, which we can study in detail in "hybrid" systems (see Section 2.4.2). The following subsections (3.2.1, 3.2.2 and 3.2.3) study the influence of various factors on "hybrid" systems.

14398 | *Phys. Chem. Chem. Phys.*, 2015, **17**, 14393–14404

This journal is © the Owner Societies 2015

Table 1 The calculated values of *relative* free energies of complexation, $(G_{M,L} - G_{M,(H_2O)_6})^{REL}$, for the six systems in the whole peptide (upper rows) and ALPHA (lower rows) representation. The relative difference of the two, $G_{M,L,ALPHA}^{2nd}$ (eqn (2) and (3)), and the corresponding average and maximum absolute deviations, $AAD_{L,ALPHA}$ and $MAD_{L,ALPHA}$ (eqn (4)) are listed[a]

System	Quantity	Mn²⁺	Fe²⁺	Co²⁺	Ni²⁺	Cu²⁺	Zn²⁺	Cd²⁺	Hg²⁺	$\dfrac{AAD_{L,ALPHA}}{(MAD_{L,ALPHA})}$[b]
CC	$(G_{M,L,peptide} - G_{M,(H_2O)_6})^{REL}$	0.0	−4.7	−4.7	−5.2	−25.7	−11.2	−20.6	−57.8	
	$(G_{M,L,APLHA} - G_{M,(H_2O)_6})^{REL}$	0.0	−5.7	−4.1	−3.9	−23.9	−11.0	−18.7	−54.6	
	$G_{M,L,APLHA}^{2nd}$	**−0.9**	**0.0**	**−1.5**	**−2.2**	**−2.7**	**−1.1**	**−2.8**	**−4.2**	**1.0** *(2.2)*
MM	$(G_{M,L,peptide} - G_{M,(H_2O)_6})^{REL}$	0.0	−2.5	−6.5	−8.6	N/A	−10.1	−21.5	−59.8	
	$(G_{M,L,APLHA} - G_{M,(H_2O)_6})^{REL}$	0.0	−3.3	−7.5	−9.6		−10.8	−21.4	−57.8	
	$G_{M,L,APLHA}^{2nd}$	**−1.0**	**−0.1**	**0.0**	**0.0**	**−1.0**	**−0.2**	**−1.1**	**−2.9**	**0.7** *(2.2)*
DHHD	$(G_{M,L,peptide} - G_{M,(H_2O)_6})^{REL}$	−8.2	−7.3	−9.8	0.0	−6.1	−10.7	−7.1	−18.8	
	$(G_{M,L,APLHA} - G_{M,(H_2O)_6})^{REL}$	−11.0	−11.8	−14.3	0.0	−10.6	−14.2	−6.2	−18.3	
	$G_{M,L,APLHA}^{2nd}$	**−1.8**	**0.0**	**−0.1**	**−4.6**	**−0.1**	**−1.0**	**−5.4**	**−5.0**	**2.1** *(3.2)*
DNDO	$(G_{M,L,peptide} - G_{M,(H_2O)_6})^{REL}$	−9.1	−6.1	−7.1	−5.7	0.0	−6.9	−9.6	−15.1	
	$(G_{M,L,APLHA} - G_{M,(H_2O)_6})^{REL}$	−5.9	0.0	−9.3	−6.6	−6.7	−8.6	−13.0	−20.5	
	$G_{M,L,APLHA}^{2nd}$	**−10.0**[c]	**−12.8**[c]	**−4.5**	**−5.9**	**0.0**	**−5.1**	**−3.4**	**−1.4**	**3.1** *(7.4)*
CHCC	$(G_{M,L,peptide} - G_{M,(H_2O)_6})^{REL}$	0.0	−10.1	−12.8	−7.2	−22.5	−9.1	−11.0	−40.7	
	$(G_{M,L,APLHA} - G_{M,(H_2O)_6})^{REL}$	0.0	−8.7	−11.2	−5.0	−20.4	−8.9	−14.2	−40.5	
	$G_{M,L,APLHA}^{2nd}$	**−3.2**	**−4.6**	**−4.8**	**−5.5**	**−5.3**	**−3.4**	**0.0**	**−3.5**	**1.3** *(3.8)*
DDSOEE	$(G_{M,L,peptide} - G_{M,(H_2O)_6})^{REL}$	−15.3	−12.4	0.0	−0.7	−1.8	−2.9	−6.1	−16.6	
	$(G_{M,L,APLHA} - G_{M,(H_2O)_6})^{REL}$	−11.8	−9.7	−2.6	0.0	−1.2	−5.2	−13.4	−21.9	
	$G_{M,L,APLHA}^{2nd}$	**−10.9**[c]	**−10.1**[c]	**−4.8**	**−8.1**	**−8.0**	**−5.1**	**0.0**	**−2.0**	**3.2** *(6.1)*

