JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

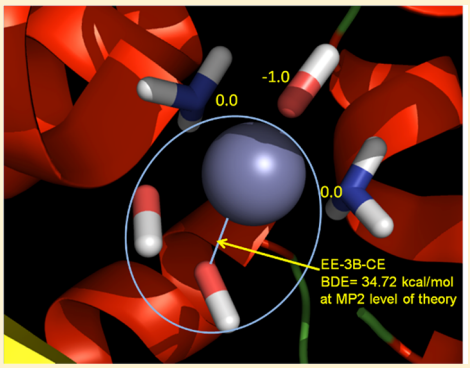
Analysis of the Errors in the Electrostatically Embedded Many-Body Expansion of the Energy and the Correlation Energy for Zn and Cd Coordination Complexes with Five and Six Ligands and Use of the Analysis to Develop a Generally Successful Fragmentation Strategy

Elbek K. Kurbanov, Hannah R. Leverentz, Donald G. Truhlar,* and Elizabeth A. Amin*

Department of Medicinal Chemistry, Department of Chemistry, and Minnesota Supercomputing Institute for Advanced Computational Research (MSI), University of Minnesota, Minneapolis, Minnesota 55414, United States

Supporting Information

ABSTRACT: In the present paper, we apply the electrostatically embedded many-body expansion of the correlation energy (EE-MB-CE) to the calculation of zinc−ligand and cadmium−ligand bond dissociation energies, and we analyze the errors due to various fragmentation schemes in a variety of neutral, positively charged, and negatively charged Zn^{2+} and Cd^{2+} coordination complexes. As a result of the analysis, we are able to present a new, simple, and unambiguous fragmentation strategy. Following this strategy, we show that both methods perform well for zinc−ligand and cadmium−ligand bond dissociation energies for all systems studied in the paper, including a model of the catalytic site of the zinc-bearing anthrax toxin lethal factor (LF), which has garnered substantial attention as a target for drug development. To draw general conclusions, we consider ten pentacoordinate and hexacoordinate zinc and cadmium containing coordination complexes, each with 10 or 15 different fragmentation schemes. By analyzing errors, we developed a prescription for the optimal fragmentation strategy. With this scheme, and using MP2 correlation energies as a test, we find that the electrostatically embedded three-body expansion of the correlation energy (EE-3B-CE) method is able to reproduce all 53 conventionally calculated bond energies with an average absolute error of only 0.59 kcal/mol. The paper also presents EE-MB-CE calculations using the CCSD(T) level of theory on an LF model system. With CCSD(T), EE-3B-CE has an average error of 0.30 kcal/mol.



1. INTRODUCTION

Enzymes containing metals (metalloenzymes) are of increasing interest and importance for drug discovery and development as their roles in a wide variety of biochemical pathways become better understood.[1,2] The largest category of metalloenzymes consists of those containing zinc, with more than 300 representatives presently known, covering all six classes of enzymes (i.e., oxidoreductases, transferases, hydrolases, lyases, isomerases, and ligases).[3] Over the past few years, zinc enzymes have been implicated in a variety of human disorders ranging from infections to cancer.[3−6] Most notably, the matrix metalloproteinases (MMPs) and the anthrax toxin lethal factor (LF), which feature a catalytic Zn^{2+} in the active site, have emerged as popular drug targets; rational design of inhibitors via molecular modeling depends on accurate rendering of the catalytic Zn^{2+} coordination environment in these enzymes. Hybrid density functional theory[7] and post-Hartree−Fock correlated wave function methods, such as second-order Møller−Plesset perturbation theory, MP2,[8] coupled cluster theory with single and double excitations, CCSD,[9,10] and CCSD with quasiperturbative connected triple excitations, CCSD(T)[11] are all able to calculate accurate energies for selected small and moderately sized systems, but they are often too computationally costly—either because the system (or model system) to be studied is too large or because a large number of calculations must be performed to achieve adequate sampling in a simulation. We note that in general the "expense" of a calculation depends on the program, the algorithm, the degree of parallelization, and the computer, and one should take account of such factors as computer time, memory requirements, communication among processors, input−output, human time, and other factors, but for discussion purposes we simply consider the number of arithmetic operations, and for convenience we call that the expense. It is well-known[12] that if one keeps the average number of basis functions per atom fixed, the expenses of, for example, hybrid density functional theory, MP2, CCSD, and CCSD(T), with conventional basis sets, nominally scale, in the large-N limit, as N^4, N^5, N^6, and N^7, respectively, where N is the number of atoms. Therefore, enabling accurate calculations on large systems including the active sites of Zn-containing enzymes with post-Hartree−Fock methods remains challenging because of the rapid scaling of the computational cost. To make the problem more tractable,

Received: March 8, 2013

dx.doi.org/10.1021/ct4001872 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

localized molecular orbitals[13−17] and fragmentation[18−36] methods have been designed. In our previous work, we developed and implemented the electrostatically embedded many-body method[37,38,40−45] (EE-MB) and the electrostatically embedded many-body expansion of the correlation energy[39,40] (EE-MB-CE), which are fragment-based approaches for calculating the energies of large systems. These methods, with pairwise additive (PA) or three-body (3B) truncation of the many-body expansion, nominally scale as N^3 or lower for EE-MB and as N^4 or lower for EE-MB-CE (where N is now the number of monomers, see below), even if the individual dimer or trimer calculations scale less favorably because the cost of the individual oligomer calculations does not increase with N.

As described in our previous work,[37,38,40−45] the EE-MB method tackles the challenge of system size by partitioning larger complexes into a series of fragments called monomers and calculating the energies of monomers, dimers, and optionally trimers or higher oligomers by embedding them in a field of point charges representing the remaining N-1, N-2, N-3, ... monomers, and running calculations in parallel. (A monomer can be defined as a single molecule, a portion of a molecule, or a collection of molecules. For example, in the systems considered here, a monomer could be an ammonia molecule or Zn^{2+} with two ammonia ligands). There are two variations of the EE-MB method: the electrostatically embedded pairwise additive method (EE-PA), which is based on the energies of monomers and dimers, and the electrostatically embedded three-body method (EE-3B), which is based on the energies of monomers, dimers, and trimers. The EE-3B method is able to predict bond energies obtained by conventional full-system calculations done at the same level of theory to within 1.0 kcal/mol for cationic, neutral, and negatively charged Zn^{2+} complexes.[44,45] In EE-MB-CE, one applies the many-body expansion only to the correlation energy, that is, to the post-Hartree−Fock part of the energy calculation. Here we apply the EE-MB-CE method to predict MP2 correlation energies for a variety of pentacoordinate and hexacoordinate Zn^{2+} and Cd^{2+} systems, and we compare its performance to the EE-MB method. Most importantly, we present a new, simple, and unambiguous fragmentation strategy that maximizes the accuracy and efficacy of both the EE-MB and the EE-MB-CE calculations for Zn^{2+} and Cd^{2+} complexes studied in this paper. The number of computational operations in the EE-PA method scales as N^2, the number for the EE-3B method scales as N^3, and the numbers for the EE-PA-CE and EE-3B-CE methods scale as N^4, where N is the number of monomers.

2. THEORY

For any level of theory (e.g., MP2 or CCSD(T) with a given basis), we can either perform full (i.e., conventional) calculations of the potential energy, E, or many-body expansions. Many-body expansion methods examined in the present study are the EE-PA approximation, the EE-3B approximation, the electrostatically embedded pairwise additive approximation of the correlation energy (EE-PA-CE), and the electrostatically embedded three-body approximation of the correlation energy (EE-3B-CE) methods. The EE-MB method approximates the energy of systems composed of monomers i, j, k, ... as

$$E = E^{(1)} + \Delta E^{(2)} + \Delta E^{(3)} + \ldots + \Delta E^{(N)} \quad (1)$$

where

$$E^{(1)} = \sum_i E_i \quad (2)$$

$$\Delta E^{(2)} = \sum_{i<j} (E_{ij} - E_i - E_j) \quad (3)$$

$$\Delta E^{(3)} = \sum_{(i<j<k)} [(E_{ijk} - E_i - E_j - E_k) - (E_{ij} - E_i - E_j) - (E_{ik} - E_i - E_k) - (E_{jk} - E_j - E_k)] \quad (4)$$

where E_i, E_{ij}, and E_{ijk}, are the energies of a monomer, dimer, and trimer, respectively, embedded in fields of point charges representing the other monomers; the EE-PA and EE-3B approximations as employed here, these energies are obtained using MP2 and CCSD(T) level theory. Then the EE-PA approximation is

