Chelator-based parameterization of the 12-6-4LJ molecular mechanics potential for more realistic metal ion-protein interactions

*Paulius Kantakevičius, Calvin Mathiah, Linus O. Johannissen and Sam Hay**

Manchester Institute of Biotechnology and Department of Chemistry, The University of Manchester, Manchester, United Kingdom.

sam.hay@manchester.ac.uk.

Keywords: Molecular Dynamics Simulations, Nonbonded Potentials, Thermodynamic Integration, Metalloproteins.

Abstract
Metal ions are associated with a variety of proteins and play critical roles in a wide range of biochemical processes. There are multiple ways to study and quantify protein-metal ion interactions, including by molecular dynamics simulations. Recently, the Amber molecular mechanics forcefield was modified to include a 12-6-4LJ potential, which allows better description of non-bonded terms through the additional pairwise C_{ij} coefficients. Here, we demonstrate a method of generating C_{ij} parameters that allows parametrization of specific metal ion-ligating groups in order to tune binding energies computed by thermodynamic integration. The new C_{ij} coefficients were tested on a series of chelators: EDTA, NTA, EGTA and the EF1 loop peptides from the proteins lanmodulin and calmodulin. The new parameters show significant improvements in computed binding energies relative to existing force fields and produce coordination numbers and ion-oxygen distances that are in good agreement with experimental values. This parametrization method should be extensible to a range of other systems and could be readily adapted to tune properties other than binding energies.

Introduction
Metal ions are thought to be associated with around 50% of all proteins.[1] They play important roles in a multitude of biologically significant processes such as enzyme catalysis and signal transduction.[2-4] Inclusion of metal ions into molecular mechanics (MM) force fields used for molecular dynamics (MD) simulations poses specific challenges due to the high polarizability of metal ions, which can take part in charge transfers, coordination number (CN) changes and ligand swapping.[5-7] Several approaches have been developed for simulating metal ions but these typically lack transferability between systems or simulate one property accurately at the expense of others. For example, explicitly bonded models treat metal-ligand coordination as immutable bonds which does not allow for ligand and CN changes.[8] Such methods therefore cannot provide insight into coordination dynamics and multiple simulations with different bonding parameters are required to determine optimal CN in a given system.[7] Another commonly used approach is to rely on nonbonded terms that treat metal-ligand coordination as nonbonded interactions facilitated by Lennard-Jones (LJ) and Coulombic terms. The 12-6LJ nonbonded model is the default in most MM force fields and defines the standard LJ potential using two pairwise parameters for each type of atom-atom interaction (**Eq. 1**, below). This approach allows for the metal ions to switch both CN and ligands but lacks charge transfer effects and polarizability. Nonetheless, this approach is commonly used for MD simulations as it is computationally efficient and does not require assumptions about the ligands coordinated to the metal ion, which usually require previous knowledge about the simulated system.

Recently an array of new 12-6LJ parameters were developed to describe divalent metal ions. The "12-6LJ HFE" parameters accurately reproduced hydration free energies (HFEs), the "12-6LJ IOD" parameters reproduced ion-oxygen distances (IODs), while the "12-6LJ CM" parameters were designed as a compromise between both.[6] As the 12-6LJ potential does not accurately reproduce both HFE and IOD experimental values with a single set of parameters, this led to the development of the new 12-6-4LJ potential.[6,9] This contains an additional pairwise parameter, which accounts for the charge induced dipole and dipole induced dipoles that are neglected in the 12-6LJ potential (**Eq. 2**, below). This addition to the force field made it possible to use a single set of parameters to adequately describe HFE, IOD and CN values in three commonly used water models, TIP3P, SPC/E and TIP4Pew.[9-11] However, the C_{ij} coefficients of the new C_{ij}/r_{ij}^4 term were

parameterized only for the interactions with the water oxygen atoms, whereas parameters for other types of atoms were adopted from molecular polarizability tensor calculations.[12] In a subsequent study of metal ion-nucleotide interactions, it was shown that the adopted C_{ij} coefficients may not produce accurate binding energies as they overestimated the binding energies to adenine, guanine and phosphate by up to 20 kJ mol^{-1}.[13] These authors were able to generate a new set of C_{ij} coefficients that increased the accuracy of estimated binding energies (± 0.5 kJ mol^{-1} relative to experimental values) to nucleotides and phosphate without affecting the metal-water interactions, as the metal ion parameters contain separate C_{ij} coefficients for each atom type. This suggests that further parameterization of these C_{ij} coefficients with amino acids offers a reasonable approach to improve the accuracy of MD simulations of metal ions.

In this study we have developed a new set of pairwise C_{ij} coefficients for the Ca^{2+}, Mg^{2+}, Y^{3+} and La^{3+} ions for the 12-6-4LJ nonbonded potential. These metals were chosen to investigate metal binding to a recently discovered lanthanide selective protein, lanmodulin (LanM), which is homologous to the calcium binding protein calmodulin (CaM).[14,15] Ethylenediaminetetraacetic acid (EDTA) and nitrilotriacetic acid (NTA) were used to determine the parameters, as these contain metal ion ligating carboxylate groups similar to those found in glutamate and aspartate (**Figure 1** and **Figure S1**) and accurate binding affinities are available (**Figure 2**, **Table S1**). A linear equation system based on the ratios of ligating oxygens and tertiary nitrogen atoms was developed to deconvolute the contribution from each ligating atom type. The new C_{ij} coefficients were tested against another chelator, egtazic acid (EGTA) as well as the EF1 loop peptides of LanM and CaM. They give binding energies that are significantly closer to experimental values than the 12-6LJ CM/IOD parameter sets or the 12-6-4LJ potential with default C_{ij} coefficients and reproduce the observed metal ion selectivity in LanM.

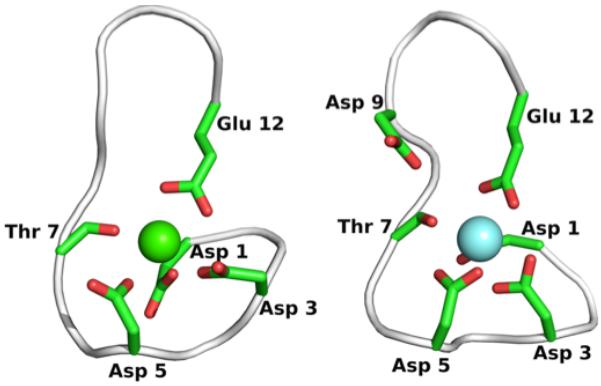


Figure 1. Metal ion coordination in the EF1 loops of CaM (left) and LanM (right). Ca^{2+} is bound to CaM (PDB 1CLL) and Y^{3+} is bound to LanM (NMR conformer 1 of PDB 6MI5).[15,16]

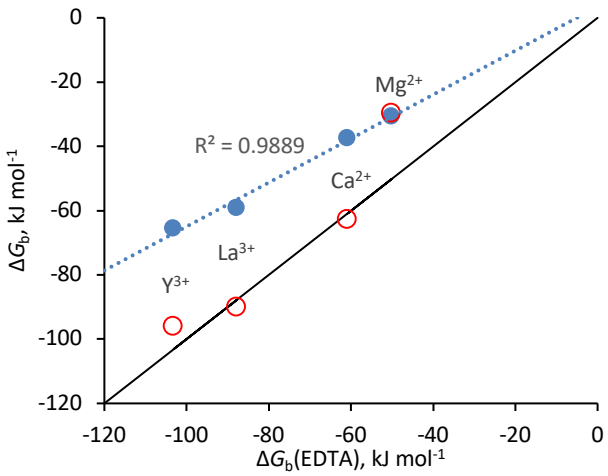


Figure 2. Experimental binding energies for metal binding to NTA (blue) and EGTA (red) plotted against those for EDTA. The black line shows the diagonal ($x = y$). The NTA data are fitted to a linear function (dotted blue line) with a slope of 0.68 and a y intercept of +3 kJ mol^{-1}. Binding affinities are given in Table S1.

Computational methods

MD Simulations and Thermodynamic Integration. The 12-6LJ and 12-6-4LJ nonbonded potentials used by the AMBER ff14SB force field are described by **Eq. 1** and **2**, respectively:[17,18]

$$U_{ij}(r_{ij}) = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{e^2 Q_i Q_j}{r_{ij}} \quad (1)$$

$$U_{ij}(r_{ij}) = \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} - \frac{C_{ij}}{r_{ij}^{4}} + \frac{e^2 Q_i Q_j}{r_{ij}} \quad (2)$$

Essentially, A_{ij}/r_{ij}^{12} is a repulsive term that prevents the attraction from becoming too strong at short distances, B_{ij}/r_{ij}^{6} is an attractive term derived from London dispersion forces, C_{ij}/r_{ij}^{4} accounts for ion-induced dipoles and the $e^2Q_iQ_j/r_{ij}$ Coulombic term accounts for the electrostatic interactions between the atoms.[9,19]

Thermodynamic integration (TI) was used to calculate Ca^{2+}, Mg^{2+}, Y^{3+} and La^{3+} ion HFE and binding energies, according to the standard thermodynamic cycle shown in **Figure S2**. All simulations were carried out using AMBER18 with the ff14SB force field and TIP3P solvation extended at least 15 Å from the solute.[17,18,20] Chelator parameters were determined using the Antechamber package from AmberTools19[17] with charges calculated using the AM1-BCC charge model.[21] Charges are given in **Table S2**. Particle Mesh Ewald was used for long-range nonbonded electrostatic interactions with a cut off distance of 10 Å.[22] The SHAKE algorithm was used to constraint all covalent bonds that involve hydrogen atoms and a 2 fs timestep was used.[23] The temperature was maintained at 300 K using the Langevin thermostat with a 2.0 ps^{-1} collision frequency and the pressure at 1 atm using the Berendsen barostat.[17,24] After initial TI testing it was observed that fewer intermediate λ states were required to reach a converged ΔG_{VdW} value than for the $\Delta G_{ele+pol}$ term, so 9 λ steps were used to calculate ΔG_{VdW} and 12 for $\Delta G_{ele+pol}$ with Gaussian integration.[17] The simulations to determine $\Delta G_{ele+pol}$ were run in triplicate to ensure sufficient sampling and to allow error estimation. The lengths of EM, NVT, NPT and MD sampling per

intermediate λ state are outlined in **Table S3**. More details on the simulation parameters and protocols are available in the Supporting Information.

We tested three existing parameter sets: 12-6LJ HFE, 12-6LJ CM/IOD and the 12-6-4LJ standard parameters. The 12-6LJ HFE parameters should provide accurate ion HFEs but are known to produce IODs that are shorter than experimental values by an average of 0.27 and 0.29 Å for divalent and trivalent ions, respectively.[10] The 12-6LJ CM parameters should provide a compromise between IOD and HFE, but these are not available for trivalent ions. The 12-6-4LJ parameters were reported to reproduce accurate IODs and HFEs.[9] Even though 12-6LJ parameters are expected to produce poorer results, they were included as a benchmark.

