Dalton Transactions

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the RSC Publishing peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, which is prior to technical editing, formatting and proof reading. This free service from RSC Publishing allows authors to make their results available to the community, in citable form, before publication of the edited article. This *Accepted Manuscript* will be replaced by the edited and formatted *Advance Article* as soon as this is available.

To cite this manuscript please use its permanent Digital Object Identifier (DOI®), which is identical for all formats of publication.

More information about *Accepted Manuscripts* can be found in the **Information for Authors**.

Please note that technical editing may introduce minor changes to the text and/or graphics contained in the manuscript submitted by the author(s) which may alter content, and that the standard **Terms & Conditions** and the **ethical guidelines** that apply to the journal are still applicable. In no event shall the RSC be held responsible for any errors or omissions in these *Accepted Manuscript* manuscripts or any consequences arising from the use of any information contained in them.

ARTICLE TYPE www. rsc.org/xxxxxx | XXXXXXXX

Structural and computational insights on the versatility of cadmium binding to proteins

Ran Friedman[*a b]

Received Xth XXXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX
First published on the web Xth XXXXXXXXXX 200X
DOI: 10.1039/b000000x

Cadmium is a highly toxic group XII metal, like zinc and mercury. Unlike zinc, which is one of the most common metal cofactors in biology, cadmium is highly toxic. Many Zn^{2+}-binding proteins can bind Cd^{2+}-ions without significantly affecting their structures. Here, the protein-data bank is analysed with respect to protein-cadmium interactions, which shows that cadmium can bind to a variety of ion binding sites in proteins. Statistical analysis for Cd^{2+}-side chain interactions is compared with a similar analysis of other ions. This analysis reveals that with respect to amino acid side-chain preference, Cd^{2+} is more similar to Mn^{2+} than to Zn^{2+} or Hg^{2+}. Finally, the interaction energies of three native metal binding proteins are calculated where Cd^{2+} binds instead of Zn^{2+}, Ca^{2+} or Cu^{2+}. The interaction energies are decomposed into individual components whose contributions are discussed.

1 Introduction

Cadmium is a group XII metal that has been discovered in 1818 as an impurity in zinc oxides[1]. Modern day uses of cadmium include batteries, pigments, metal coating and plastics[2]. ^{113}Cd is used in NMR, and hence Cd^{2+} is sometimes used in structural biology of metalloproteins where it replaces other metals. Furthermore, cadmium salts are used as precipitating agents to induce protein crystallisation[3]. Although there is a decrease in cadmium usage in the recent years, the metal continues to be a notable environmental pollutant, particularly in industrial areas where cadmium or zinc have been treated or produced[4]. Moreover, cadmium is a notable marine pollutant, e.g., in the Baltic sea[5,6].

Cadmium is carcinogenic and highly toxic to human and animals. For example, risk for lung cancer from recurrent cadmium exposure can be almost as high as that from smoking[7]. The multi-organ damage from acute or chronic cadmium poisoning[8,9] suggests a complex mode of toxicity. It has been proposed that some of the toxic effects of cadmium are due the cadmium ion binding to proteins having calcium, zinc or magnesium cofactors[10,11]. On the other hand, some metal binding proteins, most notably metallothioneins, appear to participate in protection against cadmium toxicity. Spectroscopic and potentiometric data reveals that binding of Cd^{2+} may be favoured thermodynamically over zinc[12–14], at least in some complexes.

In a comprehensive review, Holm and co-workers discussed the structural aspects and coordination of multivalent metal-binding sites in proteins[15]. Their study reveals high similarity between the XII group ions Zn^{2+} and Cd^{2+}, but lower similarity between Cd^{2+} and Hg^{2+}, which belong to the same group. Data on Zn^{2+}-protein interactions are more prevalent than those available for Cd^{2+} since zinc is a common protein cofactor, whereas cadmium is highly toxic to many organisms. One notable exception is cadmium-containing carbonic anhydrase of the marine algae *Thalassiosira weissflogi*[16]. The marine depth density profile of cadmium is linked to the prevalence of plankton[17], which may suggest that it is used by other proteins and/or organisms as well. The bacterial proteins CadC and CmtR are transcriptional repressors that bind Cd^{2+} and other toxic metal ions[18,19].

Quantum chemical calculations are becoming an integral part of coordination chemistry[20], and shed light on many aspects of metal-ligand interactions[21]. Cd^{2+}-protein interactions have been studied using quantum mechanical (QM) methods since more than a decade ago. Ryde and Hemmingsten used QM calculations to interpret experimental studies of a Cd^{2+}-bound liver alcohol dehydrogenase (LADH), which normally employs a zinc cofactor[22]. Another early work[23] dealt with calculating the affinities of metallothionein to Zn^{2+} and Cd^{2+}. One of the most interesting compilations on protein-ion interactions came from Rulíšek and Havlas, who, in a series of papers, developed a framework for the calculation of interaction energies with density function theory (DFT)[24–26]. These works formed the theoretical basis for a computer-based molecular design of metal-binding peptide sequences[27], revealing a potential for biotechnological appli-

[a] *Computational Chemistry and Biochemistry Research Group, Department of Chemistry and Biomedical Sciences, Linnæus University, 391 82 Kalmar, Sweden.* [a] *Linnæus University Centre for Biomaterials Chemistry, 391 82 Kalmar, Sweden. ; Tel: +46 480 446290; E-mail: ran.friedman@lnu.se*

cations.

Analysis of structures from the protein data bank (PDB) can provide useful information on protein-cofactor interactions[28]. Indeed, Ramos and co-workers carried out statistical analysis of zinc-binding proteins[29] and later of metal-binding proteins in general[30] to shed light on protein-ion interactions from a structural point of view. Similarly, Rarey and co-workers have independently developed a statistics-based method for modelling of metal interaction sites[31], that is also based on survey of the PDB, with the aim of assisting in computer aided drug design involving metalloproteins.

The data available in the literature provide many interesting details on Cd^{2+}-ligand interactions. Yet, a better understanding of such interactions in proteins is desired and can be used e.g. in the fields of protein-engineering and material science. Moreover, cadmium compounds have been tested in a combined therapy against cancer[32,33], further revealing the need for a thorough understanding of cadmium's chemistry in a biological context. This study therefore deals with specific protein-cadmium interactions. A list of high resolution Cd^{2+}-binding proteins is compiled based on data from the PDB and the binding sites are analysed with respect to the ligands and coordination numbers. Statistical analysis for Cd^{2+}-side chain interactions is compared with a similar analysis of other ions. Additionally, the interaction energies of Cd^{2+} and protein ligands are calculated for three proteins: LADH, parvalbumin and azurin. The results are compared with interaction energies for the native ligands, i.e., Zn^{2+}, Ca^{2+}, and Cu^{2+}, respectively. The interaction energies are further analysed by applying the Localised Molecular Orbital Energy Decomposition Analysis (LMOEDA) technique[34,35].

