SIMULATING THE CHELATE EFFECT AND THE MOLECULAR DYNAMICS OF TRANSITION METAL IONS

By

Anthony Joseph Seitz

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry—Master of Science

2020

ABSTRACT

SIMULATING THE CHELATE EFFECT
AND THE MOLECULAR DYNAMICS
OF TRANSITION METAL IONS

By

Anthony Joseph Seitz

Transition metal ions are of great significance in biological processes, with a presence in approximately one third of known human proteins. Computational modeling of transition metal ions, however, has remained limited. Using molecular dynamics simulations, very large chemical systems can be studied due to the use of a 'force field' which simplifies atomic interactions to an easily calculable form. Unfortunately, 'traditional' force fields used in molecular dynamics fail to accurately portray the properties of transition metal ions. By modifying the 12-6 Lennard-Jones potential to include an r^{-4} term, the Lennard-Jones type 12-6-4 potential has shown to be a successful replacement for the 12-6 potential, with an accurate portrayal of important properties of solvated transition metal ions (hydration free energies, coordination number, ion-oxygen distance). We have built on the foundation of the 12-6-4 potential by applying it in the field of small molecule coordination chemistry. The 12-6-4 potential is parameterized for accurate binding energies in the reaction of ethylenediamine with various divalent transition metal ions. In comparing these results with the similar reaction of the same metal ions with methylamine ligands, we demonstrate that the 12-6-4 model naturally reproduces the chelate effect. Therefore, continued use of the 12-6-4 potential is proposed for the use in modeling coordination chemistry reactions.

ACKNOWLEDGEMENTS

As much as I strive to work on my own, the one thing that graduate school has taught me more than anything else is that this simply isn't possible. First, I would like to thank my lovely wife, Kristen, who has been the center of my world, almost from the day we met, but especially now more than ever. Next, my parents, Joe and Theresa Seitz, who have been my biggest supporters throughout my scientific endeavors. Every scientist I have worked with over these years in graduate school has been an influence on my work, whether doing research in the Merz group, under my advisor, Dr. Kennie Merz, or teaching students throughout the Michigan State University general chemistry program, under Dr. Amy Pollock. I would like to thank all current members of, and recent alumni of the Merz group, as well as every teaching assistant I have had the pleasure of working with (although at this point the sheer amount of names I would list here eludes my memory). Lastly, I would like to thank our dog, Nova, and our cats, Hendrix, Vinnie, Tacocat, and Purrito, for providing a constantly entertaining home for Kristen and I. Graduate school at Michigan State University has been the most exciting time of my life, and I am ready to see what's up next.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

LJ	Lennard-Jones
en	ethylenediamine
FEP	free energy perturbation
fs	femtosecond
MD	molecular dynamics
ns	nanosecond
PME	Particle Mesh Ewald
PMF	potential of mean force
QM	quantum mechanical
TI	thermodynamic integration
WHAM	Weighted Histogram Analysis Method

CHAPTER 1: INTRODUCTION

1.1 Motivation

The significance and importance of transition metals in biological processes cannot be overstated. As I type this, the world's peoples and economies are being ravaged by the Covid-19 Novel Coronavirus, which contains several zinc-bound sites as primary structural features.[1] It is estimated that one third of known human proteins contain transition metal ions. If one virus, with structural dependence on transition metal ions, can cause such a large impact on human life, the importance of this field of study is inarguable. However, even with the breadth of computational tools available, the interactions between metal ions and biological systems have remained elusive in the realm of computational chemistry and molecular modeling, due largely to the computational expense of accurate methods, such as the various quantum mechanical tools available.[2] Computationally inexpensive methods often either fail to represent important processes involved in these interactions, such as ligand exchange, which cannot be properly modeled with bonded models that are typically used, or they simply give inaccurate results, due to the high polarizability of metal ions that cannot accurately be represented with typical Lennard-Jones potentials used in molecular dynamics simulations.[3] Therefore, it would be desirable to obtain an accurate computational model of these polarizability effects, specifically in the context of metal-ligand interactions, that does not incur the high computational cost of quantum mechanical models.

1.2 Molecular Dynamics

The core concept of molecular dynamics is a simple one, but even the simplest concepts can quickly grow to be impossibly complicated if left unchecked. Molecular dynamics simulations use computers to simulate the movement of atoms and molecules by solving Newton's equations of motion. For a simple system of one or two particles, this problem is easy. However, for anything beyond that, non-iterative solutions do not exist and approximations must be made, especially after increasing the complexity from three particles to a protein consisting of thousands of atoms. The heart of molecular dynamics lies in two fundamental ideas, being the development and sampling of an energy landscape.

The aim of molecular dynamics is to solve Newton's equations, $F(r_i) = ma(r_i)$ for a many-particle system evolving in time (where $i = 1, \dots, N$ for an *N*-particle system and r_i represents the position coordinates for any particle) . Knowing additionally that $F(r_i) = -\nabla V(r_i)$ where V is assumed to be a pairwise-additive potential, independent of velocity and time (ex: [$V_{123}(r_i) = V_{12}(r_i) + V_{13}(r_i) + V_{23}(r_i)$]) we can evolve our system in time simply by knowing initial positions, velocities, and our potential function (which will be discussed later). An important aspect of Newton's equations is that they are time-reversible, and the potential functions used will be analytic (although the hard-sphere potential is solvable). This will allow the use of a discrete finite-difference method to solve the equations. A finite difference method will estimate the infinitesimal time of evolution δt as a discrete time-step. It is important to consider the timescale of the chemical processes in question when choosing a time-step length (femtosecond scale for bond vibrations). A typical MD calculation will proceed as follows:

1. Solve for the positions, velocities, and accelerations at time $t + \delta t$
2. Solve $F(r_i) = ma(r_i)$ and $F(r_i) = -\nabla V(r_i)$ at new positions
3. Correct positions, velocities, and accelerations with new accelerations
4. Calculate interesting chemical properties
5. Repeat

This general algorithm applies for most methods in molecular dynamics, but will vary slightly between choices in integrators, being the methods for predicting and correcting the positions, velocities, and accelerations of the system in question.[3]

The most common integration algorithms are all based on the idea that the positions, velocities, and accelerations of the particles can be expressed accurately in a Taylor series expansion. For any given coordinate, this is as follows:

$$r_i(t + \delta t) = r_i(t) + v_i(t)\delta t + \frac{1}{2}a_i(t)\delta t^2 + \cdots$$

$$v_i(t + \delta t) = v_i(t) + a_i(t)\delta t + \frac{1}{2}\dot{a}_i(t)\delta t^2 + \cdots$$

$$a_i(t + \delta t) = a_i(t) + \dot{a}_\iota(t)\delta t + \cdots$$

The Verlet integration algorithm was first introduced into molecular dynamics in 1967 by Loup Verlet although the history of the algorithm goes back much further. Verlet examined a system of 864 argon atoms interacting through Lennard-Jones potentials and using periodic boundary conditions. In a single line, Verlet expressed his algorithm which has become one of the most popular integration methods in molecular dynamics.[4]

$$r_i(t+h) = -r_i(t-h) + 2r_i(t) + \sum_{i \neq j} F(r_{ij}(t))h^2$$

Where r_{ij} is the distance between any two particles and

$$m_i \frac{d^2 r_{ij}}{dt^2} = \sum_{i \neq j} F(r_{ij})$$

Noting that h is the length of the time-step, this equation can be derived from our Taylor expansion, truncating after the acceleration term.

$$r_i(t+\delta t) = r_i(t) + v_i(t)\delta t + \frac{1}{2} a_i(t)\delta t^2$$

$$r_i(t-\delta t) = r_i(t) - v_i(t)\delta t + \frac{1}{2} a_i(t)\delta t^2$$

Adding these equations will give the result:

$$r_i(t+\delta t) = 2r_i(t) - r_i(t-\delta t) + a_i(t)\delta t^2$$

Multiplying the acceleration term by the mass m_i will make this equivalent to the algorithm proposed by Verlet. The algorithm's popularity can be attributed to its simplicity. The algorithm only requires the input of positions and is quick because the force is calculated only once per iteration. Its basis in forward and backward Taylor expansions makes it reversible in time (within the realm of round-off error) and it is conservative of both energy and phase-space volume. The Verlet algorithm is fairly accurate but does allow for error in the order of δt^2 which can be significant for large time-steps.

A mathematically equivalent algorithm is the velocity Verlet method, which predicts velocities instead of positions. From the expression for the Verlet method, we can obtain:

$$r_i(t + 2\delta t) = 2r_i(t + \delta t) - r_i(t) + a_i(t + \delta t)\delta t^2$$

This will now be substituted into our expression for the velocity, given by rearrangement and addition of our forward and backward Taylor expansions.

$$v_i(t) = \frac{r_i(t + \delta t) - r_i(t - \delta t)}{2\delta t}$$

$$v_i(t + \delta t) = \frac{r_i(t + 2\delta t) - r_i(t)}{2\delta t}$$

$$v_i(t + \delta t) = \frac{r_i(t + \delta t) - r_i(t)}{\delta t} + \frac{1}{2}a_i(t + \delta t)\delta t$$

Noting the original forward Taylor expansion, we get our final result, which is simply the Verlet algorithm, predicting velocities instead of positions.

$$v_i(t + \delta t) = v(t) + \frac{1}{2}[a_i(t) + a_i(t + \delta t)]\delta t$$

Another commonly used integration scheme is the 'leap-frog' scheme, popularized by Hockney and Eastwood, which is named after the way it operates. Velocities will be calculated at half-steps, followed by position calculations at the whole step. This method eliminates some numerical error by removing the subtraction step from the velocity calculations.[5] It can be derived by defining the velocities at half time-steps.[6]

$$v_i\left(t - \frac{1}{2}\delta t\right) = \frac{r_i(t) - r_i(t - \delta t)}{\delta t}$$

$$v_i\left(t + \frac{1}{2}\delta t\right) = \frac{r_i(t + \delta t) - r_i(t)}{\delta t}$$

This gives a new expression for the positions and velocity.

$$r_i(t+\delta t) = r_i(t) + v_i\left(t+\frac{1}{2}\delta t\right)\delta t$$

$$v_i\left(t+\frac{1}{2}\delta t\right) = v_i\left(t-\frac{1}{2}\delta t\right) + a_i(t)\delta t$$

$$v_i(t) = \frac{1}{2}\left[\, v_i\left(t+\frac{1}{2}\delta t\right) + v_i\left(t-\frac{1}{2}\delta t\right)\right]$$

Issues with the leap-frog integration scheme may occur with variable time-steps in an MD calculation, as the half time-steps will not line up properly. This could quickly lead to accumulated error and the method will be 'destabilized.' Unfortunately, since the positions and velocities are not generally defined at the same time in this method, an expression for the total energy is meaningless until the end of the calculation unless an explicit $v(t)$ calculation is performed. The similarities to the Verlet method should be noted, as it is also time-reversible and conserving in energy, momentum, and phase space volume.

The last integration method to be discussed will be Beeman's algorithm. Many other methods exist, including (among many) the Euler method (and Euler-based methods), trapezoidal rule, the midpoint method, and the Runga-Kutta method, but only Verlet-based methods are discussed here due to their popularity in molecular dynamics. Beeman's algorithm is more complex than the Verlet algorithm, but ultimately gives a more accurate expression of the velocities, eliminating some round-off error, and consequently it has a more accurate conservation of energy. By continuing the Taylor expansion in time to the $\dot{a}$ term, we can obtain Beeman's Algorithm, which should result in the same particle positions as the Verlet algorithm, subject to computer error. Unfortunately, the velocities are no longer time reversible, but the expression is more physically accurate in the forward direction. [7]

$$r_i(t+\delta t) = r_i(t) + v(t)\delta t + \frac{2}{3}a_i(t)\delta t^2 - \frac{1}{6}a_i(t-\delta t)\delta t^2$$

$$v_i(t+\delta t) = v_i(t) + \frac{1}{6}[2a_i(t+\delta t) + 5a_i(t) - a_i(t-\delta t)]$$

Constraint Methods: SHAKE and RATTLE

Whether to ensure physical accuracy or for computational efficiency, it may be necessary to apply a series of constraints in an MD calculation. The most obvious constraints on a system are simply its conservation laws (conservation of energy, momentum, phase space volume, and if possible angular momentum), but these should automatically be satisfied by a good choice of an integrator.

One of the most popular methods of constraints is called SHAKE. If, for example, the bond vibrations were not of interest in a particular calculation, the calculation could be sped up significantly by increasing the time-step to longer than the vibrational frequency (but shorter than the frequency of any other important degrees of freedom). In order to keep any sort of physical accuracy in this case, the bond lengths must be defined at a fixed distance. The Newtonian equations of motion must now be solved for a system under constraints of fixed bond lengths.

Assuming a situation like the one just described, where all constraints being applied are holonomic, we can solve the equations of motion using the method of Lagrange multipliers. We will assume the situation of fixed bond lengths and express the constraint as follows.

$$\sigma[r(t)] = (r_i - r_j)^2 - d_{ij}^2 = 0$$

For bond ij of length d_{ij} we can express the force on particle i as

$$m_i \ddot{r}_i = -\nabla_i V - \sum_1^i \lambda \nabla_i \sigma$$

where the $-\nabla V$ term is the standard gradient of the potential, used to describe the force between particles, and the sum shows the Lagrange multipliers related to constraints of the particle in question.[8] In the SHAKE method, all constraints are treated as being decoupled. The Lagrange multipliers will be solved using the Newton-Raphson method, which will solve for the multipliers by approximating for them and iteratively correcting based on the constraints.

$$\lambda_k = -\frac{\sigma_k}{\partial \sigma_k / \partial \lambda_k}$$

Due to the decoupling of the constraints, the SHAKE method converges fairly quickly, with the Lagrangian multipliers being evaluated in linear time.[9] This method is built to work with the Verlet integration scheme and implementation into the Verlet algorithm gives the general result:

$$r_i(t + \delta t) = r_i'(t + \delta t) + g(r)$$

where r' is the position result without constraints and $g(r)$ is the deviation from that position due to the forces of constraints.[3] The low computational cost of SHAKE can be attributed to the decoupling of constraints, which also introduces a primary source of error. Many forms of SHAKE exist, all dealing with different aspects of the constraints, whether it be

different degrees of freedom or the iterative procedure. All versions, however, do incorporate the Newton method for solving Lagrange multipliers.

The most notable of these other methods is called RATTLE. RATTLE is based on the velocity form of the Verlet equations. From the velocity equation and our forces due to the potential and constraints, we can write:

$$v_i(t+\delta t) = v_i(t) + \frac{F[\,r(t)]\delta t + \; g_{RR}(t) + F[r(t)]\delta t + \; g_{RV}(t)}{2}$$

This equation shows the primary difference between SHAKE and RATTLE, being that RATTLE uses two separate approximations for the constraints, which allows for both the velocities and the positions to satisfy the constraints. It is notable that no prior knowledge of the system is required for RATTLE and only the current coordinates are necessary. The more complex expression does increase computational time, but RATTLE will be comparable to SHAKE in most cases because calculating forces is generally the most expensive step of any MD calculation.[10]

Boundary Conditions

A problem exists about the 'edges' of a system, where instead of being surrounded by other molecules, the environment is the vacuum. In reality, we know that the environment is rarely, if ever, vacuum and that we must simulate the environment around the edges of the system in order to give any accurate representation, especially if we wish to simulate any sort of condensed phase.

One of the simplest approaches to dealing with the boundary of the system is with periodic boundary conditions. The original system is taken as a cubic cell and copied across

many different cells. One interesting property resulting from this is that particle number is always conserved. As particles are moving throughout the system, some may leave the original cubic cell, but the periodicity guarantees that as a particle leaves the original cell, a new particle enters from the opposite side. [11]

Periodic boundary conditions allow for more physical accuracy around the boundaries, but drastically increase the computational time unless some effort is made to reduce it. For example, if there are 1000 particles in the center cell we would want some method that would allow us to calculate forces for only the 1000 particles and not any others. The minimum image convention allows for this by placing each particle at the center of its own cubic cell for its force calculations. The force applied from only the particles within that new cell will be calculated. For a system of 1000 particles, this means that each particle would have forces calculated from the nearest 'copies' of each of the 999 other particles. [12] This can lead to some inaccuracies, as charged ions (generally next to each other in the original box) may be spread to opposite sides of the new cell. However, total charge is guaranteed to be conserved, as each 'image cell' will have the same charge as the original cubic cell.[3]

In many cases, such as a protein in solution, periodic boundary conditions may not be desired, and instead one may want to simulate the direct environment of the solution. In the case of a protein, this would usually be water. However, if a solvent of many water molecules is added, the computational can increase significantly, so other methods are often sought out. One common method is implicit solvation, where the solvent molecules are represented as a continuum instead of individual molecules. In these cases, instead of finding the force applied on the system from each individual solvent molecule, the force is averaged across the implied molecules as a 'potential of mean force.'[13]

1.3 Force Fields

The energy surface is a model of the different interactions between all atoms and molecules and is the main aspect of the force calculations, ultimately determining the motion of all particles and most of the interesting properties of the system. All potentials will be assumed to be pairwise additive. The total potential energy in any system will be modelled as a sum of all bonded and non-bonded interactions in the molecules.

$$E_{total} = \left(E_{bond} + E_{angle} + E_{dihedral}\right) + \left(E_{electrostatic} + E_{van\ der\ Waals}\right)$$

Typically the vibrational and angle terms will be modelled as harmonic oscillators (where k is the force constant for the bond/angle/dihedral in question) while the dihedral angles will be represented with a periodic function (showing steric repulsion effects in an eclipsed conformation, where n is the periodicity of the angle). [14]

$$V = \frac{1}{2}\sum_{bonds} k_{ij}\left(r_{ij} - r_{eq}\right)^2 + \frac{1}{2}\sum_{angles} k_{ijk}\left(\theta_{ijk} - \theta_{eq}\right)^2$$

$$+ \frac{1}{2}\sum_{dihedrals}\sum_{n} k_{ijkl}\left[1 + \cos\left(n\emptyset_{ijkl} - \emptyset_{eq}\right)\right]$$

Although this equation is written in a decoupled form, the degrees of freedom will typically be coupled with many different cross-terms. Electrostatic terms between charged particles will be modelled with Coulomb's law, where F is the force between particles and C is Coulomb's constant.

$$F = C\frac{q_2 q_1}{r^2}$$

This equation integrates to the Coulomb potential.

$$V = C\frac{q_1 q_2}{r}$$

The $1/r$ dependence introduces certain difficulties with the periodic boundary conditions that will be discussed briefly at the end of this section. One of the most important aspects of molecular dynamics is the ability to model large systems. Many of the bulk properties of systems are determined by the long-range van der Waals effects. These effects between two particles can be modelled as two charged oscillators. The electric field for each oscillating dipole can be modeled as

$$E(r) = \frac{-2u}{4\pi\varepsilon_0\varepsilon R^3}$$

where u is the dipole moment. The interaction energy is proportional to the square of the electric field strength and therefore this can be further approximated as a two-body problem, dependent on the separation of the two oscillators as

$$-\frac{\sigma}{R^6}$$

where the σ value is dependent on the frequency of the oscillators: $\hbar\omega_0\alpha^2$. This effect is quantum in nature and would approach zero as $\hbar$ approaches zero. However, it is completely necessary for the modeling of bulk systems as this term is responsible for most of the long-range effects. The repulsive terms in this model are more difficult to evaluate (and are representative of the electron-electron repulsion terms, one of the primary difficulties in quantum calculations),

but experimental data on noble gases fits well to an inverse R^{12} potential. This allows simple evaluation of the repulsive portion of the potential, as it is the square of the dipole-attraction calculated above. Combining these terms gives the well-known Lennard-Jones potential. [15]

$$V = 4\varepsilon[\left(\frac{\sigma}{R}\right)^{12} - \left(\frac{\sigma}{R}\right)^{6}]$$

Many variations on the Lennard-Jones potential exist, correcting for various aspects of the potential. A few such alternatives are the Buckingham potential, which introduces an exponential term to replace the R^{-12} term in the Lennard-Jones potential, the Morse potential, which is highly accurate, but computationally expensive, and a '10-12' potential, which adds corrections to the Lennard-Jones form to account for hydrogen-bonding interactions.[16] Many other potential forms exist, for modelling different degrees of freedom or different interactions but these all serve different purposes and their usefulness can vary depending on the complexity of the problem at hand. The development of these potentials ultimately reduces to fitting the parameters to either experimental data or high-level QM calculations.