In order to calculate average IODs and CNs, data were taken from the TI simulations with λ=0.00922, which reproduced the known IOD and CN of fully charged Ca^{2+} and Mg^{2+} ions in water (**Table S4**). For full-length LanM, a separate 10 ns MD simulation was used. IODs and CNs were calculated from radial distribution functions (RDFs), which were calculated to a resolution of 0.01 Å using VMD.[25,26] The IODs were taken as the peak of a quadratic fit applied to ±0.1 Å of the first peak of the RDF, and CN (taken as the number of atoms within the first coordination sphere around the metal ion) was calculated by integrating the RDF from 0 to the first minimum.

Experimental benchmarking. The 12-6LJ and 12-6-4LJ parameters were obtained using experimental HFE values, so we used the same values for our comparisons.[6,9,10,27] EDTA and NTA metal ion stability constants (K_1) were taken from NIST reports.[28,29] For EGTA, no NIST values was found, so values from the Dojindo metal chelate affinity report were taken.[30] Dissociation constants (K_d) for LanM and CaM were taken from references 14,31-33. Note that these will give average binding energies for the four EF loops in each protein. All experimental values were converted to binding energies in kJ mol^{-1} at 298K (**Table S1**) assuming:

$$\Delta G_b^{exp} = RT \ln(K_1/1\ \mathrm{M}),\ \ K_1 = 1/K_d \qquad (3)$$

Results and discussion

Initially, EDTA binding energies for Ca^{2+}, Mg^{2+}, Y^{3+} and La^{3+} were computed by TI using three established parameter sets:[6,9,10] 12-6LJ HFE and 12-6-4LJ for all metals and 12-6LJ CM for Ca^{2+}, Mg^{2+} or 12-6LJ IOD for Y^{3+} and La^{3+} (12-6LJ CM parameters not available for trivalent ions).The difference between the computed EDTA-metal ion binding energies (ΔG_b^{sim}) and the experimental values in **Figure 2** (ΔG_b^{exp}) are shown in **Figure 3**. HFEs are given in **Table S5** and absolute binding energies are given in **Table S6**. In most cases all three parameter sets significantly overestimate the binding energy (producing more negative values), although they produced relatively accurate binding energies for Ca^{2+}. While 12-6-4LJ produced the best results, there is significant room available for improvement as this showed errors ranging from +4.8 to -52.8 kJ mol^{-1}.

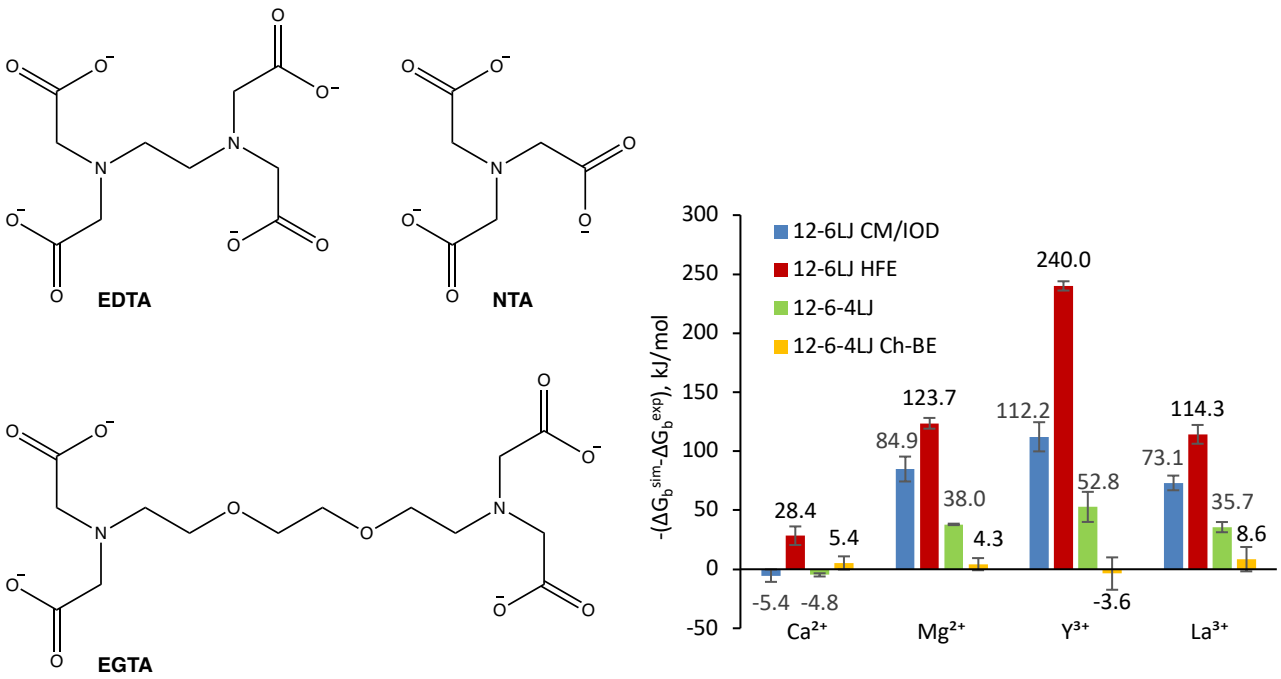


Figure 3. *Left*, the structures of EDTA, NTA and EGTA. *Right*, The difference between experimental and computed EDTA-metal ion binding energies obtained using four sets of parameters. 12-6LJ CM was used for Ca^{2+} and Mg^{2+} while 12-6LJ IOD was used for Y^{3+} and La^{3+}. "12-6-4LJ Ch-BE" denotes the parameters generated during this study.

To increase the accuracy of the TI calculations for the binding energy between chelators and metal ions, we chose to re-parameterize specific C_{ij} coefficients of the 12-6-4LJ parameter set (**Eq. 2**) as this allows the modification of the pairwise interactions between the metal ion and specific atom types (i.e. ligating atoms) without affecting other interatomic interactions and the HFE.[13] EDTA, NTA and EGTA can coordinate metal ions using both carboxylate groups and tertiary nitrogen atoms (**Figure 3**). As the ligating groups are in similar chemical environments in the three chelators, we reasoned that a single set of ligating oxygen and nitrogen parameters can be shared between these molecules. EGTA also contains ether groups, which we did not specifically parameterize in this study. C_{ij} coefficients for the carboxylate oxygen and the tertiary nitrogen were parameterized using TI simulations of EDTA and NTA using the 12-6-4LJ standard parameters. EGTA was then used for benchmarking.

Figure 4 shows ΔG_b^{sim} values for La^{3+} coordination by EDTA and NTA, which were computed at different ligating oxygen and nitrogen C_{ij} values while holding all other C_{ij} parameters to their default 12-6-4-LJ values; i.e. C_{ij}(O) was set to default when varying C_{ij}(N) and vice versa. Data for other metal ions are shown in **Figures S4-6** and for Y^{3+}, C_{ij}(N) was set to 0 while C_{ij}(O) was being varied and vice versa (see Supporting Information). These data all show linear dependences of ΔG_b^{sim} on C_{ij}, so they were fitted to a linear function, with the gradients, m given in **Table 1**. Binding free energies were also computed with C_{ij} values for both ligating oxygen and nitrogen atoms set to 0 and these $\Delta G_b^{sim}(C_{ij} = 0,0)$ values are also given in **Table 1**. In principle, the following relationship should then describe the experimental binding energy:

$$\Delta G_b^{sim}(C_{ij} = 0,0) + m(\mathrm{O})C_{ij}(\mathrm{O}) + m(\mathrm{N})C_{ij}(\mathrm{N}) = \Delta G_b^{exp} \quad (4)$$

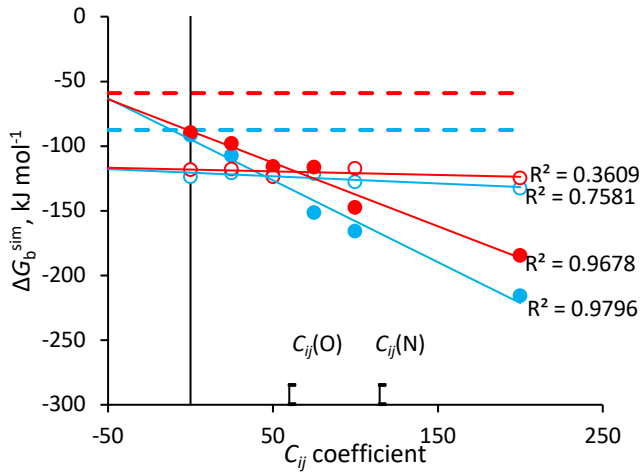


Figure 4. Computed binding energies for the chelation of La^{3+} by EDTA (blue) and NTA (red) as a function of the ligating oxygen and nitrogen 12-6-4LJ C_{ij} values. Filled symbols are for oxygen and open symbols for nitrogen. Solid lines are linear fits to the data and the horizontal dashed lines are the experimental ΔG_b^{exp} values from **Table 1**. The default C_{ij} values are labeled.

Table 1. Metal binding energies (kJ mol^{-1}) for EDTA and NTA and gradient (m) values (kJ mol^{-1} C_{ij}^{-1}) for the data in **Figures 4**, **S4-S6**.

	ΔG_b^{exp} [28,29]		ΔG_b^{sim}, $\Delta G_b^{sim}(C_{ij}=0,0)$ [a]		m(O)		m(N)	
Metal	EDTA	NTA	EDTA	NTA	EDTA	NTA	EDTA	NTA
Ca^{2+}	-60.8	-37.5	-56.0, -41.3 (-24.0)	-32.7, -17.6	-0.593	-0.442	-0.177	-0.081
Mg^{2+}	-50.1	-30.6	-88.1, -37.6 (-21.8)	-84.9, -52.5	-1.028	-0.842	-0.124	+0.077
Y^{3+}	-103.2	-65.5	-156.0, -81.2 (-76.1)	-148.9, -107.1	-0.878	-0.594	-0.032, -0.16 [b]	0
La^{3+}	-87.6	-59.1	-123.3, -85.0 (-79.0)	-115.4, -86.7	-0.632	-0.491	-0.056	-0.028

[a] Computed using standard parameters and with C_{ij} for ligating oxygen and nitrogen groups both set to 0. The EDTA values in (parenthesis) are the $\Delta G_b^{sim\prime}(C_{ij}=0,0)$ values determined using **Eq. 5**. [b] the second m(N) value was obtained with an additional C_{ij} data point as describe in **Figure S6**.

