J|A|C|S
ARTICLES
Published on Web 01/04/2006

A DFT/CDM Study of Metal−Carboxylate Interactions in Metalloproteins: Factors Governing the Maximum Number of Metal-Bound Carboxylates

Todor Dudev[†] and Carmay Lim*[,†,‡]

Contribution from the Institute of Biomedical Sciences, Academia Sinica, Taipei 115, and Department of Chemistry, National Tsing Hua University, Hsinchu 300, Taiwan

Received August 23, 2005; E-mail: carmay@gate.sinica.edu.tw

Abstract: The number of negatively charged metal-bound Asp/Glu residues determines the net charge of the carboxylate-rich metal-binding site, which has been found to play a role in enhancing the affinity and/or selectivity of a protein cavity for a given metal cofactor. Therefore, it is of interest to know the maximum number of carboxylates that could bind to a given metal (M^{q+}) of charge q and the key factors determining this upper limit in protein cavities, which are usually relatively buried. Using density functional theory combined with the continuum dielectric method to compute the $H_2O \rightarrow CH_3COO^-$ exchange free energies, the maximum number of carboxylates bound to M^{q+} in a relatively buried metal-binding site is found to depend on (i) the metal charge, q, (ii) the carboxylate-binding mode, and (iii) the first-shell carboxylate−second-shell ligand interactions. The maximum number of carboxylates bound to M^{q+} in a fully/partially solvent inaccessible protein cavity would not likely exceed $q + 2$ if (a) the metal-bound Asp/Glu side chains are hydrogen bonded to a Lys/Arg side chain or several peptide backbone amides/Asn/Gln side chains in the metal's second coordination shell or (b) at least one acidic residue binds bidentately, as opposed to monodentately, to the metal cofactor. This number is reduced to $q + 1$ in the absence of stabilizing interactions from outer-shell ligand(s) and if all the carboxylates are bound monodentately to the metal cofactor in a buried cavity. The computational results are consistent with findings from a PDB survey of uni-, di-, and trivalent metal-binding sites containing Asp/Glu residues.

Introduction

Carboxylate-containing residues, viz., deprotonated aspartates (Asp^-) and glutamates (Glu^-), are among the most common amino acid ligands found coordinated to metal cations in metalloproteins.[1−16] They are unique among the 20 amino acids in being able to bind the metal cation either monodentately (via one of the carboxylate oxygens) or bidentately (with both carboxylate oxygens). In general, they share a common role in sequestering the metal cation from physiological fluids because their interactions with the metal cofactor in a protein cavity are much more favorable than those of noncharged protein ligands.[13,16−20] In certain cases, they may also play a role in enhancing the affinity and/or selectivity of a protein cavity for a given metal cofactor by determining the net charge of the carboxylate-binding pocket. For example, the three Asp^-/Glu^- side chains lining EF-hand calcium-binding sites contribute a net ligand charge of −3, which helps to selectively bind Ca^{2+} against a much higher background concentration of monovalent cations such as Na^+ and K^+.[21−24] The maximum number of carboxylates that could bind to a metal (M^{q+}) of charge q, denoted by $Max^{COO}(M^{q+})$, is therefore of interest, but the $Max^{COO}(M^{q+})$ and the key factors determining this upper limit are generally not known (to the best of our knowledge).

Previous calculations have shown that Mg^{2+} will not exchange all its first-shell water molecules for Asp^-/Glu^- ligands.[13] In a solvent inaccessible protein cavity, sequential binding of the first three Asp/Glu carboxylates (modeled by

[†] Academia Sinica.
[‡] National Tsing Hua University.

(1) Vallee, B. L.; Auld, D. S. *Biochemistry* **1990**, *29*, 5647−5659.
(2) Christianson, D. W. *Adv. Protein Chem.* **1991**, *42*, 281−355.
(3) Coleman, J. E. *Annu. Rev. Biochem.* **1992**, *61*, 897−946.
(4) Bertini, I.; Gray, H. B.; Lippard, S. J.; Valentine, J. S. *Bioinorganic Chemistry*; University Science Books: Mill Valley, CA, 1994.
(5) Jernigan, R.; Raghunathan, G.; Bahar, I. *Curr. Opin. Struct. Biol.* **1994**, *4*, 256−263.
(6) Casareno, R. L. B.; Li, D.; Cowan, J. A. *J. Am. Chem. Soc.* **1995**, *117*, 11011−11012.
(7) Dismukes, G. C. *Chem. Rev.* **1996**, *96*, 2909−2926.
(8) Lipscomb, W. N.; Strater, N. *Chem. Rev.* **1996**, *96*, 2375−2433.
(9) Alberts, I. L.; Nadassy, K.; Wodak, S. J. *Protein Sci.* **1998**, *7*, 1700−1716.
(10) Rulisek, L.; Vondrasek, J. *J. Inorg. Biochem.* **1998**, *71*, 115−127.
(11) Bock, C. W.; Katz, A. K.; Markham, G. D.; Glusker, J. P. *J. Am. Chem. Soc.* **1999**, *121*, 7360−7372.
(12) Christianson, D. W.; Cox, J. D. *Annu. Rev. Biochem.* **1999**, *68*, 33−57.
(13) Dudev, T.; Cowan, J. A.; Lim, C. *J. Am. Chem. Soc.* **1999**, *121*, 7665−7673.
(14) Bertini, I., Sigel, A., Sigel, H., Eds. *Handbook on Metalloproteins*; Marcel Dekker: New York, 2001.
(15) Pidcock, E.; Moore, G. R. *J. Biol. Inorg. Chem.* **2001**, *6*, 479−489.
(16) Dudev, T.; Lim, C. *Chem. Rev.* **2003**, *103*, 773−787.
(17) Garmer, D. R.; Gresh, N. *J. Am. Chem. Soc.* **1994**, *116*, 3556−3567.
(18) Dudev, T.; Lim, C. *J. Phys. Chem. B* **2000**, *104*, 3692−3694.
(19) Rulisek, L.; Havlas, Z. *J. Am. Chem. Soc.* **2000**, *122*, 10428−10439.
(20) Rulisek, L.; Havlas, Z. *J. Phys. Chem. A* **2002**, *106*, 3855−3866.
(21) Falke, J. J.; Drake, S. K.; Hazard, A. L.; Peersen, O. B. *Q. Rev. Biophys.* **1994**, *27*, 219−290.
(22) Drake, S. K.; Lee, K. L.; Falke, J. J. *Biochemistry* **1996**, *35*, 6697−6705.
(23) Drake, S. K.; Zimmer, M. A.; Kundrot, C.; Falke, J. J. *J. Gen. Physiol.* **1997**, *110*, 173−184.
(24) Drake, S. K.; Zimmer, M. A.; Miller, C. L.; Falke, J. J. *Biochemistry* **1997**, *36*, 9917−9926.

$HCOO^-$) *monodentately* to Mg^{2+} (where one of the three carboxylates first binds to the metal ion, followed by the second carboxylate, then the third carboxylate) is highly favorable, characterized by a large free energy gain.[13] However, binding of a fourth negatively charged carboxylate *monodentately* to a negatively charged $[Mg(H_2O)_3(HCOO)_3]^-$ complex is unfavorable due to repulsive charge−charge interactions.[13] Thus, the calculations suggest a maximum of *three monodentately* bound carboxylates in a *buried* Mg^{2+}-binding site. However, it is not clear if this upper bound applies to other divalent cations as well and how Max^{COO} varies for uni-, tri-, or tetravalent cations.

In this study we elucidate the key factors determining the $Max^{COO}(M^{q+})$ in proteins using a combined density functional theory (DFT) and continuum dielectric method (CDM) approach in conjunction with a Protein Data Bank (PDB)[25] survey of metal-binding sites containing Asp/Glu residues. Specifically, we evaluate if and how the $Max^{COO}(M^{q+})$ depends on (i) the charge, *q*, of a metal cation, (ii) the binding mode (monodentate vs bidentate binding, and sequential vs concerted binding) of the carboxylate, and (iii) the interactions of the metal-bound carboxylate with amino acid residues/water molecules in the second coordination shell. We validate the findings from the DFT/CDM calculations by comparison with results from the PDB survey. Based on the combined results from the DFT/CDM calculations and the PDB survey, we derive guidelines for estimating the $Max^{COO}(M^{q+})$ in metalloproteins.

