Received: 15 September 2022 | Revised: 23 November 2022 | Accepted: 24 November 2022

DOI: 10.1002/jcc.27052

RESEARCH ARTICLE

Journal of COMPUTATIONAL CHEMISTRY WILEY

Benchmarking biomolecular force field-based Zn^{2+} for mono- and bimetallic ligand binding sites

Okke Melse[1,2] | Iris Antes[1,2] | Ville R. I. Kaila[3] | Martin Zacharias[1]

[1]Center for Functional Protein Assemblies (CPA), Technical University of Munich, Garching, Germany

[2]SynBiofoundry@TUM, Technical University of Munich, Straubing, Germany

[3]Department of Biochemistry and Biophysics, Stockholm University, Stockholm, Sweden

Correspondence
Okke Melse and Martin Zacharias, Center for Functional Protein Assemblies (CPA), Technical University of Munich, Ernst-Otto-Fischer-Straße 8, 85748 Garching, Germany.
Email: okke.melse@tum.de and zacharias@tum.de

Funding information
CIPSM; Deutsche Forschungsgemeinschaft; Knut och Alice Wallenbergs Stiftelse; Svenska Forskningsrådet Formas; Swedish National Infrastructure for Computing

Abstract

Zn^{2+} is one of the most versatile biologically available metal ions, but accurate modeling of Zn^{2+}-containing metalloproteins at the biomolecular force field level can be challenging. Since most Zn^{2+} models are parameterized in bulk solvent, in-depth knowledge about their performance in a protein environment is limited. Thus, we systematically investigate here the behavior of non-polarizable Zn^{2+} models for their ability to reproduce experimentally determined metal coordination and ligand binding in metalloproteins. The benchmarking is performed in challenging environments, including mono- (carbonic anhydrase II) and bimetallic (metallo-β-lactamase VIM-2) ligand binding sites. We identify key differences in the performance between the Zn^{2+} models with regard to the preferred ligating atoms (charged/non-charged), attraction of water molecules, and the preferred coordination geometry. Based on these results, we suggest suitable simulation conditions for varying Zn^{2+} site geometries that could guide the further development of biomolecular Zn^{2+} models.

KEYWORDS
binding site, dummy atom model, metal ion model, Metalloproteins, zinc

1 | INTRODUCTION

Metalloproteins play an important role in many biological processes, making them interesting targets for drug design, as well as many biotechnological applications.[1–4] In this regard, transition metals enable the catalysis of a large range of reactions, and understanding their binding modes in proteins is therefore key to the development of biocatalysts as well for drug design. However, the complex coordination chemistry render it highly challenging to accurately model their interactions with the ligating atoms in intricate (biochemical) environments.[3]

Zn^{2+} is one of the most versatile biologically available metal ions, and often found in protein binding sites. For example, the ZincBind database[5] comprises more than 42,000 Zn^{2+}-containing binding sites. Zn^{2+} can play multiple roles in proteins, by either stabilizing or enforcing a certain protein conformation to support the folding process or to bring the protein in an active conformation. Zn^{2+} can adopt multiple coordination numbers and geometries in proteins, typically 4-, 5-, and 6-coordinated Zn^{2+} complexes with tetrahedral/square planar, trigonal bipyramidal, and octahedral coordination geometries, whereas the low energy barriers to switch between these binding modes makes Zn^{2+} an ideal metal ion to fulfill a variety of roles in biocatalysis.[6–9] Biomolecular simulation of the catalysis requires the inclusion of quantum mechanical (QM) potentials, in which bond-breaking and –formation events can be described. However, these QM or hybrid QM/MM simulations need a starting pose of the substrate and an equilibrated protein-ligand complex structure, which in case of the absence of a crystal structure can be retrieved from (MD-based) molecular docking and subsequent molecular dynamics simulations. Moreover, these classical simulations can also support

Iris Antes: Initial corresponding author; passed away before submission of this work.

drug design, for example, the design of an inhibitor or antagonist via high-throughput screening.[10] These applications benefit from an accurate description of Zn^{2+} coordination, ideally without restraints to allow a suboptimal pose to further refine, and the binding site to adapt on the presence of a ligand.