To find a unique solution for C_{ij}(O) and C_{ij}(N), **Eq. 4** must be solved simultaneously for two or more different chelators with different binding energies and m(O) and m(N) values. However, when calculating the binding energies for EDTA and NTA with a range of C_{ij} values, it became apparent that $\Delta G_{\mathrm{b}}{}^{sim}(C_{ij} = 0{,}0)$ values for NTA with Mg^{2+}, Y^{3+} and La^{3+} were significantly larger than the experimental values (**Table 1**). Since larger C_{ij} coefficients increase the binding energy, this meant that in order to satisfy **Eq. 4** at least one of the C_{ij} coefficients would have to become negative, which is not physically realistic. Further, some computed NTA binding energies are larger than the EDTA values for the same metal ion, in opposition to the experimental values (**Table 1**, **Figure 2**). This suggests that the $\Delta G_{\mathrm{b}}{}^{sim}(C_{ij} = 0{,}0)$ are unreliable as computed. Instead, an equivalent binding free energy $\Delta G_{\mathrm{b}}{}^{sim\prime}(C_{ij} = 0{,}0)$ was determined by back-extrapolating from $\Delta G_{\mathrm{b}}{}^{\mathrm{sim}}$ determined using default 12-6-4-LJ values:

$$\Delta G_{\mathrm{b}}{}^{sim\prime}(C_{ij} = 0{,}0) = \Delta G_{\mathrm{b}}{}^{sim} - (m(\mathrm{O})C_{ij\ \mathrm{default}}(\mathrm{O}) + m(\mathrm{N})C_{ij\ \mathrm{default}}(\mathrm{N})) \qquad (5)$$

The EDTA $\Delta G_{\mathrm{b}}{}^{sim\prime}(C_{ij} = 0{,}0)$ values were determined using **Eq. 5** and are given in **Table 1**. Equivalent values can be determined for NTA by scaling the EDTA values by the ratio of the experimental binding constants:

$$\Delta G_{\mathrm{b}}{}^{sim\prime}(\mathrm{NTA}, C_{ij} = 0{,}0) = \frac{\Delta G_{\mathrm{b}}{}^{exp}(\mathrm{NTA})}{\Delta G_{\mathrm{b}}{}^{exp}(\mathrm{EDTA})} \Delta G_{\mathrm{b}}{}^{sim\prime}(\mathrm{EDTA}, C_{ij} = 0{,}0) \qquad (6)$$

This approach ensures that the $\Delta G_{\mathrm{b}}{}^{sim\prime}(C_{ij} = 0{,}0)$ values for different chelators follow the experimental binding energy trends.

As seen in **Table 1**, the Mg^{2+} and Y^{3+} the NTA m(N) values are not physically realistic (non-negative). Visual inspection of the relevant MD trajectories suggested that these metal ions did not interact with the tertiary nitrogen atoms in the same way as EDTA (**Figure S7**), which likely leads to underestimation of these m(N) values.[34] From the EDTA and NTA Ca^{2+} simulations, we noted that the ratios of the m(O) and m(N) values are approximately ¾ and ½ for oxygen and nitrogen, respectively, thus mirroring the maximum coordination numbers available (**Figure 3**).[35] Based on this observation, we can modify **Eq. 4** to scale the EDTA m(O) and m(N) values by ¾ and ½, respectively in order to describe the NTA binding energy. Making use of this approximation, and by substituting **Eq. 5** and **6** into **Eq. 4**, we can determine unique C_{ij} values using only the EDTA TI simulation data and the experimental binding energies for EDTA and NTA in **Eq. 7**. The resulting ligating oxygen and nitrogen C_{ij} coefficients are given in **Table 2**.

$$\begin{cases} \Delta G_{\mathrm{b}}{}^{sim\prime}(EDTA, C_{ij} = 0{,}0) + m(\mathrm{EDTA, O})C_{ij}(\mathrm{O}) + m(\mathrm{EDTA, N})C_{ij}(\mathrm{N}) = \Delta G_{\mathrm{b}}{}^{exp}(\mathrm{EDTA}) \\ \Delta G_{\mathrm{b}}{}^{sim\prime}(\mathrm{NTA}, C_{ij} = 0{,}0) + 3/4\ m(\mathrm{EDTA, O})C_{ij}(\mathrm{O}) + 1/2\ m(\mathrm{EDTA, N})C_{ij}(\mathrm{N}) = \Delta G_{\mathrm{b}}{}^{exp}(\mathrm{NTA}) \end{cases} \qquad (7)$$

Table 2. Default and newly derived "Ch-BE" 12-6-4LJ pairwise oxygen and nitrogen C_{ij} coefficients for Ca^{2+}, Mg^{2+}, Y^{3+} and La^{3+} ligation by EDTA and NTA.

Metal	default		Ch-BE	
	O	N	O	N
Ca^{2+}	34.4	65.9	29.2	110.6
Mg^{2+}	52.4	100.3	12.3	126.1
Y^{3+}	85.1	163.0	16.6 [a]	78.3 [a]
La^{3+}	59.9	114.7	9.5	46.3

[a] To obtain new C_{ij} coefficients for Y^{3+} ions two different m(N) values were used (**Table 1**). For more details see **Figure S6**.

Validation of new Cij coefficients. Our new C_{ij} coefficients for the 12-6-4LJ nonbonded potential, which we denote "12-6-4LJ Ch-BE," resulted in significant improvements to the TI-computed binding energies compared to the 12-6LJ CM/IOD, 12-6LJ HFE and default 12-6-4LJ parameter sets. These are shown in **Figure 3** and the absolute binding energy values are given in **Table S6**. Our new binding energies are all within 8.6 kJ mol^{-1} (~2 kcal mol^{-1}) of the experimental values, while the previously best results using the 12-6-4LJ default set differed by up to ~50 kJ mol^{-1} from the experimental data. The average absolute error was reduced to 5.5 kJ mol^{-1} compared to 32.8 kJ mol^{-1} for the default 12-6-4LJ set.

To ensure that the new C_{ij} coefficients were not overfitted for EDTA, they were also tested on NTA (the NTA TI values were not used for parametrization; see **Eq. 7**) and the similar chelator EGTA (**Figure 3**). The difference in computed and experimental binding energies are shown in **Figure 5** and the absolute binding energy values are given in **Table S7**. In each case a clear improvement was observed. Overall, the "12-6-4LJ Ch-BE" parameter set reduced the absolute average binding energy errors from 49.7 kJ mol^{-1} to 26.2 kJ mol^{-1} for NTA and from 30.9 kJ mol^{-1} to 14.4 kJ mol^{-1} for EGTA, relative to the default 12-6-4LJ parameters. This represents ~1.9 fold and ~2.1-fold increase in accuracy, respectively for NTA and EGTA, although it should be noted that there was significant variability in the EGTA values with Y^{3+} using both forcefields. This arose due to EGTA adopting another conformation in some simulations, which introduces additional hydrogen bonding between EGTA and solvent molecules. More detailed analysis of these results are discussed in the Supporting Information.

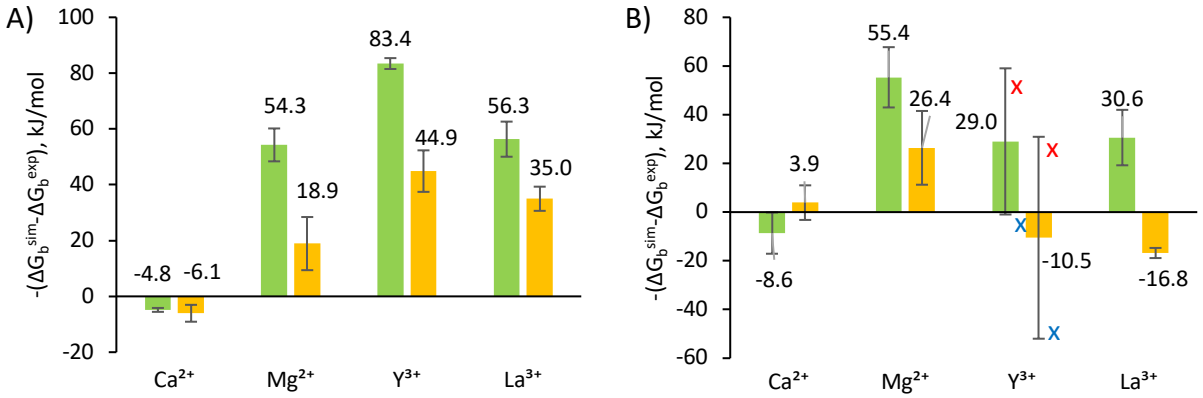


Figure 5. Difference between computed NTA (A) and EGTA (B) (ΔG_b^{sim}) and experimental (ΔG_b^{exp}) binding energies obtained using TI with default 12-6-4LJ (green) and 12-6-4-LJ Ch-BE (yellow) O and N C_{ij} coefficients for the ligating oxygen and nitrogen groups. Energies are given below or above the bar and the red and blue crosses in (B) indicate the approximate energies for EGTA Y^{3+} in two different chelator conformations (**Figure S8**).

Next, we compared the 12-6-4LJ Ch-BE parameters to the 12-6-4LJ default and 12-6LJ CM/IOD parameter using TI simulations of metal ion binding to the CaM and LanM 12-residue EF1 loop peptides (**Figures 1**, **S1**). The 12-6LJ HFE set was not selected as it was the worst performing set of parameters with EDTA (**Figure 3** and **Table S6**), and it is known to produce inaccurate IODs.[6,10] Although 12-6LJ CM/IOD produced poorer results in EDTA metal ion simulations than 12-6-4LJ (**Figure 3**) it was still included here as 12-6LJ potentials are more widely used. As the EF-hand peptides do not possess tertiary amines, only the new metal ion-oxygen pairwise C_{ij} coefficients were used. These were used for the carboxylate oxygens, while the coordinating oxygen atom from the backbone of Thr 7 retained the default C_{ij} coefficient as it is part of an amide bond which has not yet been re-parameterized. The length of the constant-pressure equilibration (NPT) and MD trajectories were optimized using LanM EF1 peptide with Ca^{2+} (**Figure S9**). 2 ns of NPT and 5 ns of MD sampling were used as they give reasonably converged binding energies. The van der Waals contribution to the chelator binding energies is relatively small and does not vary significantly; for Ca^{2+}, ΔG_{VdW} = 9.5 kJ mol^{-1} in EDTA, 9.2 kJ mol^{-1} in NTA, 9.1 kJ mol^{-1} in EGTA and 9.1 kJ mol^{-1} in LanM EF1. Consequently, we chose to use a fixed $-\Delta G_{VdW}$ contribution of 9 kJ mol^{-1} for all metal ions in both the LanM EF1 and CaM EF1 systems to reduce computational cost. The $\Delta G_{ele+pol}$ portion of the TI calculations was run in triplicate for each system to improve sampling and allow error estimation. The van der Waals contribution ($-\Delta G_{VdW}$) to ΔG_b^{sim} was not included in the error analysis. The average binding energies for LanM EF1 and CaM EF1 are shown in **Table 3**.

Table 3. Mean binding energies (kJ mol^{-1}) for metal binding to the LanM and CaM EF1 loop peptides.

Metal	exp[14,31-33]	12-6LJ CM/IOD [a]	12-6-4LJ	12-6-4LJ Ch-BE
LanM				
Ca^{2+}	-18.0	-67.4	-52.3	-55.5 ± 2.7
Mg^{2+}	NA [b]	-119.2	-84.3	-39.9 ± 8.6
Y^{3+}	-61.4	-205.6	-151.7	-70.0 ± 7.6
La^{3+}	-64.3	-159.6	-127.9	-95.8 ± 7.8
CaM				
Ca^{2+}	-33.9	-53.6	-71.5	-54.1 ± 1.9
Mg^{2+}	-22.8	-121.1	-100.3	-57.8 ± 3.2
Y^{3+}	NA [b]	-189.5	-148.3	-84.9 ± 5.6
La^{3+}	-45.3	-152.3	-126.7	-89.1 ± 3.2

[a] 12-6LJ CM was used for divalent and 12-6LJ IOD for trivalent ions; [b] not available

As seen in **Table 3**, all simulations significantly overestimate the binding energies. As the experimental values are the average affinity for the 3-4 EF loop binding sites in LanM and CaM, some of this error may reflect differences in affinity between the different EF loops in each protein. The 12-6LJ CM/IOD parameter sets performed the worst in each case, except for Ca^{2+} binding to CaM EF1. The 12-6-4LJ Ch-BE parameters performed significantly better than the default 12-6-4LJ parameters, but still significantly overestimate the binding energy for all cases. The average errors in binding energy are 84.7 kJ mol^{-1}, 63.1 kJ mol^{-1} and 28.4 kJ mol^{-1} for 12-6LJ CM/IOD, 12-6-4LJ and 12-6-4LJ Ch-BE, respectively, and there is a 3-fold and 2.2-fold increase in accuracy for the 12-6-4LJ Ch-BE parameter set compared to 12-6LJ CM/IOD and 12-6-4LJ, respectively. The observed standard deviations in $\Delta G_{\mathrm{b}}{}^{sim}$ values (between triplicate $\Delta G_{\mathrm{ele+pol}}$ simulations) are relatively low, not exceeding 9 kJ mol^{-1}, indicating that these simulations give consistent results when starting from the same input geometry.