Methods

Models Used. The metal cation and ligands forming the first and second coordination shell were treated quantum mechanically to account for electronic effects such as polarization of the participating entities and charge transfer from the ligands to the metal ion. The rest of the protein was treated as a dielectric continuum, characterized by a dielectric constant varying from 4 for totally buried metal-binding sites to 80 for fully solvent-exposed sites. The side chains of Asp^-/Glu^- and Lys are modeled by CH_3COO^- and $CH_3NH_3^+$, respectively, while those of Asn/Gln and the peptide backbone group are modeled by $HCONH_2$.

To study the effect of the cation charge on the $Max^{COO}(M^{q+})$, we studied the natural metal cofactors in metalloproteins, Mg^{2+}, Ca^{2+}, and Zn^{2+}, as well as non-natural metal cofactors, lithium (Li^+), lanthanum (La^{3+}), and zirconium (Zr^{4+}). In proteins, Mg^{2+} is predominantly six-coordinated,[5] as in aqueous solution, but Ca^{2+} is usually seven-coordinated, while Zn^{2+} is tetracoordinated;[5] thus complexes containing these three natural metal cofactors were modeled as MgL_6, CaL_7, and ZnL_4, respectively (L = H_2O or CH_3COO^-). On the other hand, complexes containing the three non-natural metal cofactors in proteins were modeled as LiL_4, LaL_8, and ZrL_6 (L = H_2O or CH_3COO^-), in line with the most common coordination number (CN) found for these metals in the Cambridge Structure Database (CSD)[26,27] and PDB surveys.[15]

DFT Calculations. Full geometry optimization for each complex studied was carried out using the Gaussian 03 program[28] employing the S-VWN[29] functional with the SDD[30] effective core potential for La^{3+} and Zr^{4+} and the 6-31+G*[31−33] basis set for all the other atoms.

Table 1. Comparison between Computed and Experimental Average Metal−Ligand Distances (in Å)

molecule	calcd[a]	expt
	Metal−O(H_2O)	
$[Li(H_2O)_4]^+$	1.93	1.94 ± 0.05[b]
$[Mg(H_2O)_6]^{2+}$	2.05	2.07 ± 0.03[c]
$[Ca(H_2O)_6]^{2+}$	2.36	2.34 ± 0.07[c]
$[Zn(H_2O)_6]^{2+}$	2.04	2.08 ± 0.03[b]
$[La(H_2O)_9]^{3+}$	2.55	2.55 ± 0.04[c]
	Metal−Cl	
$[ZrCl_6]^{2-}$	2.48	2.47 ± 0.02[b]

[a] From fully optimized S-VWN/(SDD, 6-31+G*) geometries. [b] From CSD analysis; this work. [c] From CSD analysis from Dudev et al.[38]

Scheme 1

$$\begin{array}{ccc} & \xrightarrow{\Delta G^1} & \\ \text{Reactants } (\varepsilon = 1) & \rightarrow & \text{Products } (\varepsilon = 1) \\ \Delta G_{solv}^{\,x}\text{ (Reactants)} \downarrow & & \downarrow \Delta G_{solv}^{\,x}\text{ (Products)} \\ \text{Reactants } (\varepsilon = x) & \rightarrow & \text{Products } (\varepsilon = x) \\ & \xrightarrow{\Delta G^x} & \end{array}$$

This functional/basis set combination was chosen as it reproduces the experimentally observed metal−ligand bond distances in Li^+, Mg^{2+}, Ca^{2+}, Zn^{2+}, La^{3+}, and Zr^{4+} complexes (see Table 1).

For each fully optimized structure, S-VWN/(SDD, 6-31+G*) vibrational frequencies were computed to verify that the molecule was at the minimum of its potential energy surface. No imaginary frequencies were found in any of the metal complexes. After scaling the frequencies by an empirical factor of 0.9833,[34] the thermal energy including the zero-point energy (E_T), work (PV), and entropy (S) corrections were evaluated using standard statistical mechanical formulas.[35] Based on the fully optimized S-VWN/(SDD, 6-31+G*) geometries, the electronic energies (E_{elec}'s) were then evaluated with the same (SDD, 6-31+G*) basis set using the B3-LYP[36,37] functional. The differences in ΔE_{elec}, ΔE_T, ΔPV, and ΔS between the products and reactants were used to compute the reaction free energy at room temperature, $T = 298.15$ K, according to the following expression:

$$\Delta G^1 = \Delta E_{elec} + \Delta E_T + \Delta PV - T\Delta S \quad (1)$$

Continuum Dielectric Calculations. The reaction free energy in an environment characterized by a dielectric constant $\epsilon = x$ can be calculated according to Scheme 1.

ΔG^1, the gas-phase free energy, was computed using eq 1, as described above. $\Delta G_{solv}{}^x$, the free energy for transferring a molecule in the gas phase to a continuous solvent medium characterized by a dielectric constant, *x*, was estimated by solving Poisson's equation using finite difference methods.[39,40] Thus, the reaction free energy in an environment modeled by dielectric constant *x*, ΔG^x, can be computed from

$$\Delta G^x = \Delta G^1 + \Delta G_{solv}{}^x(\text{Products}) - \Delta G_{solv}{}^x(\text{Reactants}) \quad (2)$$

(25) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. *J. Mol. Biol.* **1977**, *122*, 535−542.
(26) Allen, F. H. *Acta Crystallogr.* **2002**, *B58*, 380−388.
(27) Dudev, M.; Wang, J.; Dudev, T.; Lim, C. *J. Phys. Chem. B*, in press.
(28) Frisch, M. J. et al. *Gaussian 03*, rev. B.03; Gaussian, Inc.: Pittsburgh, PA, 2003.
(29) Vosko, S. H.; Wilk, L.; Nusair, M. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(30) Kaupp, M.; Schleyer, P. v. R.; Stoll, H.; Preuss, H. *J. Chem. Phys.* **1991**, *94*, 1360−1366.
(31) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.
(32) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294−301.
(33) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650−654.
(34) Wong, M. W. *Chem. Phys. Lett.* **1996**, *256*, 391−399.
(35) McQuarrie, D. A. *Statistical Mechanics*; Harper and Row: New York, 1976.
(36) Becke, A. D. *Phys. Rev.* **1988**, *A38*, 3098−3100.
(37) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev.* **1988**, *B37*, 785−789.
(38) Dudev, T.; Chang, L.-Y.; Lim, C. *J. Am. Chem. Soc.* **2005**, *127*, 4091−4103.
(39) Gilson, M. K.; Honig, B. *Biopolymers* **1986**, *25*, 2097−2119.
(40) Lim, C.; Bashford, D.; Karplus, M. *J. Phys. Chem.* **1991**, *95*, 5610−5620.

Table 2. Comparison between Computed and Experimental Hydration Free Energies of Metal Cations and Ligands (in kcal/mol)

	ΔG_{solv}^{80}(kcal/mol)		
metal/ligand	expt	calcd	error[a]
Zr^{4+}	−1695.4[b]	−1695.8	−0.4
La^{3+}	−791.1[c]	−791.7	−0.6
Zn^{2+}	−484.6[c]	−484.7	−0.1
Mg^{2+}	−455.5[c]	−456.3	−0.8
Ca^{2+}	−380.8[c]	−381.5[d]	−0.7[d]
		(−381.4[d])	(−0.6[d])
Li^{+}	−123.5[c]	−123.9	−0.4
CH_3COO^{-}	−82.2[e]	−82.4	−0.2
$CH_3NH_3^{+}$	−70.0[f]	−70.5	−0.5
$HCONH_2$	−10.0[g]	−10.6	−0.6

[a] Error = ΔG_{solv}^{80}(calcd) − ΔG_{solv}^{80}(expt). [b] Obtained from the experimental ΔG_{solv}^{80}(Marcus) values taken from Marcus, 1991[47] using ΔG_{solv}^{80}(Friedman) = 1.9654 + 1.0459ΔG_{solv}^{80}(Marcus), whose linear correlation coefficient, r^2, is 0.9996 (this work). [c] From Friedman, 1973.[48] [d] Numbers with and without brackets correspond to the computed hydration free energies of octa- and hexahydrated Ca^{2+}, respectively (see text). [e] From Lim et al., 1991.[40] [f] From Chambers et al., 1996.[49] [g] Experimental solvation free energy of $HCONH(CH_3)$ from Wolfenden et al., 1978.[50]

The continuum dielectric calculations employed a 71 × 71 × 71 lattice with an initial grid spacing of 1.0 Å, refined with a spacing of 0.25 Å, ab initio geometries, and natural bond orbital (NBO) atomic charges.[41] The low-dielectric region of the solute was defined as the region inaccessible to contact by a 1.4 Å radius sphere rolling over the molecular surface. This region was assigned a dielectric constant of 2 ($\epsilon_{in} = 2$) to account for the electronic polarizability of the solute. The molecular surface was defined by *effective* solute radii, which were obtained by adjusting the CHARMM (version 22)[42] van der Waals radii to reproduce the experimental hydration free energies of the metal cations and ligands (see below and Table 2). Poisson's equation was solved with $\epsilon_{in} = 2$ and ϵ_{out} equal to 1, 4, 10, 20, or 80. The difference between the computed electrostatic potentials in a given dielectric medium ($\epsilon_{out} = x$) and in the gas phase ($\epsilon_{out} = 1$) yielded the solvation free energy ΔG_{solv}^{x} of the molecule.