$$E^{\mathrm{PA}} = E^{(1)} + \Delta E^{(2)} \quad (5)$$

and the EE-3B approximation is

$$E^{\mathrm{3B}} = E^{\mathrm{PA}} + \Delta E^{(3)} \quad (6)$$

The electronic energy at the MP2 level of theory can be written as

$$E_{\mathrm{MP2}} = E_{\mathrm{HF}}^{\mathrm{total}} + \Delta E_{\mathrm{corr,MP2}} \quad (7)$$

where $E_{\mathrm{HF}}^{\mathrm{total}}$ is the Hartree−Fock energy of the system and $\Delta E_{\mathrm{corr,MP2}}$ is the MP2 correlation energy, which can be rewritten as the many-body expansion of the correlation energy

$$\Delta E_{\mathrm{corr,MP2}} = \Delta E_{\mathrm{corr,MP2}}{}^{(1)} + \Delta E_{\mathrm{corr,MP2}}{}^{(2)} + \Delta E_{\mathrm{corr,MP2}}{}^{(3)} + \ldots + \Delta E_{\mathrm{corr,MP2}}{}^{(N)} \quad (8)$$

$$\Delta E_{\mathrm{corr,MP2}}{}^{(1)} = \sum_i \Delta E_i^{\mathrm{corr}} \quad (9)$$

$$\Delta E_i^{\mathrm{corr}} = E_i^{\mathrm{MP2}} - E_i^{\mathrm{HF}} \quad (10)$$

where E_i^{HF} and E_i^{MP2} are respectively the Hartree−Fock and MP2 energies of a monomer i, and $\Delta E_i^{\mathrm{corr}}$ is the MP2 correlation energy of a monomer i

$$\Delta E_{\mathrm{corr,MP2}}{}^{(2)} = \sum_{i<j} (\Delta E_{ij}^{\mathrm{corr}} - \Delta E_i^{\mathrm{corr}} - \Delta E_j^{\mathrm{corr}}) \quad (11)$$

$$\Delta E_{ij}^{\mathrm{corr}} = E_{ij}{}^{\mathrm{MP2}} - E_{ij}{}^{\mathrm{HF}} \quad (12)$$

where E_{ij}^{HF} and E_{ij}^{MP2} are the the Hartree−Fock and MP2 energies, respectively, of a dimer ij, and $\Delta E_{ij}^{\mathrm{corr}}$ is is the MP2 correlation energy for a dimer ij; therefore, EE-PA-CE energy can be defined as

$$E^{\mathrm{PA-CE}} = E_{\mathrm{HF}}^{\mathrm{total}} + \Delta E_{\mathrm{corr,MP2}}{}^{(1)} + \Delta E_{\mathrm{corr,MP2}}{}^{(2)} \quad (13)$$

The EE-3B-CE energy can be written as

$$E^{\mathrm{3B-CE}} = E^{\mathrm{PA-CE}} + \Delta E_{\mathrm{corr,MP2}}{}^{(3)} \quad (14)$$

$$\Delta E_{\mathrm{corr,MP2}}{}^{(3)} = \sum_{i<j<k} (\Delta E_{ijk}^{\mathrm{corr}} - \Delta E_i^{\mathrm{corr}} - \Delta E_j^{\mathrm{corr}} - \Delta E_k^{\mathrm{corr}}) - (\Delta E_{ij}^{\mathrm{corr}} - \Delta E_i^{\mathrm{corr}} - \Delta E_j^{\mathrm{corr}}) - (\Delta E_{ik}^{\mathrm{corr}} - \Delta E_i^{\mathrm{corr}} - \Delta E_k^{\mathrm{corr}}) - (\Delta E_{jk}^{\mathrm{corr}} - \Delta E_j^{\mathrm{corr}} - \Delta E_k^{\mathrm{corr}}) \quad (15)$$

dx.doi.org/10.1021/ct4001872 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 1. Systems Considered in This Work and Types of Zn^{2+}- and Cd^{2+}-Containing Fragments in Each[a]

	full system	types of largest fragments[a]
1	$([Zn(NH_3)_2(OH)_3]^-)$	$Zn(OH)_2$, $[Zn(OH)(NH_3)]^+$, $[Zn(NH_3)_2]^{2+}$
2	$[Zn(NH_3)_3(OH)_2]Zn(OH)_2$	$[Zn(OH)(NH_3)]^+$, $[Zn(NH_3)_2]^{2+}$
3	$([Zn(NH_3)_5]^{2+})$	$[Zn(NH_3)_2]^{2+}$
4	$([Zn(NH_3)_5]^{2+})$	$[Zn(NH_3)_2]^{2+}$
5	$([Zn(H_2O)_5]^{2+})$	$[Zn(H_2O)_5]^{2+}$
6	$([Zn(H_2O)_4(OH)]^+)$	$[Zn(OH)(H_2O)]^+$, $[Zn(H_2O)_2]^{2+}$
7	$[Cd(H_2O)_2(SH)_2(NH_3)]$	$[Cd(H_2O)_2]^{2+}$, $[Cd(SH)_2]$, $[Cd(H_2O)(SH)]^+$ $[Cd(H_2O)(NH_3)]^{2+}$, $[Cd(SH)(NH_3)]^+$
8	*fac* isomer of $([Zn(NH_3)_3(OH)_3]^-)$	$Zn(OH)_2$, $[Zn(OH)(NH_3)]^+$, $[Zn(NH_3)_2]^{2+}$
9	*mer* isomer of $([Zn(NH_3)_3(OH)_3]^-)$	$Zn(OH)_2$, $[Zn(OH)(NH_3)]^+$, $[Zn(NH_3)_2]^{2+}$
10	$([Zn(OH)_6]^{4-})$	$Zn(OH)_2$

[a]For example, for system **1**, there are three possible fragmentations of the type $Zn(OH)_2$, one of the type $[Zn(NH_3)_2]^{2+}$, and six of the type $[Zn(OH)(NH_3)]^+$. In Table 4, we average results over all ten possible fragmentations, whereas in Table 8, we give results for the single scheme selected by our new fragmentation strategy.

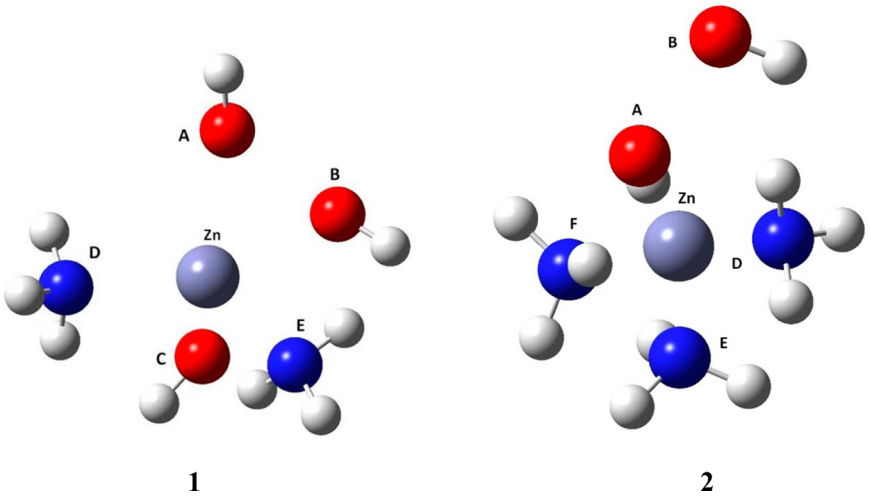


Figure 1. Structures of truncated model Zn biocenter complexes: (**1**) the anthrax toxin lethal factor active site (LF) (1PWU),[55] ($[Zn(NH_3)_2(OH)_3]^-$), and (**2**) matrix metalloproteinase-3 (MMP-3, stromelysin-1) (1SLN),[56] $[Zn(NH_3)_3(OH)_2]$. In both cases, histidine residues are represented by ammonias (NH_3), and Glu residues and zinc-binding group oxygens in the cocrystallized inhibitors are represented by hydroxyls (OH^-).

$$\Delta E_{ijk}^{\mathrm{corr}} = E_{ijk}^{\mathrm{MP2}} - E_{ijk}^{\mathrm{HF}} \tag{16}$$

where E_{ijk}^{HF} and E_{ijk}^{MP2} are the Hartree−Fock and MP2 energies, respectively, of a trimer *ijk*, and $\Delta E_{ijk}^{\mathrm{corr}}$ is is the MP2 correlation energy for a trimer *ijk*. The individual energies are obtained using the MP2 level of theory. One can obtain the EE-MB-CE approximation of the CCSD(T) energy by replacing "MP2" with "CCSD(T)" in eqs 5−16.