The $1/r$ dependence in the Coulomb term, discussed earlier, introduces difficulties when we account for periodic boundary conditions, due to an increasing particle density faster than the $1/r$ decrease in energy. The consequence of this is a long-range Coulomb energy that is non-convergent and increases with distance, instead of decreasing. The modern solution to this is the Particle Mesh Ewald Method. PME changes the Coulomb term from a sum of two-body interactions, written as

$$E = \sum_{ij} C \frac{q_i q_j}{r_{ij}}$$

to a two-part sum of the short-range two-body coulomb potentials and a long-range component, which creates a grid across the unit cell of the simulation and evaluates the coulomb potentials and charge densities in reciprocal Fourier space.

$$E = E_{sr} + E_{lr}$$

$$E_{sr} = \varphi_{sr}(r_{ij})$$

$$E_{lr} = \tilde{\varphi}_{lr}(k)|\tilde{\rho}(k)|^2$$

This approximation allows the Coulomb term to converge to a finite energy. The potentials and methods described have been the basis for molecular dynamics simulations for a good portion of the field's lifetime due to the large amount of success in modeling proteins. However, the field has historically struggled with coordination chemistry, as the LJ 12-6 potential, while being very successful in modeling the induced dipole-induced dipole interactions and repulsions between atoms, fails to model both the charge-induced dipole and dipole-induced dipole interactions. In the case of transition metal ions, these effects are significant, due to the highly polarizable electron clouds, often providing an even stronger influence on the energy than the LJ r^{-6} term.[3]

1.4 Lennard Jones 12-6-4 Potential

The 'standard' molecular dynamics force field described above describes a potential energy surface that ultimately determines the motion and all interesting properties of the chemical system being described.[14] However, in the field of coordination chemistry, this leaves a lot to be desired as some very important effects are ignored entirely. The Coulomb term will describe all charge—charge interactions, as well as dipole—dipole and charge—dipole interactions, while the LJ 12-6 potential is parameterized to describe induced dipole—induced dipole interactions and electron—electron repulsions.[15] The typical LJ potential usually suffices for proteins and organic molecules but will give terribly inaccurate predictions when used to describe systems containing transition metals. In particular, the 12-6 potential can be fit to reproduce either structural or energetic effects for transition metal ions, but never both, as their highly polarizable electron clouds will require charge—induced dipole and dipole—induced dipole interactions for proper modeling. In fact, when present, these effects are typically stronger than the induced dipole—induced dipole term described in the LJ 12-6 potential.

As a solution, Li and Merz proposed a modified Lennard Jones potential, adding an additional r^{-4} term to mimic the charge-induced dipole interaction,[17] leaving the long-range interactions in the force field as follows:

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

where

$$C_4^{ij}(atom\ type) = \frac{C_4(H_2O)}{\alpha_0(H_2O)} \times \alpha_0(atom\ type)$$

This additional attractive $\frac{C_4^{ij}}{r_{ij}^4}$ (C_4) term provides a scalable means of including the charge-induced dipole term into the force field description of a chemical system.[18] By following a parameterization process and scaling the α_0 (polarizability) value for a given metal-ligand interaction to fit experimental data, any pairwise interaction with a metal ion can be reproduced. In the initial development of this potential, Li has scaled the $C_4(H_2O)$ terms for 55 metal ions to reproduce hydration free energies, as well as structural features for ion-water interactions, in 3 different water models for the AMBER molecular dynamics software package.[18] This leaves the process of parameterizing a transition metal—ligand interaction to determining the $\alpha_0(atom\ type)$ polarizability value for a given atom type (as it occurs in the interaction). This can be accomplished through a fitting procedure to experimental data (which will implicitly allow for inclusion of typically unattainable many-body effects). The natural choice in which to reproduce for the parameterization of any force field would be the binding energy for the given metal—ligand interaction. Therefore, free energy methods must be used to calculate the binding energy for any fitting procedure to succeed.

1.5 Free Energy Methods

Free energy is the central quantity in thermodynamics, as free energy changes drive the vast majority of interesting processes in nature. It is no surprise that calculating free energy changes is a fruitful area of study in the field of statistical mechanics. Predicting accurate free energies allows for the understanding and creation of new chemical processes, especially in the development of new drugs and materials. The design of these new molecules would amount to mere luck if the understanding of the reaction pathways is ignored. Without reference, free energies are only useful for very specific applications, so the focus will be free energy differences, which are much more interesting in the context of chemical reactions. Specifically, potential of mean force (via umbrella sampling) and thermodynamic integration will be discussed.

Potential of mean force and umbrella sampling

Umbrella sampling is a method of sampling states in a molecular dynamics simulation that samples the entirety of a reaction coordinate by use of a bias potential along the reaction coordinate and "free MD" throughout the rest of the system. This bias potential will drive the reaction from the 'reactant state' to the 'product state,' all the while collecting sampling data throughout the simulation. After collecting sampling data from each 'window' of the simulation, a probability distribution of the entire reaction coordinate is obtained as a Boltzmann distribution. The Boltzmann distribution is then used to calculate a potential of mean force, using methods such as the Weighted Histogram Analysis Method.[19,20]

In a Boltzmann distribution, the probability of any state is given by the following equation, where P is the probability, k is Boltzmann's constant, T is temperature, and ε is the free energy.

$$P \alpha e^{\frac{-\varepsilon}{kT}}$$

Therefore, the difference in free energy between the 'reactant' and 'product' states of a given system is related to the probability of each state by:

$$\frac{P_r}{P_p} \alpha e^{\frac{\varepsilon_p - \varepsilon_r}{kT}}$$

This allows us to calculate the free energy of a reaction by determining the probability distribution of the reaction coordinate through the use of a bias potential, that samples every state along a reaction pathway. Then the free energy difference between each set of windows in the umbrella sampling calculation is determined by solving for the free energy difference in the Boltzmann distribution, using the obtained probability distribution of the reaction coordinate.

Thermodynamic Integration

Thermodynamic Integration is a method very similar to umbrella sampling, but in the limit of an infinite bias potential.[21] Through the usage of a nonspatial parameter defined as τ, the free energy difference between two states can be calculated by coupling τ to a known physical potential that contributes to the free energy (usually the coulomb or LJ potential).[22] The free energy difference is then calculated as:

$$\Delta\varepsilon = \int_0^1 \frac{\partial V(\tau)}{\partial \tau} d\tau$$

By varying the value of τ and integrating through the TI windows, we get the free energy contribution of a potential. This can be especially useful in binding energy calculations (such as the metal—ligand simulations performed in our work), as it allows for the determination of free energy contributions from the ligand—solvent, metal—solvent, and metal—ligand interactions. Being able to split the solvent—system interactions from the metal—ligand interactions is especially interesting, as it would normally be an enormously complicated task due to the sheer number of solvent molecules in a typical molecular dynamics simulation. Additionally, TI allows for parameterization of important properties of atoms, as it directly relates the potentials for each atom to the free energy. Therefore, the parameterization process amounts to changing the potentials and recording the free energy change.[23,24]

CHAPTER 2: SIMULATING THE CHELATE EFFECT

2.1 Purpose

Despite the rich history of experimental studies focusing on the thermochemistry and kinetics associated with the chelate effect, molecular-level computational studies on the chelate ring opening/ring closure are scarce. The challenge lies in an accurate description of both the metal ion and its aqueous environment. Herein, we demonstrate that an optimized 12-6-4 Lennard-Jones (LJ) model can capture the thermodynamics and provide detailed structural and mechanistic insights into the formation of ethylenediamine (en) complexes with metal ions. The water molecules in the first solvation shell of the metal ion are found to facilitate the chelate ring formation. The optimized parameters further simulate the formation of bis and tris(en) complexes representing the wide applicability of the model to simulate coordination chemistry and self-assembly processes.

2.2 Computational Methods

All simulations were carried out using the CUDA version of the AMBER18 molecular dynamics package, the GAFF force field, and the TIP3P water model.[25] Quantum calculations for torsion profile analyses were carried out using the Gaussian16 suite of programs.[26]

Simulation Protocols

All simulations were carried out with the AMBER 18 suite of programs while the modeling and data analyses were performed using programs contained within the AmberTools suite of programs. For all simulations, periodic boundary conditions (PBCs) were employed together with PME to model long-range interactions with a 12 Å cut off. Complexes Cd(en), $Cd(en)_2$, and $Cd(en)_3$ and other metal ions complexes were solvated in cubic truncated octahedron with a minimum 24 Å distance between the atom of the solute and the edge of the periodic box. In total, there were 3743, 4506, 4876 water molecules for $[Cd(en)(H_2O)_4]^2+$, $[Cd(en)_2(H_2O)_2]^{2+}$, and $[Cd(en)_3]^2+$ respectively for the TIP3P water model. Chloride ion counterions were also treated with the 12-6-4 LJ nonbonded model. The compromise 12-6 LJ model was used as the representative 12-6 nonbonded model representing a good balance between structure and thermodynamics. We performed minimizations involving 2000 steps of steepest descent followed by 8000 steps of conjugate gradient. A 2 ns NPT heating procedure was performed to heat the system from 0 K to 300 K using a Langevin thermostat at 1atm. This was followed by an 8 ns NVT production simulation run. An integration time step of 1 fs was used in the equilibration step and 2 fs in the production studies. Snapshots for analysis were taken after every 10ps. Langevin dynamics temperature control was employed in the heating and

the production runs with a collision rate equal to 1.0 ps. SHAKE was utilized for the water molecules for all simulations. The reaction coordinate R was between the center of mass of ethylenediamine and the metal ion and used a step size of 0.05 Å moving from 1.8 Å to 5.0 Å and 0.1A thereafter till 11.0 Å. While the simulations were performed at constant volume, the obtained Helmholtz energies can be viewed as Gibbs free energies due to the expected negligible differences between the two under the simulation conditions used.

Ethylenediamine parametrization

In our investigation we found the default GAFF force field parameters in a box of explicit water molecules overestimates the intramolecular hydrogen bonding in ethylenediamine, favoring the *gauche* conformer of ethylenediamine (en) by 1 kcal/mol. This restricted the availability of the lone-pairs on Nitrogen to interact with the metal ions, leading to inaccurate representation of metal-ligand binding. A constrained optimization of ethylenediamine was performed at the MP2/6-311+G(d,p) level of theory using the implicit solvation model SMD to obtain the torsion (N–C–C–N) profile. At an interval of every 15° the N–C–C–N dihedral was fixed, while the rest of the ligand was allowed to optimize. The MP2 energies of both *gauche* and *anti* conformers were found to be equally stable (as shown in Figure 1) with the *cis* conformer being 7 kcal/mol relative to both *anti* and *gauche* conformers.

To resolve this issue, similar to the GLYCAM force field for the glycols the 1-4 electrostatic scaling factor was optimized to access both the *anti* and *gauche* conformers at room temperature. A scaling factor of 1.1 achieved our goal (the default scaling factor of 1.2) of an equal stability of the *anti* and *gauche* conformers. Figure 1 represents the torsion profile of ethylenediamine at the MP2/6-311+G(d,p) and the scaled 1-4 model from the GAFF force field.

CM5 Charges

Since we compared the binding energies of the metal ion bound to en and the requisite number of methylamine molecules to measure the chelate effect, different charge models were assessed to provide an alpha value that is transferable across the ligands. Table 1 lists the calculated binding energies for en bound complexes of Cd^2+ using different charge models and alpha parameters optimized to reproduce the experimental binding energy of the $MeNH_2$ complex of Cd^2+. We find similar order of magnitude in error in the calculated binding energies of $[Cd(en)(H_2O)_4]^2$+ using the pairwise parameters optimized to reproduce experimental binding energies of $[Cd(MeNH_2)(H_2O)_5]^2$+ using CM5 and AM1BCC charge models. However, we find CM5 charge model resulted the lowest calculated error in solvation energy for en against experiment value of –6.1 kcal/mol (see Table 2). Hence, we used the CM5 charge model for the present study. The ligand was first optimized at the M062X/6-311+(2df,2p) and then Hirshfeld charges were calculated using Pop=CM5 in Gaussian.

The potential functional form for the 12−6−4 ion parameters consists of both electrostatic and van der Waals interactions. The former is modeled using Coulomb pair potentials, q_iq_j/r^2, where *i* and *j* represent two particles, q_i and q_j are the charges belonging to the particles, and r_{ij} is the distance between the particles. The latter expands upon the classic Lennard-Jones (12−6) potential by including an extra attractive term that falls off as r^{-4}, denoted as a 12−6−4 potential. The 12−6–4 potential for nonbonded interactions is given by:

$$U_{ij}(r_{ij}) = \frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} - \frac{C_4^{ij}}{r_{ij}^4} + \frac{e^2 Q_i Q_j}{r_{ij}}$$

where e represents charge of the proton, Q_i and Q_j are partial charges of atoms i and j. To circumvent the tedious process of the parameterization of C_4 parameters between the different metal ions and ligands, we used a scaling factor, κ against C_6^{ij} to represent the C_4 term resulting in:

$$U_{ij}(r_{ij}) = \varepsilon_{ij}\left[\left(\frac{R_{min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{min,ij}}{r_{ij}}\right)^{6} - 2\kappa R_{min,ij}^{2}\left(\frac{R_{min,ij}}{r_{ij}}\right)^{4}\right] + \frac{e^2 Q_i Q_j}{r_{ij}}$$

where $\varepsilon_{ij} = \sqrt{\varepsilon_i \times \varepsilon_j}$ represents the well-depth and r_{ij} represents intermolecular distances.

$R_{\min,ij} = \frac{R_{min,ii} + R_{min,jj}}{2}$ is the distance at which the LJ potential has its minimum value.

The κ parameter has units of $Å^{-2}$. The additional attractive term, C_{ij}/r_{ij}^4, implicitly accounts for polarization effects by mimicking the charge-induced dipole interaction.

Optimization of C4 parameters

Divalent ion model parametrization for the interaction with ethylenediamine was performed to optimize the pairwise term, C_4^{ij}, between the metal ion and the nitrogen atom sites to reproduce the experimental binding free energies. We have computed binding free energies for Ni^{2+}, Fe^{2+}, Zn^{2+}, and Cd^{2+} interacting directly with the nitrogen atoms of en. Since we are only interested in the interaction between the metal ion and the ligand, the pairwise parameters corresponding to these are optimized based on the following equation:

$$\mathrm{C_4(atom\ type)} = \frac{\mathrm{C_4\ (H_2O)}}{\alpha_0\ (\mathrm{H_2O})} \times\ \alpha_0\ (\mathrm{atom\ type})$$

where α_0 is an atom type dependent polarizability. The optimized α_0 values to obtain the experimental binding energies of en complexes of different metal ions are listed in Table 3.

2.3 Simulating the Chelate Effect

The concept of the chelate ring is a natural outgrowth of Werner's coordination theory of 1893.[27] However, even Werner took eight more years to explicitly assign a cyclic structure to metal ion complexes of acetylacetone.[28,29] Later, the initial observation of the preferential association of bidentate ligands relative to monodentate ligands to metal ions was termed the "chelate effect" derived from the Greek word chela, referring to the great claw of lobsters.[30] The thermodynamics of these ring systems was quantified in the seminal work of Schwarzenbach in 1952.[31] The stability constant of ethylenediamine (en) complexes was found to be 10 orders of magnitude higher than that of the corresponding ammonia complexes. Over the past century, inorganic chemists have harnessed the chelate effect for controlled ligand design in various fields ranging from pharmacology[32] to material science (e.g., MOFs).[33,34] However, the mechanistic details of this widely employed reaction remain fully to be elucidated.