[a] All values are in kcal mol⁻¹. Metal-dependent shifts, ΔG_{corr}, pertaining to the protocol and reference states used and shown to lead to the best computational estimates to the experimental β in our previous work (ref. 32) were applied. ΔG_{corr} = 0.2, −0.9, 1.9, −1.1, 2.5, −1.6, 0.4, −1.4 kcal mol⁻¹ for Mn²⁺, Fe²⁺, Co²⁺, Ni²⁺, Cu²⁺, Zn²⁺, Cd²⁺, and Hg²⁺, respectively. [b] Lower values of $AAD_{L,ALPHA}$ and $MAD_{L,ALPHA}$ indicate high retention of metal-ion selectivity in the model. See eqn (4) for definition. [c] These large values can be traced to a different position of a 2nd-shell charged group, resulting in significantly different interaction with a metal-ion.

The values of $G_{M,L}^{2nd,REL,hybrid}$ are collected in Tables S1–S24 (ESI†). These values exhibit a strong trend present in virtually all structures and systems, decreasing gradually across the row to reach a minimum (the most negative value; implying the strongest interaction with the 2nd shell) at Ni²⁺ or Cu²⁺, while the maximum (the weakest influence) is usually exhibited by Cd²⁺ systems. Thus, the models skew the relative free energy between *e.g.* Cu²⁺ and Cd²⁺ in favor of the latter. While this trend follows, at least in partial, the Irving–Williams series, its physico-chemical origin remains unknown to us. Bearing this trend in mind, we choose to present only the $AAD_{L,model}^{hybrid}$ (and $MAD_{L,model}^{hybrid}$).

These values for one of the structures (selected as a representative example), of all six peptide systems and for all four representations are presented in Table 2. Despite the diversity of the systems, most of the $AAD_{L,model}^{hybrid}$ values are below 1 kcal mol⁻¹ even for the *SMALL* model, which translate into $MAD_{L,model}^{hybrid}$ values below ∼2 kcal mol⁻¹ (*vide* Table 2). Examining relationship between these values and the nature of the systems provides insight into the importance of the 2nd shell for selectivity and basis for constructing a robust and well-balanced model.

In general, the $AAD_{L,model}^{hybrid}$ values decrease as the size of the model increases (Tables S1–S24, ESI†):

$$AAD_{L,FULL\_AA}^{hybrid} < AAD_{L,ALPHA}^{hybrid} < AAD_{L,SMALL}^{hybrid} < AAD_{L,TINY}^{hybrid}$$

There are a few remarkable exceptions to this progression, but the magnitude of this discrepancy is diminutive (<few tenths of kcal mol⁻¹) and they are not highlighted in further discussions.

3.2.1 Hardness of ligands. The most influential factor is the hardness of a ligand. In our set of systems, harder ligands are represented by O- and N-binding moieties, namely aspartate, glutamate, histidine, and serine side-chains, peptide bond nitrogen and oxygen, and water molecules. Softer ligands are represented by cysteinate and methionine side-chains.

The influence is most visibly exhibited by the *TINY* model. While systems with predominantly soft residues (**CC, MM, CHCC**)

Table 2 Average and maximum absolute deviations of interaction of a metal ion with 2nd shell, $AAD_{L,model}^{hybrid}$ and $(MAD_{L,model}^{hybrid})$, for Cd²⁺-optimized structures, chosen as a representative example[a]

Model	System					
	CC	MM	DHHD	DNDO	CHCC	DDSOEE
TINY	**2.2** (4.4)	**1.8** (2.6)	**3.8** (5.6)	**2.5** (5.0)	**0.6** (1.9)	**0.8** (1.1)
SMALL	**0.9** (1.3)	**0.8** (1.7)	**1.5** (4.1)	**0.4** (1.0)	**0.3** (0.7)	**0.5** (0.7)
ALPHA	**0.4** (0.6)	**0.5** (1.1)	**0.4** (1.0)	**0.1** (0.2)	**0.3** (0.9)	**0.4** (1.4)
FULL_AA	**0.2** (0.3)	**0.1** (0.3)	**0.1** (0.2)	**0.1** (0.2)	**0.1** (0.3)	**0.2** (0.3)

[a] All values in kcal mol⁻¹.

have $AAD_{L,model}^{hybrid}$ values of ~2 kcal mol^{-1}, those of harder ligands are 1 kcal mol^{-1} or less.