In the EE-PA and EE-3B calculations, one only performs MP2 or CCSD(T) calculations on the monomers, dimers, and (for 3B) the trimers. In the EE-PA-CE and EE-3B-CE approximations, one also carries out a Hartree−Fock calculation on the entire system. The electronic energy at the MP2 or CCSD(T) level of theory can be written as the Hartree−Fock energy plus the correlation energy. In the EE-PA-CE and EE-3B-CE approximations, one uses expansions only on the correlation energy, and adds the correlation energy to the directly calculated Hartree−Fock calculation of the entire system (eqs 13 and 14). Since Hartree−Fock theory formally scales as N^4, where N is the number of atoms, it is less computationally demanding to carry out a Hartree−Fock calculation on the entire system than to carry out an MP2 or other correlated wave function calculation on the entire system, and the goal here is to obtain accuracy equivalent to a full MP2 or CCSD(T) calculation without the cost of the latter.

In past work,[39,40] the EE-MB-CE method has been applied to large water clusters, ranging in size from 5 to 20 water molecules and water hexamers. Here, we examine the application of the EE-MB-CE method to metal−ligand bonding in various pentacoordinate and hexacoordinate Zn and Cd complexes and present a fragmentation scheme that can be applied to all Zn and Cd systems studied in the paper. These metal−ligand systems are more challenging than noncovalently bonded clusters because the electrostatic and induction effects are much larger, and there is some covalent character in the metal−ligand coordination bonds.

We also present an application where MP2 is replaced by CCSD(T).

3. COMPUTATIONAL METHODS

Most of the calculations were done using the MP2 level of theory. The B2[46] basis set was used for the Zn calculations, and the def2-TZVP[47] basis set was used for the Cd calculations. Both B2 and def2-TZVP are polarized valence-triple-ζ basis sets for Zn and Cd, respectively. The MP2 level was chosen because it is the least expensive of the post-Hartree−Fock methods, allowing for direct comparison of the EE-MB and the EE-MB-

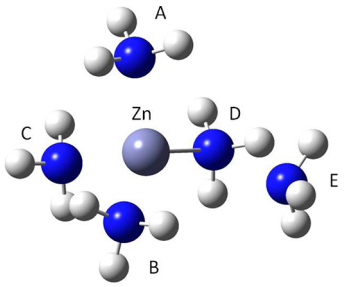
dx.doi.org/10.1021/ct4001872 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

CE energies to the full MP2 energies. Such direct comparison for all clusters and fragmentation schemes studied in this work would not be practical at a reasonable cost with more expensive post-Hartree−Fock methods such as CCSD(T). Nevertheless, we present a CCSD(T) calculation on the system **1** to test the performance of the methods with other correlated methods. Our earlier published work[46,48] showed that incorporating relativistic effects on core electrons significantly increased the accuracy of geometric and energetic calculations for Zn coordination complexes; in the current study we therefore replaced the innermost ten electrons of Zn and 28 core electrons of Cd with the relativistic effective core potential (RECP).[49−51]

Charges are calculated for each fragment at the geometry of that monomer in the overall system. For example, if we are calculating the energy of ZnABCDEF, where A, B, C, D, E, and F are ligands, and if one of the fragments is ZnBC, we calculate the partial atomic charges of ZnBC by removing A, D, E, and F from the system. Here we calculate charges using Merz−Kollman (MK) electrostatic-potential fitting,[52] as in previous work on Zn compounds.[44,45] Charges for CCSD(T) calculations on system **1** were obtained using M05-2X level theory.[53]

All benchmark calculations were performed using *Gaussian 09*.[54] All EE-MB and EE-MB-CE calculations were carried out using MBPAC 2011−5,[55] a freely available software package that allows the user to define a particular fragmentation scheme and then accesses *Gaussian 09* to perform the necessary monomer, dimer, and trimer calculations.

In the current work, we consider seven pentacoordinate and three hexacoordinate Zn systems (see Table 1). Two pentacoordinate complexes are model compounds based on experimental X-ray structures of Zn metalloenzyme active sites relevant to the drug design process: the anthrax toxin lethal factor (LF) (PDB ID 1PWU)[56] active site, and the matrix metalloproteinase-3 (MMP-3) catalytic site (PDB ID 1SLN)[57] (systems **1** and **2**, respectively, Figure 1). These model structures are the same as those reported in our previously published work.[45,48] System **3** (Figure 2) is a model compound



3

Figure 2. Structure of truncated model Zn biocenter complex: (**3**) matrix metalloproteinase-7) (MMP-7, matrilysin) (1MMR),[57] ($[Zn(NH_3)_5]^{2+}$). Histidine residues and zinc-binding group nitrogens in the cocrystallized inhibitor are represented by ammonias (NH_3).

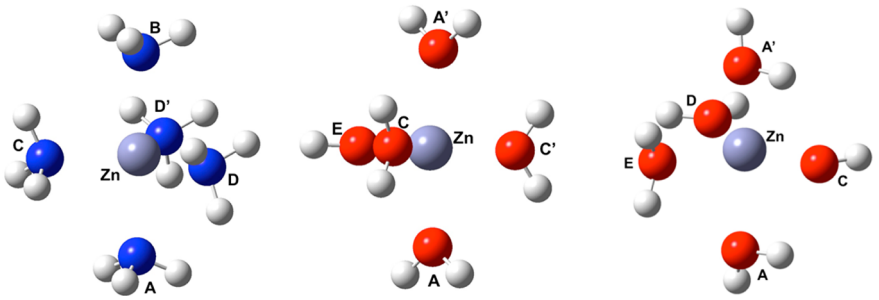
based on the X-ray structure of the matrix metalloproteinase-7 (MMP-7, also known as matrilysin) active site, cocrystallized with sulfodiimine (PDB ID 1MMR).[58] In MMP-7, the catalytic Zn is coordinated by three histidines, and in 1MMR, the two remaining coordination sites are occupied by imine nitrogens in the cocrystallized inhibitor. We use these systems to test the ability of the EE-MB-CE method to reproduce Zn-ligand bond dissociation energies of biological Zn-containing active sites. System **3** is also used to demonstrate the advantage of our newly developed fragmentation strategy over previously published fragmentation guidelines.[45] We also consider three other pentacoordinate systems: $Zn(NH_3)_5^{2+}$, $Zn(H_2O)_5^{2+}$, and $Zn(H_2O)_4(OH)^+$ (systems **4**, **5**, and **6**, respectively, Figure 3). These are used to test the applicability of the EE-MB-CE method to positively charged systems. Calculations on a Cd complex (system **7**, Figure 4) were performed to test the applicability of our methods to a transition metal other than Zn. System **7** (Figure 4) is a model compound based on the X-ray structure of the cadmium carbonic anhydrase active site (CDCA1-R2) (PDB ID 3BOB).[59] In CDCD1-R2, the catalytic Cd is coordinated by two cysteine residues, a histidine residue, and two water molecules.

Two of the three hexacoordinate complexes examined here are the *fac* and *mer* isomers of $[Zn(NH_3)_3(OH)_3]^-$ (systems **8** and **9**, respectively, Figure 5), used here to evaluate the performance of the EE-MB-CE method with larger, hexacoordinate systems. We also use hexacoordinate $[Zn(OH)_6]^{4-}$ (system **10** Figure 6) to test the applicability of the methods to symmetric Zn complexes. Zn−ligand distances in system **10** were placed at 2.10 Å. To make system **10** completely symmetric angles Zn−O−H were made equal to 180 degrees. The labeling schemes of all systems are consistent with previously published work.[44,45] In total, these systems comprise two neutral, four positively charged, and four negatively charged complexes. All structures are provided in Supporting Information.

We use the labeling scheme defined in Figures 1−6, in which A, B, and C (when present) are negatively charged hydroxyl ligands and D, E, and F (when present) are neutral ligands. Each of the coordination complexes in Figure 3 has the structure of an irregular trigonal bipyramid, with axial ligands A and B and equatorial ligands C, D, and E. If r_X denotes the distance from the non-hydrogen atom of a ligand to Zn, we label the atoms so that $r_A \leq r_B$ and $r_C \leq r_D \leq r_E$. However, if $r_B = r_A$, then B is also called A′; if $r_D = r_C$, then D is also called C′, and if $r_E = r_D$, then E is also called D′.