A quick perusal of the literature spanning over 60 years will reveal, in contrast to the exhaustive list of experimental studies, computational studies of the chelate effect are rare. With both thermodynamic and kinetic aspects of this macrocyclic effect depending on the surrounding solvent molecules, it is essential to include explicit solvent molecules in mechanistic studies.[35,36] The detailed ab initio study by Vallet et al. demonstrate the importance of microsolvated clusters in order to maintain the particle number, thus avoiding spuriously large contributions from translational entropies.[37] However, the effectiveness of using a few explicit water molecules with a continuum solvation model is often debated.[25,38] Nonetheless, when explicit water molecules are subjected to ab inito based molecular dynamics (MD) using, for example, Car−Parrinello MD

(CPMD), the steep scaling of these methods restricts them to shorter time scales as seen in the study of Buhl et al. on the binding modes of oxalate to UO2(oxalate) that can result in poor convergence.[39]

Beyond quantum based approaches classical modeling can afford an effective strategy, but the accuracy of these methods using the traditional Lennard-Jones potential (so-called 12-6 potential) has not been effective due to inaccuracies of the model in replicating the thermodynamics of solvation and structural details (e.g., ion-oxygen distances). Figure 2 shows a conventional MD calculation using the 12-6 potential simulating the formation of the chelate complex to be a thermodynamically less stable state than the separated species. Alternatively, we can employ our newly developed 12-6-4 LJ nonbonded model that can simultaneously reproduce thermodynamic and structural properties of broad range of aqueous ions.[17,40] The model includes a r^{-4} term tuned to incorporate the ion-induced dipole interactions (Figure 2, green). Prior studies demonstrated this term reproduced the induction contribution but was never incorporated in a generalized model to simulate the reactivity of metal ions with diverse ligands.[41,42] Important contributions from manybody interactions toward the solvation energy of metal ions further preclude the use of conventional 12-6 LJ nonbonded model.[43,44] While sophisticated force fields like SIBFA (sum of interactions between fragments ab initio computed),[45] NEMO (nonempirical molecular orbital)[46] and CTPOL include local polarization effects and charge transfer,[47] the lengthy parametrization and higher computational cost of the methods restrict their applicability. However, Panteva et al. found the 12-6-4 model to be a more accurate model over 12-6 models in simulating the interactions between the Mg2+ ion and nucleotide bases in aqueous solution.[48] Subsequent application of the model provided an improved understanding of metal ion transfer in the hammerhead ribozyme catalytic reaction.[49] These initial studies provide a strong rationale

toward the applicability of the 12-6-4 model to study solvent effects in metal complexation reactions.

Herein, we apply our 12-6-4 LJ nonbonded model for metal ions along with TIP3P water model to study the ion complexation of Zn^{2+}, Ni^{2+}, Cu^{2+} and Cd^{2+} with en. MD calculations over 8 ns were performed along the reaction coordinate between the metal ion and the center of mass of the ligand to study chelate ring formation. We present the reaction between Cd^{2+} and en as a representative example for the study of this process. The conventional 12-6 LJ model (Figure 2, red) suggests the chelate complex to be thermodynamically unfavorable relative to the solvated metal ion by 8.8 kcal/mol. The default 12-6-4 LJ model provides similar results with a small additional stability for the complex owing to some of the charge-induced dipole interaction being picked up by the default parameters (see Figure 3). Furthermore, both 12-6 and the default 12-6-4 depict the presence of only one minimum corresponding to the absence of any stable structure involving monocoordination of en with Cd^{2+}. This figure cautions against the use "off the shelf" ion parameters because they will likely be unable to model processes they were not designed to represent. At a minimum, careful validation is called for when using available metal ion models on new problems. Upon optimizing the 12-6-4 LJ (called m12-6-4) parameters between the metal ion and the nitrogen atoms of en to reproduce the experimental binding energies of the $[Cd(en)(H_2O)_4]^{2+}$ complex, the obtained potential of mean force (PMF) plot (Figure 2, green) captures two different minima corresponding to the ligand bound to the metal ion in monodentate (1Cd) as well as in a bidentate mode (2Cd, chelate complex). All reaction profiles with the m12-6-4 parameters can be viewed in Figure 4.

Figure 5 represents snapshots of different stationary points in the energy profile for Cd^{2+} interacting with en obtained with the m12-6-4 ion model. It should be noted that the highlighted

water molecules around the metal ion only represent a snapshot of the structural features around the ion that are, in reality, constantly undergoing exchange with the bulk water molecules. In Figure 5, we start from the solvated metal ion (at 8 Å, Figure 5a) and describe the reaction process associated with the formation of the chelate ring complex (at 2.4 Å, Figure 5f). The en ligand was parametrized to access both the gauche and anti conformers (for further details, see Figure 2). The free ligand itself can act as a hydrogen bond donor or acceptor in interactions with water molecules, while the Cd^{2+} ion is bound to six water molecules in an octahedral geometry.

The association of the metal ligand proceeds from the first minima ($\Delta Gmin1 = 7.8 \pm 0.3$ kcal/mol relative to the chelated complex) corresponding to the fully solvated hexacoordinated metal ion and unbound ligand at R = 4.80 Å (see Figure 5b). The minimum corresponds to the solvent separated ion ligand pair, wherein en engages in intermolecular hydrogen bonding with a water molecule present in the first solvation shell of Cd^{2+}. As the ligand approaches closer to the metal ion, at R = 4.35 Å we have transition state TS1 ($\Delta G1 \dagger = 8.4 \pm 0.1$ kcal/ mol) corresponding to the state where the ligand starts associating with the metal ion while concomitantly one of the water molecules begins to depart from the ion coordination sphere. Note that the small barrier height of 0.6 kcal/mol relative to the separated ion ligand pair is similar (1.1 kcal/ mol) to the one found in formation of $[Cd(MeNH_2)\text{-}(H_2O)_5]^{2+}$ when the pairwise parameter was optimized to reproduce the experimental binding energy of the $MeNH_2$ complex (Figure 6). We find the coordination number of the metal ion increases from 6 to 7 in these transient states (see Figure 5c) corresponding to an associative mechanism. The average bond lengths Cd−Owat and Cd−N distances for the outgoing water and the incoming amine are found to be 2.40 and 2.61 Å respectively. The reaction then proceeds to form 1Cd at R = 3.80 Å (see Figure 5d) corresponding to the ligand in the anti conformation coordinating with Cd^{2+} in a

monodentate fashion. The average Cd−N distances were found to be 2.36 and 5.18 Å for the bound and unbound nitrogen atoms. A further decrease in the intermolecular distance (R < 3.50 Å) results in the ligand losing its conformational flexibility and being present more in the gauche conformer allowing both of the nitrogen atoms to interact with the ion. At R = 3.23 Å (see Figure 5e), we obtain transition state TS2 (ΔG1 † = 7.9 kcal/ mol) corresponding to a ligand exchange between the second amine end of en and a water molecule. The average Cd−O_{wat} and Cd−N distances for the outgoing water and the incoming nitrogen atom are found to be 2.40 and 3.96 Å, respectively. We observe that while the lone pair of incoming nitrogen still faces the metal ion, the water molecules closer to this nitrogen end reorient themselves to hydrogen bond with this nitrogen atom and interact with the metal ion simultaneously. This bridging interaction likely reduces the free energy barrier for the formation of the second Cd−N bond. The gradual decrease in intermolecular distance to R = 2.40 Å results in chelate ring closure forming the final complex. Among the different metal ions, we find the highest barrier heights for Ni-en complexes corresponding to the stronger Ni−N bonds. Detailed information on the Cd-en reaction, including structures, can be seen in Figures 7-9. Detailed information on all metal-en reaction profiles, including comparisons in binding distance with quantum calculations, can be seen in Tables 4-10. Additionally, the Zn-en reaction scheme can be viewed in Figure 10, along with its reaction profile in Figure 11.

A 12-6-4 parameter set was then specifically built to reproduce the experimental binding energy of methylamine complexes of different metal ions. These parameters were then used to calculate the free energies of formation for bis(methylamine) and compare against prior obtained energies of en complex of Cd^{2+} (see Figure 12). The global minimum as expected for the en is deeper than the one for bis- (methylamine) where the free energy difference corresponds to the

chelate effect. Further, the constrained 2D PMF contour plots highlights the involvement of a rotational movement of the en for ring closure relative to the sole translational motion associated with the formation of Cd(II) complex of bis- (methylamine). Ab initio DLPNO-CCSD(T) calculations on model systems considering first solvation shell overestimate the chelate effect by 8 kcal/mol. Additional calculations with the pairwise parameters for Cd en interactions reproduced the experimental binding energies of bis(en) and tris(en) complexes of Cd(II) highlighting the transferability of the parameters in a different environment. The excellent agreement with experimental reaction energies is partly due to cancellation of errors in the incomplete representation of charge-transfer and many-body effects across the reactants and products. We recommend the present model for simulating metal ligand complexes with structural features (bond distances and coordination numbers) similar to that of the solvated metal ion. A comparison between the Cd-en complexation reaction scheme and the Zn-en reaction scheme, using the described level of quantum theory can be viewed in Tables 11-15, along with snapshots from the reaction scheme in Figures 13 and 14.

The present parameter sets allowed us to perform real time simulations on 10 Cd^{2+} and 30 en ligand molecules in a box of water to form mono-, bis-, and tris(en) complexes of Cd^{2+} (see Figure 15). We find that a 0.1 M system is saturated by the bis(en)Cd^{2+} after 100 ns of simulation, while higher concentrations (0.15 M) are required to observe the formation of tris(en)Cd^{2+} in a similar time-scale. We note that the present calculations are obtained from a single trajectory for each concentration but that an ensemble of such trajectories would afford a better comparison with experiment.

2.4 Conclusions

Overall, our model effectively simulates the chelate effect and provides molecular-level mechanistic details of the solvent assisted chelate ring opening/closure exchange reaction. The hydrogen bonding interactions between water molecules near Cd^{2+} and en facilitates the ring closure. Finally, the present model can be readily applied to studies of related metal ion complexation processes in condensed phases. Future work would be expected in the development of a robust set of LJ 12-6-4 nonbonded parameters for simulations of transition metals and biological ligands, based on experimental binding data. Ultimately, the goal of parameterization would be to create a metal-associated force field for organic ligands and proteins that allows for prediction of experimental properties, such as binding energies and docking sites. Many challenges will occur in the development and parameterization process, especially with the normalization of parameters between similar functional groups in different systems, and the search for accurate experimental data, but these challenges are to be expected. Just a few years ago, simulating the chelate effect seemed nearly impossible using molecular dynamics simulations, but it is done, and it is only a small step in expanding the application of the LJ 12-6-4 potential to encompass a vast amount of metal-ligand interactions.

APPENDICES

APPENDIX A: Tables

Charge	m12-6-4 (α_0)	ΔG_{calc}	Error
AM1-BCC	1.92	–2.9	–4.8
RESP	1.52	NA	–7.7
CM5	3.16	–12.7	5.0

NA= solvent separated ion ligand pair is more stable than the metal ligand complex

Table 1: The calculated energies (in kcal/mol) of $[Cd(en)(H_2O)_4]^{2+}$ using different charge models and alpha values optimized to reproduce experimental binding energies of $[Cd(MeNH_2)(H_2O)_5]^{2+}$ at the respective charge model.

The experimental Gibbs free energy for formation is –7.7 kcal/mol value.

Charge	ΔG_{calc}	Error
AM1-BCC	–12.4	–6.3
RESP	–22.0	–15.9
CM5	–6.8	–0.7

Table 2: Calculated solvation energies for en using different charge models.

Solvation energies are measured in kcal/mol. The experimental $\Delta G_{solvation}$ = –6.1 kcal/mol.

Ion	12-6-4		m12-6-4		Experimental[1] binding energy (kcal/mol)
	α_0	C_4^{ij}	α_0	C_4^{ij}	
Ni^{2+}	1.09	160.632	2.35	346.316	–10.5[a]
Fe^{2+}	1.09	123.040	2.47	278.816	–5.8[a]
Zn^{2+}	1.09	174.822	2.16	347.400	–8.1[b]
Cd^{2+}	1.09	140.099	2.59	332.897	–7.7[b]

Table 3: Comparison of default and modified α_0 values and the corresponding C_4 in the formation of ethylenediamine complexes of different divalent ions.

[1]The experimental binding energies were obtained from the database: Paoletti, P. *Pure & Appl. Chem*. **1984**, *56*, 491–522. Individual references: [a]Bjerrum, "Metal Animine Formation", P. Haase and Son, Copenhagen (**1941**). [b] Bjerrum, J. and Anderson, P. *Kgl. Danske Videnskab, Selskab . Mat-Phys. Medd* **1945**, *22*, 7.

Ion	R_{min1}	ΔG_{min1}	$R_1^\dagger$	$\Delta G_1^\dagger$	$R_{\mathbf{1M}}$	$\Delta G_{\mathbf{1M}}$	$R_2^\dagger$	$\Delta G_2^\dagger$	$R_{\mathbf{2M}}$	$\Delta G_{\mathbf{2M}}$
Fe^{2+}	5.12	5.1±0.4	3.95	7.4±0.4	3.59	5.8±0.1	3.30	6.6±0.1	2.06	0.0
Ni^{2+}	5.12	9.7±0.5	4.10	11.0±0.5	3.52	8.5±0.1	2.57	9.8±0.1	2.06	0.0
Zn^{2+}	5.09	7.6±0.4	3.98	9.4±0.2	3.53	7.4±0.1	3.30	8.1±0.0	2.03	0.0
Cd^{2+}	5.17	7.6±0.2	4.35	8.4±0.2	3.82	7.0±0.2	3.23	7.8±0.2	2.41	0.0

Table 4: Comparison of structural and energetic features in the association profiles (reverse scan, discussed in main text) of ethylenediamine complexes $[\mathbf{M}(en)(H_2O)_4]^{2+}$ for different metal (**M**) ions using the corresponding m12-6-4 model.

R_{min1} - intermolecular distance (Å) in the solvent separated ion ligand pair. ΔG_{min1} is the free energy (kcal/mol) relative to chelate complex for the solvent separated ion ligand pair. $R_1^\dagger$- intermolecular distance (Å) in transition state **TS1**. $\Delta G_1^\dagger$ - activation free energy (kcal/mol) of **TS1** relative to the chelate complex, $\Delta G_{\mathbf{1M}}$ corresponding to the free energy (kcal/mol) of $\mathbf{1_M}$. $R_{\mathbf{1M}}$ - intermolecular distance (Å) *R* in $\mathbf{1_M}$, $R_2^\dagger$- intermolecular distance (Å) *R* in transition state **TS2**, $\Delta G_2^\dagger$ - activation free energy (kcal/mol) corresponding to **TS2**, $R_{\mathbf{2M}}$ - intermolecular distance, *R* (Å) in $\mathbf{2_M}$. Standard deviations come from three independent 8 ns segments of data.

Ion	R_{min1}	ΔG_{min1}	$R_1^{\dagger}$	$\Delta G_1^{\dagger}$	R_{1M}	ΔG_{1M}	$R_2^{\dagger}$	$\Delta G_2^{\dagger}$	R_{2M}	ΔG_{2M}
Fe^{2+}	5.27	5.4±0.1	3.95	7.5±0.1	3.59	5.7±0.1	3.30	6.6±0.1	2.06	0.0
Ni^{2+}	5.27	10.3±0.3	4.02	8.5±0.3	3.52	9.1±0.2	2.57	10.0±0.2	2.06	0.0
Zn^{2+}	5.09	7.9±0.4	3.98	9.6±0.3	3.53	7.4±0.2	3.30	8.1±0.1	2.03	0.0
Cd^{2+}	5.17	7.7±0.1	4.27	8.5±0.2	3.82	7.1±0.2	3.23	7.9±0.2	2.41	0.0

Table 5: Comparison of structural and energetic features in the dissociation profiles (forward scan) of the monoethylenediamine complexes [**M**(en)$(H_2O)_4]^{2+}$ for different metal (**M**) ions using the m12-6-4 model.

The scan along the opposite direction was done to probe the possible hysteresis. For consistency, the notations used in this Table are same as the ones used in Table 4. Standard deviations come from three independent 8 ns segments of data.

Ion	R_{min1}	ΔG_{min1}	$R_1^\dagger$	$\Delta G_1^\dagger$	$R_{\mathbf{3M}}$	$\Delta G_{\mathbf{3M}}$	$R_2^\dagger$	$\Delta G_2^\dagger$	$R_{\mathbf{4M}}$	$\Delta G_{\mathbf{4M}}$
Fe^{2+}	5.41	4.1±0.1	3.88	6.8±0.1	3.59	5.6±0.1	3.30	6.6±0.1	2.13	0.0
Ni^{2+}	5.34	10.3±1.0	3.22	8.9±0.3	2.93	8.2±0.2	2.49	10.3±0.2	2.06	0.0
Zn^{2+}	5.17	5.7±0.3	3.90	8.2±0.2	3.60	6.6±0.1	3.30	7.6±0.1	2.03	0.0
Cd^{2+}	5.32	6.4±0.1	4.34	7.6±0.1	3.82	6.7±0.1	3.23	7.7±0.0	2.41	0.0

Table 6: Comparison of structural and energetic features obtained from free energy profiles of ion dissociation of bisethylenediamine complexes $[\mathbf{M}(en)_2(H_2O)_2]^{2+}$ for different metal (**M**) ions using the m12-6-4 model.

R_{min1} - intermolecular distance (Å) in solvent separated ion ligand pair. ΔG_{min1} is the free energy (kcal/mol) relative to the chelate complex for the solvent separated ion ligand pair. $R_1^\dagger$- intermolecular distance (Å) in transition state **TS3**. $\Delta G_1^\dagger$ - activation free energy (kcal/mol) of **TS3** relative to the chelate complex, $\Delta G_{\mathbf{3M}}$ corresponds to the free energy (kcal/mol) of $\mathbf{3_M}$. $R_{\mathbf{3M}}$ - intermolecular distance (Å) *R* in $\mathbf{3_M}$, $R_2^\dagger$- intermolecular distance (Å) *R* in transition state **TS4**, $\Delta G_2^\dagger$ - activation free energy (kcal/mol) corresponding to **TS4**, $R_{\mathbf{4M}}$ - intermolecular distance, *R* (Å) in $\mathbf{4_M}$. Standard deviations come from three independent 8 ns segments of data.