In biomolecular force fields, the interaction between the metal ion and its environment is solely described via the nonbonded Coulomb and Lennard–Jones (LJ) potentials. Li et al.[11] showed the difficulty of reproducing multiple experimentally measurable properties, such as hydration free energy and ion-oxygen distance with a single LJ-parameter set for divalent metal ions. Therefore, three parameter sets were developed, each optimized to reproduce hydration free energies, the ion-oxygen distances, or a combination of both. Subsequently, Li and Merz[12] presented a modified LJ-potential, constructed based on adding a $1/r^4$ term to account for ion-induced dipole interactions. Applying this new potential, called the 12-6-4 LJ-type potential, a single parameter set for each metal ion was developed that could reproduce both hydration free energies and ion-oxygen distances of the metal ion in water, as well as coordination numbers.[12,13] Several studies have shown that this model can be used to predict a wide range of properties, often after some optimization of the 12-6-4 LJ parameters. For example, Panteva et al.[14] showed that a tuned 12-6-4 LJ-type model can be applied to accurately describe nucleic acid binding, whereas Sengupta et al.[15] were able to describe chelate effects in an aqueous environment. More recently, Song et al.[16] showed that this model can also reproduce the thermodynamics of transition metal ion binding in protein binding sites, as they were able to accurately calculate absolute binding free energies of Co^{2+} and Ni^{2+} in glyoxalase I, applying an optimized 12-6-4 LJ-type potential. Since the modification of the LJ-potential requires changes in the molecular dynamics (MD) code, only a limited number of MD-engines support the use of this LJ-type. However, recent implementations has allowed exploration of the new LJ-potential via GPU-accelerated MD,[17] in combination with long-timescale dynamics. Recently, Zhang et al.[18] (re)designed the traditional 12-6 LJ model for Zn^{2+} and other divalent ions, scanning a larger LJ parameter space as compared to the design of the models described above, and leading to a model with a strongly increased well depth. Interestingly, this refined 12-6 LJ parameter reproduced both hydration free energies and ion-oxygen distances rather well. Simultaneous description of the hydration free energy and ion-oxygen distance can also be reached by the implementation of a switcher function[19] or via the application of a double exponential potential.[20] Moreover, polarized models have been developed, such as the AMOEBA[21,22] or Drude polarizable force field,[23] or the CTPOL model developed by Sakharov and Lim.[24] In addition of fitting the Zn^{2+} LJ parameters, Macchiagodena et al.[25,26] also introduced new residues for Zn^{2+} coordinating residues (cysteine, histidine, aspartate and glutamate), which were parameterized in the presence of a Zn^{2+} ion, leading to a different charge distribution and modified LJ parameters compared to the standard residues. The electronic continuum correction (ECC) approach presents another strategy to account for missing electronic polarization, which can either be incorporated via a continuum medium, or by scaling charged groups, the latter being the least intrusive method.[27–29] This leads to simulations containing metal ions and protein residues with a non-integer net charge. As a consequence of the charge scaling, the LJ parameters of the ion need to be adjusted (sometimes called ECCR because of the rescaling[30]), which was conducted for a large variety of ions, including Zn^{2+}.[31,32]

Another strategy to describe the metal coordination chemistry is via a bonded model, where the divalent coordinate bonds between the metal ion and its ligating atoms are modeled explicitly.[33,34] The bonded models generally require system-specific parameterization of the newly introduced bonds and angles with dihedral potentials generally set to zero, as well as refitting partial charges of the metal ion(s) and the coordinating residues. Multiple strategies have been proposed to parameterize the bonded parameters, including the commonly applied Seminario method,[35] in which the force constants are derived from the Cartesian Hessian matrix. However, this method requires QM calculations, which may limit the setup of high-throughput calculations, for example, in rational drug design. The metal-complex can be unstable in the QM-optimization without proper modeling of the protein environment, which makes the parameterization even more difficult. To avoid these drawbacks, Yu et al.[36] presented the Extended Zinc Amber Force Field (EZAFF), which allows a fully empirical derivation of the bonded parameters required in the bonded model.