The effective solute radii (in Å), which have been parametrized for S-VWN/(SDD, 6-31+G*) geometries and NBO charges, are as follows: $R_{Li} = 1.48$, $R_{Mg} = 1.50$, $R_{Ca} = 1.75$, $R_{Zn} = 1.40$, $R_{La} = 2.15$, $R_{Zr} = 1.50$, $R_C = 1.88$, $R_N = 1.75$, $R_O(HCONH_2) = 1.78$, $R_O(H_2O-Li/Mg/Ca/Zn/La) = 1.70$, $R_O(H_2O-Zr) = 1.64$, $R_O(CH_3COO^-) = 1.575$, $R_H = 1.468$, $R_H(N^+) = 1.35$, $R_H(H_2O-Li) = 1.28$, $R_H(H_2O-Mg) = 1.16$, $R_H(H_2O-Ca) = 1.09$, $R_H(H_2O-Zn) = 1.13$, $R_H(H_2O-La) = 0.98$, and $R_H(H_2O-Zr) = 0.70$. In computing the hydration free energies of the cations (see Table 2), Li^+ and La^{3+} were hydrated with 4 and 9 water molecules in the first shell, respectively, while Mg^{2+}, Ca^{2+}, and Zn^{2+} were hexahydrated, according to the most common CN measured in *aqueous* solution for these metal ions,[43−45] while Zr^{4+} was hexahydrated, according to its most preferred CN of 6 found in CSD complexes.[27] As recent X-ray diffraction data report a CN of eight for a 1 M $CaCl_2$ solution with the CN decreasing as the $CaCl_2$ salt concentration increases,[46] the hydration free energy of octahydrated Ca^{2+} was also computed using an $R_H(H_2O-Ca) = 0.98$ Å. These effective solute radii consistently overestimate the magnitude of the experimental hydration free energies of the metal cations and individual ligands by less than 1 kcal/mol (see Table 2).

Database Survey. The Protein Data Bank[25] was surveyed for <3.0 Å X-ray and NMR structures of proteins that contained at least one Asp or Glu coordinated to a number of univalent (Li^+, Na^+, K^+, Rb^+, Cs^+), divalent (Mg^{2+}, Ca^{2+}, Zn^{2+}, Cd^{2+}), and trivalent (lanthanides, Ln^{3+}) metals. Complexes of tetravalent metals such as Zr^{4+}, Hf^{4+}, or Ce^{4+} were not found in any PDB structures. Polynuclear metal-binding sites and sites that contain other nonaqua/nonamino acid ligands such as phosphate or sulfate groups were excluded from the survey. The protein sequences were aligned using the Modeler 4 program,[51] and those with a sequence identity higher than 30% were considered to belong to the same protein family. Only one representative from each protein family, namely, the structure solved at the highest resolution, was included in the survey.

Results

In evaluating how the metal's charge and the metal-binding site's relative solvent exposure affect the $Max^{COO}(M^{q+})$, we assume that a metal ion, M^{q+}, from aqueous solution has found a carboxylate-rich site, characterized by a dielectric constant ϵ, and is already bound to one or more Asp/Glu side chains in the site. We then compute the free energy for the successive exchange of metal-bound water molecules for Asp/Glu carboxylates, which are modeled as *monodentately* bound to the metal ion at this stage. To take into account the release of the displaced water molecule into bulk solvent, the experimentally determined ΔG_{solv}^{80} (−6.3 kcal/mol[52]) was used for the free water molecule in the reactions below.

1. Effect of Metal Cation Charge and Metal-Binding Site Solvent Accessibility. (a) Water-Acetate Exchange in *Monocationic* Complexes. We first consider reactions where acetate, modeling the side chain of Asp^-/Glu^-, competes with a first-shell water molecule for the metal ion in *monocationic* complexes:

$$[M(H_2O)_m(CH_3COO)_n]^+ + CH_3COO^- \rightarrow [M(H_2O)_{m-1}(CH_3COO)_{n+1}]^0 + H_2O \quad (3)$$

In eq 3 M = Mg^{2+}, Ca^{2+}, Zn^{2+}, La^{3+}, or Zr^{4+}; $m = 3$, 5, or 6; and $n = 1$, 2, or 3; and the computed enthalpies and free energies are summarized in Table 3 (reactions 1−5). Although the various metal complexes have different coordination geometries and different ratios of carboxylate-to-water ligands, they exhibit the following common trends in the respective thermodynamic parameters for eq 3. The water−acetate exchange reactions (eq 3) are enthalpy driven with ΔH^1 accounting for 97−99% of the gas-phase free energy, ΔG^1. This implies that the gas-phase reaction is driven primarily by favorable charge−charge interactions between the positively charged metal complex and the negatively charged carboxylate. Consequently, in fully/partially buried protein cavities, replacing a metal-bound water molecule with acetate is favorable, accompanied by a significant free energy gain (Table 3, reactions 1−5, negative ΔG^{ϵ}, $\epsilon < 20$). However, increasing the solvent exposure of the metal-binding site attenuates the free energy gain of the $H_2O \rightarrow CH_3COO^-$

(41) Reed, A. E.; Curtiss, L. A.; Weinhold, F. *Chem. Rev.* **1988**, *88*, 899−926.
(42) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187−217.
(43) Marcus, Y. *Chem. Rev.* **1988**, *88*, 1475−1498.
(44) Ohtaki, H.; Radnai, T. *Chem. Rev.* **1993**, *93*, 1157−1204.
(45) Markham, G. D.; Glusker, J. P.; Bock, C. W. *J. Phys. Chem. B* **2002**, *106*, 5118−5134.
(46) Megyes, T.; Grosz, T.; Radnai, T.; Bako, I.; Palinkas, G. *J. Phys. Chem. A* **2004**, *108*, 7261−7271.
(47) Marcus, Y. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2995−2999.
(48) Friedman, H. L.; Krishnan, C. V. Thermodynamics of ionic hydration. In *In Water: A comprehensive treatise*; Franks, F., Ed.; Plenum Press: New York, 1973; Vol. 3, pp 1−118.
(49) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385−16398.
(50) Wolfenden, R. *Biochemistry* **1978**, *17*, 201−204.
(51) Sali, A.; Blundell, T. L. *J. Mol. Biol.* **1993**, *234*, 779−815.
(52) Ben-Naim, A.; Marcus, Y. *J. Chem. Phys.* **1984**, *81*, 2016−2027.