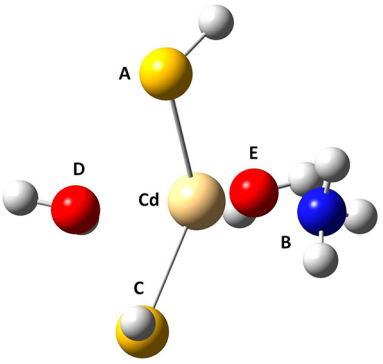
The quantity we calculate is an instantaneous bond dissociation energy, which is defined as the energy to remove one of the ligands from the coordination system at a given predefined geometry (if this were the equilibrium geometry and if the separated subsystems were reoptimized after removal of the ligand whose bond is being broken, then the instantaneous bond dissociation energy would be the equilibrium bond dissociation energy). As discussed previously,[44,45] the instantaneous bond dissociation energy is the sum of the energies of the two products (separated frozen fragments) minus the energy of the reactant, without reoptimization and without including vibrational energy. When calculating the energies of a given dissociation product, the embedding charges of the other product are not included because the other product is considered to be infinitely separated.

Our EE-MB-CE calculations on neutral, negatively, and positively charged Zn and Cd systems demonstrate, consistently with our previous findings,[44,45] that one must choose a fragmentation scheme, where one of the monomers is Zn^{2+} or Cd^{2+} coordinated to at least two ligands. We rationalize this rule in terms of partial atomic charges. In particular, the charge on unligated or monoligated Zn and even on biligated Zn is much larger than the charge on polyligated Zn; thus fragments consisting of unligated, monoligated, or to a lesser extent,



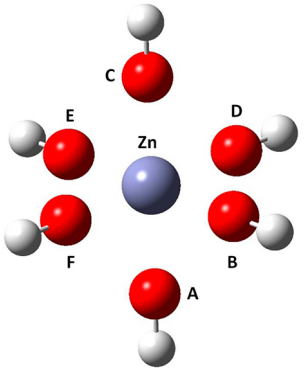
4 5 6

Figure 3. (**4**) Structure of ($[Zn(NH_3)_5]^{2+}$), (**5**) structure of ($[Zn(H_2O)_5]^{2+}$), and (**6**) structure of ($[Zn(H_2O)_4(OH)]^{+}$).



7

Figure 4. Structure of truncated model Cd biocenter complex: (**7**) cadmium carbonic anhydrase (3BOB),[58] $[Cd(H_2O)_2(SH)_2(NH_3)]$. Cysteine residues are represented by thiol groups (SH). Histidine residue is represented by ammonium (NH_3).

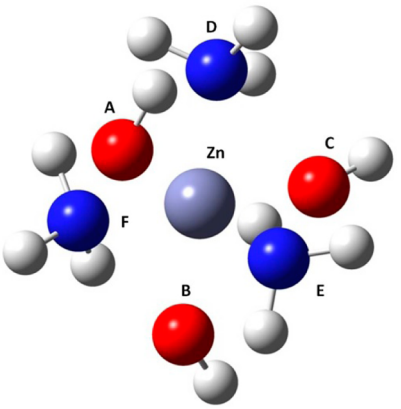


10

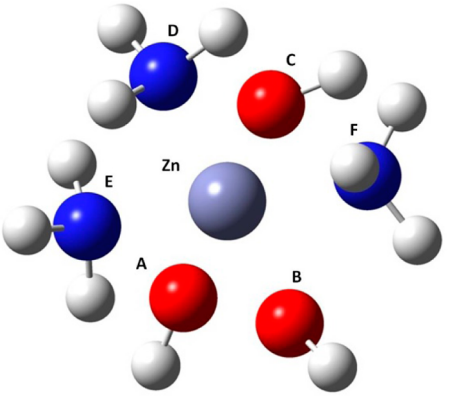
Figure 6. (**10**) Structure of ($[Zn(OH)_6]^{4-}$).

biligated Zn are not precisely representative of a portion of a larger system. But if each fragment already has two ligands on Zn, then even in dimers there are three ligands on Zn. Thus, we only consider fragmentation schemes where one of the fragments is Zn^{2+} or Cd^{2+} with two ligands and the other fragments are individual ligands. Therefore, for pentacoordinate systems, there are ten possible ways to fragment a system. For example, for pentacoordinate system **1**, the largest fragments for all ten fragmentation schemes are ZnAB, ZnAC, ZnAD, ZnAE, ZnBC, ZnBD, ZnBE, ZnCD, ZnCE, and ZnDE. However, there are fifteen different ways to fragment a hexacoordinate system. The largest fragments in each fragmentation for each hexacoordinate system are as follows: ZnAB, ZnAC, ZnAD, ZnAE, ZnAF, ZnBC, ZnBD, ZnBE, ZnBF, ZnCD, ZnCE, ZnCF, ZnDE, ZnDF, and ZnEF. Table 1 shows the possible ligands of Zn^{2+} and Cd^{2+} in a fragment for each studied system. We performed calculations utilizing all possible fragmentation schemes for all systems.

Each fragmentation scheme included all possible bond dissociation processes. For pentacoordinate systems we



8



9

Figure 5. Structures of two octahedral, hexacoordinate Zn complexes ($[Zn(NH_3)_3(OH)_3]^{-}$): (**8**) *fac* isomer and (**9**) *mer* isomer.

dx.doi.org/10.1021/ct4001872 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

consider breaking 5 bonds whereas for hexacoordinate systems we consider breaking 6 bonds. For example, consider system **1** where each fragmentation scheme was used within the EE-MB and EE-MB-CE approximations to compute each of the following instantaneous bond dissociation processes:

$$ZnABCDE^- \rightarrow ZnBCDE + A \quad (R1)$$

$$ZnABCDE^- \rightarrow ZnACDE + B \quad (R2)$$

$$ZnABCDE^- \rightarrow ZnABDE + C \quad (R3)$$

$$ZnABCDE^- \rightarrow ZnABCE^- + D \quad (R4)$$

$$ZnABCDE^- \rightarrow ZnABCD^- + E \quad (R5)$$

For systems **1**−**10**, there are a total of 53 bond dissociation processes labeled from R1 to R53. Benchmark values for R1−R53 bond dissociation energies were obtained by full single-point calculations, that is, without using the many-body approximation (see Table 2). Table 3 shows CCSD(T) benchmark bond energies for system **1**. We measure "errors" as the absolute deviation of the EE-MB and EE-MB-CE results from the full calculations with the same method.

4. RESULTS AND DISCUSSION

For each method, we consider 350 bond dissociation energy calculations for pentacoordinate systems **1**−**7** (seven systems, each with five bonds and ten different fragmentation schemes) and 270 bond dissociation energy calculations for hexacoordinate systems **8**, **9**, and **10** (three systems, each with six bonds and 15 fragmentation schemes). And we compute final errors averaged over a total of 620 bond dissociation energies for each method. Table 4 shows the EE-MB and the EE-MB-CE mean signed errors (MSEs) and mean unsigned errors (MUEs) in the instantaneous bond energies (kcal/mol) averaged over the 620 cases of bond dissociation.

The EE-3B MUEs range from 0.74 to 2.64 kcal/mol except system **2**, which has a MUE of 6.85 kcal/mol. Note that the EE-3B MUE for all ten systems is 2.13 kcal/mol. The EE-3B MUEs in bond dissociation energies of the hexacoordinate systems **8** and **9**, at 2.08 and 1.65 kcal/mol, are comparable to those of the pentacoordinate systems, which is an encouraging result. Note that the EE-3B MUE for Cd^{2+} system **7** is 1.40 kcal/mol. As expected, the EE-PA method is less accurate, resulting in MUEs in bond dissociation energies ranging from 2.69 to 18.69 kcal/mol for the systems studied here. Averaging the MUEs over all ten Zn^{2+} and Cd^{2+} systems gives an overall MUE of 6.66 kcal/mol for the EE-PA method.