The 12-6-4LJ Ch-BE parameters were the only parameters that reproduced the experimentally observed order of binding affinities of LanM EF1 with three metal ions, La^{3+}>Y^{3+}>Ca^{2+}.[14] Although there are no reported LanM EF-Mg^{2+} binding energies, the data in **Table 3** predict that the affinity is weaker than for Ca^{2+}. This is usually the case for CaM EF-hands.[31,32] For CaM EF1, all three parameter sets failed to predict the correct order of affinities for the 3 metal ions: La^{3+}>Ca^{2+}>Mg^{2+} (**Table 3**). Instead, they predict the same binding affinity order of: La^{3+}>Mg^{2+}>Ca^{2+}. Nevertheless, the 12-6-4LJ Ch-BE parameters performed significantly better

than the other parameter sets. There is no experimental data for the Y^{3+}, but the predicted binding energy is similar to that for La^{2+}.

To gain better insight into EF-hand metal ion coordination, the IODs and CNs were calculated using the same 3 parameter sets and these values are given in **Table 4**. Additionally, to compare the metal ion coordination of isolated EF-hands to the EF-hand motifs in a full-length protein, the IODs and CNs for each of the EF1, EF2 and EF3 binding sites in full-length LanM were determined with bound Ca^{2+}, Mg^{2+}, Y^{3+} and La^{3+} during 10ns MD simulations (**Tables 4** and **S8**). In general, the new 12-6-4LJ Ch-BE parameters produced very similar results to the existing 12-6-4LJ parameters.

Table 4. Ion-oxygen distances (IODs in Å) and coordination numbers (CNs) of metal ions in full-length LanM and the EF1 loop peptides of LanM and CaM.

	12-6LJ CM/IOD [a]		12-6-4LJ		12-6-4LJ Ch-BE [b]	
Metal	CN	IOD	CN	IOD	CN	IOD
LanM EF1 peptide						
Ca^{2+}	8.3-8.7 [c]	2.43	8.1	2.39	8.0-8.5 [c]	2.40
Mg^{2+}	6.0	1.94	7.0	2.08	6.0-6.7 [c]	2.07
Y^{3+}	9.0	2.28	9.0	2.30	9.0-9.3 [c]	2.32
La^{3+}	10.0	2.45	10.0	2.50	10.0	2.49
LanM EF1 in full-length LanM						
Ca^{2+}	7.7-7.8 [c]	2.41	8.0-8.1 [c]	2.39	7.8-8.0 [c]	2.38
Mg^{2+}	6.0	1.94	6.5-6.8 [c]	2.07	6.1-6.3 [c]	2.06
Y^{3+}	9.0	2.29	9.0	2.29	8.9	2.29
La^{3+}	9.9	2.44	10.0	2.46	10.0	2.48
CaM EF1 peptide						
Ca^{2+}	8.2	2.42	8.0	2.40	7.7-8.0 [b]	2.40
Mg^{2+}	6.0	1.94	7.0	2.09	6.0	2.05
Y^{3+}	9.0	2.29	9.1	2.31	9.0	2.31
La^{3+}	10.0	2.45	10.0	2.48	10.0	2.49

[a] 12-6LJ CM for divalent and 12-6LJ IOD for trivalent ions; [b] IODs were calculated as the average from triplicate simulations, whereas for CNs if different values were obtained between simulations the full observed range is shown; [c] no clear minima was observed so the range is given.

As seen in **Table 4**, Ca^{2+} and Mg^{2+} ion coordination in LanM and CaM EF-hands is not represented by a single binding mode as the CN values are non-integer. Y^{3+} and La^{3+} ions displayed

consistent CN numbers in both LanM EF1 and CaM EF1 being ~9 and ~10, respectively. Inspection of the simulations revealed that some coordinating residues were fluctuating between monodentate and bidentate ligating modes (**Figure S10**). This is likely to be physically realistic (e.g. as is observed in CaM, PDB 1CFF). However, if a fixed integer coordination number is required while retaining the ability to switch ligands this could be achieved by applying the cationic dummy atom model (CDAM) in combination with re-parameterization of the ligating atom C_{ij} terms. CDAM partitions the metal ion charge between itself and dummy atoms surrounding the central metal ion in a predefined geometry, which interacts with the surrounding atoms and can freely exchange ligands.[36] Recently, CDAM was combined with 12-6-4LJ potential for Mg^{2+}, Fe^{3+}, Al^{3+} and Cr^{3+} ions in a predefined octahedral geometry and was shown to reproduce HFE, IOD and CN values in a water solution.[37]

The computed IODs from the Ca^{2+}, Mg^{2+}, Y^{3+} and La^{3+} simulations with the EF1 peptides and full-length LaM do not significantly differ between the 12-6-4LJ and 12-6-4LJ Ch-BE parameters. These IODs varied within ± 0.02 Å in cases where CN remained the same (**Table 4**). The 12-6LJ CM/IOD parameter set usually gave rise to lower IODs for Mg^{2+}, Y^{3+} and La^{3+} while the Ca^{2+} IODs was relatively similar to those computed with the 12-6-4LJ forcefield. All three parameter sets produced relatively accurate average IODs for Y^{3+} in LanM EF1 (~2.3 Å) compared to experimental structures in PDB 6MI5 (2.2-2.4 Å depending on ligating group).[15] The Ca^{2+} IODs were also consistent with experimental CaM structures. Experimental IODs of 2.42 Å for monodentate and 2.41 Å for bidentate ligands in EF1 of PDB 1CLL are in good agreement with the 12-6-4LJ Ch-BE average IODs of 2.40 Å for 2 bidentate and 4 monodentate ligands.[16] For Mg^{2+} the 12-6-4LJ Ch-BE average IOD of 2.05 Å is comparable with the experimental monodentate ligation distance of 2.11 Å in EF1 of PDB 3UCW.[38] No crystal structures of LanM or CaM with La^{3+} are available for comparison.

Conclusions

We have demonstrated a method of generating C_{ij} parameters for 12-6-4LJ molecular mechanics forcefields that allows parametrization of specific ligating groups in order to tune binding energies computed by TI. The new C_{ij} coefficients were tested on a series of chelators: EDTA, NTA, EGTA and EF1 loop peptides from LanM and CaM proteins and showed significant improvements in computed binding energies relative to existing forcefields. The new parameters also produce CN and IOD values that are in good agreement with experimental values. The parametrization method should be extensible to a range of other systems and could be readily adapted to tune properties other than binding energies.

Acknowledgements

This work was supported in part by BBSRC grant: BB/M017702/1 and a BBSRC DTP studentship to CM. The authors acknowledge the assistance given by IT Services and the use of the Computational Shared Facility at The University of Manchester.

References

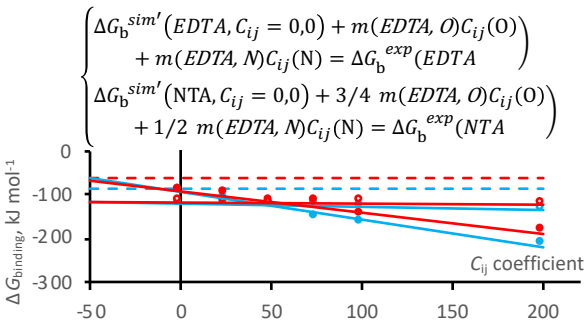
1. Thomson, A. J., Gray, H. B. Bio-inorganic chemistry. *Curr. Opin. Chem. Biol.* **1998**, *2*, 155– 158, DOI: 10.1016/S1367-5931(98)80056-2
2. Andreini, C., Bertini, I., Cavallaro, G., Holliday, G. L., Thornton, J. M. Metal ions in biological catalysis: from enzyme databases to general principles. *J. Biol. Inorg. Chem.* **2008**, *13*, 1205– 1218, DOI: 10.1007/s00775-008-0404-5
3. Clapham, D. E. Calcium signaling. *Cell* **2007**, *131*, 1047– 1058, DOI: 10.1016/j.cell.2007.11.028
4. Maret, W. Zinc in Cellular Regulation: The Nature and Significance of "Zinc Signals". *Int. J. Mol. Sci.* **2017**, *18*, 2285, DOI: 10.3390/ijms18112285
5. Marcus, R. A. Chemical and Electrochemical Electron-Transfer Theory. *Annu. Rev. Phys. Chem.* **1964**, *15*, 155– 196, DOI: 10.1146/annurev.pc.15.100164.001103
6. Li, P. F., Roberts, B. P., Chakravorty, D. K., Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733– 2748, DOI: 10.1021/ct400146w
7. Li, P. F., Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117*, 1564– 1686, DOI: 10.1021/acs.chemrev.6b00440
8. Peters, M. B., Yang, Y., Wang, B., Fusti-Molnar, L., Weaver, M. N., Merz, K. M. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6*, 2935– 2947, DOI: 10.1021/ct1002626
9. Li, P. F., Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10*, 289– 297, DOI: 10.1021/ct400751u
10. Li, P. F., Song, L. F., Merz, K. M. Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883– 895, DOI: 10.1021/jp505875v
11. Li, P. F., Song L. F., Merz, K. M. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory Comput.* **2015**, *11*, 1645– 1657, DOI: 10.1021/ct500918t
12. Miller, K. J. Calculation of the molecular polarizability tensor. *J. Am. Chem. Soc.* **1990**, *112*, 8543– 8551, DOI: 10.1021/ja00179a045
13. Panteva, M. T., Giambasu, G. M., York, D. M. Force Field for Mg2+, Mn2+, Zn2+, and Cd2+ Ions That Have Balanced Interactions with Nucleic Acids. *J. Phys. Chem. B* **2015**, *119*, 15460– 15470, DOI: 10.1021/acs.jpcb.5b10423
14. Cotruvo, J. A., Featherston, E. R., Mattocks, J. A., Ho, J. V., Laremore, T. N. Lanmodulin: A Highly Selective Lanthanide-Binding Protein from a Lanthanide-Utilizing Bacterium. *J. Am. Chem. Soc.* **2018**, *140*, 15056– 15061, DOI: 10.1021/jacs.8b09842
15. Cook, E. C., Featherston, E. R., Showalter, S. A., Cotruvo, J. A. Structural Basis for Rare Earth Element Recognition by Methylobacterium extorquens Lanmodulin. *Biochemistry* **2019**, *58*, 120– 125, DOI: 10.1021/acs.biochem.8b01019
16. Chattopadhyaya, R., Meador, W. E., Means, A. R., Quiocho, F. A. Calmodulin structure refined at 1.7 angstrom resolution. *J. Mol. Biol.* **1992**, *228*, 1177– 1192, DOI: 10.1016/0022-2836(92)90324-D
17. Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghoreishi, D., Giambasu, G., Giese, T., Gilson, M. K., Gohlke, H., Goetz, A. W., Greene, D., Harris, R., Homeyer, N., Huang, Y., Izadi, S., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C.,