2 Computational methods

2.1 Identification of Cd^{2+} binding sites

A list of PDB structures of proteins that bind Cd^{2+} was prepared by use of PDBEmotif[36]. The structures were sorted according to the resolution, and structures which were solved at resolutions worse than 2.5Å were discarded, leaving 151 structures to deal with. Cadmium has been widely used to aid in resolving crystal structures and in many cases it binds only at the surface and does not form a metal complex. Therefore, binding sites in which the metal binds to one or two protein residues at the protein surface were not considered. In case of duplicates or of the same protein (including mutants or proteins that differ only in a cofactor or drug molecule which binds them), only the structure with the best resolution was maintained. The remaining 26 structures contain at least one interesting Cd^{2+}-binding site although few are not known to be metal-binding proteins, and it is likely that the binding was due to the high concentration of Cd^{2+} ions in the crystal liquor. Such sites were considered relevant to shed light on Cd^{2+}-protein interactions as long as they fulfilled the criteria described above. Solution NMR structures were not considered, because of the uncertainty in the location of atoms in the NMR ensemble, which does not represent an energy minimum.

The number of ligands and the overall Cd^{2+}-coordination number of each ion were extracted through visual examination of all residues within 3.5Å of the ions. Visual inspection was deemed necessary because the proteins bind to ions in solution, and small deviations from the X-ray structure are expected[37]. In the case of carboxylate ligands, the binding was considered bidentate (by two oxygens) if their distances to the Cd^{2+}-ion were similar (within a tolerance of 0.2Å). Note that this pertains only to the calculation of the coordination number, and not for interaction energies.

2.2 Side chain preference for various ions in the PDB

The binding preferences of Ca^{2+}, Cu^{+}, Cu^{2+}, Zn^{2+}, Cd^{2+}, Hg^{2+} and Mn^{2+} to amino-acid side chains were extracted from the PDB by use of PDBeMotif[36] and Perl scripts written in-house.

2.3 Interaction energies and energy decomposition analysis

2.3.1 Preparation of the ion binding sites Interaction energies were calculated for models of the ion-binding sites, covering the ion and immediate ligands. Amino acid residues were substituted by smaller chemical groups to enable the calculations using QM methods. Thus, cys residues were replaces by CH_3S^-, asp and glu by acetate ions, his by monoprotonated imidazole, ser by CH_3O^- and backbone carbonyl by CH_3CHO. Hydroxyl residues of ser were unprotonated to avoid the formation of hydrogen bonds with other ligands if such hydrogen bonds were not present in the original structure. Note that similar (and even smaller) models of side chains have been used successfully for calculations of metal-ion selectivity[24–26,38].

The coordinates of LADH (pdb codes 2jhf, 2jhg[39]), parvalbumin (1cdp and 5cpv[40]), and azurin (1aiz and 1azc[41]) were downloaded from the protein data bank. Hydrogens were not resolved in the crystal structures, and have therefore been added using openbabel (www.openbabel.org) or ghemical[42]. The coordinates of the hydrogens were optimised using GAMESS-US[43], while keeping the heavy atoms fixed unless otherwise stated, because trial calculations revealed that the interaction energies were similar or more favourable in the original structures (results not shown). The wave functions were calculated by use of the M06 DFT functional[44] with a standard grid and the def2TZVP basis set[45], which involves

core potentials for effective treatment of relativistic effects in cadmium and mercury, except for LADH where MP2 was used with the aug-cc-pVDZ basis set (aug-cc-pVDZ-PP with effective core potential was used for Cd atoms)[46,47]. M06 was found to perform better than other DFT functionals in a set of calculations of interaction energies between Cd^{2+} or Zn^{2+} and a set of biologically relevant ligands[48].

The def2 basis sets were downloaded from the EMSL basis set exchange server[49].

2.3.2 Interaction energy calculations for models of protein-ion complexes Interaction energies and their decomposition were calculated with the LMOEDA code in GAMESS. Energy decomposition analysis (EDA) is an analysis method that aims to break down the contribution of interaction energy between monomers (or sets of atom) in a complex to terms that yield physical insights into the interaction.[50]. Here, we apply the LMOEDA technique, that decomposes the interaction energy ΔE^{int} to five components, namely electrostatic, exchange, repulsion, polarisation and dispersion interaction energies. Solvent effects were included through the use of the polarisable continuum model (PCM)[51]. Atomic radii of 1.39Å, 1.58Å, 1.99Å and 1.40Å were used for Zn, Cd, Ca and Cu, respectively. Radii for other atoms were GAMESS default. The radii for Zn, Cd and Cu are from the work of Bondi[52]. The radius of Ca is not given in Bondi's compilation and was therefore calculated as $R_{Ca} = 1.15R_{Mg}$ where the 1.15 ratio was calculated according to a survey of radii in crystals and molecules[53].

A fine grid (126 radial points and 1202 angular points in the expansion) was used for the DFT calculations, which were carried out with the M06 functional and def2TZVP basis set. Counterpoise correction to the basis set superposition error[54] has been applied in all but the Cu^{2+} complexes, where the SCF could not converge with the larger basis set.

3 Results and discussion

3.1 Cd^{2+}-binding proteins in the PDB

More than 150 proteins whose structures (of resolution 2.5Å or better) contain Cd^{2+} ions were identified in the PDB. The vast majority of Cd^{2+}-binding sites, however, occur at the protein surface and seem to be unspecific. Accounting only for interactions with three protein ligands at least, and removing identical proteins resulted in 26 Cd^{2+}-binding protein structures, as summarised in Table 1.

The proteins presented in Table 1 have different biological roles, e.g., metabolic enzymes, specific ion binding proteins, redox proteins, toxins and chelating agents. Cd^{2+} replaces other ions in 20 of the 51 Cd^{2+}-binding sites reported here. These are Zn^{2+} (7 sites), Ca^{2+} (9 sites), Cu^{2+} (one site), Mg^{2+} (2 sites) and Mn^{2+} (one site). Cd^{2+} is the native protein cofactor of one enzyme (diatom carbonic anhydrase). In all other cases it binds to sites that did not necessarily evolve as metal-ion ligands, or sites which are not highly specific to a certain ion.

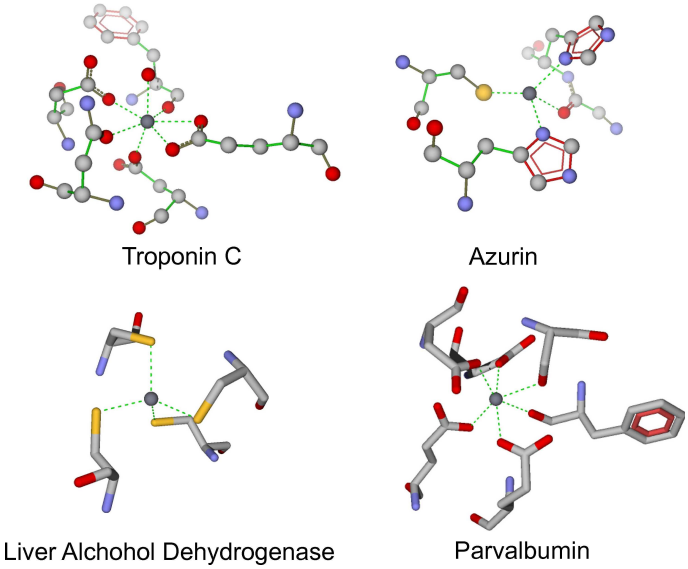

Fig. 1 Cd^{2+} binding sites. Examples of Cd^{2+} binding sites in proteins.