Table 3. Calculated Free Energies (ΔG^{ϵ}) for Water (W) → Acetate Exchange in Li^+, Mg^{2+}, Ca^{2+}, Zn^{2+}, La^{3+}, and Zr^{4+} Complexes for Media of Different Dielectric Constant, ϵ[a]

rxn no.	reactant metal complex + CH_3COO^-	product metal complex + H_2O	ΔH^1	ΔG^1	ΔG^4	ΔG^{10}	ΔG^{20}	ΔG^{80} [b]
1	$[MgW_5(CH_3COO)]^+$	$[MgW_4(CH_3COO)_2]^0$	−119.3	−115.3	−29.8	−9.4	−2.0	3.7
2	$[CaW_6(CH_3COO)]^+$	$[Ca\,W_5\,(CH_3COO)_2]^0$	−109.8	−106.0	−25.0	−6.4	0.0	4.7
3	$[ZnW_3(CH_3COO)]^+$	$[ZnW_2(CH_3COO)_2]^0$	−143.1	−141.3	−47.7	−24.9	−16.7	−10.4
4	$[LaW_6(CH_3COO)_2]^+$	$[LaW_5(CH_3COO)_3]^0$	−114.8	−112.7	−31.2	−11.1	−3.7	2.1
5	$[ZrW_3(CH_3COO)_3]^+$	$[ZrW_2(CH_3COO)_4]^0$	−134.5	−133.4	−46.8	−23.1	−13.9	−6.2
6	$[LiW_3(CH_3COO)]^0$	$[LiW_2(CH_3COO)_2]^-$	−31.6	−28.9	−12.2	−6.3	−4.1	−2.6
7	$[MgW_4(CH_3COO)_2]^0$	$[MgW_3(CH_3COO)_3]^-$	−35.6	−33.3	−13.5	−6.8	−4.2	−2.3
8	$[CaW_5(CH_3COO)_2]^0$	$[CaW_4(CH_3COO)_3]^-$	−41.7	−40.8	−18.4	−10.4	−7.2	−4.5
					(−17.7[c])	(−9.3[c])	(−5.8[c])	(−2.9[c])
9	$[ZnW_2(CH_3COO)_2]^0$	$[ZnW(CH_3COO)_3]^-$	−30.3	−26.7	−10.1	−3.7	−1.0	1.2
10	$[LaW_5(CH_3COO)_3]^0$	$[LaW_4(CH_3COO)_4]^-$	−42.7	−41.1	−16.8	−8.4	−4.9	−2.0
11	$[ZrW_2(CH_3COO)_4]^0$	$[ZrW(CH_3COO)_5]^-$	−26.8	−24.7	−4.7	1.9	4.5	6.5
12	$[LiW_2(CH_3COO)_2]^-$	$[LiW(CH_3COO)_3]^{2-}$	56.2	55.6	11.3	3.7	1.3	−0.5
13	$[MgW_3(CH_3COO)_3]^-$	$[MgW_2(CH_3COO)_4]^{2-}$	44.2	44.4	9.1	3.2	1.3	−0.2
14	$[CaW_4(CH_3COO)_3]^-$	$[CaW_3(CH_3COO)_4]^{2-}$	41.9	41.2	7.8	2.2	0.5	−0.9
					(8.2[c])	(2.8[c])	(1.2[c])	(−0.1[c])
15	$[ZnW(CH_3COO)_3]^-$	$[Zn(CH_3COO)_4]^{2-}$	50.6	53.9	23.8	20.1	19.2	18.4
16	$[LaW_4(CH_3COO)_4]^-$	$[LaW_3(CH_3COO)_5]^{2-}$	45.2	50.0	24.5	21.4	20.6	20.0
17	$[ZrW(CH_3COO)_5]^-$	$[Zr(CH_3COO)_6]^{2-}$	36.8	37.9	18.1	17.0	17.0	16.9

[a] All thermodynamic quantities in kcal/mol; $\epsilon = 1$ corresponds to gas-phase values evaluated at the B3-LYP/6-31+G*//S−VWN/6-31+G* level, whereas $\epsilon = 4$, 10, 20, and 80 represent increasing solvent exposure of the binding site with $\epsilon = 80$ for a fully solvent-exposed site. [b] Calculated using experimental values for the solvation free energies of CH_3COO^- (−82.2 kcal/mol[40]) and water (−6.3 kcal/mol[52]). [c] The number in parentheses differs from the respective number without parentheses in using $R_H(H_2O-Ca) = 0.98$ Å instead of $R_H(H_2O-Ca) = 1.09$ Å in computing the solvation free energies of the calcium complexes (see Methods).

exchange because the desolvation penalty of the charged reactants, CH_3COO^- and $[M(H_2O)_m(CH_3COO)_n]^+$, exceeds the solvation free energy gain of the noncharged products, H_2O and $[M(H_2O)_{m-1}(CH_3COO)_{n+1}]^0$.

(b) Water−Acetate Exchange in *Neutral* Complexes. We next consider reactions where acetate binds to a neutral metal complex Table 3 (reactions 6−11):

$$[M(H_2O)_{m-1}(CH_3COO)_{n+1}]^0 + CH_3COO^- \rightarrow [M\,(H_2O)_{m-2}\,(CH_3COO)_{n+2}]^- + H_2O \quad (4)$$

In eq 4, M = Li^+, Mg^{2+}, Ca^{2+}, Zn^{2+}, La^{3+}, or Zr^{4+}; $m = 3$, 4, 5 or 6; and $n = 0$, 1, 2 or 3. As for eq 3, the $H_2O \rightarrow CH_3COO^-$ exchange reactions (eq 4) are favorable in the gas phase and in a buried protein cavity (Table 3, reactions 6−11, negative ΔG^{ϵ}, $\epsilon < 10$), although the respective free energy gains are smaller than those for eq 3 due to weaker interactions with a neutral metal complex, as compared to a charged metal complex in eq 3. Also in analogy to eq 3, increasing the solvent accessibility of the metal-binding site disfavors the $H_2O \rightarrow CH_3COO^-$ exchange because the desolvation penalty of CH_3COO^- outweighs the solvation free energy gain of the much bulkier $[M(H_2O)_{m-2}(CH_3COO)_{n+2}]^-$ product even though both (reactant and product) are negatively charged. Consequently, in a relatively solvent-exposed site, replacing a Zr-bound water molecule with acetate may be unfavorable as the free energy gain upon $H_2O \rightarrow CH_3COO^-$ exchange in the $[Zr(H_2O)_2(CH_3COO)_4]^0$ complex is smaller than that for the other neutral metal complexes in the gas phase and diminishes rapidly with increasing ϵ (Table 3, reaction 11, positive ΔG^{ϵ}, $\epsilon \geq 10$). However, the $H_2O \rightarrow CH_3COO^-$ exchanges for the other neutral metal complexes remain favorable even when the metal-binding site is only partially buried (Table 3, reactions 6−10, negative ΔG^{ϵ}, $\epsilon \leq 20$).

(c) Water−Acetate Exchange in *Monoanionic* Complexes. Table 3 (reactions 12−17) lists the free energies for $H_2O \rightarrow CH_3COO^-$ exchange in *mono*anionic metal complexes:

$$[M(H_2O)_{m-2}(CH_3COO)_{n+2}]^- + CH_3COO^- \rightarrow [M(H_2O)_{m-3}(CH_3COO)_{n+3}]^{2-} + H_2O \quad (5)$$

In eq 5, M = Li^+, Mg^{2+}, Ca^{2+}, Zn^{2+}, La^{3+}, or Zr^{4+}; $m = 3$, 4, 5, or 6; and $n = 0$, 1, 2, or 3. Unlike eqs 3 and 4, the gas-phase reactions (eq 5) are unfavorable (Table 3, reactions 12−17, positive ΔG^1) due to repulsive charge−charge interactions between two negatively charged species, CH_3COO^- and $[M(H_2O)_{m-2}(CH_3COO)_{n+2}]^-$.[13] Consequently, in a totally/partially buried protein cavity, a carboxylate group may not be able to replace a metal-bound water molecule (Table 3, reactions 12−17, positive ΔG^{ϵ}, $\epsilon < 20$). Although the *di*anionic $[M(H_2O)_{m-3}(CH_3COO)_{n+3}]^{2-}$ product complexes are better solvated than the respective *mono*anionic counterparts or CH_3COO^-, the respective free energies remain positive even in a relatively solvent-exposed metal-binding site ($20 < \epsilon < 80$), indicating the unlikelihood of forming a *di*anionic metal complex in protein binding sites.

In summary, the results in Table 3 imply that, in a relatively buried protein cavity (with $\epsilon < 20$), a *monoanionic* metal complex is unlikely to exchange one of its water molecules for another monodentately bound carboxylate to form the corresponding *di*anionic metal complex (Table 3, reactions 12−17, positive ΔG^{ϵ}, $\epsilon < 20$). Thus, in a relatively buried metal-binding site containing Asp/Glu side chains that bind successively to the metal ion via one of its carboxylate oxygen atoms only, $Max^{COO}(M^{q+})$ equals $q + 1$ in the absence of stabilizing interactions from the second shell; i.e., univalent Li^+ can bind one or two carboxylates; divalent Zn^{2+}, Mg^{2+}, and Ca^{2+} can bind at most three acidic residues; trivalent La^{3+}, a maximum of four carboxylates; and tetravalent Zr^{4+}, no more than five carboxylates.