Liu, J., Luchko, T., Luo, R., Man, V., Mermelstein, D. J., Merz, K. M., Miao, Y., Monard, G., Nguyen, C., Nguyen, H., Onufriev, A., Pan, F., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C. L., Smith, J., Swails, J., Walker, R. C., Wang, J., Wei, H., Wilson, L., Wolf, R. M., Wu, X., Xiao, L., Xiong, D. M., York D. M., Kollman P. A. AMBER 2019, University of California, San Francisco, **2019**

18. Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696– 3713, DOI: 10.1021/acs.jctc.5b00255
19. Lennard-Jones, J. E. Cohesion. *Proc. Phys. Soc.* **1931**, *43*, 461– 482, DOI: 10.1088/0959-5309/43/5/301
20. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926– 935, DOI: 10.1063/1.445869
21. Jakalian, A., Bush, B.L., Jack, B.D., Bayly, C.I., Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comp. Chem.* **2000**, *21*, 132-146, DOI: 10.1002/(SICI)1096-987X(20000130)21:2<132::AID-JCC5>3.0.CO;2-P
22. Darden, T., York, D., Pedersen, L. Particle Mesh Ewald - an n.log(n) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089– 10092, DOI: 10.1063/1.464397
23. Ryckaert, J-P., Ciccotti, G., Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327– 341, DOI: 10.1016/0021-9991(77)90098-5
24. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., Haak, J. R. Molecular-Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684– 3690, DOI: 10.1063/1.448118
25. Humphrey, W., Dalke, A., Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics Modell.* **1996**, *14*, 33– 38, DOI: 10.1016/0263-7855(96)00018-5
26. Levine, B. G., Stone, J. E., Kohlmeyer, A. Fast analysis of molecular dynamics trajectories with graphics processing units-Radial distribution function histogramming. *J. Comput. Phys.* **2011**, *230*, 3556– 3569, DOI: 10.1016/j.jcp.2011.01.048
27. Marcus, Y. Thermodynamics of solvation of ions. Part 5.–Gibbs free energy of hydration at 298.15K. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2995– 2999, DOI: 10.1039/ft9918702995
28. Anderreg, G. Critical Survey of Stability Constants of EDTA Complexes. IUPAC chemical data series, no 14. Pergamon Press, Oxford, U.K. **1977**.
29. Anderegg, G. Critical Survey of Stability-Constants of NTA Complexes. *Pure Appl. Chem.* **1982**, *54*, 2693– 2758, DOI: 10.1351/pac198254122693
30. Dojindo Molecular Technologies, Inc. Chelate table of stability constants. Retrieved from: https://www.dojindo.co.jp/technical/pdf/Chelate_Table_of_Stability_Constants.pdf
31. Linse, S., Helmersson, A., Forsen, S. Calcium-binding to calmodulin and its globular domains. *J. Biol. Chem.* **1991**, *266*, 8050– 8054, DOI: 10.1016/S0021-9258(18)92938-8
32. Tsai, M. D., Drakenberg, T., Thulin, E., Forsen, S. Is the binding of magnesium (II) to calmodulin significant? An investigation by magnesium-25 nuclear magnetic resonance. *Biochemistry* **1987**, *26*, 3635– 3643, DOI: 10.1021/bi00386a057
33. Xu, K., Yang, X. D., Wang, K. Metal binding discrimination of the calmodulin Q41C/K75C mutant on Ca2+ and La3+. *Sci. China Chem.* **2010**, *53*, 797– 806, DOI: 10.1007/s11426-010-0059-2

34. Kremer, C., Torres, J., Dominguez, S. Lanthanide complexes with oda, ida, and nta: From discrete coordination compounds to supramolecular assemblies. *J. Mol. Struct.* **2008**, *879*, 130– 149, DOI: 10.1016/j.molstruc.2007.08.024
35. Barnett, B. L., Uchtman, V. A. Structural Investigations of Calcium-Binding Molecules .4. Calcium-Binding to Aminocarboxylates - Crystal-Structures of Ca(CaEDTA)-7H20 and Na(CaNTA). *Inorg. Chem.* **1979**, *18*, 2674– 2678, DOI: 10.1021/ic50200a007
36. Aqvist, J., Warshel, A. Free-Energy Relationships in Metalloenzyme-Catalyzed Reactions - Calculations of the Effects of Metal-Ion Substitutions in Staphylococcal Nuclease. *J. Am. Chem. Soc.* **1990**, *112*, 2860– 2868, DOI: 10.1021/ja00164a003
37. Liao, Q. H., Pabis, A., Strodel, B., Kamerlin, S. C. L. Extending the Nonbonded Cationic Dummy Model to Account for Ion-Induced Dipole Interactions. *J. Phys. Chem. Lett.* **2017**, *8*, 5408– 5414, DOI: 10.1021/acs.jpclett.7b02358
38. Senguen, F. T., Grabarek, Z. X-ray Structures of Magnesium and Manganese Complexes with the N-Terminal Domain of Calmodulin: Insights into the Mechanism and Specificity of Metal Ion Binding to an EF-Hand. *Biochemistry* **2012**, *51*, 6182– 6194, DOI: 10.1021/bi300698h

Table of Contents Graphic:



Supporting information

Chelator-based parameterization of the 12-6-4LJ molecular mechanics potential for more realistic metal ion-protein interactions

*Paulius Kantakevičius, Calvin Mathiah, Linus O. Johannissen and Sam Hay**

Manchester Institute of Biotechnology and Department of Chemistry, The University of Manchester, Manchester, United Kingdom. sam.hay@manchester.ac.uk.

Additional details: Molecular dynamics simulations

For chelators: EDTA, NTA and EGTA the atom types were generated using the Antechamber package from AmberTools19,[1,2] with charges calculated using the AM1-BCC charge model.[3] The obtained charges of all atom types are displayed in Table S2.

For full LanM MD simulations that were used in CN and IOD calculations State 1 from NMR structure (PDB 6MI5) with one metal ion in each: EF1, EF2 and EF3 hand was used as initial structure.[4] EF4 was left without a metal ion as it contains the lowest affinity towards metal ions and it did not contain a metal ion in the original structure. To produce full LanM MD trajectory for IOD and CN calculations three step EM was done with 2000/3000, 2000/3000, 20000/30000 steepest and conjugate descent cycles and 500 kJ mol^{-1}, 10 kJ mol^{-1}, 0 kJ mol^{-1}, restraints applied on the protein and metal ion in each step, respectively. Two step NVT equilibration of 500 ps each was done with 10 kJ mol^{-1} restraints applied on the protein and metal ion during the first step. Then 2 ns NPT equilibration and 10 ns MD sampling with no constraints were carried out.

From Figure S9 it was observed that 2 ns of NPT and 5 ns of sampling gives reasonably converged binding energies. Sampling lengths above 5 ns did not improve the results as the plateau was already reached, whereas NPT longer than 2 ns gives a small ~2 kJ mol^{-1} improvement when the NPT length is extended to 5 ns. However, this improvement is not large enough to justify the increase in processing time as the majority of convergence (~10 kJ mol^{-1}) happened during the first 2 ns of NPT. The same length NPT and MD sampling was applied to CaM EF1 as the geometry and amino acid sequence (and thus the system size) is highly similar to that of Lan EF1 as seen in Figures 1 and S1, respectively

Y^{3+} parameterization

When computing binding energies (ΔG_b^{sim}) with different C_{ij}(O) values the C_{ij}(N) was fixed to 0 instead of the default value and vice versa as simulations with C_{ij}(O) = 0 and C_{ij}(N) = 0 become the same as $\Delta G_b^{sim}(C_{ij} = 0{,}0)$. However, as computed $\Delta G_b^{sim}(C_{ij} = 0{,}0)$ simulations were deemed unreliable and it was decided to back-extrapolate $\Delta G_b^{sim\prime}(C_{ij} = 0{,}0)$ using Eq. 5 fixing non-varied C_{ij}(O) or C_{ij}(N) coefficient to 0 became unnecessary. Another set of EDTA simulations were run with Cij(O) and Cij(N) coefficients fixed to the default values (Figure S6 b), which did not change the calculated 12-6-4LJ Ch-BE coefficients significantly (C_{ij}(O) changed from 16.6 to 14.9 and C_{ij}(N) from 78.3 to 77.6).

Y^{3+} binding to EGTA

Visual inspection of the relevant MD trajectories indicated that with both default and new C_{ij} coefficients Y^{3+} ions were coordinated inconsistently by EGTA. In some simulations EGTA would use most of its ligating groups and one water molecule for coordination of metal ion, whereas in other cases the Y^{3+} ion would move outwards and become coordinated by a higher number of water molecules as shown in Figure S8. Interestingly, EGTA systems with more water molecules coordinating the Y^{3+} ion were more energetically favorable as TI calculated higher binding energies. The energy difference between these states varied by over 80 kJ mol^{-1} between replicate runs with new 12-6-4LJ C_{ij} coefficients producing a lowest binding energy of -41.8 kJ mol^{-1} ("normal" state) and highest of -124.3 kJ mol^{-1} ("EGTA-solvent hydrogen bonded" state). The energetic difference may be explained by a metal ion-coordinating water molecule forming two hydrogen bonds to the ether oxygens of EGTA and another water molecule stabilizing metal ion coordinating carboxylate groups, which likely stabilized this geometry thus increasing the binding energy (Figure S8). To further investigate this, EGTA-Ca^{2+} simulations were inspected and no such phenomenon was observed as in all cases Ca^{2+} ions used ligands provided by EGTA as observed in the crystal structure.[5]

Table S1. Experimental metal ion hydration free energies (HFE) and binding energies (in kJ mol^{-1}) for EDTA, NTA, EGTA, LanM and CaM. Experimental logK_1 values for EDTA, NTA and EGTA and K_d values (in mol L^{-1}) for LanM and CaM are included in (parenthesis).

Metal	Ion HFE[6]	Binding energy EDTA[7]	NTA[8]	EGTA[9]	LanM EF1[10]	CaM EF1[11-13]
Ca^{2+}	-1505	-60.8, (10.65)	-37.5, (6.57)	-62.8, (11.00)	-18.0, (7.1×10^{-4})	-39.9, (1×10^{-7})
Mg^{2+}	-1830	-50.1, (8.79)	-30.6, (5.36)	-29.7, (5.21)	NA [a]	-22.8, (1×10^{-4})
Y^{3+}	-3450	-103.2, (18.08)	-65.5, (11.48)	-96.0, (16.82)	-61.4, (1.7×10^{-11})	NA [a]
La^{3+}	-3145	-87.6, (15.36)	-59.1, (10.36)	-90.1, (15.79)	-64.3, (5.3×10^{-12})	-45.3, (1.17×10^{-8})

[a] Not available.