The major disadvantage of the bonded model is the lack of ligand exchange during the simulation. Since all metal ligands are explicitly bonded to the metal ion, a change of the metal environment cannot be sampled within this approach. Already in 1990, Åqvist and Warshel[37] described a dummy atom model (also called multisite model) for Mn^{2+}, whereas in 1999, Pang[38] parameterized this type of model for a tetrahedral Zn^{2+} ion. To this end, dummy atoms carrying a portion of the mass and charges of the metal ion are placed around the ion, either in a tetrahedral or octahedral geometry. The dummy atoms can freely rotate around the metal ion without an energy penalty, but bonds between the metal ion and the dummy atoms are kept at the correct distance and geometry around the metal ion. Therefore, this alternative approach allows exchange of metal-ligating atoms, but still enforces/supports a certain coordination geometry due to the charge delocalization in the dummy atoms. More recently, Duarte et al.[39] (re)parameterized the dummy atom model for octahedral Zn^{2+} for the OPLS-AA force field, with the main focus to reproduce metal-oxygen distances, solvation free energies, and coordination numbers. Jiang et al.[40] further refined the dummy atom model parameters, using revised reference solvation free energies with Tissandier's[41] proton hydration free energies.

The presence of the different ion binding models illustrates the complexity of a classical description of a metal ion in proteins. Moreover, the nonbonded models, including the dummy atom models, are parameterized in bulk water, but there is little information about their performance in a protein environment. Yu et al.[36] compared their EZAFF-derived bonded model in several Zn^{2+}-containing systems with their bonded and nonbonded models and a selection of popular semi-empirical QM methods, based on geometry-optimized structures. However, in MD simulations, as well as in MD-based docking

simulations, it is of great value to simulate around an equilibrium state, requiring analysis of the preferred conformational state of these Zn^{2+} models. Therefore, we performed atomistic MD simulations of non-polarizable Zn^{2+} models relying on the Lennard Jones potential to analyze the characteristics of the Zn^{2+} models in a protein environment. The models benchmarked here are widely applicable and because of their simplicity easily transferrable, also to MD-based docking algorithms such as DynaDock.[42] The bonded models were solely included as reference simulations in this benchmarking study. We analyze the strengths and weaknesses of these Zn^{2+} models to reproduce experimental protein conformations around the Zn^{2+} ions, to study which Zn^{2+} models may be most suitable for simulations in which a nonbonded description of Zn^{2+} is desired. This systematic benchmarking was performed for Zn^{2+} ions in ligand binding sites, since the proper modeling of the protein environment around the ion can be challenging due to the high flexibility and presence of multiple possible Zn^{2+} ligating atoms, including solvent-mediated binding poses. Additionally, a large amount of metalloproteins, such as metallo-β-lactamases, contain two Zn^{2+} ions in the binding site. Therefore, the performance of these Zn^{2+} models was benchmarked in a monometallic as well as a bimetallic ligand binding site, using carbonic anhydrase II (CAII) and the verona integron-encoded metallo-β-lactamase (VIM-2) as model systems.

2 | MATERIALS AND METHODS

2.1 | Force fields and potential function forms

The nonbonded model in the AMBER force field is described in the following form:

$$\begin{aligned} U_{ij}^{\text{nonbonded}}(r_{ij}) &= U_{ij}^{\text{LJ}}(r_{ij}) + U_{ij}^{\text{el}}(r_{ij}) \\ &= \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \\ &= \frac{C_{12}}{r_{ij}^{12}} - \frac{C_6}{r_{ij}^{6}} + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}, \end{aligned} \tag{1}$$

where r_{ij} is the distance between atoms i and j, and $R_{\min,ij}$ and ε_{ij} are the distance between this particle pair at which the LJ potential reaches its minimum and the LJ-well depth, respectively, obtained by applying the Lorentz-Berthelot combination rules, whereas q_i and q_j are the partial charges of the respective atoms.