The structures of four Cd^{2+} binding sites are displayed in Figure 1. These examples and the data in Table 1 reveal the versatility of Cd^{2+} binding to proteins. The Cd^{2+}-binding coordination numbers range from 3 to 8, with four being the most prevalent (see Figure 2). With four binding sites, the structure can be either symmetric or distorted tetrahedral (compare the sites for LADH and azurin). The variation in the coordination number appears to be larger than in other transition or group XII metals[15].

3.2 Binding of Cd^{2+} to amino acid side chains and comparison with other metal ions

As other cations, Cd^{2+} is capable of binding to negatively charged and polar amino acid side-chains. Analysis of Cd^{2+}-bound protein structures reveals preference for glu (27% of the sites), his (26%), asp (24%) and cys (10%) residues. Interestingly, Zn^{2+} and Hg^{2+} show different preferences (Figure 3). Zn^{2+} tends to bind to his (36%) and cys (23%) side chains over glu (15%) and asp (17%). Hg^{2+} clearly favours cys (44%) and his (11%), and has no particular affinity to carboxylate residues in proteins. The smaller alkali-earth cations Ca^{2+} and Mg^{2+} have high affinities towards glu (27% / 22%) and asp (36% / 35%), which is also the case for Cd^{2+}. This

Table 1 Cadmium binding proteins in the PDB. Only proteins whose structures are solved at a resolution of 2.5Å and which bind Cd^{2+} in sites that are fully or partially shielded from the solvent are discussed.

PDB code	Protein	No. of Cd^{2+} [a]	Ligands [b]	Coordination number(s) [c]	Native cofactor(s)
2jhf	liver alcohol dehydrogenase	2	Cys(4)	4	Zn^{2+}
			Cys(3)His [S=O]	4	Zn^{2+}
3boe	carbonic anhydrase[d]	1	Cys(2)HisWat	4	
1cdp	parvalbumin	2	Asp(2)Glu(2)Ser[C=O]	6	Ca^{2+}
			Asp(2)Glu(2)Wat[C=O]	6	Ca^{2+}
1lne	thermolysine	4	AspWat(3)[C=O]	6	
			Asp(2)Wat(3)[C=O]	7	Ca^{2+}
			Asp(2)GluWat(2)[C=O]	6	Ca^{2+}
			AspGluHis(2)Wat(2)	8	Zn^{2+}
1aiz	azurin	1	His(2)Cys[C=O]	4	Cu^{2+}
1ncx	troponin c	2	Asp(2)GluAsnWat[C=O]	7	Ca^{2+}
			Asp(2)GluAsnWat[C=O]	7	Ca^{2+}
3g7d	hydroxyethylphosphonate dioxygenase	3	AspHis(2)Wat(3)	6	
			Asp(2)Wat(3)[C=O]	6	
			Asp(2)His(2)Wat(3)	7	
3kbs	D-xylose isomerase	2	Asp(2)Glu(2)Wat	5	Mg^{2+}
			Asp(2)GluHisWat	6	Mg^{2+}
1esf	staphylococcal enterotoxin	1	AspHis(2)[N=H]	4	Zn^{2+}
3lkw	Dengue virus 1 NS2B/NS3 protease	1	GluHis(2)Wat	5	
4mt2	metallothionein	5	Cys(4)	4	
			Cys(4)	4	
			Cys(4)	4	
			Cys(4)	4	
			Cys(4)	4	
1z98	aquaporin SoPIP2	1	Glu(2)Wat(2)	5	
1gm6	lipocalin	1	Glu(2)His	3	
1cfz	endopeptidase HYBD	1	AspGluHis	5	
1vqo	large ribosomal subunit	4	Cys(4)	4	
			Cys(4)	4	
			Cys(4)	4	
			Cys(4)	4	
3jqx	colH collagenase	1	Asp(2)Glu(2)ThrWat	8	Ca^{2+}
3kxd	regulatory domain of calcium-gated potassium channel	2	AspGluWat	4	Ca^{2+}
			AspGlu	3	
3mmu	endoglucanase Cel5A	3	Glu(2)Wat(3)	6	
			Glu(2)Wat(4)	6	
			GluHisWat(2)	4	
1zji	KDO8P synthase	2	AspGluHis	4	
			AspGluHisSer	4	
1feu	ribosomal protein TL	1	GluHisWat[NH2]	4	
2enr	concanavalin A	2	Asp(2)GluHisWat(2)	6	Zn^{2+}
			Asp(2)Wat[C=O]	4	Ca^{2+}
2x7w	endonuclease iv	2	AspGlu(2)HisWat	6	Zn^{2+}
			AspHis(2)[COO-][C=O]	5	Zn^{2+}
3ggf	ste20-like kinase	2	GluHisSer	4	
			GluHisWat[C=O]	4	
1ii0	arsenite-translocating ATPase	3	Cys(3)Wat	4	
			CysHisSerWat	4	
			AspHis(3)Wat	5	
1p9e	parathion hydrolase	1	AspHis(3)Wat(2)	6	Zn^{2+}
1hk7	hsp90 middle domain	1	Glu(4)Wat	8	

[a] Ions bound loosely at the protein surface and coordinated mostly to water are excluded.
[b] Amino acid residues are listed if they bind Cd^{2+} through their side-chains, otherwise a functional group is given.
[c] The coordination numbers are given per ion, and may be higher than the number of metal-ion ligands depending on the functional group (carboxylates may bind Cd^{2+} through one or two of their oxygens).
[d] Structure from diatom, Cd^{2+} is the native cofactor.

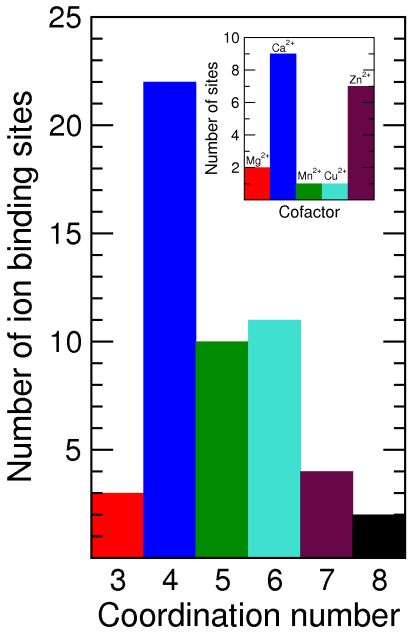

Fig. 2 Properties of the Cd^{2+} binding sites. The total number of sites in PDB structures is displayed as a function of the coordination number. The inset presents the number of binding sites in which Cd^{2+} substitutes another cofactor.