Table S2. Atom types and charges generated by Antechamber for chelators EDTA, NTA and EGTA. Atom types/numbering are shown in Figure S3.

Atom type	Charge		
	EDTA	NTA	EGTA
O	-0.8922	-0.9003	-0.8797
C	0.9123	0.8996	0.8976
C3	0.0711	0.0758	0.0648
H1	0.0186	0.0282	0.0282
N3	-0.6946	-0.6926	-0.7106
C3a	0.1493		0.1668
H1a	0.0367		0.0552
C3b			0.1254
H1b			0.0189
C3c			0.1019
H1c			0.041
Os			-0.4311

Table S3. Summary of energy minimization (EM), NVT, NPT and molecular dynamics (MD) sampling lengths per each λ step used for thermodynamic integration in different systems.

System	λ steps (Ele+pol, VdW)	EM steepest, conjugate descent (number of cycles)	NVT (ps)	NPT (ps)	MD sampling (ps)
Water box	9, 9	500, 500 (1000) [b]	20	100	900
EDTA, NTA, EGTA	12 (19), 9 [a]	500, 500 (1000) [b]	20-50 [c]	100-500 [c]	900-1900 [c]
EF-hand	12, 9	2000, 3000 (5000) [b]	500	2000	5000

[a] in several cases 19 λ steps with 0.05 step size were used with rectangular integration, however 12 λ steps with Gaussian integration is more computationally efficient; [b] for VdW removal steepest descent cannot be used, only conjugate descent was used with the number of steps indicated in parentheses; [c] there was some variation in NVT, NPT and MD sampling lengths between simulations in chelators – ranges are given in the table.

Table S4. Coordination number (CN) and ion-oxygen distance (IOD in Å) values for metal ions in water solution calculated from λ=0.00922 thermodynamic integration (TI) simulations.

	exp[14,15]		12-6LJ CM/IOD [a, b]		12-6LJ HFE [b]		12-6-4LJ [b]	
Metal	CN	IOD	CN	IOD	CN	IOD	CN	IOD
Water box								
Ca^{2+}	8	2.46	8/8 [c]	2.50/ 2.49 [c]	7.3-7.4/ 7.4 [c]	2.33/ 2.33 [c]	8/8 [c]	2.47/ 2.46 [c]
Mg^{2+}	6	2.09	6/6 [c]	2.04/ 2.03 [c]	6/6 [c]	1.95/ 1.95 [c]	6/6 [c]	2.09/ 2.09 [c]
Y^{3+}	8	2.36	8.9-9	2.36	~8-9	2.16	9/9 [c]	2.36/ 2.36 [c]
La^{3+}	8-9.1	2.52	9	2.49	9	2.40	9.4-9.6/ 9.7 [c]	2.52/ 2.53 [c]

[a] 12-6LJ CM parameter set was used for divalent ions and 12-6LJ IOD for trivalent ions; [b] IODs were calculated as an average of 3 runs whereas for CNs if different values were obtained between runs the full observed range is shown; [c] first number for CN and IOD for Ca^{2+} and Mg^{2+} ions are calculated from the λ=0.00922 TI simulations, while the second number are taken from the parameterization studies. Trivalent ions did not contain values in parameterization studies for 12-6LJ IOD and HFE parameter sets.[16-18]

Table S5. Hydration free energies (HFE, in kJ mol^{-1}) calculated using thermodynamic integration with three different metal parameter sets.

Metal	exp[6]	12-6LJ CM/IOD [a]	12-6LJ HFE	12-6-4LJ
		HFE		
Ca^{2+}	-1505	-1402.0 ± 0.9	-1510.0 ± 0.1	-1507.5 ± 0.5
Mg^{2+}	-1830	-1727.0 ± 1.2	-1833.0 ± 0.6	-1829.1 ± 0.3
Y^{3+}	-3450	-3180.9 ± 1.7	-3416.2 ± 0.2	-3451.2 ± 0.9
La^{3+}	-3145	-3007.0 ± 2.7	-3143.7 ± 2.9	-3137.4 ± 1.5

[a] 12-6LJ CM parameter set for divalent ions and 12-6LJ IOD for trivalent ions.

Table S6. Ethylenediaminetetraacetic acid (EDTA) metal ion binding energies (kJ mol^{-1}) calculated using thermodynamic integration with four parameter sets.

Metal	exp[7]	12-6LJ CM/IOD [a]	12-6LJ HFE	12-6-4LJ	12-6-4LJ Ch-BE [b]
Ca^{2+}	-60.8	-55.4 ± 5.2	-89.2 ± 7.9	-56.0 ± 1.3	-66.2 ± 5.6
Mg^{2+}	-50.1	-135.0 ± 10.5	-173.8 ± 4.6	-88.1 ± 0.6	-54.4 ± 5.1
Y^{3+}	-103.2	-215.4 ± 12.4	-343.2 ± 3.9	-156.0 ± 12.7	-99.6 ± 13.7
La^{3+}	-87.6	-160.7 ± 6.2	-201.9 ± 8.0	-123.3 ± 4.3	-96.2 ± 10.3

[a] 12-6LJ CM used for divalent ions and 12-6LJ IOD for trivalent ions; [b] 12-6-4LJ new is the parameter set developed in this study.

Table S7. Absolute nitriloacetic acid (NTA) and egtazic acid (EGTA) binding energies with default and new C_{ij} coefficients of 12-6-4LJ potential in kJ mol^{-1}.

Metal	exp[8,9]	12-6-4LJ	12-6-4LJ Ch-BE
NTA			
Ca^{2+}	-37.5	-32.7 ± 0.7	-31.4 ± 3.0
Mg^{2+}	-30.6	-84.9 ± 5.9	-49.5 ± 9.5
Y^{3+}	-65.5	-148.9 ± 1.9	-110.4 ± 7.4
La^{3+}	-59.1	-115.4 ± 6.3	-94.1 ± 4.3
EGTA			
Ca^{2+}	-62.8	-54.2 ± 8.5	-66.7 ± 7.1
Mg^{2+}	-29.7	-85.1 ± 12.4	-56.1 ± 15.1
Y^{3+}	-96.0	-125.0 ± 30.0	-85.5 ± 41.5
La^{3+}	-90.1	-120.7 ± 11.4	-73.3 ± 2.1

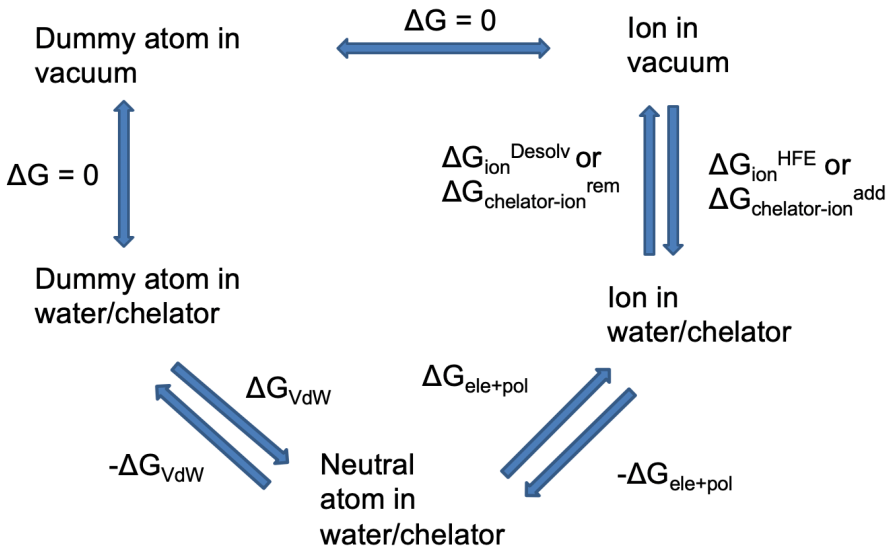
Table S8. Ion-oxygen distances (IODs in Å) and coordination numbers (CNs) of metal ions in EF2 and EF3 loops of full-length LanM.

	12-6LJ CM/IOD [a]		12-6-4LJ		12-6-4LJ Ch-BE	
Metal	CN	IOD	CN	IOD	CN	IOD
EF2						
Ca^{2+}	8.0	2.41	8.3-8.4 [b]	2.39	7.8-8.0 [b]	2.39
Mg^{2+}	6.0	1.95	7.0	2.09	6.0	2.06
Y^{3+}	8.8-8.9 [b]	2.26	9.0	2.28	9.0	2.30
La^{3+}	9.8	2.43	10.0	2.47	10.0	2.47
EF3						
Ca^{2+}	8.0	2.41	8.0	2.38	8.0	2.38
Mg^{2+}	6.0	1.95	6.9	2.08	6.0	2.05
Y^{3+}	9.0	2.27	9.0	2.28	9.0	2.28
La^{3+}	10.0	2.45	10.0	2.46	10.0	2.49

[a] 12-6LJ CM for divalent and 12-6LJ IOD for trivalent ions; [b] no clear minima was observed so the range is given.

Calmodulin			Lanmodulin		
EF1	20	DKDGDGTI TTKE 31	EF1	35	DPDKDGTI DLKE 46
EF2	56	DADGNGTI DFPE 67	EF2	59	DPDKDGTLDAKE 70
EF3	93	DKDGNGYI SAAE 104	EF3	84	DPDNDGTLDKKE 95
EF4	129	DI DGDGQVNYEE 140	EF4	108	NPDNDGTI DARE 119

Figure S1. Calmodulin (PDB 1CLL) and lanmodulin (PDB 6MI5) EF-hand motif sequence comparison.[4,19]



$$\Delta G_{ion}{}^{HFE}\left(or\ \Delta G_{chelator-ion}{}^{add}\right) = \Delta G_{ele+pol} + \Delta G_{VdW}$$
$$\Delta G_b = \Delta G_{chelator-ion}{}^{add} - \Delta G_{ion}{}^{HFE} = -(\Delta G_{chelator-ion}{}^{rem} - \Delta G_{ion}{}^{Desolv})$$
$$\Delta G = G(\lambda = 1) - G(\lambda = 0) = \int_0^1 \langle \frac{\delta V}{\delta \lambda} \rangle_i \, d\lambda \ = \sum w_i \langle \delta V / \delta \lambda \rangle_i$$

Figure S2. Thermodynamic cycle for ion hydration free energy and chelator-ion binding energy calculations.