2. Effect of the Carboxylate-Binding Mode. (a) Mono- vs Bidentate Binding. To establish the extent the carboxylate-

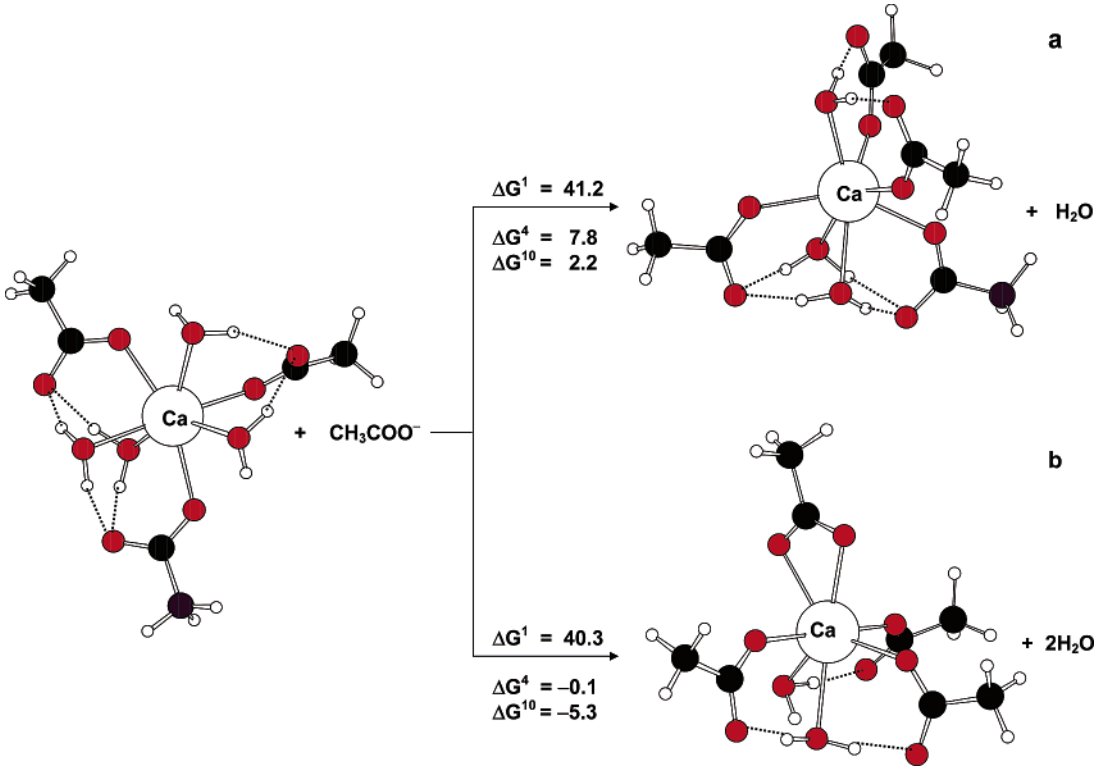

Figure 1. Ball and stick diagrams for the fully optimized Ca^{2+} carboxylate complexes and calculated free energies (in kcal/mol) for (a) $[CaW_4(CH_3COO)_3]^- + CH_3COO^- \rightarrow [CaW_3(CH_3COO)_4]^{2-} + H_2O$ and (b) $[CaW_4(CH_3COO)_3]^- + CH_3COO^- \rightarrow [CaW_2(CH_3COO)_3(CH_3C\tilde{O}\tilde{O})]^{2-} + 2H_2O$, where $CH_3C\tilde{O}\tilde{O}$ denotes a bidentately bound acetate.

binding mode (mono- vs bidentate) affects the thermodynamics of metal−carboxylate interactions in protein cavities, which are usually relatively buried, we compared the ΔG^1, ΔG^4, and ΔG^{10} for replacing a water molecule in a heptacoordinated monoanionic Ca^{2+} complex with a bidentately bound acetate (Figure 1b) with those for reaction 14 in Table 3 (Figure 1a) where all four acetates are *mono*dentately bound to the metal ion. As shown in Figure 1, switching the carboxylate-binding mode from *mono*dentate to *bi*dentate favors the $H_2O \rightarrow CH_3COO^-$ exchange in relatively buried metal-binding sites in proteins (negative ΔG^4 and ΔG^{10}). This is mainly because the release of an extra water molecule in Figure 1b, as compared to Figure 1a, results in a gain of the gas-phase entropy and solvation free energy (−6.3 kcal/mol). The change in the ΔG^{10} sign upon *mono*dentate → *bi*dentate binding in Figure 1 implies that, in a partially buried protein cavity, a fourth carboxylate could coordinate *bi*dentately but not *mono*dentately to a monoanionic Ca^{2+} complex. This is in line with previous PDB analyses showing that polycarboxylate Ca^{2+}-binding sites in proteins have at least one Asp/Glu side chain bound *bi*dentately to the metal cation.[53]

(b) Concerted vs Sequential Binding. So far, we have studied the successive/sequential as opposed to *simultaneous*/concerted binding of acetates to the metal complexes assuming that the respective preformed binding sites are relatively flexible. Here, we examine the binding of two or three acetates *simultaneously* to a given metal complex, modeling Asp/Glu residues flanking *rigid* metal-binding sites that act as polydentate chelating superligands. To explore the chelate effect on the water → carboxylate exchange (where the metal-bound water molecules are replaced by all the Asp/Glu carboxylates lining the rigid metal-binding site simultaneously), we constructed polycarboxylate models containing two and three COO^- groups separated by a polymethylene spacer of an appropriate length, $^-OOC-(CH_2)_5-COO^-$ (denoted by $L2^{2-}$) and $^-OOC-(CH_2)_4-(CHCOO^-)-(CH_2)_4-COO^-$ (denoted by $L3^{3-}$), and evaluated their binding to heptacoordinated Ca complexes:

$$[Ca(H_2O)_5(CH_3COO)_2]^0 + L2^{2-} \rightarrow [Ca(H_2O)_3(CH_3COO)_2(L2)]^{2-} + 2H_2O \quad (6)$$

$$[Ca(H_2O)_6(CH_3COO)]^+ + L3^{3-} \rightarrow [Ca(H_2O)_3(CH_3COO)(L3)]^{2-} + 3H_2O \quad (7)$$

In both cases the reaction products are dianionic complexes where each carboxylate group of the L2 or L3 ligand binds monodentately to the metal cation (Figure 2b and 2c). The free energies evaluated for eqs 6 and 7 are shown in Figure 2b and 2c, respectively, and compared with those calculated for reaction 14 in Table 3, which also yields a dianionic complex with all four carboxylates *mono*dentately bound (Figure 2a).

Unlike the sequential (successive) water → acetate substitution (Table 3 and Figure 2a), the concerted (simultaneous) binding of two or three acidic residues to the metal cation is favorable both in the gas phase and in a protein cavity (Figure 2b and 2c, negative ΔG^ϵ, $\epsilon \leq 10$), resulting in a stable dianionic product. This is because the desolvation penalty of the reactants is compensated by the following factors: (i) favorable electrostatic interactions between the dianionic $L2^{2-}$ and trianionic $L3^{3-}$ with the neutral $[Ca(H_2O)_5(CH_3COO)_2]^0$ and cationic $[Ca(H_2O)_6(CH_3COO)]^+$ complexes, respectively; (ii) release of two

(53) Dudev, T.; Lim, C. *J. Phys. Chem. B* **2004**, *108*, 4546−4557.