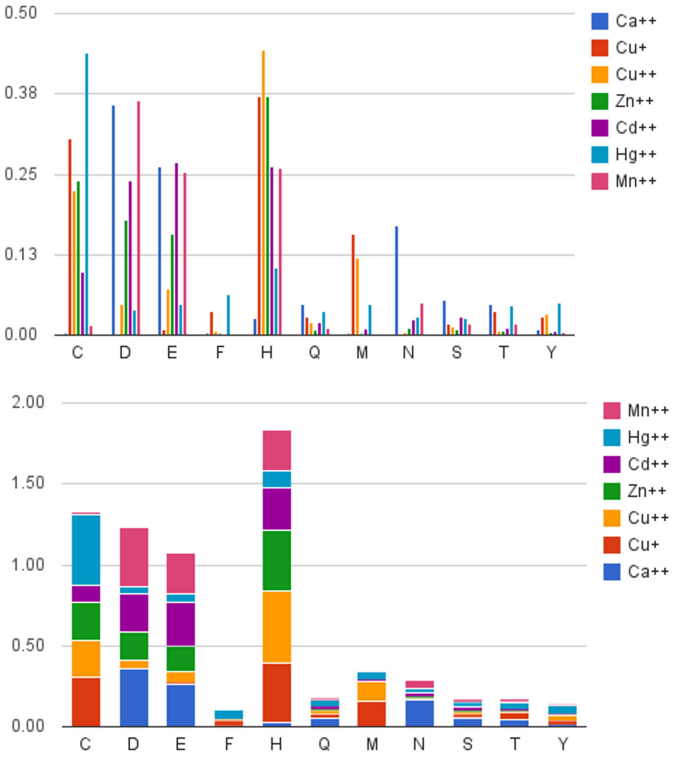

Fig. 3 Preference of Cd^{2+} and other ions to amino acid side chains. The distribution of contacts with amino acid side chains is presented for Ca^{2+} (3492 sites), Mn^{2+} (1027 sites), Cu^{+} (108 sites), Cu^{2+} (408 sites), Zn^{2+} (4529 sites), Cd^{2+} (627 sites), and Hg^{2+} (378 sites). The data is calculated from the PDB and extracted by use of PDBeMotif. Amino acid residues are represented by their single letter code. The ordinate displays the relative distribution (which is cumulative on the lower frame).

can explain why Cd^{2+} can substitute these ions in metalloproteins. Interestingly, Mn^{2+} has amino acid side-chain binding preferences that are similar to Cd^{2+}, with the most common amino acid side chain residues that bind it being asp (37%), glu (25%) and his (26%). Mn^{3+} seems to have the same preferences as Mn^{2+}, except that its affinity towards imidazole (his) residues is higher, but no conclusive evidence on its binding affinity can be drawn due to the paucity of data. The similarity between Cd^{2+} and Mn^{2+} with respect to peptide binding has also been noticed in a survey of metal binding proteins[30]. A comprehensive summary on the energetics of Mn^{2+}-binding in proteins has recently appeared[55].

The similarity between the side-chain preferences of Cd^{2+} and Mn^{2+} in the PDB may seem surprising. Indeed, analysis of the stability constants reveals that Zn^{2+} binds better than Cd^{2+} to amino acid side chains except where the ligand is thiolate (cysteine)[56], whereas Mn^{2+} binds N and O ligands even better than Zn^{2+}[57]. A possible explanation to this discrepancy is that the binding sites of proteins have evolved to bind natural cofactors such as Zn^{2+} and Fe, not Cd^{2+} or Mn. Any binding preferences of the ions are based on their ability to bind to pre-existing sites. However, it should also be stated that Cd^{2+} is more prevalent in high-coordination complexes (coordination number CN=6) than Zn^{2+}[56]. Mn^{2+} is even more prone to form CN=6 complexes[58]. Thus, both the specific protein environment (which is different than that of a free ligand in water) and the tendency towards higher CN may contribute to the surprising similarity between Cd^{2+} and Mn^{2+} with respect to amino acid side chain preference.

3.3 The energetics of Cd^{2+}-binding compared with other ions

The interaction energies between cofactor binding sites and Cd^{2+} were calculated for three proteins and compared with the corresponding energies for binding of the original metal cofactors. All of the surveyed proteins evolved to have ions other than Cd^{2+} in their binding site, yet, they do bind Cd^{2+} when it is in excess. A binding site mimic was used in all cases, with the side chains of cys, asp/glu, his and ser modelled as CH_3S^-, CH_3COO^-, mono-protonated imidazole and CH_3O^-. Backbone carbonyl was modelled as CH_3CHO.

Binding free energies of ions to proteins can in principle be obtained by NMR, potentiometry or isothermal titration

calorimetry. For example, titration spectrometry was used to infer on the binding of Cd^{2+}, and the dissociation constants were in the nM range[59]. Cd-113 NMR and titration can also be used to infer on Cd^{2+} binding affinities[60]. Further studies were carried out with metallothioneins, employing several methods[61]. Such measurements, however, are not routinely used, and cannot be compared with calculated values. When a protein forms a complex with a multivalent ion, the formation of the protein-ion complex often involves significant structural changes. Moreover, changes in the protonation of side chains and dehydration are often part of the process. Calculations of binding energies, however, can only take into account the binding of an ion to a binding site whose shape is already formed (see e.g,.[62]). Even if the calculations cannot be directly compared with the experiment, they can lead to a better understanding on the preference for certain ions. Our own calculations have identified a DFT functional that can be useful for discriminating between Zn and Cd interaction energies[48]. Even if calculations of binding affinities *per se* are challenging, it can be possible to calculate the free energy gain or loss upon binding of one metal instead of another[63]. It is perhaps more difficult, but still possible, to account for partial desolvation of ions such as Mg^{2+} and Ca^{2+} upon binding to proteins[64,65]. Moreover, the calculations can further shed light on the forces that govern the interactions for the different ions by applying energy decomposition analysis methods[50]. Here, I used the Localised Molecular Orbital Energy Decomposition Analysis (LMOEDA) method[34] and its recent extension to free energies in solvent[35]. The calculated components of the energy include electrostatic, exchange, repulsion, polarisation and dispersion. The electrostatic energy is due to Coulomb interactions. Exchange refers to the quantum-mechanical exchange of electrons, that does not have a classical analogue. Repulsion refers to the difference between the total exchange energy and an approximate energy expression for the *supermolecule* (the complex in LMOEDA terminology) calculated with the monomer orbitals forming a single-determinant wave function (E_x^3 in Ref.[34]). The repulsion term arises from the fact that the electron densities of the monomers partially overlap in the supermolecules, and can be somewhat compensated by the favourable exchange. The exchange and repulsion terms are given together in other EDA schemes, as "exchange repulsion"[66] or the closely related ΔE^{Pauli}[67]. The polarisation interaction is due to reshaping of the distribution of electrons upon binding. For example, the effective charge on a multivalent ion which is bound to a protein will be smaller than its formal charge, due to interactions with the ligands (the metal-ligand bond is partially covalent). Consequently, the reshaping of the electric distribution yields a favourable energy contribution $\Delta E^{pol} < 0$. Finally, the DFT dispersion interaction is defined as the difference between the energies calculated by applying the DFT correlation functional on the supermolecule and the monomers.