Now let us consider the EE-MB-CE method. Table 4 also shows the EE-MB-CE MSEs and MUEs for bond-breaking energies for all possible fragmentation schemes for all ten Zn^{2+} and Cd^{2+} systems. As expected, the EE-3B-CE approximation is more accurate, resulting in MUEs in bond dissociation ranging from 0.42 to 1.54 kcal/mol for all pentacoordinate and hexacoordinate systems except **2**, which has a MUE of 3.94 kcal/mol. The EE-3B-CE MUE for Cd^{2+} system **7** is 0.65 kcal/mol, which is comparable to those for the Zn-containing pentacoordinate systems. The EE-3B-CE MUEs in bond dissociation energies for the hexacoordinate systems **8** and **9** are comparable to those for the pentacoordinate systems, at 1.54 and 0.98 kcal/mol, respectively. As expected, the EE-PA-CE method is less accurate, resulting in MUEs in bond dissociation energies ranging from 1.10 to 3.16 kcal/mol for the systems studied here except system **2**, which has a MUE of 8.06 kcal/mol. Averaging the MUEs over all ten Zn^{2+} and Cd^{2+} systems including **2** gives an overall MUE of 2.73 kcal/mol for the EE-PA-CE method, but only 1.10 kcal/mol for the EE-3B-CE method.

On the basis of the results in Table 4, the many-body methods can be ranked in order of increasing MUE as EE-3B-

Table 2. MP2 Benchmark Bond Dissociation Energies (kcal/mol) for Bonds in Each Model Complex

reaction	system	dissociated bond	bond energy
R1	**1**	Zn−A	33.70
R2	**1**	Zn−B	28.14
R3	**1**	Zn−C	67.19
R4	**1**	Zn−D	17.06
R5	**1**	Zn−E	14.12
R6	**2**	Zn−A	116.45
R7	**2**	Zn−B	53.31
R8	**2**	Zn−D	7.66
R9	**2**	Zn−E	−11.54
R10	**2**	Zn−F	8.38
R11	**3**	Zn−A	51.27
R12	**3**	Zn−B	51.19
R13	**3**	Zn−C	45.12
R14	**3**	Zn−D	27.18
R15	**3**	Zn−E	2.03
R16	**4**	Zn−A	33.06
R17	**4**	Zn−B	32.86
R18	**4**	Zn−C	45.35
R19	**4**	Zn−D	45.35
R20	**4**	Zn−D′	45.28
R21	**5**	Zn−A	34.34
R22	**5**	Zn−A′	33.38
R23	**5**	Zn−C	45.27
R24	**5**	Zn−C′	39.52
R25	**5**	Zn−E	43.87
R26	**6**	Zn−A	23.25
R27	**6**	Zn−A′	23.25
R28	**6**	Zn−C	260.83
R29	**6**	Zn−D	26.05
R30	**6**	Zn−E	8.66
R31	**7**	Cd−A	147.65
R32	**7**	Cd−B	12.51
R33	**7**	Cd−C	144.68
R34	**7**	Cd−D	1.08
R35	**7**	Cd−E	9.11
R36	**8**	Zn−A	9.67
R37	**8**	Zn−B	9.81
R38	**8**	Zn−C	9.67
R39	**8**	Zn−D	−14.69
R40	**8**	Zn−E	−22.45
R41	**8**	Zn−F	−20.91
R42	**9**	Zn−A	25.89
R43	**9**	Zn−B	11.66
R44	**9**	Zn−C	32.37
R45	**9**	Zn−D	−14.96
R46	**9**	Zn−E	−16.91
R47	**9**	Zn−F	−8.85
R48	**10**	Zn−A	−257.79
R49	**10**	Zn−B	−257.79
R50	**10**	Zn−C	−257.79
R51	**10**	Zn−D	−257.79
R52	**10**	Zn−E	−257.79
R53	**10**	Zn−F	−257.79

dx.doi.org/10.1021/ct4001872 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 3. CCSD(T) Benchmark Bond Dissociation Energies (kcal/mol) for Every Bond in System 1

reaction	system	dissociated bond	bond energy
R1	**1**	Zn−A	34.98
R2	**1**	Zn−B	29.20
R3	**1**	Zn−C	67.48
R4	**1**	Zn−D	16.60
R5	**1**	Zn−R	14.08

CE, EE-3B, EE-PA-CE, and EE-PA. Overall, both the EE-MB and the EE-MB-CE methods perform well, even for the hexacoordinate Zn complexes, indicating their usefulness for larger metalloprotein systems for which high-level electronic structure calculations on the entire system might not be feasible.

Table 5 provides a more detailed view of the results for system **2**, with MSEs and MUEs in bond energies for system **2** for all ten possible fragmentations. The ZnFE, ZnDE, and ZnDF fragmentation schemes have the largest errors, with MUEs of 12.47, 10.94, and 9.07 kcal/mol respectively for EE-3B and 6.71, 6.19, and 5.23 kcal/mol respectively for EE-3B-CE. Thus, any combination of Zn with ammonia ligands in a fragment yields poor results. In contrast, the ZnAB fragmentation scheme, where Zn is combined with two hydroxyl ligands in a fragment, yields the best results; note that for this fragmentation, the MUE of EE-3B-CE is only 0.84 kcal/mol, whereas that of the EE-3B is a reasonable 3.28 kcal/mol (Table 5). Our analysis suggests that the ZnAB fragmentation yields better results than the other fragmentation schemes because there is a large interatomic Coulomb interaction between the O atom of A and the O atom of B, and the ZnAB fragmentation treats these interactions entirely by quantum mechanics without using point charges because it places the A and B groups in a single fragment. This observation was corroborated by additional calculations (not reported in detail here) that resulted in a lower error when the distance between ligands A and B was increased in system **2** without altering any other distances when the ZnAB fragmentation scheme was not used, and that resulted in an larger error when decreasing the distance between the ligands A and B when the ZnAB fragmentation scheme was not used. This is consistent with the major part of the error coming from the A−B interaction when A and B are not in a single fragment.

Following these observations, we developed a fragmentation strategy designed to minimize the error. To obtain accurate energies for Zn systems with the EE-MB and EE-MB-CE methods, one should locate the two ligands in a system that have the strongest Coulomb interaction with each other, and combine them with Zn^{2+} or Cd^{2+} in a single fragment. To quantify "the strongest Coulomb interaction", we used the monomer embedding charges to calculate Coulomb interactions of the partial atomic charges in one ligand with those in another. For example, consider ligands A, B, D, E, and F of system **2**. First, we calculate the monomeric partial atomic charges using the method described above (that is, we perform an MK partial charge analysis on each isolated ligand in the geometry it has in the cluster). Then those charges are used to calculate absolute maximum atom−atom Coulomb interaction between each pair of ligands. We define the absolute atom−atom Coulomb interaction (V_{ij}^{AB}) between atom i of ligand A and atom j of ligand B as

$$V_{ij}^{AB} = \left| \frac{q_i^A q_j^B}{r_{ij}} \right|, \; i \in \text{A and } j \in \text{B, A} \neq \text{B}$$

where q_i^A is the charge on i, and q_j^B is the charge on j, and r_{ij} is the distance from i to j.

The maximum Coulomb interaction between two fragments is defined as the largest of all of the absolute atom−atom Coulomb interaction energies between a partial atomic charge in one fragment and a partial atomic charge on an atom of a different fragment. For system **2**, ligands A and B have the largest absolute maximum Coulomb interaction out of all of the possible pairs of ligands in the complex. Therefore, our strategy places these two ligands together with Zn^{2+} in a fragment; this leaves the other ligands as individual fragments. This strategy produces a preferred fragmentation scheme for each of the ten complexes. Note that our prescription can be stated in a way that suggests even greater generality: of all possible fragmentation schemes that combine two ligands in the same fragment with the transition metal, one should apply the fragmentation scheme that minimizes the absolute maximum Coulomb interaction between any two remaining ligands that are treated as individual fragments.