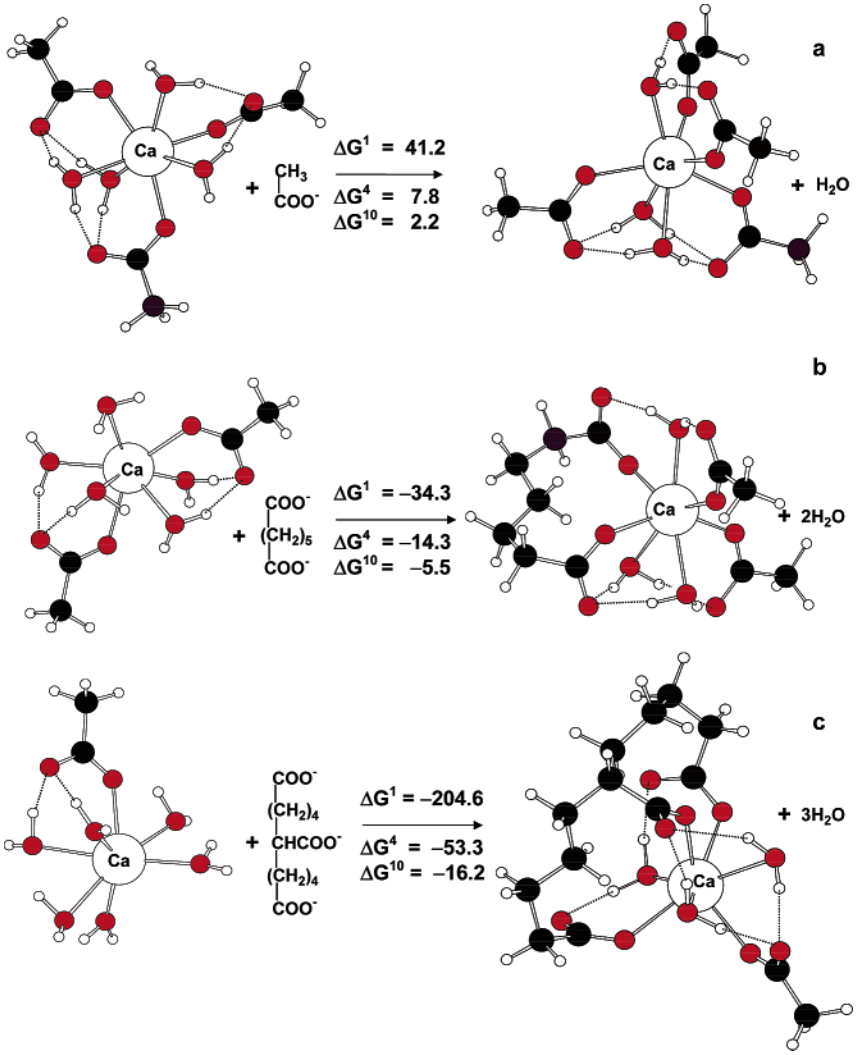

Figure 2. Ball and stick diagrams for the fully optimized Ca^{2+} carboxylate complexes and calculated free energies (in kcal/mol) for (a) $[CaW_4(CH_3COO)_3]^- + CH_3COO^- \rightarrow [CaW_3(CH_3COO)_4]^{2-} + H_2O$; (b) $[CaW_5(CH_3COO)_2]^0 + L2^{2-} \rightarrow [CaW_3(CH_3COO)_2L2]^{2-} + 2H_2O$, $L2^{2-} = {}^-OOC-(CH_2)_5-COO^-$; and (c) $[CaW_6CH_3COO]^+ + L3^{3-} \rightarrow [CaW_3CH_3COOL3]^{2-} + 3H_2O$, $L3^{3-} = {}^-OOC-(CH_2)_4-(CHCOO^-)-(CH_2)_4-COO^-$.

or three water molecules, which make favorable contributions to both the entropy and solvation free energy of the respective reactions; and (iii) a gain in the solvation free energy of the dianionic product.

3. Effect of the Second-Shell Ligands. In proteins, metal-bound Asp/Glu are often found hydrogen-bonded to backbone peptide groups or Asn, Gln, Lys, and Arg side chains in the metal's second coordination sphere.[54] To evaluate the effect of the second-shell ligands on the binding of a carboxylate to the metal cofactor, we computed the free energies for replacing a water molecule with an acetate in (i) a monoanionic Ca^{2+} complex interacting with a second-shell $CH_3NH_3^+$ (modeling Lys, Figure 3b) and (ii) a monoanionic Zn^{2+} complex whose metal-free carboxylate oxygen atoms are hydrogen-bonded to two (Figure 4b) or four (Figure 4c) second-shell $HCONH_2$ ligands (modeling a backbone peptide group or Asn/Gln side chain). These are compared with the respective ΔG^{ϵ} ($\epsilon = 1$, 4, or 10) in Table 3 (reactions 14 and 15) where only first-shell ligands are considered (Figures 3a and 4a).

The results in Figure 3 show that, in a fully/partially buried cavity, the presence of a single positively charged Lys side chain (modeled by $CH_3NH_3^+$) interacting with the Ca^{2+}-bound carboxylates enables a negatively charged acetate to displace a water molecule from the metal's first coordination shell (Figure 3b, negative ΔG^4 and ΔG^{10}). This is because a second-shell positively charged ligand effectively neutralizes the negatively charged $[Ca(H_2O)_4(CH_3COO)_3]^-$ complex, thus enabling it to interact favorably with the acetate anion. It also helps to stabilize the product complex $\{[Ca(H_2O)_3(CH_3COO)_4]\cdot CH_3NH_3\}^-$ by alleviating the electrostatic repulsion within the cluster. The change in the sign of the $H_2O \rightarrow CH_3COO^-$ ΔG^4 and ΔG^{10} when a second-shell Lys interacts with the monoanionic Ca^{2+} complex implies that, in a fully/partially buried protein cavity, a monoanionic metal complex could bind a fourth carboxylate monodentately provided a second-shell cationic residue, such as Lys or Arg, can interact with the first-shell carboxylate(s).

The results in Figure 4 show that, in a fully/partially buried cavity, the presence of second-shell amide ligands (modeled by $HCONH_2$) interacting with the Zn^{2+}-bound carboxylates favors the water → acetate exchange, and the greater the number of second-shell amides, the more favorable the respective free

(54) Dudev, T.; Lin, Y. L.; Dudev, M.; Lim, C. *J. Am. Chem. Soc.* **2003**, *125*, 3168−3180.

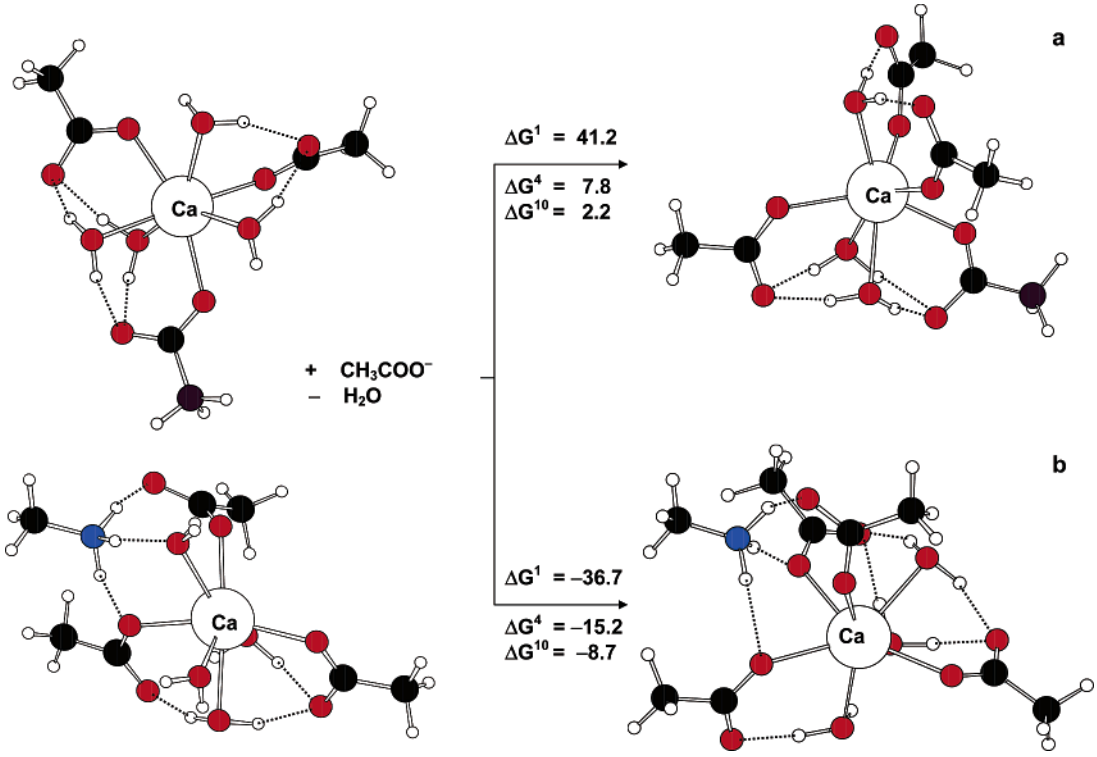

Figure 3. Ball and stick diagrams for the fully optimized Ca^{2+} carboxylate complexes, and calculated free energies (in kcal/mol) for (a) $[CaW_4(CH_3COO)_3]^- + CH_3COO^- \rightarrow [CaW_3(CH_3COO)_4]^{2-} + H_2O$ and (b) $\{[CaW_4(CH_3COO)_3]\cdot CH_3NH_3\}^0 + CH_3COO^- \rightarrow \{[CaW_3(CH_3COO)_4]\cdot CH_3NH_3\}^- + H_2O$.