3.3.1 Liver alcohol dehydrogenase: Zn^{2+} binding Liver alcohol dehydrogenase (LADH), is the first metal-binding protein studied here. The enzyme naturally binds two Zn^{2+} ions, one at the catalytic site and one at a $(cys)_4$ structural site. The calculations reported here refer to the structural site because of the promiscuity of the $(cys)_4$ binding domain and because an inhibitor is coordinated to the Cd^{2+} ion in the catalytic site (see Ref.[68] for EDA of the catalytic site in Zn^{2+}-bound LADH). The tetrahedral metal-binding site (Figure 1) differs slightly between the Zn^{2+} and Cd^{2+} binding proteins. Distances are shorter in the first case, and the organisation is less symmetric (Table 2).

	d_1	d_2	d_3	d_4	a_1	a_2	a_3	a_4
Zn^{2+}	2.31	2.34	2.35	2.35	103°	106°	117°	119°
Cd^{2+}	2.54	2.55	2.55	2.56	104°	106°	106°	107°

Table 2 Differences in the structural metal binding site of LADH between the native Zn^{2+} and Cd^{2+}. The four Zn-S distances (d_1-d_4, in Å) and S-Zn-S angles (a_1-a_4) are presented. The data is for the PDB structures 1JHG and 1JHF
.

Interaction energies were calculated for the Cd^{2+} and Zn^{2+} binding sites both in the gas and aqueous phases (see Table 3). In addition, tetrahydrofuran (THF) was used to mimic binding in a protein environment. THF is a polar solvent with a dielectric constant of 7.58 that is compatible with that of proteins[69]. Interactions with the solvent were approximated by use of the polarisable continuum model (PCM)[51].

The gas phase interaction energy is dominated by a very favourable electrostatic interaction between the doubly charged cation and the negatively charged binding site. Polarisation plays an additional role, as the excess positive charge of the cation is distributed between the ligands. In both cases, the smaller Zn^{2+} ion is preferred over the larger Cd^{2+}, whereas the weaker exchange, repulsion and dispersion are more favourable with Cd^{2+} than Zn^{2+}. In the solvent phase, desolvation is also a major contribution. The unfavourable desolvation free energy is larger for Zn^{2+}, but the binding of Zn^{2+} is still favoured by some 46–49 kcal/mol in water or THF according to the MP2 calculation.

3.3.2 Parvalbumin: Ca^{2+} binding Parvalbumin is a calcium binding protein which is involved in calcium signalling. It binds two Ca^{2+} ions in octahedron-shaped binding sites, which are termed EF and CD sites. The structures considered for the calculations reported here are for parvalbumin from *Cyprinus carpis* (carp)[40] and the ligands are four carboxyls, one carbonyl backbone oxygen and one water or hydroxyl oxygen. Examination of the ion-ligand distances (Table 4) reveals slight differences of the ion-ligand interactions. First,

	ΔE^{elec}	ΔE^{ex}	ΔE^{rep}	ΔE^{pol}	ΔE^{disp}		ΔE^{int}
MP2:							
Cd^{2+}	-824.4	-92.6	202.6	-254.4	-20.1		-988.9
Zn^{2+}	-853.9	-78.9	207.0	-308.7	-7.3		-1041.9
	ΔG^{elec}	ΔG^{ex}	ΔG^{rep}	ΔG^{pol}	ΔG^{disp}	ΔG^{desol}	ΔG^{int}
aqueous phase (PCM), MP2:							
Cd^{2+}	-930.5	-94.0	203.9	-145.3	-36.1	657.8	-344.1
Zn^{2+}	-1005.4	-82.4	213.5	-160.5	-30.6	675.1	-390.3
aqueous phase (THF), MP2:							
Cd^{2+}	-918.6	-93.1	202.2	-156.4	-35.3	575.7	-425.4
Zn^{2+}	-990.8	-81.9	212.2	-174.3	-29.7	590.3	-474.2

Table 3 Energy decomposition analysis for the interaction between Cd^{2+} or Zn^{2+} and the structural ion binding sites. LMOEDA interaction energies are given in kcal/mol. The energies were calculated for the crystal coordinates (with added hydrogens optimised) or for the optimised binding site. MP2 calculations were carried out with aug-cc-pVDZ.

metal-ligand distances tend to be slightly smaller for Cd^{2+} compared to Ca^{2+}. Second, calcium binds to the two oxygens of glutamate residues Glu62 and Glu101 at similar distances (within approx. 0.1Å or less) but Cd^{2+} binds in an asymmetric fashion, where the distance to one of the oxygens is larger by some 0.4Å.

Interestingly, Cd^{2+} is preferred over Ca^{2+} in the absence or presence of solvent. The softness of cadmium leads to smaller ion-ligand distances and therefore more favourable contributions from electrostatics, polarisation and dispersion, whereas the desolvation interactions are comparable in magnitude (Table 5). Overall, Cd^{2+} binding is favoured by some 61-67 kcal/mol in THF.

Interestingly, ΔE^{ex} is found here to be positive, which is unlike the previous case (LADH) and the reactions reported in Ref.[34]. The calculation of ΔE^{ex} is carried out by operating the exchange functional on the supermolecule and the monomers, and subtracting the monomeric exchange energies from those of the supermolecule; it follows that this interaction will strongly depend on the functional form.

3.3.3 Azurin: Cu^{2+} binding Azurin is a copper binding protein characterised by its distinctive blue colour. The colour arises from redistribution of charges between the copper ion and a thiol ligand. The copper ion binds to cys, his, his and a backbone carbonyl oxygen in a distorted tetrahedral orientation (Figure 2). In some structures only three ligands are present (excluding the carbonyl) whereas a fifth ligand (met) may also be present in others. The structure of the binding site is to a large extent independent of the copper oxidation state, i.e., metal ion coordination is almost the same when Cu^{+} is bound rather than Cu^{2+} [70]. The structure of Cd^{2+}-substituted azurin (PDB code 1aiz, [71]) reveals some variations in the binding site, which are presented in Table 6. Most notably, the distances between the ligands and Cd^{2+} are larger. It should also be mentioned that the structure of azurin was resolved as a dimer in the crystallographic unit, in which one of the Cu^{+} binding sites has only three ligands. Cd^{2+} has four ligands in both protein chains, but there is some variation in the metal-ligand distances.