The hardness characteristic can be thought of as the distance at which the ligand can be polarized by its environment. Thus, while all systems converge to the original peptide as we increase the size of the model, the soft ligands will be more sensitive to this change than the hard ligands. Moving from *TINY* to *SMALL* model, the $AAD_{L,model}^{hybrid}$ values drop below 1 kcal mol^{-1} in all cases, but the decrease continues upon moving further to *ALPHA* model only for soft ligands; while being almost non-existent for hard ligands. In general, a similar level of selectivity is achieved by larger model (*ALPHA*) for sulfur based ligands, while smaller model (*SMALL* or even *TINY*) is sufficient for harder ligands.

3.2.2 Hydrogen bonding to ligands. Binding atom is influenced not only by the rest of the ligand, *i.e.* its covalent partners but also by non-covalent interactions, most notably hydrogen bonding. To investigate the importance of this effect, a "control" system is required. We consider hypothetical systems where the N–H hydrogen bond donors are replaced with isoelectronic O atom (*i.e.* turning the amide into an ester), and the NH_2 groups are replaced with chemically similar Cl with no optimization of the molecular geometry. In our opinion, these substitutions (illustrated in Fig. 3) represent the minimum conceivable perturbation of the system, thus allowing for maximum comparability of $AAD_{L,model}^{hybrid}$ values of substituted and original systems.

It should be noted that only hydrogen bonds to atoms directly ligating the central metal ion are considered, as these are the ones that can be expected to chemically (and possibly selectively) influence the binding of a metal ion. Other hydrogen bonds and interactions are unlikely to depend on the identity of the metal ion and can influence the binding only indirectly through structural change (not addressed in the "hybrid" systems studied here).

The following analysis is performed for each relevant peptide (bearing at least some hydrogen bonds to binding atoms) on one of the structures, selected as a representative example. It comes at no surprise (Table S25, ESI†) that there are almost none hydrogen bonds in the smaller representations of the full peptide (up to *FULL_AA* model). Thus, the effect of a hydrogen bond, if any, is present in the full peptide and *FULL_AA* model but usually not in the smaller models – which can contribute to different selectivity of a peptide and its model. The "ester" systems, on the other

Table 3 Average absolute deviations of interaction of a metal ion with 2nd shell, $AAD_{L,model}^{hybrid}$, evaluated for Zn^{2+}-optimized peptide (**upper rows**) systems and their "ester" (*lower rows*) analogues[a]

Model	System	CC	DHHD	DNDO	CHCC	DDSOEE
TINY	**Peptide**	**2.3**	**0.4**	**0.7**	**2.1**	**0.8**
	Ester	*1.9*	*0.3*	*0.6*	*1.9*	*1.0*
SMALL	**Peptide**	**1.0**	**0.3**	**0.6**	**0.9**	**0.6**
	Ester	*0.4*	*0.3*	*0.6*	*0.7*	*0.6*
ALPHA	**Peptide**	**0.7**	**0.3**	**0.4**	**0.9**	**0.6**
	Ester	*0.2*	*0.4*	*0.4*	*0.7*	*0.4*
FULL_AA	**Peptide**	**0.2**	**0.3**	**0.3**	**0.3**	**0.1**
	Ester	*0.2*	*0.2*	*0.2*	*0.4*	*0.2*

[a] All values in kcal mol^{-1}.

hand, do not possess these hydrogen bonds in either a model or its parent, which results in models being a more faithful representation of its parent system. The *FULL_AA* model (of a peptide system) usually contains these hydrogen bonds, thus we expect little differences in values of $AAD_{L,FULL\_AA}^{hybrid}$ based on peptides and those based on their "ester" analogue.