Ion	R_{min1}	ΔG_{min1}	R_{5M}	ΔG_{5M}	$R_1^\dagger$	$\Delta G_1^\dagger$	R_{6M}	ΔG_{6M}	$R_2^\dagger$	$\Delta G_2^\dagger$	R_{7M}	ΔG_{7M}
Fe^{2+}	5.58	1.9±0.1	3.59	4.7±0.2	3.88	5.9±0.1	NA	NA	3.29	5.0±0.1	2.13	0.0
Ni^{2+}	5.58	7.8±0.1	3.59	7.1±0.6	3.30	8.4±0.6	3.00	7.4±0.6	2.49	10.8±0.5	2.06	0.0
Zn^{2+}	5.17	4.3±0.5	3.60	6.3±0.2	3.30	7.5±0.2	2.93	6.4±0.2	2.56	7.5±0.1	2.11	0.0
Cd^{2+}	5.24	4.0±0.2	3.90	5.4±0.2	4.35	5.9±0.2	NA	NA	3.30	7.0±0.1	2.41	0.0

Table 7: Comparison of structural and energetic features obtained from free energy profiles for the formation of the trisethylenediamine complexes $[\mathbf{M}(en)_3]^{2+}$ for different metal (**M**) ions using the m12-6-4 model.

R_{min1} - intermolecular distance (Å) in the solvent separated ion ligand pair. ΔG_{min1} is the free energy (kcal/mol) relative to the chelate complex for the solvent separated ion ligand pair. $R_1^\dagger$- intermolecular distance (Å) in transition state **TS5**. $\Delta G_1^\dagger$ - activation free energy (kcal/mol) of **TS5** relative to the chelate complex, ΔG_{5M} corresponds to the free energy (kcal/mol) of $\mathbf{5_M}$. R_{5M} - intermolecular distance (Å) *R* in $\mathbf{5_M}$, $R_2^\dagger$- intermolecular distance (Å) *R* in transition state **TS6**, $\Delta G_2^\dagger$ - activation free energy (kcal/mol) corresponding to **TS6**, R_{6M} - intermolecular distance, *R* (Å) in $\mathbf{6_M}$. R_{7M} - intermolecular distance, *R* (Å) in $\mathbf{7_M}$. Standard deviations come from three independent 8 ns segments of data.

	Fe	Cd	Zn
DFT	2.12	2.42	2.22
12-6-4	2.15	2.37	2.13
Expt.	2.21	2.37	2.18

Table 8: Comparison of calculated 12-6-4 and DFT structural features against experimental[a,b] observations of bond lengths (in Å). DFT = PBE0-D3BJ/def2TZVPP.

[a] Pham, D. N. K.; Roy, M.; Golen, J. A. and Manke, D. R. *Acta Cryst*. **2017,** *C73*, 442–446. [b] Ma, G.; Fischer, A.; Nieuwendaal, R.; Ramaswamy, K.; Hayes, S. *Inorg. Chim. Acta* **2005**, 358, 3165–3173.

Ion	R_{min1}	ΔG_{min1}	$R_1^{\dagger}$	$\Delta G_1^{\dagger}$	R_{min2}	ΔG_{min2}
Fe^{2+}	4.76	0.8±0.0	NA	NA	NA	NA
Ni^{2+}	4.25	2.0±0.2	2.64	9.8±0.2	2.13	0.0
Zn^{2+}	4.27	1.6±0.1	2.71	6.9±0.1	2.11	0.0
Cd^{2+}	4.65	1.2±0.0	3.23	3.2±0.0	2.33	0.0

Table 9: Comparison of structural and energetic features obtained from free energy profiles for ion dissociation of methylamine complexes $[\mathbf{M}(CH_3NH_2)(H_2O)_5]^{2+}$ for different metal (**M**) ions using the m12-6-4 model.

NA= solvent separated ion ligand pair is more stable than the metal ligand complex. R_{min1} intermolecular distance in solvent separated ion ligand pair, $R_1^{\dagger}$- intermolecular distance (Å) in transition state corresponding to formation of methylamine complex. $\Delta G_1^{\dagger}$ - activation free energy (kcal/mol) corresponding to transition state, R_{min2} - equilibrium direct contact distance (Å), Standard deviations come from three independent 8 ns segments of data. Note the 12-6-4 pairwise parameters used here are the ones optimized to reproduce experimental binding energies of metal ion complexes of en.

Ion	R_{min1}	ΔG_{min1}	$R_1^\dagger$	$\Delta G_1^\dagger$	R_{min2}	ΔG_{min2}
Fe^{2+}	4.32	0.3±0.0	2.71	6.2±0.0	2.13	0.0
Ni^{2+}	4.25	1.5±0.9	2.64	10.2±0.7	2.13	0.0
Zn^{2+}	4.27	1.8±0.2	2.71	8.1±0.2	2.11	0.0
Cd^{2+}	4.65	0.8±0.0	3.00	2.6±0.0	2.33	0.0

Table 10: Comparison of structural and energetic features obtained from free energy profiles for ion dissociation of bis(methylamine) complexes $[\mathbf{M}(CH_3NH_2)_2(H_2O)_4]^{2+}$ to $[\mathbf{M}(CH_3NH_2)(H_2O)_5]^{2+}$ for different metal (**M**) ions using the m12-6-4 model.

R_{min1} intermolecular distance in solvent separated ion ligand pair, $R_1^\dagger$- intermolecular distance (Å) in transition state corresponding to formation of bis(methylamine) complex. $\Delta G_1^\dagger$ - activation free energy (kcal/mol) corresponding to transition state, R_{min2} - equilibrium direct contact distance (Å), Standard deviations come from three independent 8 ns segments of data. Note the 12-6-4 pairwise parameters used here are the ones optimized to reproduce experimental binding energies of metal ion complexes of en.

Reaction	**Model A**	**Model B**	Expt.
R1	–9.0	–1.6	–3.75
R2	–7.0	0.7	NA
R3	–8.2	2.0	0.9
R4	–25.6	–9.3	–7.7
R5	–17.1	0.4	–2.8
R6	–8.4	–9.6	–1.4

Table 11: Binding free energies (kcal/mol) calculated at DLPNO-CCSD(T)/DKH-TZVP//PBE0-D3BJ/def2-TZVP of reactions involved in various step in formation of bis(methylamine) and ethylenediamine bound complex of Cd^{2+}.

$Cd(H_2O)_6]^{2+} + L\ ! \longrightarrow Cd(H_2O)_5L]^{2+} + H_2O\ !$ **Model A**

$Cd(H_2O)_6]^{2+} + L\ ! \longrightarrow Cd(H_2O)_5L]^{2+} H_2O\ !$ **Model B**

$Cd(H_2O)_6]^{2+} + MeNH_2 \longrightarrow [Cd(H_2O)_5(MeNH_2)]^{2+} + H_2O$ **R1**

$[Cd(H_2O)_6]^{2+} + en \longrightarrow [Cd(H_2O)_5(en\text{-}uni)]^{2+} + H_2O$ **R2**

$[Cd(H_2O)_5(MeNH_2)]^{2+} + MeNH_2 \longrightarrow [Cd(H_2O)_4(MeNH_2)_2]^{2+} + H_2O$ **R3**

$[Cd(H_2O)_6]^{2+} + en \longrightarrow [Cd(H_2O)_4(en)]^{2+} + 2\,H_2O$ **R4**

$[Cd(H_2O)_6]^{2+} + 2\ MeNH_2 \longrightarrow [Cd(H_2O)_4(MeNH_2)_2]^{2+} + 2\ H_2O$ **R5**

$[Cd(H_2O)_4(MeNH_2)_2]^{2+} + en \longrightarrow [Cd(H_2O)_4(en)]^{2+} + 2\ MeNH_2$ **R6**

Table 12: Reaction Schemes for DLPNO-CCSD(T) calculations for the chelate effect with the Cd^{2+} ion.

Herein we perform some high level QM calculations considering model systems of $[Cd(MeNH_2)_2(H_2O)_4]^{2+}$ and $[Cd(en)(H_2O)_4]^{2+}$. Additional calculations with $[Cd(MeNH_2)_2(H_2O)_4]^{2+}(H_2O)_2$ and $[Cd(en)(H_2O)_4]^{2+}(H_2O)_2$ with same number of water molecules as that in $[Cd(H_2O)_6]^{2+}$ consistent to the models used by Vallet et. al. [*J. Am. Chem. Soc.*, **2003**, *125*, 14941].

Reaction	**Model A**	**Model B**	Expt.
R7	–10.3	–3.1	NA
R8	–8.8	–1.6	NA
R9	–19.0	–2.7	NA
R10	–27.7	–14.6	NA
R11	–8.7	0.3	NA
R12	–8.7	–11.9	NA

Table 13. Binding free energies (kcal/mol) calculated at DLPNO-CCSD(T)/DKH-TZVP//PBE0-D3BJ/def2-TZVP of reactions involved in various step in formation of bis(methylamine) and ethylenediamine bound complex of Zn^{2+}.

$Zn(H_2O)_6]^{2+} + L\ ! \longrightarrow Zn(H_2O)_5L]^{2+} + H_2O\ !$	**Model A**
$Zn(H_2O)_6]^{2+} + L\ ! \longrightarrow Zn(H_2O)_5L]^{2+} H_2O\ !$	**Model B**
$Zn(H_2O)_6]^{2+} + MeNH_2 \longrightarrow [Zn(H_2O)_5(MeNH_2)]^{2+} + H_2O$	**R7**
$[Zn(H_2O)_6]^{2+} + en \longrightarrow [Zn(H_2O)_5(en\text{-}uni)]^{2+} + H_2O$	**R8**
$[Zn(H_2O)_5(MeNH_2)]^{2+} + MeNH_2 \longrightarrow [Zn(H_2O)_4(MeNH_2)_2]^{2+} + H_2O$	**R9**
$[Zn(H_2O)_6]^{2+} + en \longrightarrow [Zn(H_2O)_4(en)]^{2+} + 2\ H_2O$	**R10**
$[Zn(H_2O)_6]^{2+} + 2\ MeNH_2 \longrightarrow [Zn(H_2O)_4(MeNH_2)_2]^{2+} + 2\ H_2O$	**R11**
$[Zn(H_2O)_4(MeNH_2)_2]^{2+} + en \longrightarrow [Zn(H_2O)_4(en)]^{2+} + 2\ MeNH_2$	**R12**

Table 14: Reaction Schemes for DLPNO-CCSD(T)/DKH-TZVP//PBE0-D3BJ/def2-TZVP calculations for the chelate effect with the Zn^{2+} ion.

$[Cd(H_2O)_6]^{2+}$				$[Zn(H_2O)_6]^{2+}$			
G_{corr} = 0.105670				G_{corr} = 0.109463			
$E_{DLPNO\text{-}CCSD(T)}$ = – 6048.4175016				$E_{DLPNO\text{-}CCSD(T)}$ = –2252.803331			
Cd	0.000002	-0.000005	-0.000058	Zn	0.000030	0.000021	-0.000162
O	1.607322	0.606510	1.552589	O	1.487474	0.497265	1.432056
O	-0.666608	2.166001	0.355858	O	-0.684905	1.964818	0.206623
O	0.666699	-2.166004	-0.355861	O	0.684818	-1.964923	-0.206544
O	-1.607443	-0.606809	-1.552556	O	-1.487516	-0.497236	-1.432096
O	-1.559325	-0.608959	1.579994	O	-1.397389	-0.605727	1.457967
O	1.559242	0.609064	-1.580189	O	1.397419	0.605542	-1.458212
H	2.232219	0.061910	2.044517	H	2.094530	-0.109995	1.870630
H	-1.455941	2.395852	0.860531	H	-1.433699	2.233840	0.751458
H	0.447369	-2.933841	0.184882	H	0.411838	-2.728736	0.314421
H	-2.232325	-0.062182	-2.044477	H	-2.094543	0.110027	-1.870705
H	-1.463670	-0.689455	2.535750	H	-1.290164	-0.680871	2.412912
H	2.401105	1.016222	-1.344394	H	2.256220	0.980193	-1.229885
H	-0.447251	2.933892	-0.184798	H	-0.411952	2.728823	-0.314078
H	1.463590	0.689516	-2.535948	H	1.290188	0.680892	-2.413140
H	-1.545302	-1.450205	-2.016252	H	-1.490593	-1.318569	-1.937329
H	1.456051	-2.395897	-0.860482	H	1.433520	-2.234179	-0.751389
H	1.545091	1.449875	2.016329	H	1.490578	1.318577	1.937323
H	-2.401173	-1.016129	1.344165	H	-2.256205	-0.980406	1.229748
$[Cd(H_2O)_5(MeNH_2)]^{2+}$				$[Zn(H_2O)_5(MeNH_2)]^{2+}$			
G_{corr} = 0.146268				G_{corr} = 0.149439			
$E_{DLPNO\text{-}CCSD(T)}$ = –6067.774915				$E_{DLPNO\text{-}CCSD(T)}$ = –2272.16226622			
Cd	-0.107000	0.065041	-0.025175	Zn	-0.136875	-0.001104	0.000535
O	-1.390998	1.476398	1.357553	O	-1.158519	1.370432	1.321610
O	-2.213387	-0.291443	-0.822737	O	-2.070659	-0.208963	-0.850188
O	0.135022	1.880636	-1.570433	O	0.083488	1.709367	-1.325219
O	-0.846812	-1.623688	1.522596	O	-0.842800	-1.561579	1.334208
H	-1.691424	1.214645	2.235198	H	-1.473798	1.175906	2.211058

Table 15: Cartesian Coordinates of the optimized geometries of various species in the formation of mono(en) and bis(methylamine) complexes of Cd(II) and Zn(II) using Model A and Model B. Single point electronic energy in the solvent phase (in a.u) calculated at DLPNO-CCSD(T) with SMD solvation model along with thermal corrections to the Gibbs free energy at PBE-0D3BJ are provided for each species.

Table 15 (cont'd)

H	-2.462812	-0.784592	-1.612454	H	-2.301456	-0.726626	-1.629749
H	0.908178	2.419150	-1.772071	H	0.872092	2.211491	-1.560367
H	-0.408230	-2.219591	2.139566	H	-0.411899	-2.148681	1.965124
H	-1.727282	-1.992554	1.383861	H	-1.740118	-1.895340	1.217336
H	-2.967007	0.252448	-0.566915	H	-2.838835	0.317236	-0.600911
H	-1.439458	2.438352	1.321891	H	-1.164122	2.328599	1.216783
H	-0.548963	2.157983	-2.189872	H	-0.581284	1.924878	-1.989179
N	1.890464	0.171385	0.979047	N	1.720189	0.177221	0.884689
H	1.907623	-0.524854	1.719610	H	1.784614	-0.482704	1.656116
H	1.984566	1.062100	1.459858	H	1.762431	1.088183	1.335793
C	3.064434	-0.041276	0.099399	C	2.926789	0.005819	0.037092
H	3.002265	-1.030925	-0.350433	H	2.953922	-1.010834	-0.350831
H	3.066058	0.701065	-0.696895	H	2.890542	0.695969	-0.804397
H	3.995965	0.039738	0.659644	H	3.837262	0.194016	0.605670
O	0.433046	-1.548276	-1.662687	O	0.419810	-1.475598	-1.484938
H	0.394187	-2.505670	-1.557671	H	0.388807	-2.429516	-1.349359
H	0.809716	-1.385505	-2.534928	H	0.868566	-1.327607	-2.324926
$[Cd(H_2O)_5(MeNH_2)]^{2+}\bullet H_2O$				$[Zn(H_2O)_5(MeNH_2)]^{2+}\bullet H_2O$			
G_{corr} = 0.168519				G_{corr} =0.1741825			
$E_{DLPNO\text{-}CCSD(T)}$ = – 6144.1755774				$E_{DLPNO\text{-}CCSD(T)}$ = –2348.5655780290			
Cd	-0.164375	0.157013	0.032397	Zn	-0.250375	-0.201092	0.161589
O	-1.473138	1.617068	1.312781	O	-1.453775	1.465122	0.822110
O	-2.048183	0.057062	-1.341810	O	-2.096715	-1.201628	-0.228080
O	0.201570	1.804273	-1.610175	O	-0.248598	0.853760	-1.675258
O	-0.716288	-1.432995	1.674851	O	-0.518311	-1.464821	1.942761
H	-1.833144	1.427784	2.185530	H	-1.858787	1.631969	1.678812
H	-2.242791	-0.715736	-1.883787	H	-2.155437	-1.882938	-0.908176
H	0.941358	2.366373	-1.861724	H	-0.275255	0.493509	-2.566568
H	0.009153	-1.714146	2.274067	H	-0.038858	-1.623067	2.762741
H	-1.541814	-1.791324	2.014722	H	-1.331526	-1.980268	1.985533
H	-2.895329	0.448433	-1.099171	H	-2.908384	-0.683644	-0.274744
H	-1.572328	2.564876	1.169307	H	-1.518649	2.278863	0.285694
H	-0.455916	1.881536	-2.310413	H	-0.579798	1.774589	-1.716516
N	1.852933	0.462291	0.972046	N	1.498140	0.548325	0.969659
H	1.952522	-0.257914	1.689663	H	1.963656	-0.179430	1.506059
H	1.855731	1.346057	1.473589	H	1.225697	1.247088	1.656125
C	3.025324	0.406308	0.072101	C	2.497854	1.150564	0.054325

Table 15 (cont'd)