energies. Thus, the water → acetate exchange free energies, ΔG^{ϵ}, in buried/partially buried metal-binding sites ($\epsilon = 4$ or 10) decrease from 24 or 20 kcal/mol in the tricarboxylate monoanionic Zn^{2+} complex (Figure 4a) to 15 kcal/mol for the corresponding complex containing two second-shell amides (Figure 4b), and become negative (−3.2 or −1.3 kcal/mol) for binding sites with four second-shell amides hydrogen bonded to the first-shell carboxylates (Figure 4c). This is partly due to stronger dipole-charge $\mu-Q^{2-}$ interactions between the second-shell amide dipoles and the *di*anionic Zn^{2+} product complex, as compared to $\mu-Q^{-}$ interactions between the second-shell amide dipoles and the respective *mono*anionic reactant counterpart. Thus, the calculations imply that in a relatively buried protein cavity the metal dication could bind up to four carboxylates if several (four in the case of the monoanionic Zn^{2+} complex) backbone peptide groups or Asn/Gln side chains from the metal's second shell hydrogen bond to the first-shell acidic residues.

Metal−Carboxylate Complexes in Proteins. The PDB search (see Methods) produced 3 Li^+-, 103 Na^+-, 37 K^+-, 2 Rb^+-, 3 Cs^+-, 82 Mg^{2+}-, 235 Ca^{2+}-, 180 Zn^{2+}-, 162 Cd^{2+}-, and 66 Ln^{3+}-binding sites that contained at least one metal-bound Asp/Glu side chain. These were grouped according to the metal charge (univalent M^+, divalent M^{2+}, and trivalent M^{3+}) and the number of acidic amino acid residues in the binding site, as shown in Figure 5. The number of metal-bound carboxylates does not exceed three for M^+ binding sites (Figure 5a) and four for M^{2+} and M^{3+} binding sites (Figure 5b and 5c, respectively). For all three groups of metal ions (M^+, M^{2+}, and M^{3+}), binding sites containing the highest number of metal-bound acidic residues are the least populated in proteins totaling 2.0% (3 sites), 3.0% (20 sites), and 7.6% (5 sites), respectively. More detailed information about the dianionic tricarboxylate M^+ and tetracarboxylate M^{2+}-binding sites are given in Table 4, which lists, for each PDB entry, the carboxylate-binding mode (mono- or bidentate) of the first-shell Asp/Glu, the second-shell Lys or Arg residue, if present, and the number of Asn, Gln, or backbone groups hydrogen bonding to the first-shell residues.

The data in Table 4 show that, in most cases, the first-shell Asp/Glu carboxylates are bidentately bound to the metal cation (17 out of 23 cases) and are encapsulated by a layer of second-shell backbone peptide groups or Asn/Gln side chains (22 out of 23 cases). In six metal-binding sites, a cationic residue (Lys/Arg) in the second coordination sphere interacts with the metal-bound Asp/Glu side chain(s). Notably, each binding site listed in Table 4 contains at least one of the factors (bidentate Asp/Glu, second-shell Lys/Arg or second-shell backbone/Asn/Gln) predicted to favor the formation of a dianionic metal complex.

Discussion

The results of the DFT/CDM calculations indicate that, in a relatively buried protein cavity, the $Max^{COO}(M^{q+})$ depends on (i) the metal cation charge, q, (ii) the carboxylate-binding mode, and (iii) the first-shell carboxylate−second-shell ligand interactions. Below, we describe how the $Max^{COO}(M^{q+})$ depends on these three variables and compare the findings with those from the PDB structures.

(1) Metal Cation Charge. The results in Table 3 show that the $Max^{COO}(M^{q+})$ can be expressed as a linear function of the metal's formal charge, q. In a relatively buried protein cavity where charge−dipole interactions between a negatively charged carboxylate group and a neutral metal complex favors formation of a monoanionic metal complex, a metal cation of charge q can generally bind as many as $q + 1$ deprotonated Asp/Glu residues *monodentately* (Table 3); i.e., $Max^{COO}(M^{q+}) = q+1$; thus univalent, divalent, trivalent, and tetravalent cations can

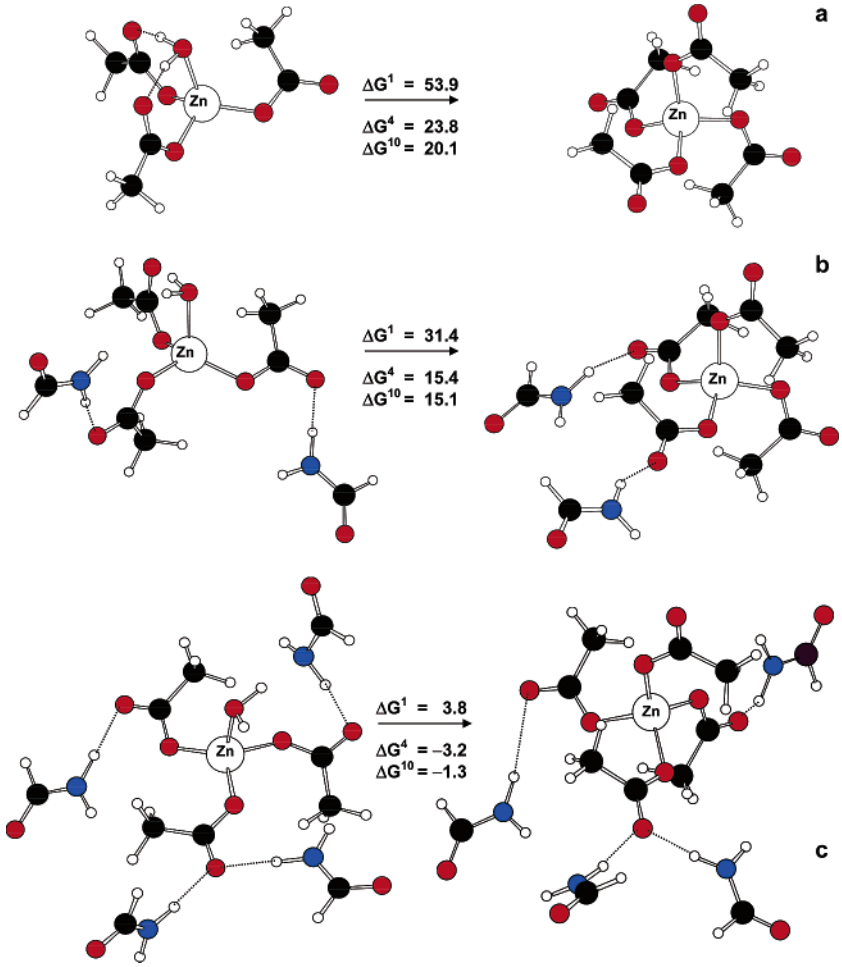

Figure 4. Ball and stick diagrams for the fully optimized Zn^{2+} carboxylate complexes, and calculated free energies (in kcal/mol) for (a) $[ZnW(CH_3COO)_3]^- + CH_3COO^- \rightarrow [Zn(CH_3COO)_4]^{2-} + H_2O$, (b) $\{[ZnW(CH_3COO)_3]\cdot(HCONH_2)_2\}^- + CH_3COO^- \rightarrow \{[Zn(CH_3COO)_4]\cdot(HCONH_2)_2\}^{2-} + H_2O$, and (c) $\{[ZnW(CH_3COO)_3]\cdot(HCONH_2)_4\}^- + CH_3COO^- \rightarrow \{[Zn(CH_3COO)_4]\cdot(HCONH_2)_4\}^{2-} + H_2O$.