The $AAD_{L,model}^{hybrid}$ values (Table 3) are almost identical in all but few cases, the differences ranging from −0.2 to 0.5 kcal mol^{-1}. The smallest differences are observed in the **DHHD**, **DNDO** and **DDSOEE** systems all of which include hydrogen bonds to hard ligands (carboxyl groups/water molecule/serine alcohol group). On the other hand, $AAD_{L,model}^{hybrid}$ differ by 0.2–0.5 kcal mol^{-1} in case of **CC** and **CHCC** systems, where the hydrogen bond acceptor is amide bond carbonyl and cysteine thiolate groups, respectively. This amount constitutes 20–70% of the total $AAD_{L,model}^{hybrid}$. Moreover, in **CHCC** system the hydrogen bonds are far from the ideal orientation and in case of strong hydrogen bond, the influence can be expected to be much higher. We thus conclude that the effect of hydrogen bond is potentially significant for strong hydrogen bonds to softer ligands.

A ligand can also act as a hydrogen bond donor, as exemplified by terminal glutamate of **DNDO** peptide, which can interact with one of the metal-bound water molecules. The validity of attributing the differences in selectivity to the single carboxyl moiety is confirmed by comparing $AAD_{L,model}^{hybrid}$ values of
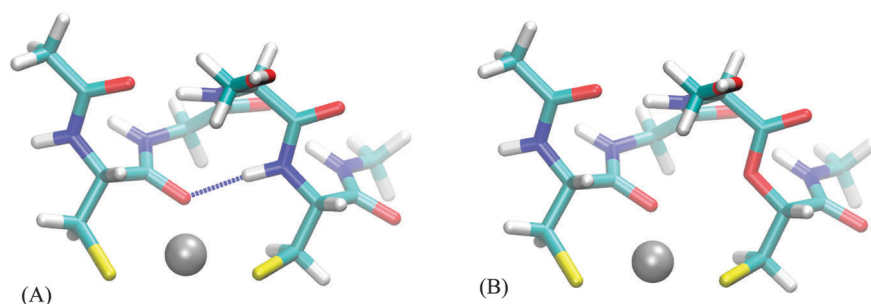

Fig. 3 Esther analogues of peptides. An example of substitution of N–H hydrogen bond donor (A) with O (B) in **CC** peptide. The substitution does not perturb geometry, but does eliminate the hydrogen bond.

Table 4 Influence of proximal charged groups on 2nd-shell metal-induced selectivity of two **DNDO** structures

Structure	Glutamate leaning in[a]		Glutamate leaning out[b]	
C-terminal residue	Glutamate	Norvaline	Glutamate	Norvaline
d(M–OOC)[c]	4.22	—	5.29	—
	AAD$_{L,model}^{hybrid}$ [d]			
TINY	1.2	0.9	0.6	0.7
SMALL	1.1	0.8	0.6	0.5
ALPHA	0.9	0.6	0.4	0.4
FULL_AA	0.5	0.2	0.2	0.2

[a] Fe^{2+}-optimized structure, see ESI for coordinates. [b] Co^{2+}-optimized structure, see ESI for coordinates. [c] Distance between a metal-ion and proximal oxygen of the terminal glutamate side-chain carboxyl group in Ångströms. [d] Values in kcal mol^{-1}.

Table 5 Influence of screening of the 2nd shell (due to addition/removal of water molecules) on its interaction with a metal ion

System	CC	DNDO			DNDO		
Structure	Mn^{2+}-optimized	Co^{2+}-optimized			Hg^{2+}-optimized		
Water molecules	0	2[a]	2	0[b]	0[c]	2	0[b]	0[c]
	AAD$_{L,model}^{hybrid}$ [d]						
TINY	1.7	1.5	0.6	0.9	1.1	0.8	1.3	3.1
SMALL	0.8	0.8	0.6	0.7	0.9	0.6	1.1	2.3
ALPHA	0.6	0.9	0.4	0.5	0.4	0.7	0.4	1.5
FULL_AA	0.2	0.1	0.2	0.4	0.3	0.3	0.4	0.9