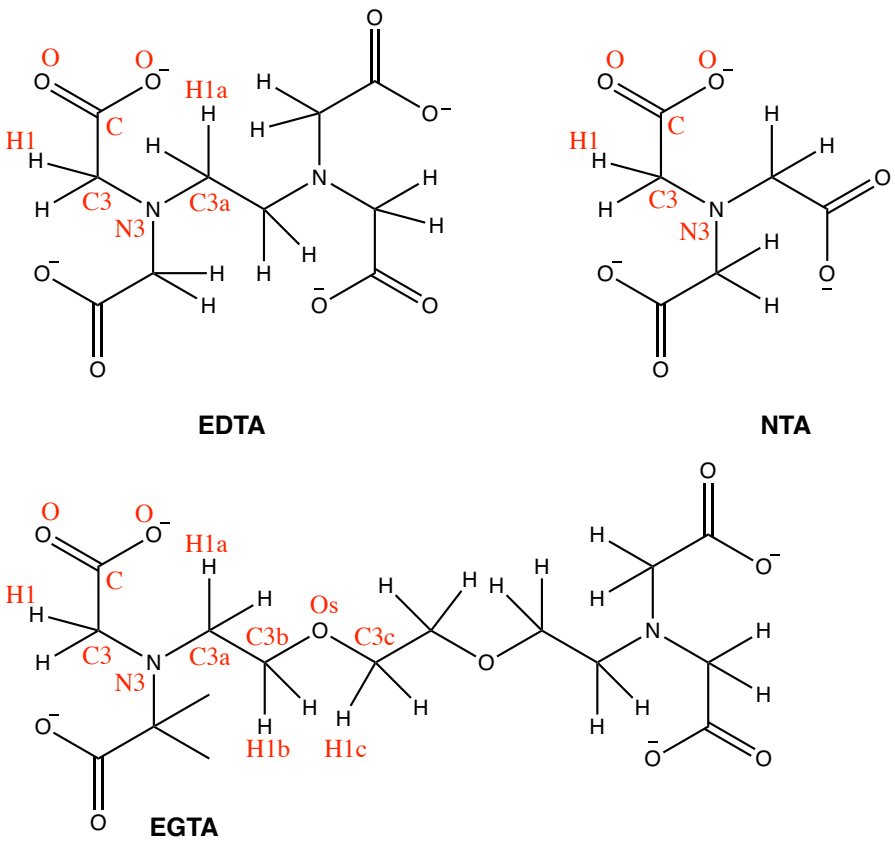


Figure S3. Atom types used in Table S2.

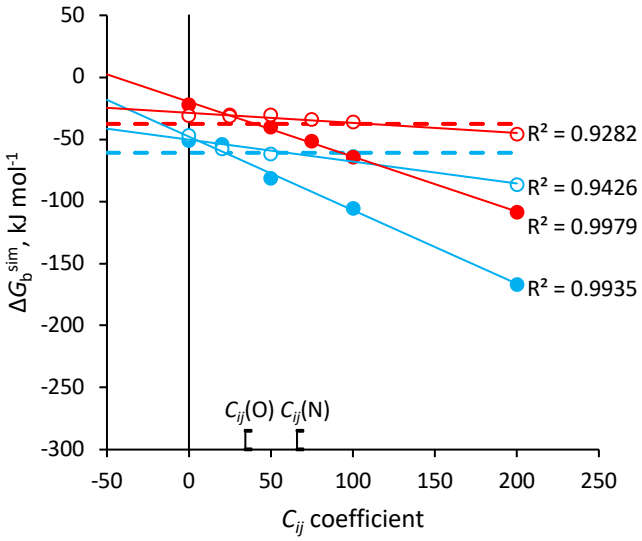


Figure S4. Computed binding energies for the chelation of Ca^{2+} by EDTA (blue) and NTA (red) as a function of the ligating oxygen and nitrogen C_{ij} values and the 12-6-4LJ force field. Filled symbols are for oxygen and open symbols for nitrogen. Solid lines are linear fits to the data and the horizontal dashed lines are the experimental values from Table 1. The default C_{ij} values are labeled.

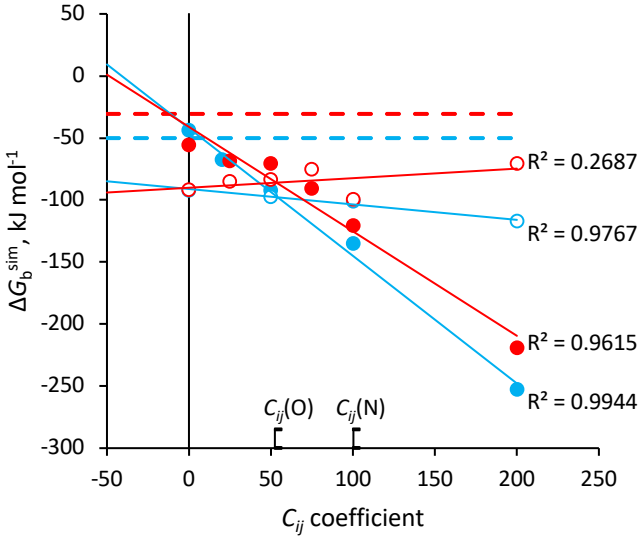


Figure S5. Computed binding energies for the chelation of Mg^{2+} by EDTA (blue) and NTA (red) as a function of the ligating oxygen and nitrogen C_{ij} values and the 12-6-4LJ force field. Filled symbols are for oxygen and open symbols for nitrogen. Solid lines are linear fits to the data and the horizontal dashed lines are the experimental values from Table 1. The default C_{ij} values are labeled.

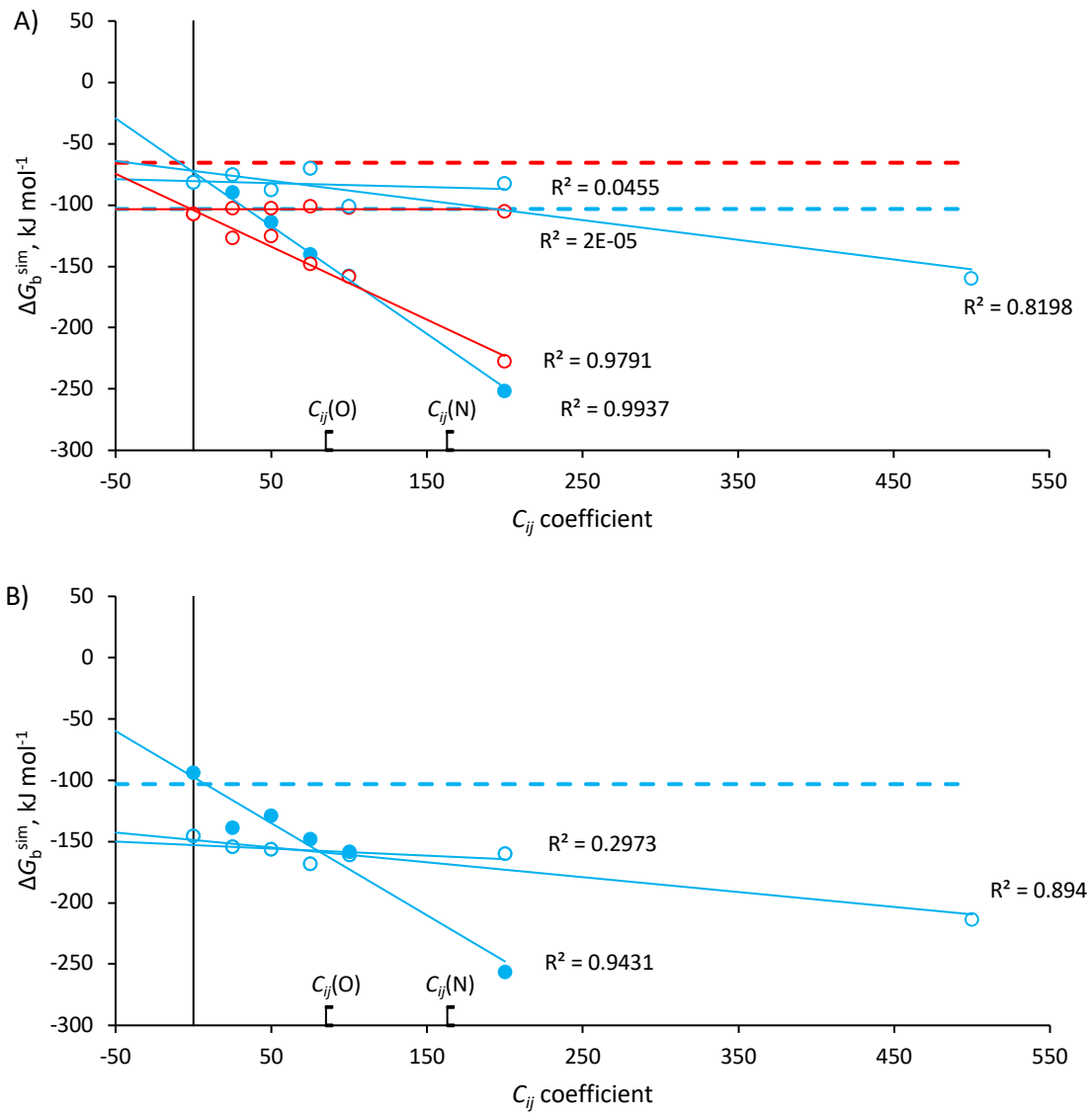


Figure S6. Computed binding energies for the chelation of Y^{3+} by EDTA (blue) and NTA (red) as a function of the ligating oxygen and nitrogen C_{ij} values and the 12-6-4LJ force field. Filled symbols are for oxygen and open symbols for nitrogen. Solid lines are linear fits to the data and the horizontal dashed lines are the experimental values from Table 1. The default C_{ij} values are labeled. Gradients obtained with other non-varied C_{ij} values fixed to default apart from C_{ij}(N) fixed to 0 when C_{ij}(O) was varied and vice versa (A) and gradients obtained while all non-varied C_{ij} values were fixed default values (B). If the gradient determined with a maximum C_{ij}(N) = 200 point was used in Eq. 7 the new C_{ij}(N) value unreasonably large at 391.3. Therefore, another point (C_{ij}(N) = 500) was added to calculate m(N), which was used with Eq. 7 to determine the reported C_{ij}(N) value. However, for consistency a gradient with a maximum point of 200 was still used in the Eq. 5 to back-extrapolate the $\Delta G_b^{sim\prime}(C_{ij} = 0{,}0)$ value.

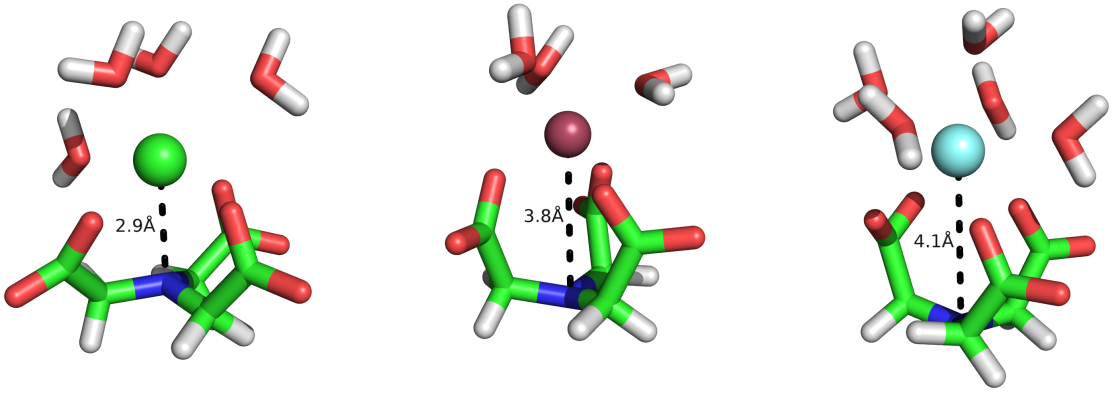


Figure S7. NTA coordination of metal ions. Snapshots were taken from 12-6-4LJ with default C_{ij} coefficients λ=0.00922 thermodynamic integration simulations where 99% of the charge are present. Metal ions: Ca^{2+} (green), Mg^{2+} (raspberry) and Y^{3+} (aqua).