	Gly45O	His46Nδ1	Cys112Sγ	His117Nδ1
Cd^{2+}, chain A	2.80	2.27	2.34	2.22
Cd^{2+}, chain B	2.72	2.23	2.43	2.20
Cu^{2+}, chain A	-	2.02	2.18	1.87
Cu^{2+}, chain B	2.66	1.84	2.31	1.76

Table 6 Links between the structural metal binding site of azurin and Cu^{2+} or Cd^{2+}. Distances are in Å. The data is for the PDB structures 1AIZ and 1AZC.

When calculating the interaction energies between the ions and the metal, only the four-coordinated copper site has been used, to allow a comparison between the ions. The interaction energy calculations (Table 7) reveal that the azurin has a much higher affinity towards its native copper ligand. Cu^{2+} is preferred over Cd^{2+} with respect to all internal interactions except the repulsion. Interactions in solvent could not be obtained for the Cu complex. Note that the polarisation energy is also more favourable for Cu^{2+} by more than 80 kcal/mol. This effect may be unique to azurin, as the polarisation of the cysteine ligand yields the distinctive absorbance of the protein around 600 nm.

The availability of two crystallographic units but the same binding site enables the comparison between them. Interestingly, whereas the overall binding (free) energy is similar to 1.5-2.9 kcal/mol, larger differences (up to 12 kcal/mol for ΔG^{elec} and 16 kcal/mol for ΔG^{rep}) are observed for the different energy components, revealing the sensitivity of LMOEDA to small modifications of the geometry.

	Asp51Oδ1	Asp53Oδ1	Ser55Oγ	Phe57O	Glu59Oε1	Glu62Oε1	Glu62Oε2
Ca^{2+}	2.23	2.36	2.60	2.27	2.48	2.55	2.43
Cd^{2+}	2.12	2.36	2.58	2.25	2.34	2.77	2.36
	Asp90δ2	Asp92Oδ1	Asp94Oδ1	Lys96O	Glu101Oε1	Glu101Oε2	HOH
Ca^{2+}	2.25	2.42	2.44	2.29	2.51	2.49	2.51
Cd^{2+}	2.26	2.40	2.17	2.30	2.26	2.66	2.37

Table 4 Links between the structural metal binding site of parvalbumin and Ca^{2+} or Cd^{2+}. Distances are in Å. The data is for the PDB structures 1CDP and 5CPV[40].

	ΔE^{elec}	ΔE^{ex}	ΔE^{rep}	ΔE^{pol}	ΔE^{disp}		ΔE^{int}
ion binding site 1 (CD site):							
Cd^{2+}	-894.3	6.6	161.0	-196.0	-86.9		-1009.6
Ca^{2+}	-866.1	14.4	118.8	-145.9	-69.0		-947.7
ion binding site 2 (EF site):							
Cd^{2+}	-1024.5	12.0	139.6	-202.2	-85.6		-1160.6
Ca^{2+}	-998.7	14.8	117.6	-161.6	-69.4		-1097.2
	ΔG^{elec}	ΔG^{ex}	ΔG^{rep}	ΔG^{pol}	ΔG^{disp}	ΔG^{desol}	ΔG^{int}
Aqueous phase (PCM)							
ion binding site 1 (CD site):							
Cd^{2+}	-967.3	6.0	163.2	-128.8	-82.8	601.8	-407.8
Ca^{2+}	-941.1	14.1	121.7	-77.7	-64.8	600.1	-347.6
ion binding site 2 (EF site):							
Cd^{2+}	-1124.7	11.1	142.9	-108.0	-82.0	707.8	-452.9
Ca^{2+}	-1108.7	14.1	122.7	-60.7	-64.5	712.6	-384.6
THF phase (PCM):							
ion binding site 1 (CD site):							
Cd^{2+}	-957.4	6.1	162.9	-137.8	-83.4	527.4	-482.3
Ca^{2+}	-930.8	14.2	121.3	-87.0	-65.4	525.9	-421.9
ion binding site 2 (EF site):							
Cd^{2+}	-1111.6	11.3	142.3	-120.2	-82.6	619.8	-540.8
Ca^{2+}	-1094.5	14.2	122.0	-73.7	-65.2	623.8	-473.5

Table 5 Energy decomposition analysis for the interaction between Cd^{2+} or Ca^{2+} and the ion binding sites of carp parvalbumin. LMOEDA interaction energies are given in kcal/mol. The energies were calculated using M06/def2TZVP for the crystal coordinates (with added hydrogens optimised)

3.4 Limitations of the model and calculations

LMOEDA calculations were carried out with models of the binding site, where calculations with a triple zeta valence basis set could be made for all atoms, rather than using smaller basis sets and larger binding site models or combination of QM and molecular mechanics (QM/MM)[72]. This allows for comparison between the ions and their binding to immediate ligands, but neglects the effects of second shell ligands and binding site flexibility which cannot be accurately accounted for.

It should also be pointed out that LMOEDA calculations based on DFT energies strongly depend on the form of the DFT exchange and correlation functionals, as discussed above (see also[48]). More experimental data on ion binding, and assessment of their accuracy[73,74] are necessary to improve the quality of the calculations. New methods and applications of energy decomposition analysis are also the subject if ongoing research[50,75,76].

Another limitation of the calculation is the necessity to rely on crystallographic data. QM refinement of the structures was performed in the case of LADH, and the results were similar with respect to the interaction energies and EDA. However, the structure of the binding site is clearly influenced by constraints due to non-binding residues, which cannot be accounted for by QM optimisation. For example, the Zn^{2+} binding site in LADH deviates from the almost tetrahedral structure that results from geometry optimisation. Calculations carried out for the two structurally equivalent binding sites in Cd^{2+}-bound azurin suggest that the individual components of the interaction energy are more sensitive to small changes in the location of ligands.

	ΔE^{elec}	ΔE^{ex}	ΔE^{rep}	ΔE^{pol}	ΔE^{disp}		ΔE^{int}
Cd^{2+}, chain A	-421.4	-23.4	192.7	-221.1	-63.7		-536.9
Cd^{2+}, chain B	-430.2	-29.0	209.1	-224.2	-63.1		-537.4
Cu^{2+}	-519.4	-47.6	338.6	-319.9	-71.5		-619.8
	ΔG^{elec}	ΔG^{ex}	ΔG^{rep}	ΔG^{pol}	ΔG^{disp}	ΔG^{desol}	ΔG^{int}
Aqueous phase (PCM)							
Cd^{2+}, chain A	-465.7	-24.9	195.5	-180.9	-60.9	254.3	-282.6
Cd^{2+}, chain B	-474.3	-30.4	211.6	-184.0	-60.3	257.7	-279.7
THF phase (PCM)							
Cd^{2+}, chain A	-459.1	-24.6	194.8	-186.7	-61.4	222.7	-314.3
Cd^{2+}, chain B	-467.5	-30.1	210.9	-190.0	-60.7	225.6	-311.8

Table 7 Energy decomposition analysis for the interaction between Cd^{2+} or Cu^{2+} and the ion binding site of azurin. LMOEDA interaction energies are given in kcal/mol. M06/def2TZVP energies were calculated for the crystal coordinates (with added hydrogens optimised). BSSE is account for only with Cd^{2+}.