H	3.060718	-0.569753	-0.410454	H	2.900405	0.378656	-0.599767
H	2.941465	1.167444	-0.701535	H	2.018387	1.907361	-0.561996
H	3.954898	0.567026	0.619153	H	3.319945	1.603284	0.609029
O	0.314259	-1.863442	-1.058670	O	0.643389	-1.903209	-0.873565
H	0.081894	-2.680411	-0.604718	H	0.827278	-2.730976	-0.414701
H	0.740818	-2.106789	-1.887244	H	1.235226	-1.875824	-1.633607
O	1.629977	-1.863591	2.981083	O	-1.275976	3.378220	-1.206162
H	2.109837	-2.690046	2.850088	H	-2.070017	3.692986	-1.656253
H	1.728989	-1.664116	3.919939	H	-0.692314	4.145192	-1.158050
$[Cd(H_2O)_5(en\text{-}uni)]^{2+}$				$[Zn(H_2O)_5(en\text{-}uni)]^{2+}$			
G_{corr} =0.188765				G_{corr} = 0.1923894			
$E_{DLPNO\text{-}CCSD(T)}$ = –6162.32055945				$E_{DLPNO\text{-}CCSD(T)}$ = –2366.709076			
Cd	-0.245705	-0.005670	-0.087980	Zn	-0.224755	-0.006050	-0.040388
O	-0.910788	1.666476	1.443913	O	-1.058877	1.446234	1.330398
O	-2.441134	-0.567434	-0.673824	O	-2.225275	-0.178557	-0.700132
O	-0.318297	1.686873	-1.839076	O	0.003594	1.604849	-1.467981
O	-0.876866	-1.612477	1.572845	O	-0.892327	-1.547028	1.371812
H	-1.300657	1.473175	2.304119	H	-1.411109	1.255693	2.206597
H	-2.580915	-1.098006	-1.466421	H	-2.502920	-0.678019	-1.476509
H	0.355634	2.333937	-2.075210	H	0.809686	2.062198	-1.732104
H	-0.390399	-2.111520	2.238139	H	-0.471323	-2.060421	2.069647
H	-1.718789	-2.068186	1.461082	H	-1.778416	-1.910094	1.259047
H	-3.294991	-0.209822	-0.407088	H	-2.956482	0.386262	-0.426795
H	-1.104349	2.591561	1.256490	H	-1.044966	2.404262	1.227226
H	-1.137293	1.998086	-2.240991	H	-0.691464	1.902863	-2.064893
N	1.805493	0.149362	0.821748	N	1.661344	-0.012257	0.754399
H	1.847681	-0.503614	1.601867	H	1.783142	-0.884677	1.265721
H	1.856189	1.065937	1.263595	H	1.707597	0.706518	1.474976
C	3.027262	-0.054882	-0.004821	C	2.854390	0.137562	-0.126990
H	2.958853	-1.041875	-0.467610	H	2.780812	-0.603796	-0.925106
H	3.023382	0.688224	-0.805133	H	2.818259	1.129840	-0.583162
C	4.313471	0.055326	0.823309	C	4.165671	-0.034922	0.647921
O	-0.024124	-1.236294	-2.085929	O	0.122545	-1.473238	-1.594821
H	0.319728	-2.128847	-2.210422	H	0.194394	-2.424904	-1.459460
H	0.208479	-0.736974	-2.876996	H	0.505642	-1.284495	-2.458871
N	5.530309	-0.130124	0.101351	N	5.361082	0.092682	-0.116441
H	4.269649	-0.680807	1.635957	H	4.157322	-1.022692	1.126272
H	4.337137	1.038789	1.307814	H	4.187444	0.696194	1.465154

Table 15 (cont'd)

H	5.666795	-1.052629	-0.282078	H	5.513455	-0.624617	-0.807667
H	5.735218	0.580746	-0.583220	H	5.542502	1.009948	-0.492418
$[Cd(H_2O)_5(en\text{-}uni)]^{2+} \bullet H_2O$				$[Zn(H_2O)_5(en\text{-}uni)]^{2+} \bullet H_2O$			
$G_{corr} = 0.21063746$				$G_{corr} = 0.216159$			
$E_{DLPNO\text{-}CCSD(T)} = -6238.720335$				$E_{DLPNO\text{-}CCSD(T)} = -2443.111548$			
Cd	-0.166139	-0.173713	0.287471	Zn	-0.195315	-0.074758	0.215555
O	-0.717098	1.473640	1.897559	O	-1.081085	0.835323	1.998229
O	-2.107349	-0.158439	-0.817933	O	-2.175421	-0.416920	-0.338990
O	-0.013688	1.810334	-1.064992	O	-0.441448	1.791397	-0.824843
O	-0.814552	-1.786964	1.980957	O	-0.524049	-1.985273	1.270154
H	-1.069686	1.344556	2.784320	H	-1.435508	0.303769	2.718574
H	-2.405014	-0.952534	-1.271087	H	-2.284583	-0.901602	-1.164504
H	0.690633	2.430320	-1.280436	H	0.157787	2.457387	-1.174303
H	-0.290016	-2.145892	2.704667	H	0.027329	-2.586571	1.782045
H	-1.721285	-2.066688	2.149957	H	-1.409215	-2.366724	1.248140
H	-2.868041	0.485494	-0.763747	H	-2.768675	0.361866	-0.372684
H	-0.972632	2.360400	1.621332	H	-1.022940	1.741275	2.318547
H	-0.682869	1.914273	-1.749746	H	-1.349012	2.146133	-0.862310
N	1.988185	0.020437	0.879342	N	1.757309	0.148764	0.850691
H	2.263878	-0.726823	1.512118	H	1.905477	-0.441803	1.667039
H	2.059238	0.870576	1.435240	H	1.864803	1.093675	1.216354
C	2.974589	0.075980	-0.233577	C	2.869875	-0.103366	-0.107098
H	2.901593	-0.861841	-0.788414	H	2.770782	-1.125749	-0.477560
H	2.678239	0.879283	-0.911561	H	2.740483	0.569124	-0.958445
C	4.409668	0.291982	0.252247	C	4.252862	0.096523	0.521877
O	-0.055712	-2.264163	-0.912504	O	0.152116	-1.204766	-1.629779
H	-0.095748	-3.091953	-0.421873	H	0.437052	-2.125696	-1.633306
H	0.313052	-2.475405	-1.777231	H	0.467783	-0.815390	-2.452655
N	5.408599	0.322487	-0.769723	N	5.359346	-0.134023	-0.347717
H	4.657857	-0.503103	0.965443	H	4.342023	-0.566469	1.391473
H	4.453878	1.229651	0.818412	H	4.318187	1.118843	0.914271
H	5.515287	-0.534715	-1.289805	H	5.480802	-1.085131	-0.657974
H	5.362742	1.115277	-1.391156	H	5.459626	0.518724	-1.108430
O	-4.001092	1.639085	-0.633168	O	-3.273588	2.095563	-0.606930
H	-4.687948	1.600619	0.040588	H	-3.723400	2.563495	0.105967
H	-4.421588	2.033797	-1.404789	H	-3.798000	2.283415	-1.396000

Table 15 (cont'd)

$[Cd(H_2O)_4(MeNH_2)_2]^{2+}$				$[Zn(H_2O)_4(MeNH_2)_2]^{2+}$			
G_{corr} = 0.1873059				G_{corr} = 0.188861			
$E_{DLPNO-CCSD(T)}$ = –6087.131487				$E_{DLPNO-CCSD(T)}$ = –2291.518042			
Cd	-0.275715	-0.070817	-0.161426	Zn	-0.284612	0.036561	-0.150656
O	-0.995847	1.624821	1.505421	O	-1.035198	1.524584	1.306020
N	-2.210261	-0.227165	-1.254087	N	-2.061901	-0.295979	-1.121585
O	0.416051	1.634398	-1.804388	O	0.333851	1.628487	-1.568148
O	-0.806615	-1.821193	1.491913	O	-0.804384	-1.614685	1.451911
H	-1.369957	1.392385	2.363102	H	-1.875613	1.488138	1.776019
C	-3.468806	-0.421640	-0.498643	C	-3.344320	-0.408868	-0.386132
H	1.197021	1.614146	-2.367565	H	1.232616	1.798517	-1.871542
H	-0.274142	-2.595749	1.705635	H	-0.966578	-1.445022	2.387097
H	-1.663310	-1.992362	1.899300	H	-1.051034	-2.536180	1.313624
H	-2.080829	-1.002715	-1.899818	H	-1.891225	-1.164671	-1.623617
H	-1.211635	2.555504	1.378634	H	-0.736122	2.439475	1.364977
H	0.024750	2.503755	-1.945501	H	-0.199373	2.358444	-1.902403
N	1.614180	0.136868	0.992416	N	1.511532	-0.048330	0.842131
H	1.667246	-0.628024	1.659145	H	1.488951	-0.954844	1.302936
H	1.466266	0.967824	1.560391	H	1.467478	0.624359	1.604558
C	2.901111	0.249027	0.273627	C	2.811947	0.104926	0.150144
H	3.026204	-0.606081	-0.388088	H	2.859659	-0.578054	-0.695171
H	2.902549	1.163301	-0.316687	H	2.920432	1.128352	-0.207367
H	3.738137	0.284346	0.971869	H	3.641135	-0.109807	0.824874
O	0.428863	-1.782438	-1.794520	O	0.431220	-1.541602	-1.755048
H	0.516087	-2.729856	-1.640484	H	0.794386	-2.422170	-1.606365
H	0.859356	-1.621621	-2.641620	H	0.674095	-1.313394	-2.659687
H	-4.334858	-0.368755	-1.159523	H	-4.146598	-0.729578	-1.050659
H	-3.561019	0.345289	0.267853	H	-3.610061	0.559121	0.036688
H	-3.457839	-1.402258	-0.024758	H	-3.240836	-1.131925	0.419286
H	-2.301868	0.593268	-1.846769	H	-2.174331	0.398404	-1.856454
$[Cd(H_2O)_4(MeNH_2)_2]^{2+}\bullet(H_2O)_2$				$[Zn(H_2O)_4(MeNH_2)_2]^{2+}\bullet(H_2O)_2$			
G_{corr} = 0.2337067				G_{corr} = 0.236968			
$E_{DLPNO-CCSD(T)}$ = –6239.930246				$E_{DLPNO-CCSD(T)}$ = –2444.320513			
Cd	-0.285767	-0.037425	-0.074939	Zn	-0.299771	-0.251914	-0.103341
O	-1.396936	1.023844	1.912391	O	-0.987935	1.283841	1.592810
N	-2.065254	0.196023	-1.400529	N	-2.127952	-0.073339	-1.029109
O	0.472178	1.710804	-1.848941	O	0.266255	1.271406	-1.850668

Table 15 (cont'd)

O	-0.797706	-1.693477	1.612617	O	-0.762290	-1.556087	1.588866
H	-1.864902	1.829732	2.153744	H	-1.358460	0.790842	2.332787
C	-3.418493	0.437382	-0.862836	C	-3.392096	-0.304402	-0.300050
H	1.020360	1.404763	-2.579574	H	0.902650	0.930121	-2.487070
H	-0.026433	-1.979550	2.148064	H	-0.007047	-1.908945	2.110514
H	-1.372335	-2.459010	1.519870	H	-1.475938	-2.198064	1.651441
H	-2.098728	-0.604751	-2.032779	H	-2.111042	-0.661078	-1.863960
H	-1.742976	0.320564	2.471924	H	-1.322172	2.183726	1.672423
H	0.565814	2.669855	-1.842877	H	0.311287	2.231046	-1.924609
N	1.523938	0.511085	1.125280	N	1.523956	0.140811	0.756824
H	1.692987	-0.259530	1.773450	H	1.705079	-0.582197	1.454109
H	1.186377	1.273874	1.706733	H	1.358231	0.985110	1.299251
C	2.807718	0.899015	0.499893	C	2.740411	0.312777	-0.061463
H	3.214758	0.054003	-0.053415	H	2.863357	-0.541503	-0.724878
H	2.650110	1.721581	-0.192949	H	2.651403	1.212429	-0.665396
H	3.533329	1.208787	1.253054	H	3.626184	0.404507	0.568149
O	0.344039	-1.669702	-1.729403	O	0.349304	-1.701406	-1.558373
H	-0.350958	-2.058498	-2.298457	H	0.922025	-2.465660	-1.453092
H	1.142993	-2.188009	-1.860109	H	-0.192075	-1.826174	-2.369181
H	-4.147297	0.535518	-1.668760	H	-4.255233	-0.092510	-0.932565
H	-3.423984	1.350627	-0.272028	H	-3.434988	0.337723	0.576626
H	-3.713832	-0.391521	-0.222584	H	-3.448507	-1.344056	0.021773
H	-1.784592	0.982395	-1.981824	H	-2.131134	0.872507	-1.401264
O	1.566983	-2.017913	2.948358	O	-1.438690	-1.621240	-3.576422
O	-1.924206	-2.321834	-3.156843	H	-1.889975	-2.414848	-3.888284
H	-1.930798	-2.204387	-4.114011	H	-1.269020	-1.101788	-4.370520
H	-2.445347	-3.116148	-2.996320	O	1.582746	-2.106969	2.848377
H	2.167784	-2.758087	2.808031	H	1.671947	-1.836689	3.770185
H	1.581813	-1.857415	3.899342	H	2.079882	-2.929514	2.775502
$[Cd(H_2O)_4(en)]^{2+}$				$[Zn(H_2O)_4(en)]^{2+}$			
$G_{corr} = 0.170176$				$G_{corr} = 0.174694$			
$E_{DLPNO\text{-}CCSD(T)} = -6085.943403$				$E_{DLPNO\text{-}CCSD(T)} = -2290.333389$			
Cd	0.289071	0.176861	-0.017453	Zn	0.247997	0.035881	0.004896
N	-1.538632	0.313660	-1.414735	N	-1.332834	0.257093	-1.381176
N	-1.462340	-0.206503	1.450255	N	-1.316811	-0.243424	1.393953
C	-2.760163	0.320996	-0.579496	C	-2.582941	0.412923	-0.599548
C	-2.612052	-0.603851	0.611449	C	-2.547949	-0.495049	0.609895
H	-1.532300	1.121017	-2.030279	H	-1.227241	1.036966	-2.023428

Table 15 (cont'd)

H	-1.273992	-0.933056	2.133847	H	-1.184107	-0.995399	2.063442
H	-2.929027	1.344912	-0.237228	H	-2.654941	1.456100	-0.281509
H	-3.542647	-0.600323	1.183135	H	-3.445957	-0.345180	1.212087
H	-1.554406	-0.490405	-2.036356	H	-1.406062	-0.560984	-1.980331
H	-1.696794	0.624248	1.987691	H	-1.429447	0.602223	1.947820
O	1.973957	-0.313228	1.548816	O	1.896844	-0.289102	1.385675
O	2.079488	0.315836	-1.520032	O	1.905769	0.318551	-1.342218
O	0.853976	2.411657	0.560379	O	0.546753	2.120174	0.578835
O	0.628422	-2.168486	-0.555321	O	0.619379	-2.070909	-0.501947
H	2.908034	-0.346436	1.312725	H	2.803920	-0.227612	1.064660
H	2.502345	1.100399	-1.884546	H	2.223516	1.151374	-1.708072
H	0.505383	3.259518	0.265091	H	0.162394	2.947953	0.270141
H	0.198814	-2.778303	-1.165085	H	0.130433	-2.704381	-1.039101
H	2.375720	-0.437216	-2.042795	H	2.183268	-0.382141	-1.943220
H	1.149496	-2.719556	0.039884	H	1.193474	-2.592389	0.071220
H	1.531390	2.605774	1.216753	H	1.225625	2.351283	1.222210
H	1.926884	-0.425401	2.504809	H	1.937145	-0.470976	2.331198
H	-3.635596	0.023291	-1.160258	H	-3.464409	0.194961	-1.204977
H	-2.439004	-1.628973	0.273582	H	-2.537790	-1.541502	0.294325
$[Cd(H_2O)_4(en)]^{2+}\bullet(H_2O)_2$				$[Zn(H_2O)_4(en)]^{2+}\bullet(H_2O)_2$			
G_{corr} = 0.21662147				G_{corr} = 0.21823073			
$E_{DLPNO\text{-}CCSD(T)}$ = – 6238.742148				$E_{DLPNO\text{-}CCSD(T)}$ = –2443.134320			
Cd	0.779507	0.172443	0.009310	Zn	0.625292	0.233334	-0.022265
N	-1.184668	0.732791	-1.077424	N	-1.030169	0.591916	-1.300425
N	-0.892502	-0.569767	1.498909	N	-0.869932	-0.281026	1.392836
C	-2.292616	0.659122	-0.110995	C	-2.240413	0.572949	-0.449815
C	-2.166774	-0.579632	0.755237	C	-2.110373	-0.495656	0.615793
H	-1.122342	1.676571	-1.454530	H	-0.960832	1.492411	-1.767589
H	-0.676805	-1.506674	1.831656	H	-0.636676	-1.118750	1.918146
H	-2.251477	1.553231	0.517327	H	-2.334523	1.554169	0.021725
H	-3.024959	-0.643186	1.428439	H	-2.995642	-0.492257	1.255211
H	-1.359804	0.119651	-1.867534	H	-1.116040	-0.099223	-2.039605
H	-0.976293	0.014662	2.324957	H	-1.014847	0.456148	2.076901
O	2.428460	0.534091	1.644266	O	2.260601	0.167446	1.425245
O	1.943238	0.351039	-2.036645	O	2.170294	0.371380	-1.541991
O	1.372642	2.502868	-0.316861	O	0.793537	2.316734	0.312683
O	1.656668	-2.000039	0.090638	O	0.997741	-1.878405	-0.370320

Table 15 (cont'd)

H	3.000409	-0.196956	1.900969	H	3.161125	0.024863	1.115766
H	2.304771	1.238407	-2.139781	H	2.494803	1.165756	-1.977239
H	0.683254	3.135020	-0.620071	H	0.319314	3.032404	-0.175287
H	2.297567	-2.403034	-0.500707	H	0.808945	-2.392075	-1.160964
H	2.402544	-0.224541	-2.654914	H	2.624462	-0.381213	-1.933278
H	1.181781	-2.723212	0.564684	H	0.974998	-2.509426	0.390271
H	2.043700	3.009582	0.149211	H	1.436357	2.706985	0.910124
H	2.509424	1.203173	2.331309	H	2.246183	-0.048060	2.362784
H	-3.263919	0.659788	-0.611250	H	-3.145314	0.405795	-1.037705
H	-2.185265	-1.472772	0.125286	H	-2.047167	-1.481279	0.148414
O	0.151381	-3.737655	1.435974	O	0.808515	-3.417223	1.813009
H	0.506025	-4.215527	2.194265	H	1.592190	-3.782740	2.237729
H	-0.389911	-4.378225	0.961486	H	0.144634	-4.112870	1.878936
O	-0.741661	3.878412	-1.300646	O	-0.623546	3.968187	-1.195171
H	-0.645026	4.367589	-2.125996	H	-0.213987	4.510367	-1.878470
H	-1.287076	4.442812	-0.740781	H	-1.319256	4.519387	-0.819664