Table 4. EE-MB and EE-MB-CE Mean Signed and Unsigned Errors in Bond Energies (kcal/mol) for All Ten Zn and Cd Complexes Averaged over 10-to-15 Fragmentation Schemes and Averaged over All Five or Six Bonds Being Broken[a]

	EE-PA		EE-3B		EE-PA-CE		EE-3B-CE	
system	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
1	−1.26	7.28	0.64	2.07	−1.35	3.16	0.34	0.95
2	0.94	14.97	0.70	6.85	−1.86	8.06	1.14	3.94
3	−2.88	4.10	0.49	0.78	−0.39	1.48	0.02	0.42
4	−3.71	4.20	0.47	0.74	−0.68	1.75	0.05	0.47
5	−0.06	2.69	−0.33	1.30	−0.17	1.10	−0.09	0.45
6	0.31	5.31	−0.85	2.64	0.16	2.33	−0.34	0.82
7	−17.07	18.69	−0.89	1.40	−1.99	2.39	−0.53	0.65
8	2.41	4.82	−0.53	2.08	0.12	2.64	−0.24	1.54
9	0.95	2.73	0.42	1.65	−0.48	1.80	0.17	0.98
10	6.50	6.52	−1.74	2.16	3.12	3.12	−0.59	0.82
mean[b]	−0.48	6.66	−0.25	2.13	−0.08	2.73	−0.05	1.10

[a]MSE is mean signed error; MUE is mean unsigned error. [b]Averaged over the 620 combinations of system, bond, and fragmentation scheme, not over the ten rows.

dx.doi.org/10.1021/ct4001872 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

Table 5. EE-MB and EE-MB-CE Mean Signed and Unsigned Errors in Bond Energies (kcal/mol) for System 2, for All Ten Fragmentation Schemes

	EE-PA		EE-3B		EE-PA-CE		EE-3B-CE	
largest fragment	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
ZnFE	−7.38	17.38	1.22	12.47	−4.21	10.56	1.63	6.71
ZnDE	−5.81	18.19	0.70	10.94	−3.58	11.77	0.84	6.19
ZnDF	−7.10	17.21	2.70	9.07	−4.26	13.17	2.42	5.23
ZnAD	−11.12	12.36	4.88	6.87	−7.30	7.84	3.31	4.54
ZnBD	17.45	17.45	−3.31	3.31	3.81	5.51	−0.46	1.37
ZnAE	−7.93	11.90	1.99	7.92	−4.41	6.35	1.29	4.77
ZnBE	5.74	10.21	0.87	5.54	1.11	8.28	0.17	3.79
ZnAF	−8.05	10.99	3.62	6.62	−5.03	6.32	2.44	4.21
ZnBF	14.62	14.98	−2.43	2.43	4.09	6.85	−0.75	1.76
ZnAB	19.01	19.01	−3.28	3.28	1.15	3.99	0.49	0.84
mean[a]	0.95	14.97	0.70	6.85	−1.86	8.06	1.14	3.94

[a]Mean signed and unsigned errors for the ten rows in the given column, where each row contains an average over five bond dissociation energies.

Table 6 presents maximum Coulomb interactions between every ligand in all ten systems. For example, ligands A and B in system **1** have the maximum absolute Coulomb interaction for that system, which is 0.56 *e* (where *e* is the charge on a proton). Thus, ligands A and B in system **1** should be combined with Zn^{2+} in a fragmentation. However, in hexacoordinate system **8**, there are three identical maximum Coulomb interactions: A and C, A and B, and B and C. Any of those ligands can be combined with Zn^{2+} in a fragment. In Table 8, we report the average of the three (ZnAB, ZnAC, and ZnBC) fragmentations for system **8**. In system **9**, ligands A and B, and B and C have identical maximum Coulomb interactions. Thus, the average of the two (ZnAB and ZnBC) fragmentations was reported in Table 8 for system **9**.

System **10** is a unique symmetric system where only two fragmentation types are possible: trans and cis. Trans fragmentation schemes are those where Zn^{2+} is paired with ligands that are directly across from each other, and cis fragmentations are those where Zn^{2+} is paired with ligands that are adjacent to one another. In system **10**, ZnAC, ZnBE, and ZnDF are trans fragmentations, whereas ZnAB, ZnAD, ZnAE, ZnAF, ZnBC, ZnBD, ZnBF, ZnCD, ZnCE, ZnCF, ZnDE, and ZnEF are cis fragmentations. Table 7 presents MUEs for trans fragmentation ZnAC and cis fragmentation ZnAB. The EE-3B-CE MUEs for trans and cis fragmentations are 0.58 and 0.88 kcal/mol, respectively.

Table 8 shows results for all ten systems where, for each complex, we apply the new strategic fragmentation scheme rather than averaging over all possible fragmentations. We see that the MUE for the EE-3B method decreases from 2.13 to 1.42 kcal/mol, and the MUE for the EE-3B-CE method decreases from 1.10 to 0.59 kcal/mol, when we use the preferred fragmentation. Note that the EE-PA-CE MUE decreases from 2.73 to 1.93 kcal/mol. The EE-PA MUE decreases from 6.66 to 6.14 kcal/mol.

Table 9 shows MSEs and MUEs for both EE-MB and EE-MB-CE methods obtained by CCSD(T) level calculations on system **1**. As expected, the EE-3B-CE approximation is more accurate, resulting in a MUE for bond dissociation of 0.30 kcal/mol, whereas the EE-3B MUE is 1.26 kcal/mol.

Previously,[45] we reported four fragmentation guidelines for Zn systems. According to those rules, Zn should be combined with at least two ligands in a fragment to have a representative charge distribution at the Zn center, and three other rules were proposed: one should not dissociate Zn−ligand bonds that are in the same fragment as Zn, one should not include more than one charged fragment, and Zn should not be in a fragment with ligands that are *trans* to each other. In the present work, the final three of these rules are superseded by one new, more general guideline, that is, of all possible fragmentation schemes under consideration, one should use the one that minimizes the maximum Coulomb interaction between atoms in two different fragments. Since here we consider only schemes that combine two ligands with Zn^{2+} or Cd^{2+} in one of the fragments and then treat all other ligands as individual fragments, the rule reduces for the present case to: to calculate accurate metal−ligand bond dissociation energies, one should find two ligands that exhibit the strongest Coulomb interaction with each other and combine them with Zn or Cd into a single fragment.

In addition to yielding lower errors, the fragmentation prescription advanced here allows the treatment of systems for which our previous guidelines were ambiguous. For example, consider system **3** (see Figure 2), which features five ammonia ligands and for which none of our previous guidelines offers an optimal fragmentation strategy. Applying our more general rule as described above results in an EE-3B MUE of 0.58 kcal/mol and an EE-3B-CE MUE of 0.70 kcal/mol.

5. COMPARISON TO OTHER FRAGMENTATION METHODS

In this section, we discuss other fragmentation methods to place the EE-MB-CE method and the present work in perspective. Such methods may be classified in various ways,[60−64] each of which focuses on a different aspect. In discussing these classifications, the language we use is that the smallest subsystems considered (the groups of atoms that are together at all stages of the calculation) are the monomers, the fragments are any group of one or more monomers on which a calculation is carried out, and the extended system is called the entire system (other notation sometimes encountered in the literature is to call the monomers groups, to call the fragments monomers, and to call the entire system the supersystem).

One possible classification is into single-level methods and multilevel methods. In the former, all fragment calculations are carried out at the same level. In the latter, either fragment calculations are carried out with more than one level or, as in the present case,fragment calculations carried out at a higher level are combined with entire-system calculations carried out at a lower level.

Table 6. Absolute Coulomb Interactions between Fragments in All Ten Systems (in e)

system	ligand	B or A or A′	C or C′	D or C	E or D′	F
1	A	0.56	0.33	0.36	0.33	
	B		0.48	0.28	0.40	
	C			0.35	0.35	
	D				0.27	
2	A	0.70		0.40	0.45	0.43
	B			0.37	0.27	0.32
	D				0.35	0.34
	E					0.35
3	A	0.30	0.29	0.29	0.25	
	B		0.30	0.26	0.28	
	C			0.33	0.21	
	D				0.41	
4	A	0.22	0.3221	0.3178	0.3180	
	B		0.3174	0.3218	0.3219	
	C			0.27	0.27	
	D				0.27	
5	A′	0.1125	0.1593	0.1593	0.1614	
	A		0.1657	0.1654	0.1626	
	C′			0.1352	0.1356	
	C				0.1354	
6	A	0.14	0.316	0.19	0.16	
	A′		0.320	0.19	0.16	
	C			0.22	0.30	
	D				0.18	
7	A	0.31	0.27	0.23	0.18	
	B		0.32	0.14	0.21	
	C			0.18	0.18	
	D				0.15	
8	A	0.57	0.57	0.45	0.32	0.45
	B		0.57	0.45	0.45	0.32
	C			0.32	0.45	0.45
	D				0.36	0.36
	E					0.36
9	A	0.57	0.40	0.45	0.45	0.45
	B		0.57	0.32	0.45	0.45
	C			0.45	0.45	0.45
	D				0.36	0.36
	E					0.25
10	A	0.52	0.36	0.52	0.52	0.52
	B		0.52	0.52	0.36	0.52
	C			0.52	0.52	0.52
	D				0.52	0.36
	E					0.52

Another possible classification is to distinguish fragment approximations from divide-and-conquer methods. In the latter, one uses fragments as an intermediate part of the calculation, but the goal is to converge to a particular unfragmented calculation. In such methods fragmentation is an algorithmic choice, not an approximation. In fragment approximations, the results converge to a high-level, entire-system result only at the limit where one of the fragments treated at the high level is so large that it is identical to the entire system. The EE-MB-CE method is a fragment approximation. The rest of this discussion is concerned only with fragment approximations, that is, with attempts to make useful calculations on larger systems feasible by introducing safe levels of approximation rather than by breaking the full calculation into smaller, more manageable parts without approximations.