[a] Water molecules added, structure optimized. [b] Water molecules removed, structure after optimization. [c] Water molecules removed, structure before optimization. [d] Values in kcal mol^{-1}.
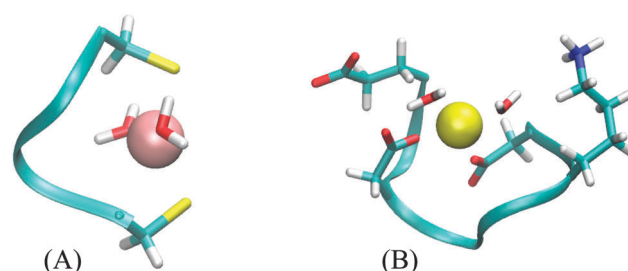

Fig. 4 Screening of peptide parts by water molecules. Backbone and charged side-chains are shown. (A) **CC** peptide – no part of the system is screened by the water molecules. (B) **DNDO** peptide – glutamate and lysine side-chains are screened by water molecules, resulting in increased 2nd-shell selectivity upon their removal.

original systems to a "mutated" peptide (glutamate → norvaline), presented in Table 4. However, a water molecule is the smallest of the studied ligands, which allows for relative proximity of the charged moiety to a metal ion, which is probably the cause of the observed differences, rather than the existence of the hydrogen bond.

3.2.3 Overall screening of metal-ion and coordination number. The non-covalent interactions of 2nd shell with the 1st-shell ligands are certainly not limited to hydrogen-bonding. In a broader sense, any lack of screening of the metal ion from the 2nd shell will contribute to the increased influence of the latter on selectivity. Higher coordination numbers can be expected to provide more "crowded" binding-sites, burying the metal ion and limiting the influence of the 2nd shell.

As in the previous section, we strive to compare systems that differ significantly in this respect (*i.e.* in the screening of metal-ion), while being as similar as possible in other respects. The most straightforward procedure is to change the number of water molecules in the systems, as these constitute ligands that are not covalently bound to the rest of the system and, thus, no other changes are necessary. Representative structures of peptides **CC** and **DNDO** that demonstrate the range of this influence vividly were chosen for discussion. In case of **CC** peptide, 2 water molecules were added. In case of **DNDO**, 2 water molecules were removed. In both cases the structures were subsequently optimized, however, in the latter case both the non-optimized and optimized structures were used for selectivity analysis, *i.e.* for construction and calculation of "hybrid" systems. The results are presented in Table 5.

The **CC** system mode of binding is a distorted plane with the ligands forming a triangle. The two water molecules are added in a line approximately perpendicular to this plane, resulting in a square pyramidal formation (Fig. 4A). There is no part of the peptide being screened by these water molecules. Correspondingly, the values of AAD$_{L,model}^{hybrid}$ for the two systems are almost identical (see Table 5).

The **DNDO** system, on the other hand, has an octahedral binding mode with 2 water molecules in *cis*-positions that "eclipse" two charged groups – a glutamate (already discussed in Section 3.2.2) and a lysine side-chain (Fig. 4B). The removal

of the water molecules exposes these charged moieties – which are *not* part of any of the models – to the metal ion. In all studied cases, this leads to a significant increase in 2nd-shell selectivity. The situation is largely remedied upon optimizing the structures, which distort to almost tetrahedral geometry. The screening is nevertheless deteriorated, compared to the original octahedral structure, and the values of AAD$_{L,model}^{hybrid}$ are thus higher.

Results for two **DNDO** structures are shown in Table 5. The original "2-water" systems and even optimized "0-water" systems show similar 2nd-shell selectivity, but there are immense differences in the selectivity of the "0-water" non-optimized structures. The fact that both of these structures were equally submitted to the selectivity analysis (as performed in the previous cases) shows that the difference must pertain to the structure of the peptide. This is alarming, as the two structures are virtually identical. An extensive investigation suggests that the difference is not due to coordination geometry, ligand–metal distances, nor solvent cavity construction. Despite all our efforts, we have not been able to pinpoint the source of this discrepancy. Although the cause of this phenomenon can be artificial, its disappearance in optimized structure does manifest the importance of a properly screened binding site; the absence of which will severely hamper the accuracy regardless of the model used.

A different point of view on the addition/removal of water molecules is to think of it as an exchange of an explicit water

molecule for an implicit one or *vice versa* (it can be reminded that despite advanced treatment of solvation in the COSMO-RS method, the method still belongs to a class of polarized continuum models, PCM). While in the case of the **CC** system, where the water molecules do not screen a part of the peptide, this seems to leave the selectivity undisturbed, the values of $G_{M,L}$ should still be compared only among systems of identical 1st shell composition.