APPENDIX B: Figures

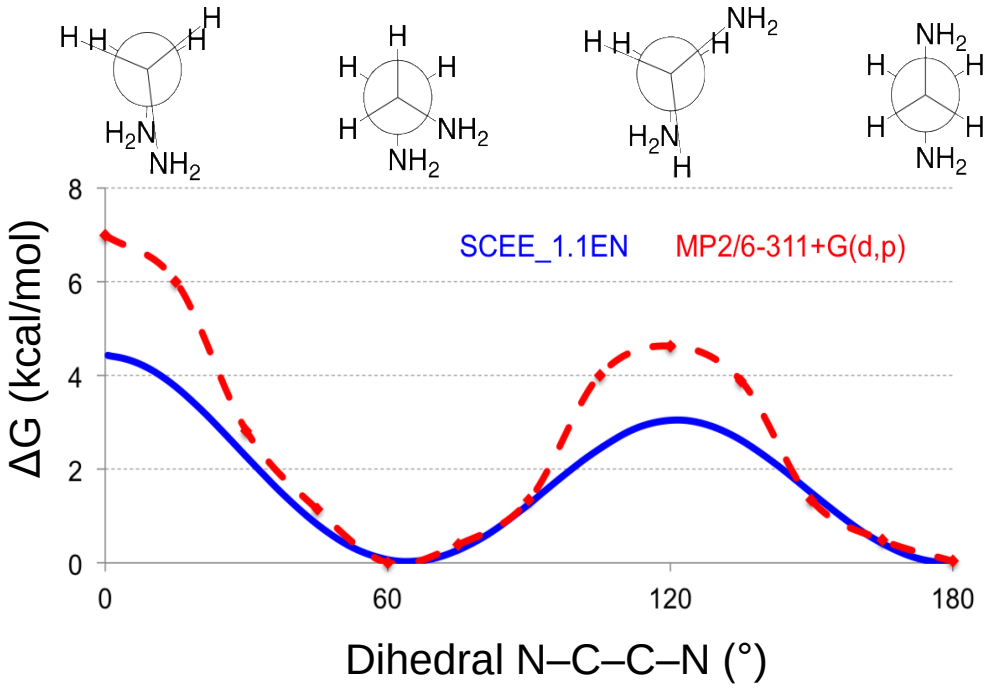

Figure 1: The torsion profiles for the N–C–C–N dihedral angle in en at MP2/6-311+G(d,p) and revised SCEE_1.1.

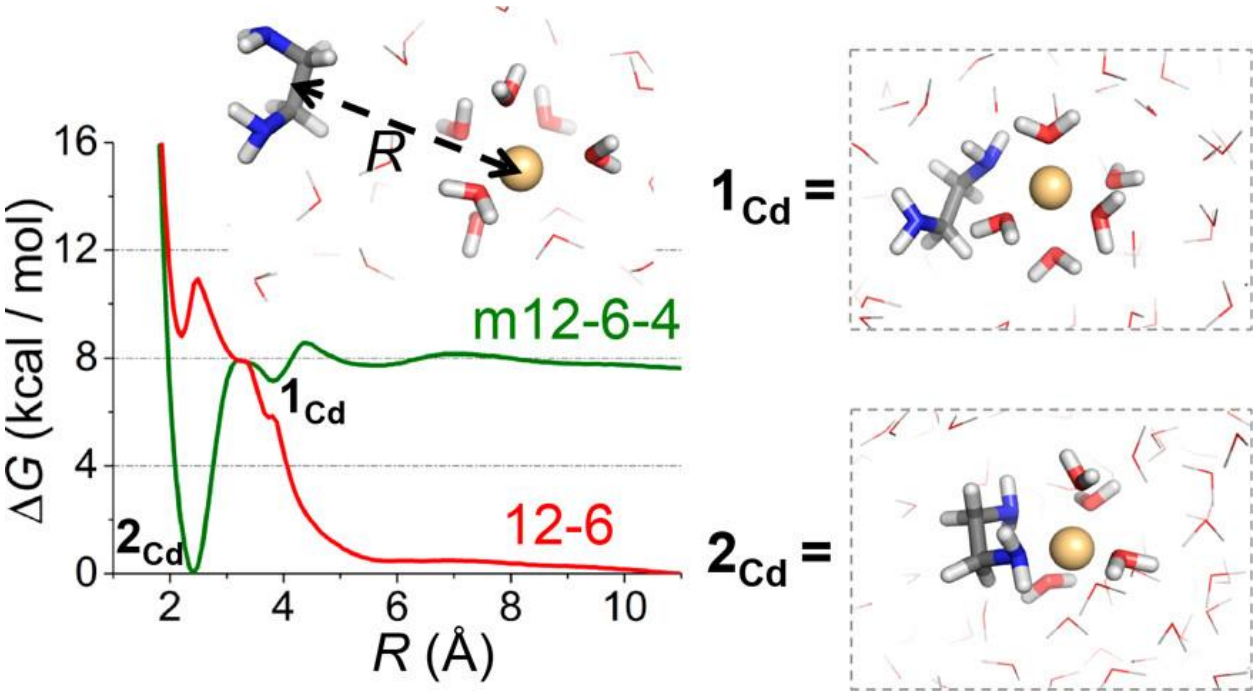

Figure 2: Comparison of PMF profiles for 12-6 (red) and m12-6-4 (green) Cd^{2+} ion parameters interacting with ethylenediamine.

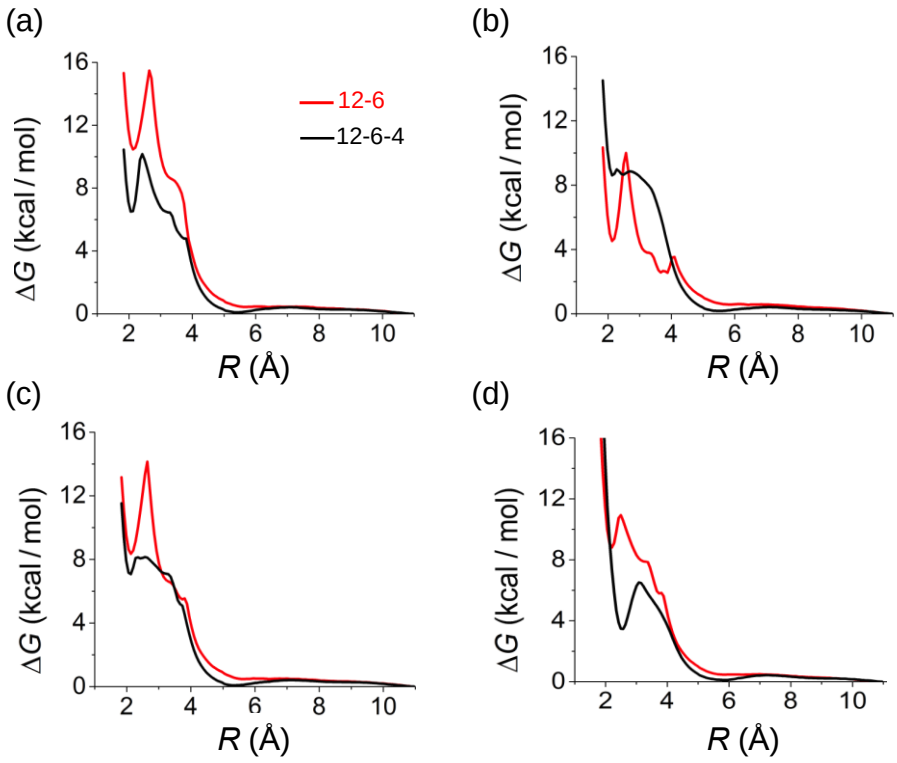

Figure 3: Comparison of binding energies of en complexes of divalent metal ions (a) Ni^{2+}, (b) Fe^{2+}, (c) Zn^{2+}, and (d) Cd^{2+} calculated using the compromise 12-6 non-bonded model and the default 12-6-4 model.

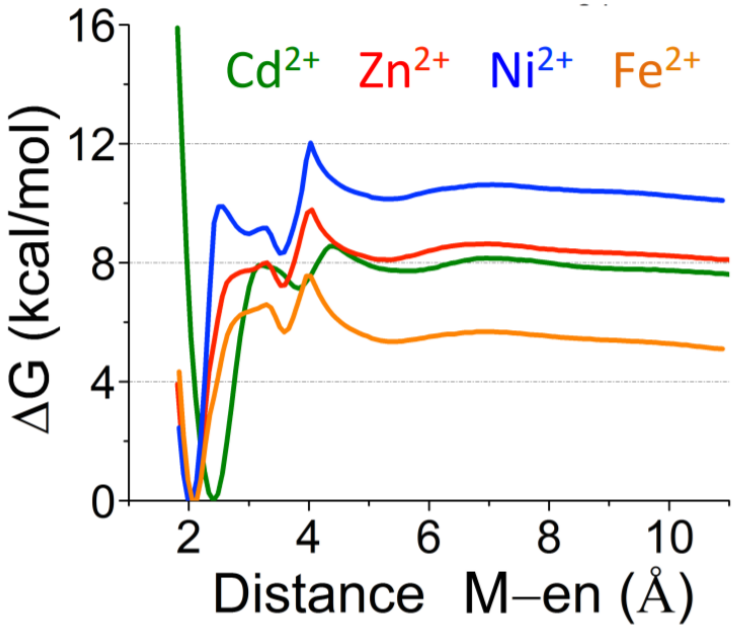

Figure 4: Potential of mean force profiles for the m12−6−4 Ni^{2+}, Fe^{2+}, Zn^{2+}, and Cd^{2+} ion parameters interacting with en.

The profiles shown here are obtained from umbrella sampling. Figure 4 represents PMF plots corresponding to the formation of mono(ethylenediamine) complexes of different divalent ions using the m12−6−4 LJ non-bonded model. The C_4 terms were optimized till the binding energies were within 0.3 kcal/mol of the experimental binding energies. The larger metal ion, Cd was found to have a longer intermolecular distance with the ligand in the chelate complex ($\mathbf{2_{Cd}}$). Among the different metal en complexes, we find the barrier heights to be the highest for Ni^{2+}. Prior studies with divalent metal ion ammonia complexes have reported the most stable complex corresponds to the Ni complex,[1,2] indicating the strongest **M**–N bond when **M** = Ni. Hence the dissociation of each Ni–N bond in the en complex was significantly higher than the other metal en complexes. As the en bound to metal ion in monodentate fashion ($\mathbf{1_M}$) approaches closer to the solvated metal ion, the unbound amine end starts interacting with one of metal bound water molecules. This intermolecular hydrogen bond between the ligand and solvent reduces the barrier height and lead to a smoother peak (barrierless process for Fe^{2+} and Zn^{2+}) corresponding to **TS2** rather than a sharp one as seen in **TS1**. The hydrogen bond (Figure 8) compensates for

the rotational penalty and the loss of configurational entropy in the ligand towards formation of the chelate complex (**2M**). This feature is more prevalent in the $[Ni(en)(H_2O)_4]^{2+}$ complex, wherein the bridged hydrogen bond structure corresponds to a minimum before chelate ring closure. Moreover, we observe a sharper peak for **TS2** in the $[Ni(en)(H_2O)_4]^{2+}$ complex relative to the other metal ion complexes owing to the stronger Ni–N bond.

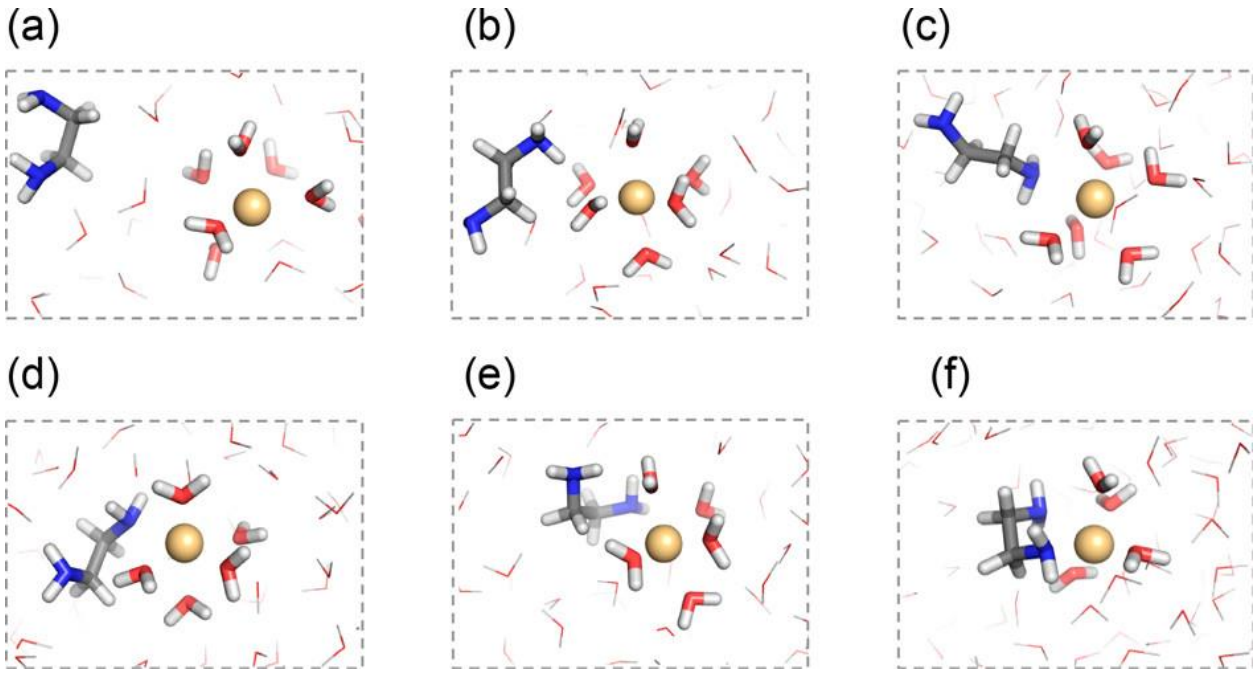

Figure 5: Snapshots at (a) intermolecular distance R = 8.00 Å, (b) R = 4.80 Å, (c) R = 4.35 Å (TS1), (d) R = 3.80 Å (1Cd), (e) R = 3.23 Å (TS2), and (f) 2.40 Å (2Cd) in the formation of $[Cd(en)(H_2O)_4]2+$. R is the intermolecular distance between the center of mass of the ligand and ion.

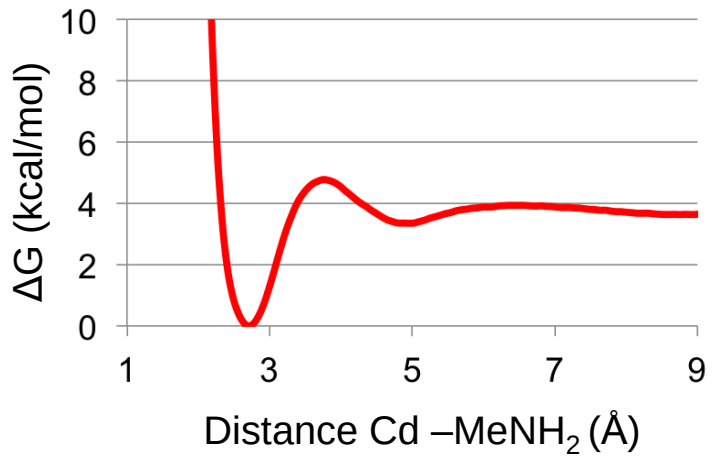

Figure 6: PMF plot for the formation of $[Cd(MeNH_2)(H_2O)_5]^{2+}$ obtained with pairwise parameters optimized to reproduce the experimental binding energies.

The optimized alpha values to reproduce the experimental binding energies of en complex of Ni^{2+}, Fe^{2+}, Zn^{2+}, and Cd^{2+} were used to derive the binding energies of methylamine complexes. Among the different metal complexes, only experimental binding energy of $[Cd(MeNH_2)(H_2O)_5]^{2+}$ is known and we find the calculated binding energy (1.2 kcal/mol) is underestimated compared to experimental binding energy (3.7 kcal/mol).[56] Furthermore, the calculations predict the thermodynamic unstability of Fe^{2+} $MeNH_2$ complex (Table S9), with the solvent separated ion ligand pair to be more stable. We should note here that the alpha was optimized to reproduce the en complexation energies. In contrast to the translational motion involved in methylamine binding, additional terms corresponding to a rotational penalty and loss of configurational entropy are involved in en binding. Hence, the underestimations in the binding energies of methylamine binding using alpha parameters of en binding are justified. While the transferability of alpha parameters across different ligands is a subject of our future investigations, in such cases where the binding processes of different ligands differs with each other we recommend optimization of alpha parameter for each ligand.