Another possible classification is between inclusion−exclusion methods and many-body methods, but this is not unique since an inclusion−exclusion method can be thought of as an untruncated many-body method (with a distance cutoff), and many-body methods can also be thought of as inclusion−exclusion based methods.[62] A more unique classification is to ask whether a given part of the system occurs in one and only fragment (fragments do not overlap) or whether it occurs in more than one fragment (fragments do overlap); this is one way to classify fragment methods into overlapping and nonoverlapping methods (also sometimes called intersecting and nonintersecting). In this sense, many-body methods are overlapping methods because a given monomer occurs in many dimers, many trimers, and so forth. This contrasts with methods like the fragment molecular orbital (FMO) method[65] or the variational explicit polarization (X-Pol) method,[66] which have nonoverlapping fragments, except for a boundary atom in X-Pol. It is easily understood[64] that overlapping methods have faster convergence than nonoverlapping methods with respect to the size of fragments needed to obtain accurate results.

Another possible classification is whether a given fragment is embedded in the electrostatic field of the rest of the entire system. Including this embedding effect leads to more accurate results or to equally accurate results with smaller fragments.

In terms of the above classifications, the EE-MB-CE method is a multilevel, overlapping, embedded fragment approximation, and, since it dates to 2007,[39] it may be the first fragment approximation to actually combine all three of these advantages into a single algorithm.[67] Combining these three features makes it very powerful. We note however that the EE-MB method may be considered to be a special case of the electrostatically embedded[64] molecular tailoring[68] approach (EE-MTA), which allows more flexibility in the choice of fragments. We also note that the general energy-based fragment (GEBF) method[69] also shares some of the advantages of EE-MTA. Thus a multilevel version of EE-MTA or GEBF could be more flexible than EE-MB-CE, and a multilevel version of EE-MTA could include EE-MB-CE as a special case. One interesting possible way to make a multilevel extension of the GEBF method has been proposed recently by Li.[70]

Further classification is possible if one uses capping methods on bonds broken at the boundaries of fragments,[71] but this is not addressed here since no caps are employed in the present

Table 7. EE-MB and EE-MB-CE Mean Signed and Unsigned Errors in Bond Dissociation Energies (kcal/mol) for System 10

		EE-PA		EE-3B		EE-PA-CE		EE-3B-CE	
system	largest fragment	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
10	ZnAB (cis)	7.14	7.14	−2.44	2.44	3.45	3.45	−0.88	0.88
10	ZnAC(trans)	3.92	4.02	1.06	1.06	1.80	1.80	0.58	0.58
mean[a]		6.50	6.52	−1.74	2.16	3.12	3.12	−0.59	0.82

[a]Mean signed and unsigned errors for the 15 fragmentation schemes, where 12 are cis and 3 are trans fragmentations.

Table 8. EE-MB and EE-MB-CE Mean Signed and Unsigned Errors in Bond Energies (kcal/mol) for Zn and Cd Systems 1−10 Using the Fragmentation Strategy Selected by Our New Criterion

		EE-PA		EE-3B		EE-PA-CE		EE-3B-CE	
system	largest fragment	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
1	ZnAB	0.44	4.63	0.31	1.41	−0.56	1.74	0.24	0.50
2	ZnAB	19.01	19.01	−3.28	3.28	1.15	3.99	0.49	0.84
3	ZnDE	−3.77	3.99	0.07	0.58	0.47	1.34	−0.54	0.70
4	ZnAC	0.15	2.06	0.40	0.52	−0.07	1.57	−0.14	0.55
5	ZnAC′	0.04	2.69	−0.53	1.32	−0.09	1.04	−0.16	0.47
6	ZnA′C	0.37	1.34	−0.03	0.62	−0.19	0.65	0.03	0.18
7	CdBC	−9.24	11.15	0.09	0.24	−1.45	1.75	0.03	0.03
8	Zn(AB,AC,BC	5.79	5.79	−1.77	2.11	0.82	2.11	−0.28	0.97
9	Zn(AB,BC	3.70	3.80	−1.14	1.83	0.46	1.65	−0.27	0.82
10	ZnAB (*cis*)	7.14	7.14	−2.44	2.44	3.45	3.45	−0.88	0.88
mean[a]		2.36	6.14	−0.86	1.42	0.40	1.93	−0.15	0.59

[a]Mean signed and unsigned errors for the ten systems in the given column.

Table 9. EE-MB and EE-MB-CE Mean Signed and Unsigned Errors in Bond Energies (kcal/mol) for Zn System 1 Using CCSD(T) Level Theory and the Fragmentation Strategy Selected by Our New Criterion

		EE-PA		EE-3B		EE-PA-CE		EE-3B-CE	
system	largest fragment	MSE	MUE	MSE	MUE	MSE	MUE	MSE	MUE
1	ZnAB	8.54	8.73	0.31	1.26	7.52	7.52	0.21	0.30

work. In particular, even though we do break bonds in forming fragments, these are coordination bonds, and one of the goals of the current work is to show that we can obtain good results for coordination bonds without capping them.

One could also make classifications at a finer level; for example, we do not iterate the background charges to self-consistency (which makes the method simpler and less expensive[26]) whereas several other proposed methods do iterate the background electrostatics.

Discovering the most efficient way to carry out fragmentation calculations is certainly one of the most important challenges in current quantum chemical research, and the answer will almost surely depend on the problem. Further explanation of the accuracies that can be achieved on various kinds of problems by various fragmentation approaches is therefore very important. The problem treated here is a very difficult one. We have fragments that contain only one side of a metal−ligand bond. This shows the advantage of overlapping fragment methods, in that the bond can be split in this way, and yet represented intact in other fragments. Furthermore the electrostatics of the systems treated here are very challenging because the partial atomic charge on the metal atom depends on the number of ligands attached to it in the fragment, but we treat background electrostatics without iterations. The good results achieved here for this very challenging problem are encouraging.

6. CONCLUSIONS

The EE-MB and the EE-MB-CE calculations were carried out on pentacoordinate and hexacoordinate Zn and Cd complexes, and the results were compared to full calculations at the same correlated level of theory. Both the EE-MB and the EE-MB-CE methods perform well. By using our new prescription for cluster fragmentation, the MUE for the EE-3B method for all 53 bond dissociation energies in all ten complexes is only 1.42 kcal/mol. Also, notably, the MUE for the EE-3B-CE method for all bond energies in all ten complexes was reduced to 0.59 kcal/mol. The average absolute error for the EE-3B-CE method is only 0.93% of the average absolute bond dissociation energy, which is 63.29 kcal/mol. These results show that our newly developed fragmentation strategy can be used for various Zn-containing systems representing other zinc-dependent enzyme active sites as well. The fact that EE-3B-CE MUE for Cd^{2+} system **7** is 0.03 kcal/mol supports our fragmentation strategy and shows the transferability of the findings to other metals such as Cd. The EE-3B-CE MUE of 0.30 kcal/mol obtained by CCSD(T) calculations on system **1** shows the transferability of the findings to other correlated methods. The new fragmentation prescription improves upon our previously published work by replacing four fragmentation guidelines with one simple and unambiguous rule, namely, to calculate accurate Zn-ligand or Cd-ligand bond dissociation energies with EE-MB and EE-MB-CE methods, one should find the two ligands that have the strongest Coulomb interaction with each other and combine them with Zn or Cd in one fragment.