3.3. Structure-induced selectivity

As discussed in Section 2.4.2, the overall metal-ion selectivity is determined by combined effect of metal-induced and structure-induced selectivity. The overall selectivity has been studied by analyzing fully optimized structures (Table 1). The metal-induced selectivity (thoroughly discussed in Section 3.2.) has been analyzed by examining interactions of a specific structure with a series of metal ions ("hybrid" systems, Fig. 2). We now return to the question of what is the lower limit of structure relaxation on selectivity (step II in Fig. 2).

Table 6 compares the selectivity of one of the hybrid systems, a representative example of metal-induced selectivity, with the fully relaxed structures, representative of the overall selectivity.

In some of the studied systems (**DHHD, DNDO, DDSOEE**), the disparity is destructive for the utility of the models. The most contributing structural changes are flexible charged groups which cause significantly altered interaction with a metal-ion. In the other cases (**CC, MM, CHCC**), the structure and selectivity of 2nd shell is virtually unchanged, or even lower after full relaxation. This results from interplay of metal-induced and structure-induced selectivity. While these are comprehensible and practical concepts for discussion, they are not independent but tend to partially cancel each other in these scenarios of diminutive changes of 2nd-shell structure.

As no conformational sampling has been performed, this merely shows that there exists a local minimum where the effect of structure-induced selectivity is minimal. A proper sampling is likely to reveal energetically more relevant structures that do not possess this property. However, this does demonstrate that the assumption of low structure-induced

selectivity on the overall selectivity is not inherently incorrect, signaling a green light to the start of search for systems with this characteristics.

This section, together with previous discussion, points out that in cases where the 2nd-shell structure is not sensitive to the identity of bound metal ion, the selectivity is to a large extent confined to the 1st shell. The rigidity condition mentioned in Section 2.4.2 has its origins in this discussion.

For these rigid systems, using a single 1st-shell model structure for each metal ion is sufficient for reproducing selectivity of the parent systems with reasonable accuracy. Looking back at Tables 1 and 2, the satisfactory performance of the models in these cases should now be fully comprehensible.

For 'bottom-up' design procedures, the structures of 1st-shell models have to be obtained without prior knowledge of the parent structure. This can be aided by constrained optimization reflecting structural support of the 2nd shell and should consider specifics of the system.

The factors relevant for selectivity identified in this study are of 'chemical' and structural nature. Their containment into the described models, that we believe has been sufficiently demonstrated herein, renders them largely independent of the nature of the 2nd shell. As long as the structural and conformational aspects are treated properly, the utility of the models can be expanded outside of the peptide context examined here. Thus, systems such as polymeric scaffolds, cluster models, zeolites, *etc.* can all be expected to be equally well described by these models. From the methodological perspective, both 'top-down' and 'bottom-up' approaches to studying and designing metal-binding systems can benefit from the models. The protocol and the models could conceivably be used for *e.g.* predicting differences in redox potentials of bound metal ions. Caution and further testing is advised, however, as the higher oxidation states can significantly affect not only the quality of treatment but also *e.g.* the required size of the model.

4. Summary

The overall metal-ion selectivity of a system is defined by different ligand(peptide)–metal interactions and by structural response of the system (ligand) to these interactions. A 1st shell model of a system can only retain information about selectivity if this response, *i.e.* free energy of relaxation upon exchanging a metal ion, is low. Existence of local minima with this property suggests this condition is achievable, yet overall rigidity of a peptide may be inevitable in order to ensure comparable conformational freedom.

The metal-induced selectivity of a system is a complex function of multiple factors, that define measure of its localization to a binding site, *i.e.* the possibility of capturing it using a 1st shell model of the system. It is essential that this model includes all of the ligands bound to a metal-ion. This requirement alone usually ensures that the binding is described reasonably well by the model and that the metal-ion is properly "shielded" from parts of a system further away, bringing the average absolute error to

Table 6 Influence of structure relaxation on 2nd-shell selectivity. Values of $\text{AAD}_{\text{L,model}}^{\text{hybrid}}$ (**upper rows**) of hybrid (*cf.*, Fig. 2) Cd^{2+}-optimized systems, taken as representative example, and $\text{AAD}_{\text{L,model}}$ (*lower rows*), *i.e.* fully optimized systems, in kcal mol^{-1}