The optimized alpha value of 3.16 was used to derive m12-6-4 pairwise parameters for Cd^{2+} and $MeNH_2$ ligand to reproduce the experimental binding energy of $[Cd(MeNH_2)(H_2O)_5]^{2+}$. The same parameter was used to model bis(methylamine) complex of Cd^{2+}. A bound state of Cd^{2+}

with two methylamine ligands was considered as the starting point and using a step size of 0.05 Å each of the methylamine was moved from 1.8 Å to 5.0 Å away from the metal ion. Energies at each of the points were calculated to derive a 2D PMF plot of bis(methylamine) binding (Figure 5b). The reaction coordinate *R* was between the methylamine nitrogen and the metal ion. Subsequently for the en complex, distances between each nitrogen and Cd^{2+} were varied from 1.8 Å to 5.0 Å and energies at each point were calculated. We find in en complex, with one nitrogen end already bound to metal ion, the ion can recognize the second nitrogen end from a farther distance than in bis(methylamine) complex. The intermolecular distances between the ligands and Cd^{2+} for transition states leading to chelate ring closure (Table S3) and formation of bis(methylamine) complex from monomethylamine complex (Table S10) are 3.30Å and 3.00Å respectively. The thermodynamic stability of the chelate complex over the bis(methylamine) complex and the presence of less solvent molecules between the metal ion and second nitrogen atom contributes towards this difference.

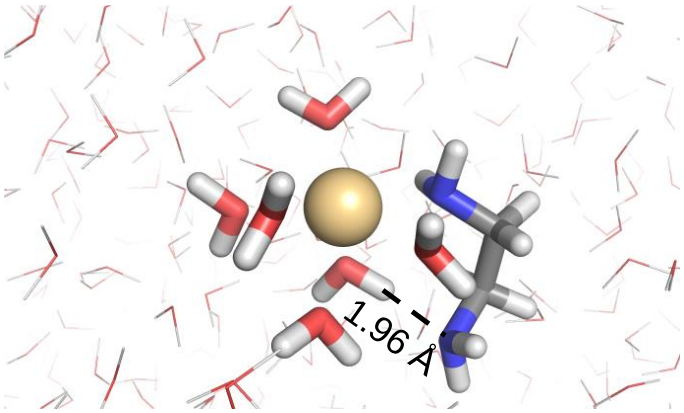

Figure 7: One of the snapshot at R = 3.30 Å (**TS2**) in formation of $[Cd(en)(H_2O)_4]^{2+}$ representing the intermolecular hydrogen bonding between one of bound water molecules and the unbound nitrogen end of en, reducing the barrier height for chelate closure.

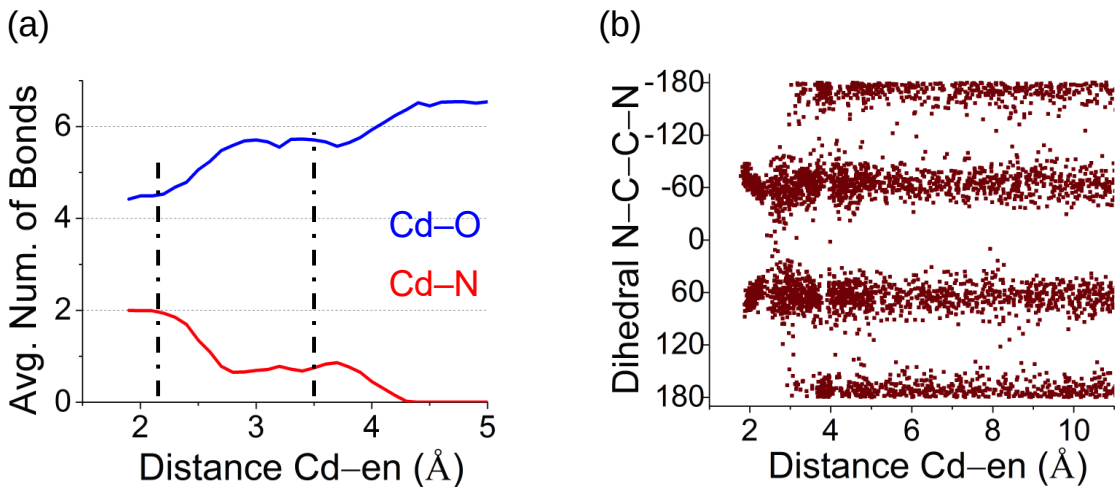

Figure 8: Variation of (a) average of number of bonds (with bond lengths <2.40 Å) and (b) dihedral N–C–C–N (°) during the formation of en complex of Cd^{2+}. Note the conformational flexibility in the ligand increases with the increase in Cd–N bond length.

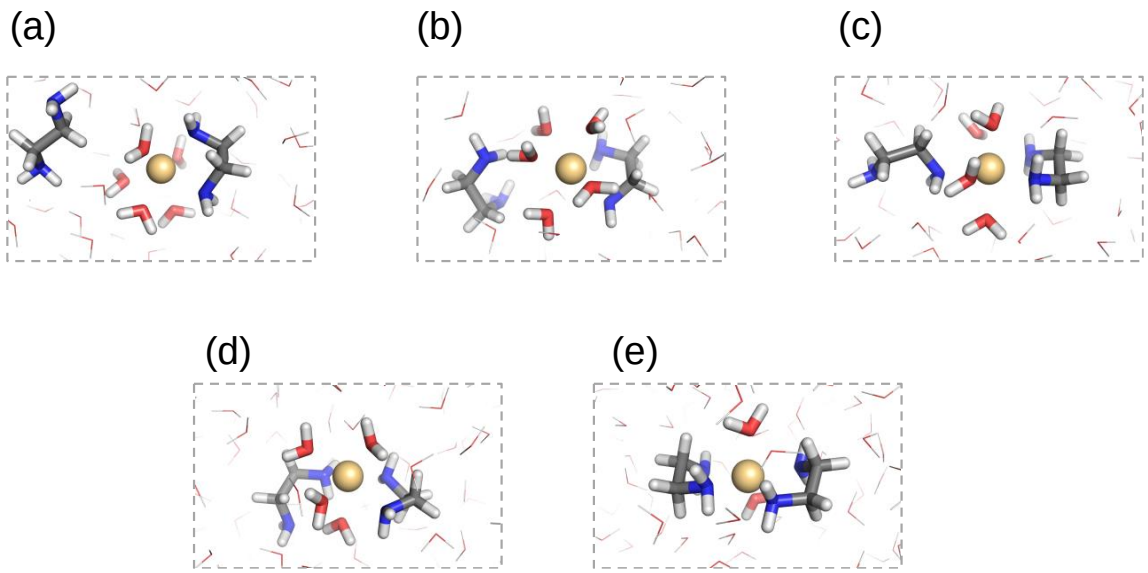

Figure 9: Snapshots of various stationary points (a) R = 5.50 Å (c) R = 4.35 Å (**TS3**) (d) R = 3.80 Å, (e) R = 3.25 Å (**TS4**) , and (f) 2.40 Å in the formation of $[Cd(en)_2(H_2O)_2]^{2+}$. R is the intermolecular distance between the center of mass of the ligand and ion.

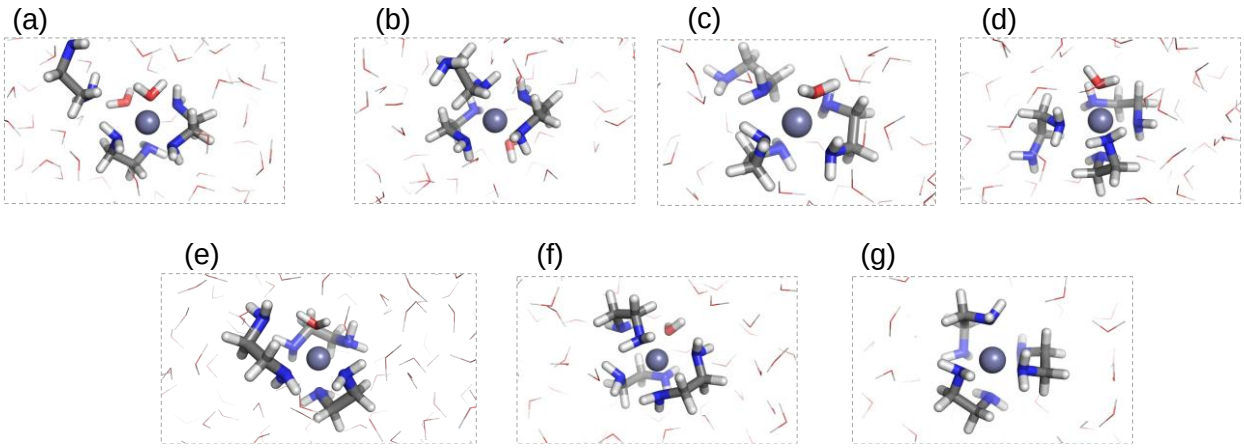

Figure 10: Snapshots at (a) intermolecular distance R = 5.10 Å (b) R = 3.80 Å, (**TS5**) (c) R = 3.60 Å (**5M**), (d) R = 3.30 Å (**TS6**), (e) R = 2.95 Å (**6M**), and (f) 2.55 Å (**TS7**), (g) 2.11 Å (**7M**) in the formation of $[Zn(en)_3]^{2+}$. R is the intermolecular distance between the center of mass of the ligand and ion.

The pairwise parameters corresponding to the ligand en binding to each of the four respective metal ions were used in the reaction of bis(en) metal complex with en towards formation of tris(en) complex. The first minimum (**7M**) at 2.11 Å corresponds to the chelate complex where both the nitrogen ends of the en ligand are bound to the metal ion. The second minimum at 2.95 Å (**6M**) corresponds to structure where one of the en nitrogen atoms engages in intermolecular hydrogen bonding with one of amine hydrogens of a bound en. The third minimum (**5M**) corresponds to the intermediate where one of the nitrogen atoms of the outgoing en ligand is still bound to the metal ion. Figure 10 represent snapshots of various minima and transition states along the reaction coordinate. The additional minimum at 2.95 Å (**6M**) is only seen in the formation of the tris(en) complexes of Zn^{2+} and Ni^{2+} . With Cd^{2+}, as a larger metal ion, the intermolecular distances between the metal ion and the nitrogen of en is 2.40 Å compared to ~2.10 Å for the other three metal ions. Hence, the en ligands gets farther away from each other in the formation of $[Cd(en)_3]^{2+}$ and fails to interact with each other to stabilize the transition state leading to the chelate ring closure (**TS7** in Figure 10b). We find that the calculated reaction energy for the formation of the $[Cd(en)_3]^{2+}$ complex from $[Cd(en)_2(H_2O)_2]^{2+}$ and en is -4.0 kcal/mol compared to the experimental value of –2.8 kcal/mol. Furthermore, we find excellent

agreement between the calculated and experimental structural features of the tris(en) complexes of metal ions as shown in Table 8.

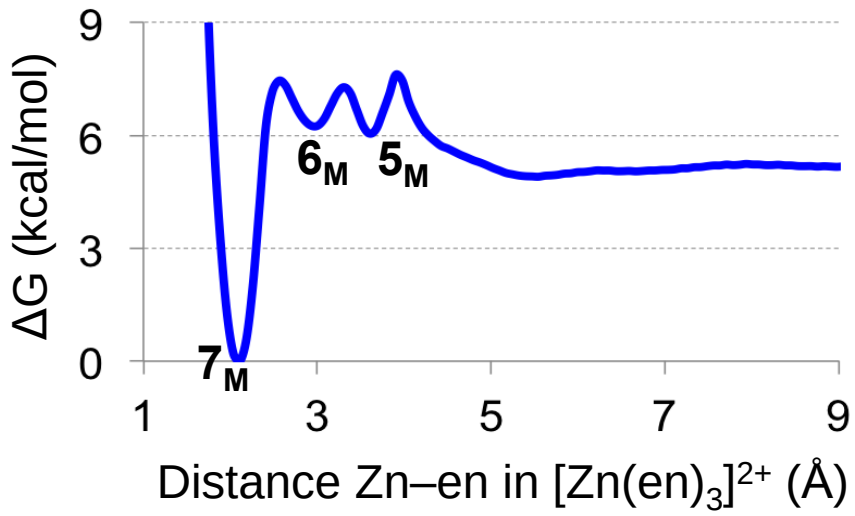

Figure 11: PMF plot for the reaction of $[Zn(en)_2(H_2O)_2]^{2+}$ with en towards formation of $[Zn(en)_3]^{2+}$.

Note the minimum at 2.95 Å corresponds to structure where one of the en amine engages in intermolecular hydrogen bonding with one of bound en. Such feature is observed only in the formation of Zn^{2+} and Ni^{2+} complexes of tris(en).

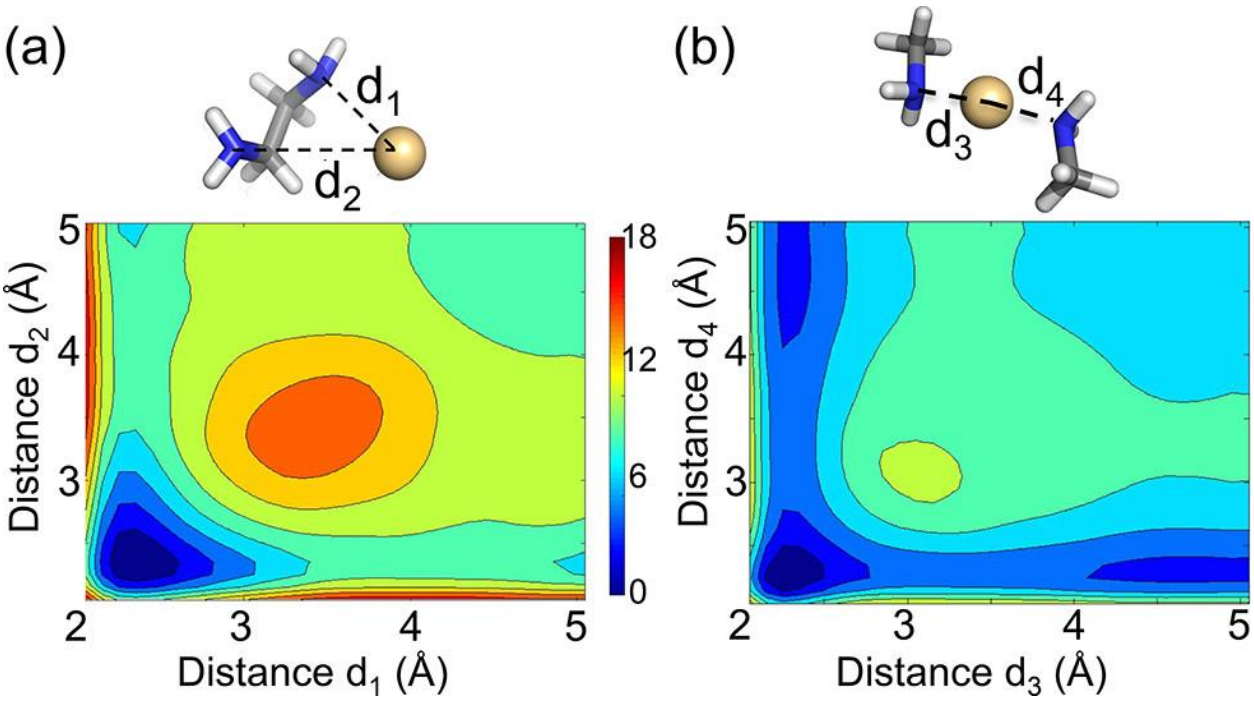

Figure 12: Comparison of 2D PMF contour plots toward formation of (a) en and (b) bis(methylamine) complexes of Cd^{2+}.

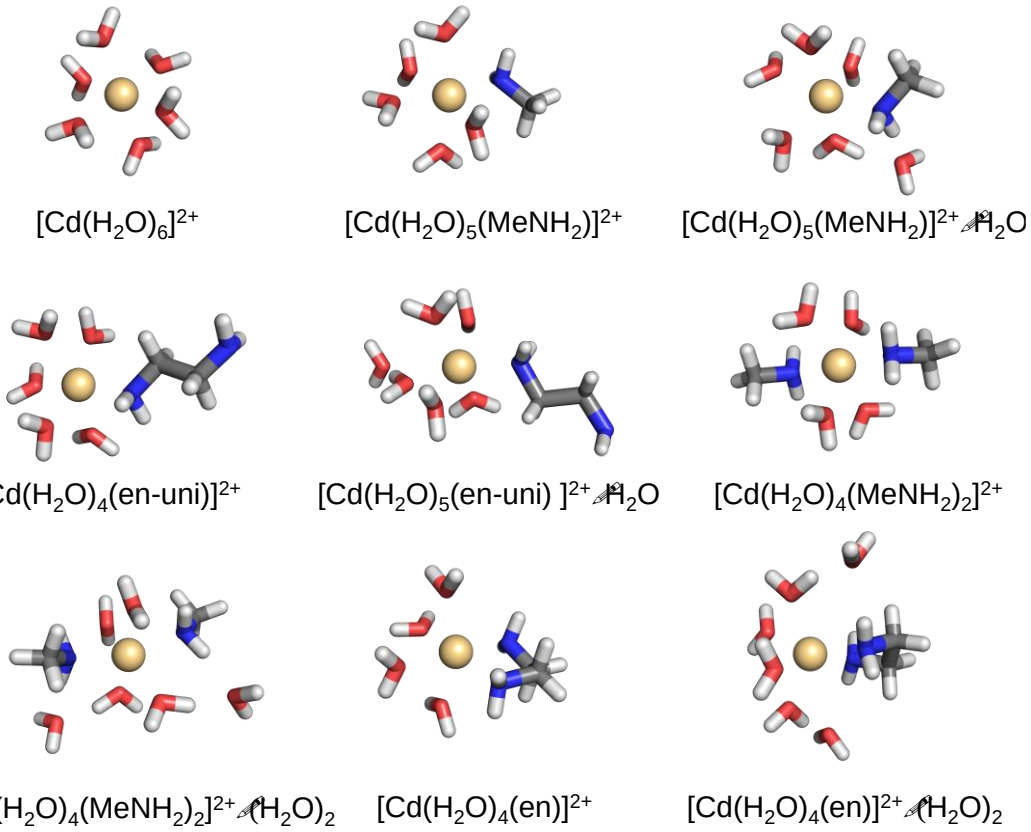

Figure 13. Optimized geometries of different species involved in the formation of mono(ethylenediamine) and bis(methylamine) complexes.