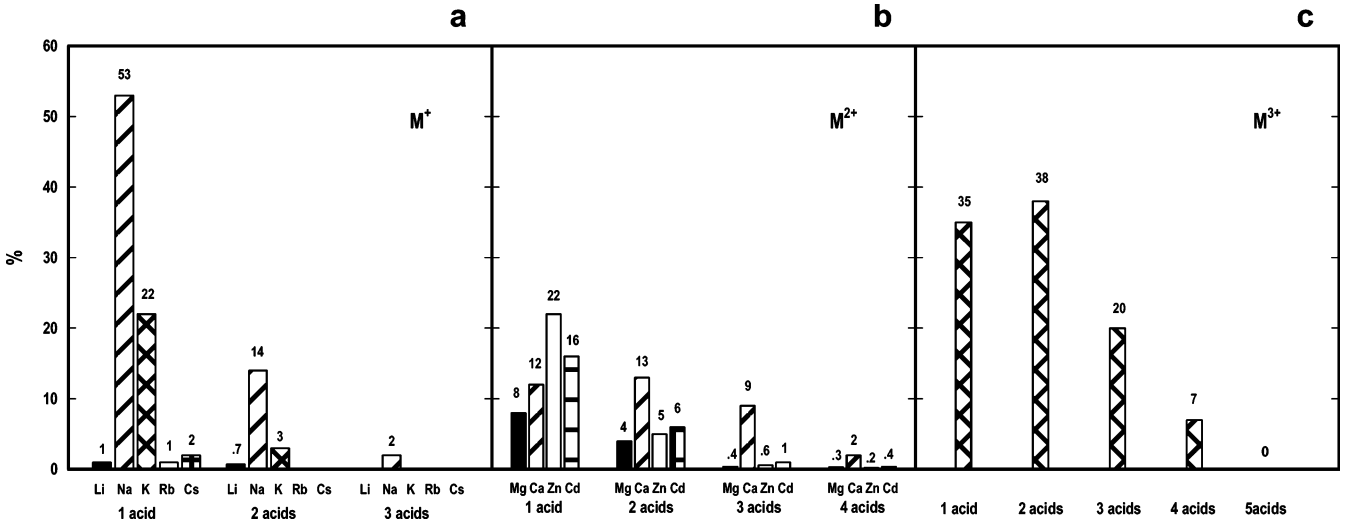

Figure 5. Percentage frequency distribution of carboxylate-binding sites in the PDB for (a) univalent cations (148 sites), (b) divalent cations (659 sites), and (c) trivalent lanthanide cations (66 sites).

bind up to 2, 3, 4, and 5 carboxylates, respectively. Indeed, 98%, 97%, and 100% of the PDB binding sites containing univalent, divalent, and trivalent metal cations contain up to 2, 3, and 4 acidic residues, respectively (Figure 5). The few PDB binding sites containing $q + 2$ acidic residues (Table 4) are discussed below.

The results in Table 3 also reveal that the balance between the positive charge of the metal and the negative charges of the constituent acidic residues, which determines the overall charge of the metal complex, governs the thermodynamics of the water → carboxylate exchange in a relatively buried protein cavity. Although the CN and specific chemical properties of the metal

Table 4. Tricarboxylate Binding Sites for Univalent Metals and Tetracarboxylate Binding Sites for Divalent Metals Found in the PDB

PDB code	metal cation	first-shell Asp/Glu		second-shell	no. of second-shell
		monodentate	bidentate	Lys/Arg	backbone/Asn/Gln
1EJA	Na^+	E70, E77, E80		R66	5
1L0Z	Na^+	E59, D66, E69		R9	4
1LTM	Na^+	D237, D241	D251		7
1WDC	Mg^{2+}	D28, D30, D32, D39			5
4PAL	Mg^{2+}	D90, D92, D94, E101			5
1AJJ	Ca^{2+}	D25, D29, D35, E36			4
1H2G	Ca^{2+}	D76B, E152A, D252B	D73B	R199B	5
1M63	Ca^{2+}	I. D140, D142, D144	E151	R146	2
		II. D99, D101, D103	E110		4
		III. D62, D64, D70	E73		5
1O88	Ca^{2+}	D129, E166, D170	D131	K190, R134	2
1OSA	Ca^{2+}	I. D20, D22, D24	E31		4
		II. D129, D131, D133	E140		6
1PMZ	Ca^{2+}	D56, D58, D60	E67		5
1RWY	Ca^{2+}	I. D90, D92, D94	E101		6
		II. D51, D53, E59	E62		4
1SRA	Ca^{2+}	D257, D259, D261	E269		4
2SCP	Ca^{2+}	D16, D18, D20	D27		5
1SW1	Zn^{2+}	E238A, E238B, D239A, E239B			4
1CDP	Cd^{2+}	I. D90, D92, D94	E101		4
		II. D51, D53, E59	E62		5
1HK7	Cd^{2+}	E406A	E406B, E409A, E409B		
3NG1	Cd^{2+}	E46, D50	E11, E33	R15	1

ion may play a role in the water → carboxylate exchange, they do not appear to dictate the $Max^{COO}(M^{q+})$ in buried protein cavities. This is in line with the observation that, in proteins, the most common CNs for Mg^{2+}, Ca^{2+}, and Zn^{2+} are 6, 7, and 4, respectively, but the maximum number of metal-bound carboxylates in Mg^{2+}-, Ca^{2+}-, and Zn^{2+}-binding sites is the same (Figure 5).

(2) Mode of Carboxylate Binding. In a relatively buried protein cavity, the $Max^{COO}(M^{q+})$ can be increased from $q + 1$ to $q + 2$ if (a) the Asp/Glu side chains can coordinate simultaneously to the metal ion (Figure 2) or (b) one of the Asp/Glu side chains can bind bidentately to the metal ion (Figure 1). The latter is especially the case for metal cations such as Ca^{2+} that can reorganize their ligation sphere at a relatively low free energy cost,[53] indicating that Ca^{2+} can bind as many as four Asp/Glu side chains if one of them is bidentately bound. These results are consistent with the finding that 16 out of the 20 tetracarboxylate dication-binding sites in the PDB structures contain at least one Asp/Glu bound bidentately to the metal ion (Table 4).

(3) Second-Shell Ligand Interactions with First-Shell Asp/Glu. In a fully/partially buried protein cavity, the $Max^{COO}(M^{q+})$ can also be increased from $q + 1$ to $q + 2$ if the metal-bound carboxylates are stabilized via (a) charge−charge interactions with a positively charged second-shell Lys/Arg side chain (Figure 3) or (b) multiple charge−dipole interactions with several second-shell peptide backbone groups or Asn/Gln side chains (Figure 4c). This is consistent with the finding in the PDB structures that univalent and divalent metal cations that bind three and four carboxylates, respectively, contain either positively charged second-shell Lys/Arg side chains hydrogen bonded to first-shell Asp/Glu side chains and/or several second-shell amide groups encapsulating the first-shell carboxylates (Table 4).

Conclusions

In summary, the results presented here show that the metal cation charge and hydrogen bonding interactions between the metal-bound Asp/Glu carboxylates with second-shell residues govern the $Max^{COO}(M^{q+})$ in metal-binding sites that are usually relatively buried in proteins. The carboxylate-binding mode could also affect the $Max^{COO}(M^{q+})$ especially for metal-binding sites with a high propensity to bind a carboxylate bidentately rather than monodentately. The combined results from the DFT/CDM calculations and the PDB survey suggest the following guidelines for estimating the $Max^{COO}(M^{q+})$ in metalloproteins:

(1) $Max^{COO}(M^{q+}) = q + 1$ in a relatively buried cavity where Asp/Glu are monodentately bound to the metal ion and lack interactions with the second-shell protein ligands, whereas

(2) $Max^{COO}(M^{q+}) = q + 2$ in a relatively buried protein cavity where the first-shell carboxylates are either (a) hydrogen bonded to a cationic ligand or several amide ligands in the second shell and/or (b) bidentately bound to the metal cation.

Acknowledgment. We are grateful to Drs. D. Bashford, M. Sommer, and M. Karplus for the program to solve the Poisson equation. This work was supported by the National Science Council, Taiwan (NSC Contract No. 91-2311-B-001), the Institute of Biomedical Sciences, and the National Center for High-Performance Computing, Taiwan.

Supporting Information Available: Complete ref 28. This material is available free of charge via the Internet at http://pubs.acs.org.

JA055797E