Model	System	CC	MM	DHHD	DNDO	CHCC	DDSOEE
TINY	**Hybrid**	2.2	3.8	0.6	0.8	2.1	1.0
	Fully relaxed	2.3	3.7	1.1	2.7	2.6	2.6
SMALL	**Hybrid**	0.9	1.5	0.3	0.5	0.8	0.7
	Fully relaxed	0.9	1.0	1.8	2.9	1.4	2.8
ALPHA	**Hybrid**	0.4	0.4	0.3	0.4	0.7	0.7
	Fully relaxed	1.0	0.7	2.1	3.1	1.3	3.2
FULL_AA	**Hybrid**	0.2	0.1	0.1	0.2	0.3	0.2
	Fully relaxed	0.4	0.3	1.7	3.2	0.7	2.1

14402 | *Phys. Chem. Chem. Phys.,* 2015, **17**, 14393–14404

This journal is © the Owner Societies 2015

values no higher than few units kcal mol^{-1}. However, some binding geometries can prove to be problematic due to insufficient shielding, *e.g.* linear or square planar geometry.

Achieving more satisfactory accuracy requires consideration of the nature of ligands. Softer ligands, represented in the realm of metalloproteins by cysteine and methionine residues, require bulkier representation, *e.g.* the whole side-chain of a residue. Hard ligands, like carboxyl groups, alcohols, amines, can be described at a similar level of accuracy more economically by the functional group plus single methyl group. Imidazole is sufficient to represent histidine side-chain, *N*-methylmethylamide to represent backbone amide group, and methylamide to represent glutamine and asparagine ligands. This level of description can be expected to result in average absolute deviations below 1 kcal mol^{-1}, which translates to maximum relative error (among studied metal ions) below 3 kcal mol^{-1}. Moreover, most of this relative error derives from bias that appears consistently in the model and can be corrected for heuristically, although this may not be desirable, due to insufficient understanding of this bias.

Non-covalent partners of ligands usually influence retention of selectivity in a negative way. The presence of charged groups can be influential, especially in case of insufficient screening or proximity to metal ion due to small 1st shell ligand, *e.g.* water molecule. In case of hydrogen bonding the effect is much less significant, influencing average absolute deviation by tenths of kcal mol^{-1}. The strength of the hydrogen bond as well as softness of the acceptor can increase this value.

The overall error in reproducing metal-ion selectivity of a system will stem from inaccuracy of the protocol (~ 2 kcal mol^{-1} in relative free energies[32]), approximation introduced by the model (< 1 kcal mol^{-1}) and assumption of low structure-induced selectivity (the error can be arbitrarily large but, as argued in Section 3.3, conceivably negligible). While some cancellation of errors can be expected, the average errors in the order of units of kcal mol^{-1} (translating to individual relative errors of roughly up to 5 kcal mol^{-1}) must be expected. As the ranges of free energies of binding are in the order of several tens of kcal mol^{-1}, the protocol holds a promise as a tool for fast estimation of selectivity.