4 Conclusions

4.1 Cadmium binding is versatile

Cd^{2+}, like Zn^{2+} has a filled d electron configuration and is therefore not a transition metal according to the definition of the International Union of Pure and Applied Chemistry (IUPAC)[77]. Indeed, it can bind to zinc-binding proteins and enzymes with little structural and catalytic effects. The analysis of Cd^{2+}-containing protein structures, however, revealed that it can also replace other cofactors, most notably Ca^{2+} that has very different binding preferences. Moreover, Cd^{2+} can bind to 3-8 ligands, whereas other ions show a more narrow distribution. This may explain some of the toxic effects that Cd^{2+} exerts through protein-binding and suggest that a combination of ions may be used to partially relieve its toxic effects[11]. Interestingly, the calculations of binding free energies reveal that Cd^{2+} is favoured over Ca^{2+} in a calcium-binding protein, which may explain why Cd^{2+} replace Ca^{2+} more than any other ion in protein structures (Figure 2). Discussion of binding preference for ligands often follows the ideas developed originally by Irving and Williams, who analysed the stability constants of many metal complexes and found them to be independent of the ligand to a wide extent[78]. Even if the binding affinity agree with the Irving-Williams series, preferences to some ligands (e.g., O- or N-) is also important[57]. The complexity of biomolecules have made them suitable for binding of specific ions. Sulphur-containing groups, for example, are preferred for Cd^{2+} binding[56]. Furthermore, the binding of ions is influenced by the exact geometry of the binding site as well as the first and second coordination shells[79]. EDA calculations, as performed here, may in the future be used for a better understanding of biomolecular metal-binding preferences.

4.2 LMOEDA calculation explain ion preferences

LMOEDA calculations could explain some of the difference between the affinities of metalloproteins towards specific ligands, in spite of the simplicity of the structural models as presented here,. As expected, the main contribution in all cases is electrostatic. As all ions here are divalent, this contribution is affected by the ion's size. The preference to smaller ions (Cu^{2+} and Zn^{2+}) is somewhat offset by the higher cost of desolvating them, even in a low dielectric solvent. On the other hand, Cd^{2+}, which is softer than Ca^{2+}, is favoured mostly because is polarises the binding site better than the similarly sized Ca^{2+}.

4.3 LMOEDA is sensitive to the geometry

The calculations show that different contributions to the energy vary to a greater extent that the total energy (or free energy). This is due to the partitioning scheme, and should not be viewed as a particular strength or weakness of LMOEDA. Likewise, the relative contribution of the free energy components may depend on the QM methods used in the calculations[48] although this should not be expected to modify the conclusions when the differences between the individual contributions is as large as in bioinorganic metal complexes.

Acknowledgements

The author would like to thank Dr. Peifeng Su for his assistance with the LMOEDA calculations with PCM and anonymous referees for comments on the manuscript. Research at the Computational Chemistry and Biochemistry research group (CCBG) at the Linnæus university is supported by the Linnus University Centre of Excellence Biophysical Chemistry. The computations were performed on resources provided by the Swedish National Infrastructure for Computing

(SNIC) at the center for scientific and technical computing for research at Lund University (Lunarc) and the center for high-performance computing at the Royal Institute of Technology in Stockholm (PDC). Rossen Apostolov and Jonathan Vincent at PDC are acknowledged for assistance concerning technical and implementation aspects.