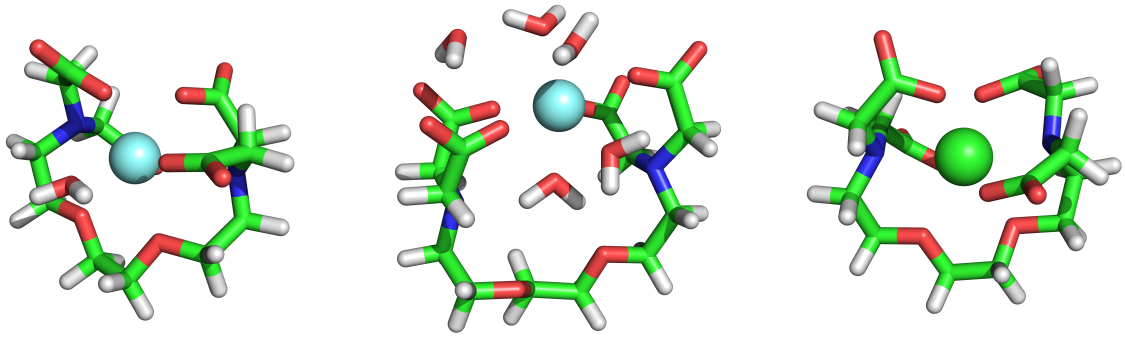


Figure S8. Different EGTA metal ion ligating geometries. Snapshots were taken from three thermodynamic integration λ=0.00922 simulations where 99% of the charge are present. Y^{3+} displayed two major coordination modes – "normal" (left) and "EGTA-solvent hydrogen bonded" (middle), which differed in coordinating water molecule and EGTA ligating group number. Metal ions like Ca^{2+} showed one major coordination state – "normal" (right), which only used ligating groups provided by EGTA.

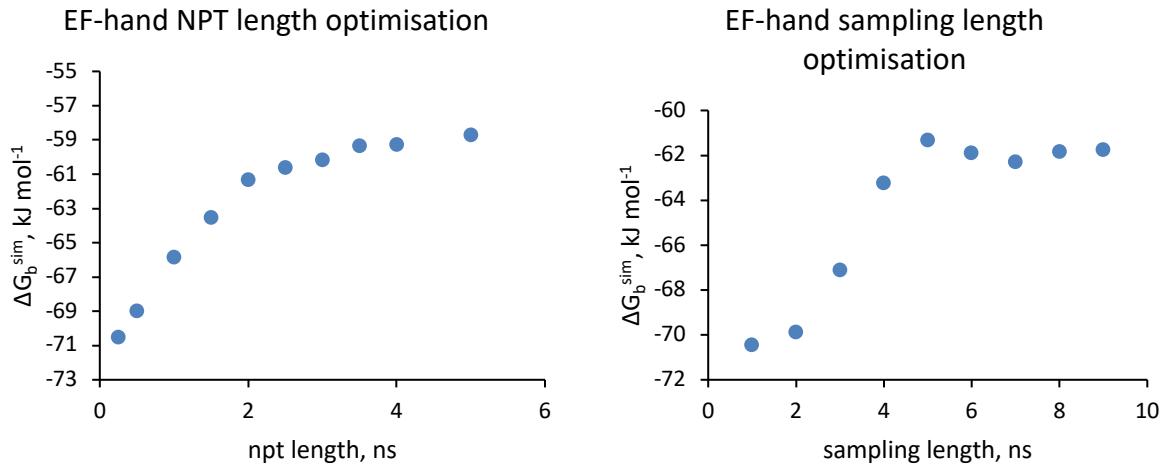


Figure S9. EF-hand and metal ion NPT and sampling length optimization. Lanmodulin EF-hand 1 with Ca^{2+} ion were used as a test system to determine NPT and sampling lengths per λ step of TI required to produce consistent binding energies. If NPT was varied then sampling length was fixed at 5 ns, while if sampling length was optimized then NPT was fixed at 2 ns.

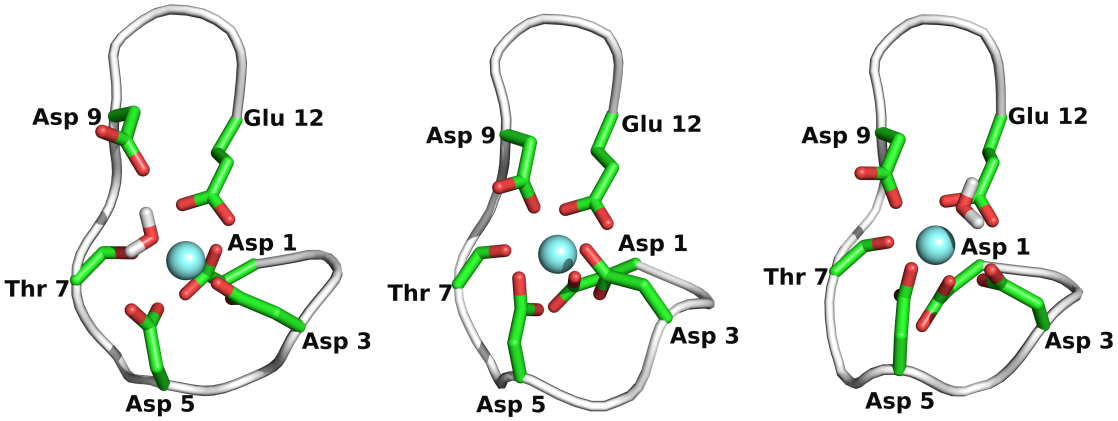


Figure S10. Comparison of computed lanmodulin EF-hand 1 Y^{3+} (LanM EF1) coordination frameworks. 12-6LJ IOD (left), 12-6-4LJ (middle) and 12-6-4LJ Ch-BE (right) representative structures were taken from MD trajectories of full-length LanM with Y^{3+} ions. The peptide backbones are shown as ribbons and the coordinating residues ligating groups as sticks. The 12-6-4LJ default and 12-6-4LJ Ch-BE parameters produced ligation frameworks in which Asp 9 directly coordinated the Y^{3+} ion, while in the 12-6LJ IOD simulation Asp 9 hydrogen bonded to a water molecule that coordinated Y^{3+} instead. Both of these Asp 9 coordination modes were seen in the experimental structures of LanM EF1 as in 3 out of 12 states Asp 9 was used for direct Y^{3+} coordination and in 9 out of 12 states Asp 9 was not used in direct coordination.[4] Interestingly, in 12-6-4LJ Ch-BE simulation Asp 3 instead of Asp 9 hydrogen bonded to the water molecule that coordinated Y^{3+}, which was not observed in the experimental states.

Additional References

1. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E.; III, Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Giambasu, G.; Giese, T.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Huang, Y.; Izadi, S.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Onufriev, A.; Pan, F.; Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiao, L.; Xiong, D. M.; York D. M.; Kollman P. A. AMBER 2019, University of California, San Francisco, **2019**
2. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graphics Modell.* **2006**, *26*, 247–260, DOI: 10.1016/j.jmgm.2005.12.005
3. Jakalian, A., Bush, B.L., Jack, B.D., Bayly, C.I., Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *J. Comp. Chem.* **2000**, *21*, 132-146, DOI: 10.1002/(SICI)1096-987X(20000130)21:2<132::AID-JCC5>3.0.CO;2-P
4. Cook, E. C.; Featherston, E. R.; Showalter, S. A.; Cotruvo, J. A. Structural Basis for Rare Earth Element Recognition by Methylobacterium extorquens Lanmodulin. *Biochemistry* **2019**, *58*, 120– 125, DOI: 10.1021/acs.biochem.8b01019
5. Schauer, C. K.; Anderson, O. P. Highly polydentate ligands. 5. Structures of alkaline-earth complexes of the calcium-selective ligand EGTA4- (H4EGTA = 3,12-Bis(carboxymethyl)-6,9-dioxi-3,12-diazatetradecanedioic acid). *Inorg. Chem.* **1988**, *27*, 3118– 3130, DOI: 10.1021/ic00291a015
6. Marcus, Y. Thermodynamics of solvation of ions. Part 5.–Gibbs free energy of hydration at 298.15K. *J. Chem. Soc., Faraday Trans.* **1991**, *87*, 2995– 2999, DOI: 10.1039/ft9918702995
7. Anderreg, G. Critical Survey of Stability Constants of EDTA Complexes. IUPAC chemical data series, no 14. Pergamon Press, Oxford, U.K. **1977**.
8. Anderegg, G. Critical Survey of Stability-Constants of NTA Complexes. *Pure Appl. Chem.* **1982**, *54*, 2693– 2758, DOI: 10.1351/pac198254122693
9. Dojindo Molecular Technologies, Inc. Chelate table of stability constants. Retrieved from: https://www.dojindo.co.jp/technical/pdf/Chelate_Table_of_Stability_Constants.pdf
10. Cotruvo, J. A.; Featherston, E. R.; Mattocks, J. A.; Ho, J. V.; Laremore, T. N. Lanmodulin: A Highly Selective Lanthanide-Binding Protein from a Lanthanide-Utilizing Bacterium. *J. Am. Chem. Soc.* **2018**, *140*, 15056– 15061, DOI: 10.1021/jacs.8b09842
11. Linse, S.; Helmersson, A.; Forsen, S. Calcium-binding to calmodulin and its globular domains. *J. Biol. Chem.* **1991**, *266*, 8050– 8054, DOI: 10.1016/S0021-9258(18)92938-8
12. Tsai, M. D.; Drakenberg, T.; Thulin, E.; Forsen, S. Is the binding of magnesium (II) to calmodulin significant? An investigation by magnesium-25 nuclear magnetic resonance. *Biochemistry* **1987**, *26*, 3635– 3643, DOI: 10.1021/bi00386a057
13. Xu, K.; Yang, X. D.; Wang, K. Metal binding discrimination of the calmodulin Q41C/K75C mutant on Ca2+ and La3+. *Sci. China Chem.* **2010**, *53*, 797– 806, DOI: 10.1007/s11426-010-0059-2
14. Marcus, Y. Ionic radii in aqueous solutions. *Chem. Rev.* **1988**, *88*, 1475– 1498, DOI: 10.1021/cr00090a003

15. Jalilehvand, F.; Spangberg, D.; Lindqvist-Reis, P.; Hermansson, K.; Persson, I.; Sandstrom, M. Hydration of the Calcium Ion. An EXAFS, Large-Angle X-ray Scattering, and Molecular Dynamics Simulation Study. *J. Am. Chem. Soc.* **2001**, *123*, 431– 441, DOI: 10.1021/ja001533a
16. Li, P. F.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733– 2748, DOI: 10.1021/ct400146w
17. Li, P. F.; Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10*, 289– 297, DOI: 10.1021/ct400751u
18. Li, P. F.; Song, L. F.; Merz, K. M. Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-Type Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 883–895, DOI: 10.1021/jp505875v
19. Chattopadhyaya, R.; Meador, W. E.; Means, A. R.; Quiocho, F. A. Calmodulin structure refined at 1.7 angstrom resolution. *J. Mol. Biol.* **1992**, *228*, 1177– 1192, DOI: 10.1016/0022-2836(92)90324-D