Acknowledgements

References

1 A. Warshel, *Computer Modeling of Chemical Reactions in Enzymes and Solutions*, John Wiley & Sons, Inc., New York, 1997.
2 J. J. R. Fraústo da Silva and R. J. P. Williams, *The Biological Chemistry of the Elements*, Oxford University Press, Inc., New York, 2nd edn, 2001.
3 *Handbook of Metal–Ligand Interactions in Biological Fluids*, ed. G. Berthon, Marcel Dekker, New York, 1995, vol. 2.
4 R. K. O. Sigel and H. Sigel, *Acc. Chem. Res.*, 2010, **43**, 974–984.
5 T. Dudev and C. Lim, *Chem. Rev.*, 2003, **103**, 773–787.
6 T. Dudev and C. Lim, *Chem. Rev.*, 2014, **114**, 538–556.
7 R. J. P. Williams, *BioMetals*, 2007, **20**, 107–112.
8 H. F. Lodish, *Molecular Cell Biology*, Scientific American Books, New York, 1999.
9 The full set of metal ion concentrations with the respective references can be found in Table 2 of ref. 6.
10 H. Irving and R. J. P. Williams, *Nature*, 1948, **162**, 746.
11 F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry*, John Wiley & Sons, New York, 1988.
12 R. J. P. Williams, *Coord. Chem. Rev.*, 2001, **216–217**, 583.
13 B. Hille, *Ionic Channels of Excitable Membranes*, Sinauer Associates, Sunderland, MA, 3rd edn, 2001.
14 J. Aqvist and V. Luzhkov, *Nature*, 2000, **404**, 881.
15 X. S. Wu, H. D. Edwards and W. A. Sather, *J. Biol. Chem.*, 2000, **275**, 31778.
16 R. J. P. Williams, *Coord. Chem. Rev.*, 2001, **216–217**, 583.
17 K. A. McCall and C. A. Fierke, *Biochemistry*, 2004, **43**, 3979.
18 Y. Li and D. B. Zamble, *Chem. Rev.*, 2009, **109**, 4617.
19 L. Rulíšek and Z. Havlas, *J. Phys. Chem. A*, 1999, **103**, 1634–1639.
20 L. Rulíšek and Z. Havlas, *J. Am. Chem. Soc.*, 2000, **122**, 10428–10439.
21 L. Rulíšek and Z. Havlas, *J. Phys. Chem. A*, 2002, **106**, 3855–3866.
22 L. Rulíšek and Z. Havlas, *Int. J. Quantum Chem.*, 2003, **91**, 504–510.
23 L. Rulíšek and Z. Havlas, *J. Phys. Chem. B*, 2003, **107**, 2376–2385.
24 T. Dudev and C. Lim, *J. Phys. Chem. B*, 2001, **105**, 10709–10714.
25 T. Dudev and C. Lim, *Acc. Chem. Res.*, 2007, **40**, 85–93.
26 O. Gutten, I. Beššeová and L. Rulíšek, *J. Phys. Chem. A*, 2011, **115**, 11394–11402.
27 T. Dudev and C. Lim, *J. Phys. Chem. B*, 2009, **113**, 11754.
28 T. Dudev, J. A. Cowan and C. Lim, *J. Am. Chem. Soc.*, 1999, **121**, 7665.
29 L. Rulíšek and J. Vondrášek, *J. Inorg. Biochem.*, 1998, **71**, 115–127.
30 J. P. Glusker, *Adv. Protein Chem.*, 1991, **42**, 1–76.
31 H. Zheng, M. Chruszcz, P. Lasota, L. Lebioda and W. Minor, *J. Inorg. Biochem.*, 2008, **102**, 1765–1776.
32 O. Gutten and L. Rulíšek, *Inorg. Chem.*, 2013, **52**, 10347–10355.
33 T. Dudev and C. Lim, *J. Am. Chem. Soc.*, 2010, **132**, 2321.
34 T. Dudev and C. Lim, *Phys. Chem. Chem. Phys.*, 2012, **14**, 12451.
35 T. Dudev, Y. L. Lin, M. Dudev and C. Lim, *J. Am. Chem. Soc.*, 2003, **125**, 3168.
36 A. T. Maynard and D. G. Covell, *J. Am. Chem. Soc.*, 2001, **123**, 1047.
37 T. Dudev and C. Lim, *J. Am. Chem. Soc.*, 2006, **128**, 1553.
38 K. Hsin, Y. Sheng, M. M. Harding, P. Taylor and M. D. Walkinshaw, *J. Appl. Crystallogr.*, 2008, **41**, 963–968.
39 M. Kožíšek, A. Svatoš, M. Buděšínský, A. Muck, M. C. Bauer, P. Kotrba, T. Ruml, Z. Havlas, S. Linse and L. Rulíšek, *Chem. – Eur. J.*, 2008, **14**, 7836–7846.
40 A. D. Becke, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1988, **38**, 3098–3100.

41 J. P. Perdew, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1986, **33**, 8822–8824.

42 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

43 A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829–5835.

44 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571.

45 K. Eichkorn, O. Treutler, H. Öhm, M. Häser and R. Ahlrichs, *Chem. Phys. Lett.*, 1995, **240**, 283–290.

46 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

47 D. Andrae, U. Haeussermann, M. Dolg, H. Stoll and H. Preuss, *Theor. Chim. Acta*, 1990, **77**, 123–141.

48 A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224–2235.

49 A. Klamt, V. Jonas, T. Buerger and J. C. W. Lohrenz, *J. Phys. Chem.*, 1998, **102**, 5074–5085.

50 F. Eckert and A. Klamt, *COSMOtherm, version C3.0, release 12.01*, COSMOlogic GmbH & Co. KG.