References

1 M. Stromeyer, *Annals of Phil*, 1819, **14**, 269–274.
2 T. Hawkins, H. Matthews and C. Hendrickson, *Int J LCA*, 2006, **11**, 38–48.
3 S. Trakhanov, D. I. Kreimer, S. Parkin, G. F. Ames and B. Rupp, *Protein Sci*, 1998, **7**, 600–604.
4 G. Nordberg, *Lancet Oncol*, 2006, **7**, 99–101.
5 HELCOM, 2010. Hazardous substances in the Baltic Sea - An integrated thematic assessment of hazardous substances in the Baltic Sea, 2010.
6 HELCOM, 2011. The Fifth Baltic Sea Pollution Load Compilation (PLC-5), 2011.
7 T. Nawrot, M. Plusquin, J. Hogervorst, H. A. Roels, H. Celis, L. Thijs, J. Vangronsveld, E. Van Hecke and J. A. Staessen, *Lancet Oncol*, 2006, **7**, 119–126.
8 P. Joseph, *Toxicol Appl Pharmacol*, 2009, **238**, 272–279.
9 N. Johri, G. Jacquillet and R. Unwin, *Biometals*, 2010, **23**, 783–792.
10 D. Sutoo, K. Akiyama and S. Imamiya, *Arch Toxicol*, 1990, **64**, 161–164.
11 V. Matović, A. Buha, Z. Bulat and D. Đukić-Ćosić, *Arh Hig Rada Toksikol*, 2011, **62**, 65–76.
12 M. Rowinska-Zyrek, D. Witkowska, S. Bielinska, W. Kamysz and H. Kozlowski, *Dalton Transactions*, 2011, **40**, 5604–5610.
13 K. Krzywoszynska, M. Rowinska-Zyrek, D. Witkowska, S. Potocki, M. Luczkowski and H. Kozlowski, *Dalton Transactions*, 2011, **40**, 10434–10439.
14 S. Potocki, M. Rowinska-Zyrek, D. Valensin, K. Krzywoszynska, D. Witkowska, M. Luczkowski and H. Kozlowski, *Inorganic Chemistry*, 2011, **50**, 6135–6145.
15 R. H. Holm, P. Kennepohl and E. I. Solomon, *Chem Rev*, 1996, **96**, 2239–2314.
16 T. W. Lane, M. A. Saito, G. N. George, I. J. Pickering, R. C. Prince and F. M. Morel, *Nature*, 2005, **435**, 42–42.
17 E. Boyle, F. Sclater and J. Edmond, *Nature*, 1976, **263**, 42–44.
18 G. Endo and S. Silver, *J Bacteriol*, 1995, **177**, 4437–4441.
19 J. S. Cavet, A. I. Graham, W. Meng and N. J. Robinson, *J Biol Chem*, 2003, **278**, 44560–44566.
20 F. Neese, *Coord Chem Rev*, 2009, **253**, 526–563.
21 P. Comba and M. Kerscher, *Coord Chem Rev*, 2009, **253**, 564 – 574.
22 U. Ryde and L. Hemmingsen, *J Biol Inorg Chem*, 1997, **2**, 567–579.
23 C. Chang and P. Huang, *Protein Eng*, 1996, **9**, 1165–1172.
24 L. Rulíšek and Z. Havlas, *J Am Chem Soc*, 2000, **122**, 10428–10439.
25 L. Rulíšek and Z. Havlas, *J Phys Chem A*, 2002, **106**, 3855–3866.
26 L. Rulíšek and Z. Havlas, *J Phys Chem B*, 2003, **107**, 2376–2385.
27 M. Kožíšek, A. Svatoš, M. Buděínskỳ, A. Muck, M. C. Bauer, P. Kotrba, T. Ruml, Z. Havlas, S. Linse and L. Rulíšek, *Chem Eur J*, 2008, **14**, 7836–7846.
28 R. Friedman, *J Phys Chem B*, 2011, **115**, 9213–9223.
29 S. F. Sousa, A. B. Lopes, P. A. Fernandes and M. J. Ramos, *Dalton Trans*, 2009, 7946–7956.
30 J. A. C. Tamames and M. J. Ramos, *J Mol Model*, 2011, **17**, 429–442.
31 B. Seebeck, I. Reulecke, A. Kämper and M. Rarey, *Proteins*, 2008, **71**, 1237–1254.
32 V. V. Kursenko, Euromedica 2010 - abstracts, 2010.
33 M. Litvak, Euromedica 2010 - abstracts, 2010.
34 P. Su and H. Li, *J Chem Phys*, 2009, **131**, 014102.
35 P. Su, H. Liu and W. Wu, *J Chem Phys*, 2012, **137**, 034111.
36 A. Golovin and K. Henrick, *BMC Bioinformatics*, 2008, **9**, 312–312.
37 R. Friedman and A. Caflisch, *ChemMedChem*, 2009, **4**, 1317–1326.
38 B. Roux, *J Phys Chem B*, 2012, **116**, 6966–6979.
39 R. Meijers, H. W. Adolph, Z. Dauter, K. S. Wilson, V. S. Lamzin and E. S. Cedergren-Zeppezauer, *Biochemistry*, 2007, **46**, 5446–5454.
40 A. L. Swain, R. H. Kretsinger and E. L. Amma, *J Biol Chem*, 1989, **264**, 16620–16628.
41 W. E. Shepard, R. L. Kingston, B. F. Anderson and E. N. Baker, *Acta Crystallogr D Biol Crystallogr*, 1993, **49**, 331–343.
42 T. Hassinen and M. Perakyla, *J Comput Chem*, 2001, **22**, 1229–1242.
43 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. Elbert, M. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, *J. Comput. Chem.*, 1993, **14**, 1347–63.
44 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–214.
45 F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.
46 N. B. Balabanov and K. A. Peterson, *J. Chem. Phys.*, 2005, **123**, 064107.
47 K. A. Peterson and C. Puzzarini, *Theoretical Chemistry Accounts*, 2005, **114**, 283–296.
48 E. Ahlstand, D. Spångberg, K. Hermansson and R. Friedman, *Binding of Zn^{2+} and Cd^{2+} to biologically-relevant ligands (amino acid side chains and water): interaction energies calculated with DFT and ab initio methods*, 2013, In preparation.
49 K. L. Schuchardt, B. T. Didier, T. Elsethagen, L. Sun, V. Gurumoorthi, J. Chase, J. Li and T. L. Windus, *J Chem Inf Model*, 2007, **47**, 1045–1052.
50 M. v. Hopffgarten and G. Frenking, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2011, n/a–n/a.
51 S. Miertus, E. Scrocco and J. Tomasi, *Chem. Phys.*, 1981, **55**, 117–129.
52 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441.
53 S. Batsanov, *Inorg. Mater.*, 2001, **37**, 871–885.
54 S. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.
55 M. H. Khodabandeh, H. Reisi, M. D. Davari, K. Zare, M. Zahedi and G. Ohanessian, *Chemphyschem*, 2013, **14**, 1733–1745.
56 I. Sóvágó and K. Várnagy, *Met Ions Life Sci*, 2013, **11**, 275–302.
57 H. Sigel and D. B. McCormick, *Acc Chem Res*, 1970, **3**, 201–208.
58 M. Dudev, J. Wang, T. Dudev and C. Lim, *J Phys Chem B*, 2006, **110**, 1889–1895.
59 J. L. Michalek, S. J. Lee and S. L. Michel, *J Inorg Biochem*, 2012, **112**, 32–38.
60 O. Iranzo, T. Jakusch, K.-H. Lee, L. Hemmingsen and V. L. Pecoraro, *Chem Eur J*, 2009, **15**, 3761–3772.
61 E. Freisinger and M. VaÅąák, *Met Ions Life Sci*, 2013, **11**, 339–371.
62 M. Enescu, J.-P. Renault, S. Pommeret, J.-C. Mialocq and S. Pin, *Phys Chem Chem Phys*, 2003, **5**, 3762–3767.
63 T. Y. Yang, T. Dudev and C. Lim, *J Am Chem Soc*, 2008, **130**, 3844–3852.
64 T. Dudev and C. Lim, *J Am Chem Soc*, 2006, **128**, 1553–1561.
65 B. Sharma, J. S. Rao and G. N. Sastry, *J Phys Chem A*, 2011, **115**, 1971–1984.
66 K. Morokuma, *J. Chem. Phys.*, 1971, **55**, 1236.
67 G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders and T. Ziegler, *J Comput Chem*, 2001, **22**, 931–967.
68 B. de Courcy, J.-P. Dognon, C. Clavaguéra, N. Gresh and J.-P. Piquemal, *International Journal of Quantum Chemistry*, 2011, **111**, 1213–1221.
69 R. Friedman, E. Nachliel and M. Gutman, *Biophys. J.*, 2005, **89**, 768–81.
70 W. E. B. Shepard, B. F. Anderson, D. A. Lewandoski, G. E. Norris and E. N. Baker, *J Am Chem Soc*, 1990, **112**, 7817–7819.

71 K. A. Blackwell, B. F. Andersen and E. N. Baker, *Acta Crystallogr D Biol Crystallogr*, 1994, **50**, 263–270.
72 A. Warshel and M. Levitt, *J Mol Biol*, 1976, **103**, 227–249.
73 G. Berthon, *Pure Appl. Chem.*, 1995, **67**, 1117–1240.
74 T. Nakashima, H. Waki and M. Toh, *J. Solut. Chem.*, 2010, **39**, 51–56.
75 M. Isegawa, J. Gao and D. G. Truhlar, *J. Chem. Phys.*, 2011, **135**, 084107.
76 S. N. Steinmann, C. Corminboeuf, W. Wu and Y. Mo, *J. Phys. Chem. A*, 2011, **115**, 5467–5477.
77 A. D. McNaught and W. A, *IUPAC. Compendium of Chemical Terminology, 2nd ed.*, Blackwell Scientific Publications, Oxford, 1997.
78 H. Irving and R. J. P. Williams, *J. Chem. Soc.*, 1953, 3192–3210.
79 T. Dudev and C. Lim, *Annu Rev Biophys*, 2008, **37**, 97–116.