The EE-3B-CE method is able to capture most of the correlation energy, requiring only a Hartree−Fock calculation for the full system and correlated calculations only for dimers and trimers of the fragments. It can also be used to obtain instantaneous bond dissociation energies, which are an important indicator of how well a method can capture the wide variety of energetic interactions that occur in a coordination complex, including both electrostatic interactions and electron−electron correlation. The fact that the EE-3B-CE approximation captures bond dissociation energies to within 0.93% of conventionally calculated values when the new fragmentation scheme is used demonstrates that it is possible to apply a relatively simple fragment-based method to obtain an accurate picture of the potential energy surface of a quite complicated system involving coordinate covalent bonds. One could therefore use the EE-MB-CE approximation to parametrize inorganic reactive force fields for use to study macromolecular drug targets, or one could use it directly without force fields to obtain more accurate results that would otherwise be impossible for large systems (for example, systems with large ligands) where full calculations on the whole complex with a reliable method are unaffordable.

dx.doi.org/10.1021/ct4001872 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

■ ASSOCIATED CONTENT

Supporting Information

Cartesian coordinates for all Zn and Cd coordination complexes studied in this article. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: eamin@umn.edu (E.A.A.); truhlar@umn.edu (D.G.T.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors express their appreciation to Bo Wang for helpful discussions. This work was supported in part by the National Institutes of Health (R01 AI083234 to E.A.A.), by the University of Minnesota Department of Medicinal Chemistry, by the Minnesota Supercomputing Institute for Advanced Computational Research (MSI), and by the National Science Foundation under grant no. CHE09-56776 to D.G.T.

■ REFERENCES

(1) Bertini, I., Sigel, A., Sigel, H., Eds. *Handbook on Metalloproteins*; Marcel Dekker: New York, 2001.
(2) Ragsdale, S. W. *Chem. Rev.* **2006**, *106*, 3317.
(3) Supuran, C. T.; Winum, J.-Y. *Drug Design of Zinc-Enzyme Inhibitors*; John Wiley & Sons, Inc.: Hoboken, NJ, 2009; pp 1.
(4) Supuran, C. T.; Scozzafava, A.; Casini, A. *Med. Res. Rev.* **2003**, *23*, 146.
(5) Jacobsen, F. E.; Lewis, J. A.; Cohen, S. M. *ChemMedChem* **2007**, *2*, 152.
(6) Anzellotti, A. I.; Farrell, N. P. *Chem. Soc. Rev.* **2008**, *37*, 1629.
(7) Cramer, C. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2009**, *11*, 10757.
(8) Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618.
(9) Cizek, J. *Adv. Chem. Phys.* **1969**, *14*, 35.
(10) Purvis, G. D.; Barlett, R. J. *J. Chem. Phys.* **1982**, *76*, 1910.
(11) Raghavachari, K.; Anderson, J. B. *Chem. Phys. Lett.* **1989**, *157*, 479.
(12) Raghavachari, K.; Anderson, J. B. *J. Phys. Chem.* **1996**, *100*, 12960.
(13) Nakao, Y.; Hirao, K. *J. Chem. Phys.* **2004**, *120*, 6375.
(14) Saebø, S.; Pulay, P. *J. Chem. Phys.* **1987**, *86*, 914.
(15) Galli, G.; Parrinello, M. *Phys. Rev. Lett.* **1992**, *69*, 3547.
(16) Murphy, R. B.; Beachy, M.; Ringnalda, M.; Friesner, R. *J. Chem. Phys.* **1995**, *103*, 1481.
(17) Nielsen, I. M. B.; Janssen, C. L. *J. Chem. Theory Comput.* **2007**, *3*, 71.
(18) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100.
(19) Sugiki, S.-I.; Kurita, N.; Sengoku, Y.; Sekino, H. *Chem. Phys. Lett.* **2003**, *382*, 611.
(20) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.
(21) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2004**, *120*, 6832.
(22) Li, S.; Li, W.; Fang, T. *J. Am. Chem. Soc.* **2005**, *127*, 7215.
(23) Jiang, N.; Ma, J.; Jiang, Y. *J. Chem. Phys.* **2006**, *124*, 114112.
(24) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.
(25) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104.
(26) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
(27) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.
(28) Kobayashi, M.; Imamura, Y.; Nakai, H. *J. Chem. Phys.* **2007**, *127*, No. 074103.
(29) Hirata, S.; Yagi, K. *Chem. Phys. Lett.* **2008**, *464*, 123.
(30) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, No. 234108.
(31) Gordon, M. S.; Mullin, J. M.; Pruitt, S. R.; Roskop, L. B.; Slipchenko, L. V.; Boatz, J. A. *J. Phys. Chem. B* **2009**, *113*, 9646.
(32) Söderhjelm, P.; Aquilante, F.; Ryde, U. *J. Phys. Chem. B* **2009**, *113*, 11085.
(33) Li, W.; Piecuch, P. *J. Phys. Chem. A* **2010**, *114*, 6721.
(34) Řezáč, J.; Salahub, D. R. *J. Chem. Theory Comput.* **2010**, *6*, 91.
(35) Rahalkar, A. P.; Katouda, M.; Gadre, S. R.; Nagase, S. *J. Chem. Theory Comput.* **2010**, *31*, 2405.
(36) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory. Comput.* **2011**, *7*, 1336.
(37) Sorkin, A.; Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 683.
(38) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 46.
(39) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 1342.
(40) Dahlke, E. E.; Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 33.
(41) Dahlke, E. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1.
(42) Leverentz, H. R.; Truhlar, D. G. *J. Chem. Theory Comput.* **2009**, *5*, 1573.
(43) Speetzen, E. D.; Leverentz, H. R.; Lin, H.; Truhlar, D. G. In *Accurate Condensed Phase Electronic Structure Theory*; Manby, F., Ed.; CRC Press: Boca Raton, FL, 2010.
(44) Hua, D.; Leverentz, H. R.; Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2011**, *7*, 251.
(45) Kurbanov, E.; Leverentz, H. R.; Truhlar, D. G.; Amin, E. A. *J. Chem. Theory Comput.* **2012**, *8*, 1.
(46) Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 75.
(47) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
(48) Sorkin, A.; Truhlar, D. G.; Amin, E. A. *J. Chem. Theory Comput.* **2009**, *5*, 1254.
(49) Dolg, M.; Wedig, U.; Stoll, H.; Preuß, H. *J. Chem. Phys.* **1987**, *86*, 866.
(50) Kaupp, M.; Stoll, H.; Preuß, H. *J. Comput. Chem.* **1990**, *11*, 1029.
(51) Andrae, D.; Häußermann, U.; Dolg, M.; Stoll, H.; Preuß, H. *Theor. Chim. Acta* **1990**, *77*, 123.
(52) Besler, B. H.; Merz, K. M., Jr.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431.
(53) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215; **2008**, *119*, 525.
(54) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Revision A.01; Gaussian, Inc.: Wallingford, CT, 2009.
(55) Dahlke, E. E.; Lin, H.; Leverentz, H.; Truhlar, D. G. *MBPAC 2011−5*; University of Minnesota: Minneapolis, MN, 2011.
(56) Turk, B. E.; Wong, T. Y.; Schwarzenbacher, R.; Jarrell, E. T.; Leppla, S. H.; Collier, J.; Liddington, R. C.; Cantley, L. C. *Nat. Struct. Mol. Biol.* **2003**, *11*, 60.
(57) Becker, J. W.; Marcy, A. I.; Rokosz, L. L.; Axel, M. G.; Burbaum, J. J.; Fitzgerald, P. M. D.; Cameron, P. M.; Esser, C. K.; Hagmann, W. K.; Hermes, J. D.; Springer, J. P. *Protein Sci.* **1995**, *4*, 1966.
(58) Browner, F. M.; Smith, W. W.; Castelhano, L. A. *Biochemistry* **1995**, *34*, 6602.

(59) Xu, Y.; Feng, L.; Jeffrey, P. D.; Shi, Y.; Morel, F. M. M. *Nature* **2008**, *452*, 56.
(60) Suárez, E.; Díaz, N.; Suárez, D. *J. Chem. Theory Comput.* **2009**, *5*, 1667.
(61) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *7*, 1336.
(62) Mayhall, N. J.; Raghavachari, K. *J. Chem. Theory Comput.* **2012**, *8*, 2669.
(63) Richard, R. M.; Herbert, J. M. *J. Chem. Phys.* **2012**, *137*, 064113.
(64) Isegawa, M.; Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 1381.
(65) Fedorov, D. G.; Kitaura, K. In *Modern Methods for Theoretical Physical Chemistry of Biopolymers*; Starikov, E. P., Lewis, J. P.; Tanaka, S., Eds.; Elsevier: Amsterdam, 2006; p 3.
(66) Xie, W.; Song, L.; Truhlar, D. G.; Gao, J. *J. Chem. Phys.* **2008**, *128*, 234108.
(67) For an interesting, more recent multilevel method, see ref 34.
(68) Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484.
(69) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem.* **2007**, *111*, 2193.
(70) Li, W. *J. Chem. Phys.* **2013**, *138*, 14106.
(71) B. Wang, B.; Truhlar, D. G. *J. Chem. Theory Comput.* **2013**, *9*, 1036.