The Model A and Model B reactions for Cd(II) were used here. Geometry optimizations of all species were carried out at the PBE0-D3BJ/def2-TZVP level. All geometries were confirmed to be minima. Single-point DLPNO-CCSD(T) calculations were done with DKH-TZVP basis set using continuum solvation SMD model. The calculated free energies for the different reaction involving formation of bis(methylamine) and ethylenediamine complexes of Cd^{2+} and Zn^{2+} are listed in Table 11 and Table 13 respectively. The calculations are carried out with limited conformers so they exclude the true conformational contributions. The lack of explicit water molecules beyond first solvation shell restricts the accuracy. However, we find the presence of equal number of water molecules associated within the complexes across the reactants and products (Model B) results in energies that have better agreement with experiment. The loss of water in Model A results in large errors associated with translational entropy.

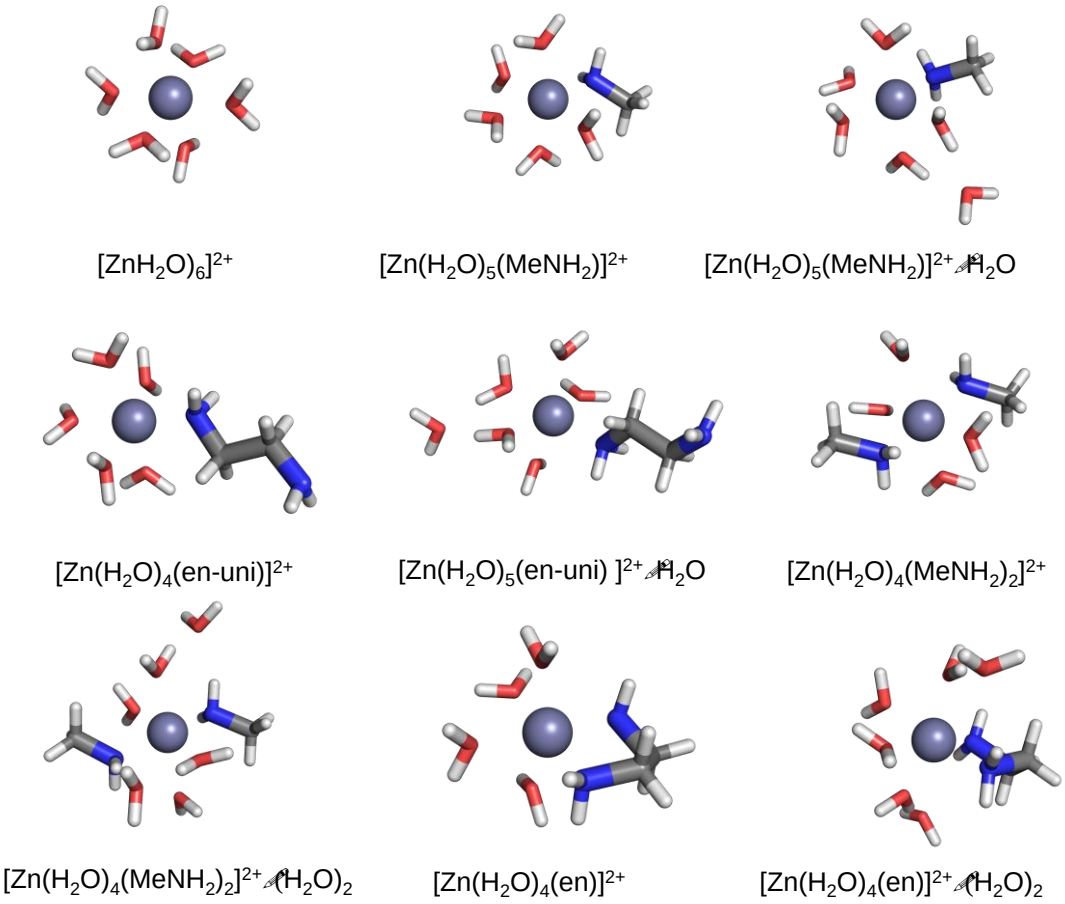

Figure 14. Optimized geometries of different species involved in the formation of mono(ethylenediamine) and bis(methylamine) complexes.

The Model A and Model B reactions for Zn(II) were used here.

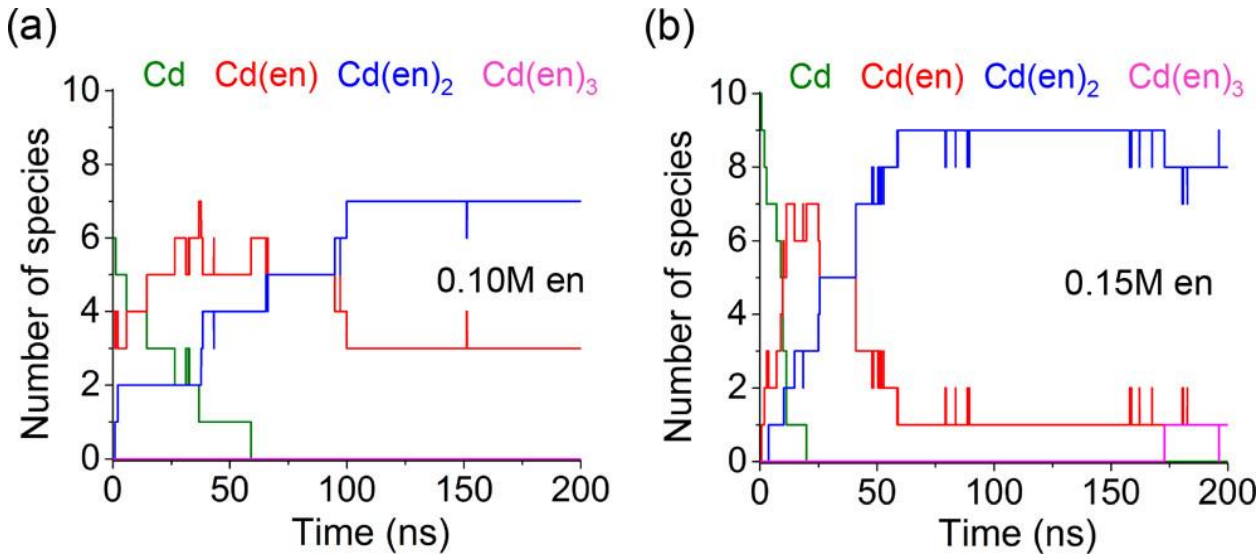

Figure 15: Variation of number of molecules of different species in real time simulation for (a) 0.10 M and (b) 0.15 M ethylenediamine (en) solutions.

The 30 en molecules were arranged in a rectangular box with 3, 5 and 2 en molecules along x, y and z axis respectively. The molecules were in a periodic arrangement with each en separated by a distance of 8.11 Å, 9.77 Å and 8.176 Å along x, y and z axis respectively. The Cd(II) ions were added in random positions by the tleap tool in AmberTool16. The systems were solvated with subsequent additions of chloride ions to neutralize the system.

For all simulations, periodic boundary conditions (PBCs) were employed together with PME to model long-range interactions with a 12 Å cut off. The ligand molecules, metal ions and counterions Cl^- were solvated in cubic truncated octahedron with total volume of 489.1$Å^3$ and 334.0$Å^3$ corresponding to form the 0.1M and 0.15M en solutions. In total, there were 14,431 and 9600 water molecules respectively in 0.1M and 0.15M en solutions with the same number of en ligands and metal ions. Chloride ion counterions were also treated with the 12-6-4 LJ nonbonded model. The compromise 12-6 LJ model was used as the representative 12-6 nonbonded model representing a good balance between structure and thermodynamics. We performed minimizations involving 12000 steps of steepest descent followed by 8000 steps of conjugate gradient. A 125ps NPT heating procedure was performed to heat the system from 0 K to 300 K followed by a 25 ps equilibration at 300K with constant NPT conditions using a Langevin

thermostat at 1atm. The equilibrated geometries were read for the NPT production simulation runs each of 200ns. An integration time step of 1 fs was used in the heating step and 2 fs in the production studies. Langevin dynamics temperature control was employed in the heating and the production runs with a collision rate equal to 1.0 ps. SHAKE was utilized for the water molecules for all simulations.

APPENDIX C: Copyright Notice

REFERENCES

REFERENCES

1. Sargsyan, K.; Chen, T.; Grauffel, C.; Lim, C. *Identifying COVID-19 Drug-Sites Susceptible To Clinically Safe Zn-ejector Drugs Using Evolutionary/Physical Principles* Int J Mol Med. 2020, [Epub ahead of print]

2. Li, P.; Merz Jr., K. M. *Metal Ion Modeling Using Classical Mechanics* Chem. Rev., 2017, 117, pp 1564–1686

3. Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*, Oxford Clarendon Pr. 1987.

4. Verlet, Loup *Computer "Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules,* Phys. Rev. 1987, 159, 98-103.

5. Elimelech, Menachem; Jia, Xiadong *Particle Deposition and Aggregation: Measurement, Modelling and Simulation,* Butterworth-Heinemann 2013.

6. Frenkel, Daan; Smit, Berend *Understanding Molecular Simulation,* Academic Pr. 2001.

7. Beeman, D. *Some Multistep Methods for Use in Molecular Dynamics Calculations*, Journal of Computational Physics 1973, 20, 130-139.

8. Ryckaert, Jean-Paul; Ciccotti, Giovanni; Berendsen, Herman JC *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes,* Journal of Computational Physics 1977, 23, 327-341.

9. Dubbeldam, David; Oxford, Gloria A. E.; Krishna, Rajamani; Broadbelt, Linda J.; Snurr, Randall Q. *Distance and angular holonomic constraints in molecular simulations,* The Journal of Chemical Physics 2010, 133, 034114.

10. Andersen, Hans C. Rattle *: A "Velocity" Version of the Shake Algorithm for Molecular Dynamics Calculations*, Journal of Computational Physics 1983, 52, 24-34.

11. Steinhauser MO, Hiermaier S. *A Review of Computational Methods in Materials Science: Examples from Shock-Wave and Polymer Physics*, International Journal of Molecular Sciences 2009, 10, 5135-5216.

12. Thijssen, J.M. *Computational Physics*, Cambridge University Press 2007.

13. Roux, B.; Simonson, T. *Implicit solvent models*, Biophysical Chemistry 1999, 78 1-20.

14. Allen, M.P. *Introduction to Molecular Dynamics Simulation*, 2004.

15. Kittel, C. *Introduction to Solid State Physics*, Wiley 2004.

16. Jensen, F. *Introduction to Computational Chemistry*, John Wiley & Sons 2006.

17. Li, P.; Merz Jr., K. M. *Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions* J. Chem. Theory Comput., 2014, 10 (1), pp 289–297

18. J. N. Israelachvili, *Intermolecular and Surface Forces*, Academic Press, San Diego, 3rd edn, 2011, pp. 23–51

19. Marc Souaille, Benoît Roux, *Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations*, Computer Physics Communications, Volume 135, 2001, Pages 40-57

20. Kästner, J. *Umbrella Sampling*. WIREs Comput Mol Sci, 1, 2011, 932-942.

21. Kästner, J., Senn, H. M., Thiel, S., Otte, N., Thiel, W. *QM/MM Free-Energy Perturbation Compared to Thermodynamic Integration and Umbrella Sampling: Application to an Enzymatic Reaction* J. Chem. Theory Comput. 2006, 2, 452-461

22. Kirkwood, J. *Statistical Mechanics of Fluid Mixtures*, The Journal of Chemical Physics 3:5, 1935, 300-313

23. Shivakumar, D., Harder, E., Damm, W., Friesner, R. A., Sherman, W. *Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field* J. Chem. Theory Comput. 2014, 10, 3570-3577

24. Li, P., Song, L. F., Merz Jr., K. M., *Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJ-Type Nonbonded Model in Explicit Water* J Phys Chem B. 119, 2015, 883–895.

25. Jensen, J. H. *Predicting Accurate Absolute Binding Energies in Aqueous Solution: Thermodynamic Considerations for Electronic Structure Methods* Phys. Chem. Chem. Phys. 2015, 17, 12441−12451.

26. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar,

S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.

27. Werner, A. Z. *Beitrag zur Konstitution Anorganischer Verbindungen* Anorg. Allg. Chem. 1893, 267−330.

28. Werner, A. *Ueber Acetylacetonverbindungen des Platins.* Ber. Dtsch. Chem. Ges. 1901, 34, 2584−2593.

29. For a nice review of history of chelate effect, see: (a) Diehl, H. *The Chelate Rings.* Chem. Rev. 1937, 21, 39−111. (b) Martell, A. E.; Hancock, R. D. *Theories of Coordination Compounds. In Coordination Chemistry A century of Progess*; Kauffman, G. B., Ed.; ACS Symposium Series 565; American Chemical Society: Washington, DC, 1993; pp 240−254.

30. Morgan, G. T.; Drew, H. D. K. *Researches on Residual Affinity and Coordination. Part II. Acetylacetones of Selenium and Tellurium*. J. Chem. Soc., Trans. 1920, 117, 1456−1465.

31. Schwarzenbach, G. *Der Chelateffekt*. Helv. Chim. Acta 1952, 35, 2344−2359.

32. Flora, S. J. S.; Pachauri, V. *Chelation in Metal Intoxication*. Int. J. Environ. Res. Public Health 2010, 7, 2745−2788.

33. Chen, C. H.; Shi, J. *Metal Chelates as Emitting Materials for Organic Electroluminescence* Coord. Chem. Rev. 1998, 171, 161−174.

34. Cook, T. R.; Zheng, Y.-R.; Stang, P. *Metal-organic Frameworks and Self-assembled Supramolecular Coordination Complexes: Comparing and Contrasting the Design, Synthesis, and Functionality of Metal-Organic Materials* Chem. Rev. 2013, 113, 734−777.

35. Carter, M. J.; Beattie, J. K. *The Kinetic Chelate Effect. Chelation of Ethylenediamine on Platinum(II)* Inorg. Chem. 1970, 9, 1233−1238.

36. Bates, G. W.; Billups, C.; Saltman, P. *The Kinetics and Mechanism, of Iron(II 1) Exchange between Chelates and Transferrin I. The Complexes of Citrate And Nitrilotriacetic Acid* J. Bio. Chem. 1967, 242, 2810−2815.

37. Vallet, V.; Wahlgren, U.; Grenthe, I. *Chelate Effect and Thermodynamics of Metal Complex Formation in Solution: A Quantum Chemical Study* J. Am. Chem. Soc. 2003, 125, 14941−14950.

38. Basdogan, Y.; Keith, J. A. *A Paramedic Treatment For Modeling Explicitly Solvated Chemical Reaction Mechanisms* Chem. Sci. 2018, 9, 5341−5346.

39. Bühl, M.; Grenthe, *I. Binding Modes of Oxalate in UO2(oxalate) in Aqueous Solution Studied with First-Principles Molecular Dynamics Simulations. Implications for the chelate effect* Dalton Trans. 2011, 40, 11192−11199.

40. Li, P.; Song, L. F.; Merz Jr., K. M. *Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model* J. Chem. Theory Comput. 2015, 11, 1645−1657.

41. Roux, B.; Karplus, M. *Potential Energy Function for Cation−Peptide Interactions: An Ab Intio Study* J. Comput. Chem. 1995, 16, 690−705.

42. Minoux, H.; Chipot, C. *Cation-π Interactions in Proteins: Can Simple Models Provide an Accurate Description?* J. Am. Chem. Soc. 1999, 121, 10366−10372.

43. Lybrand, T. P.; Kollman, P. A. *Water-water and Water-ion Potential Functions Including Terms for Many Body Effects* J. Chem. Phys. 1985, 83, 2923−2933.

44. Curtiss, L. A.; Jurgens, R. *Nonadditivity of Interaction in Hydrated Copper(1+) and Copper(2+) Clusters* J. Phys. Chem. 1990, 94, 5509.

45. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. *Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy* J. Chem. Theory Comput. 2007, 3, 1960−1986.

46. Hermida-Ramon, J. M.; Brdarski, S.; Karlstrom, G.; Berg, U. *Inter- and Intramolecular Potential for the N-Formylglycinamide Water System. A Comparison between Theoretical Modeling and Empirical Force Fields* J. Comput. Chem. 2003, 24, 161−76.

47. Sakharov, D. V.; Lim, C. *Force Fields Including Charge Transfer and Local Polarization Effects: Application to Proteins Containing Multi/Heavy Metal Ions* J. Comput. Chem. 2009, 30, 191−202.

48. Panteva, M. T.; Giambasu, G. M.; York, D. M. *Comparison of¸ Structural, Thermodynamic, Kinetic and Mass Transport Properties of Mg2+ Ion Models Commonly used in Biomolecular Simulations* J. Comput. Chem. 2015, 36, 970−982.

49. 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.

50. Lee, T. -S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A. Case, D. A. Walker, R. C. and York D. M. *GPU-Accelerated Molecular Dynamics*

and Free Energy Methods in Amber18: Performance Enhancements and New Features J. Chem. Inf. Model, 2018, 58, 2043-2050.

51. Case, D. A.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, III, T. E.; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Ghoreishi, D.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Greene, D.; Harris, R.; Homeyer, N.; Izadi, S.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Mermelstein, D. J.; Merz, K. M.; Miao, Y.; Monard, G.; Nguyen, C.; Nguyen, H.; Omelyan, I.; Onufriev, A.; Pan, F. Qi, R.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Smith, J.; Salomon-Ferrer, R.; Swails, J.; Walker, R. C.; Wang, J;. Wei, H.; Wolf, R. M.; Wu, X.; Xiao, L.; York, D. M. and Kollman, P. A. (2018) AMBER 2018, University of California, San Francisco.

52. Hancock, R. D. and Bartolott, L. J. *Density Functional Theory-Based Prediction of the Formation Constants of Complexes of Ammonia in Aqueous Solution: Indications of the Role of Relativistic Effects in the Solution Chemistry of Gold (I)* Inorg. Chem. 2005, 44, 7175.

53. Martell, A. E.; Smith, R. M. *Critical Stability Constants*, Plenum Press: New York, 1974-1985; Vol. 1-6.

54. Spike, C. G. and Parry, R. W. *Thermodynamics of Chelation. II. Bond Energy Effects in Chelate Ring Formation* J. Am. Chem. Soc., 1953, 75, 2726–2729.