To improve the description of highly charged systems, Li and Merz[12] presented a modified LJ potential for all interactions with metal ions that included a $1/r^4$ term, describing the charge-induced dipole interactions. This potential is known as the 12-6-4 LJ-type potential because of its mathematical form:

$$\begin{aligned} U_{ij}^{\text{LJ}} &= \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6} - 2\kappa {R_{\min,ij}}^2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{4}\right] \\ &= \frac{C_{12}}{r_{ij}^{12}} - \frac{C_6}{r_{ij}^{6}} - \frac{C_4}{r_{ij}^{4}}, \end{aligned} \tag{2}$$

C_4 is defined as $C_4 = \kappa C_6$, for interactions between the metal ion and the oxygen of water molecules, with κ being a scaling factor with unit $Å^{-2}$ that is parameterized for a large set of divalent ions.[12] For the interactions of other atom types with metal ions, C_4 is defined as follows:

$$C_4^{\text{metal-atomtype}} = \frac{C_4^{\text{metal}}(H_2O)}{\alpha_0^{\text{metal}}(H_2O)} \times \alpha_0(\text{atom type}), \tag{3}$$

where α_0 is the polarizability of the metal ion or respective atom type.

2.2 | Zn^{2+} models

We benchmarked six classical nonbonded (NB) parameter sets: the parameters from Merz,[43] which is the standard parameter set in the AMBER ff03 force field,[44,45] as well as the three parameter sets from Li et al.[11] The latter were optimized to reproduce the hydration free energy (NB-Li^{HFE}), ion-oxygen distance (NB-Li^{IOD}) or a compromise of both (NB-Li^{CM}), where NB-Li^{CM} is the default parameter set in AMBER ff14SB. Furthermore, Zn^{2+} parameters from recent studies by Zhang et al.[18] (NB-Zhang) and Macchiagodena et al.[25,26] (NB-Macchiagodena) were also included in our benchmarking set, where the latter model contains additionally force field parameters for the coordinating residues. The largest difference between these parameter sets lies in the value of the LJ well-depth, which ranges from 7.16×10^{-4} to 2.95×10^{3} kcal mol^{-1}. Two ECC models were benchmarked as well, with parameters retrieved from either Nikitin et al.[31] or Duboué-Dijon et al.[32] Besides scaling the charge on the Zn^{2+} ions, the partial charges of the coordinating residues including the hydroxide ion and inhibitor were scaled as well, in line with Reference 46. Only the partial charges of the side chains were rescaled such that the total charge of the respective residue is scaled by factor 0.75 (Table S1). The effect of alternative protein scaling approaches on ECC simulations was studied as well. We also evaluated three available Zn^{2+} dummy atom models (DU), containing one tetrahedral and two octahedral models.[38–40,47] The dummy atoms for the dummy atom models were placed around the Zn^{2+} ions at a distance matching the equilibrium distance of the dummy atom model applied. For the tetrahedral dummy atom model by Pang,[47] the revised force constant from Park et al.[48] was used for the bonds involving dummy atoms. For this DU-Pang* model, the binding site was prepared according to the simulation setups as described in References 47 and 48. To this end, metal-coordinating histidine residues were treated in their deprotonated form (histidine anion), and in the bimetallic binding site of VIM-2, Asp118 was protonated to form a hydrogen bond with the hydroxide ion. Additionally, all residues forming a hydrogen bond with a histidine anion were protonated, with parameters for the histidine anion and hydroxide ion obtained from Reference 47. The protonation state of the remaining residues either matched experimentally determined protonation states, or were predicted, as described below in more detail.

TABLE 1 $R_{min}/2$ (Å), ε (kcal·mol^{-1}) and κ (Å^{-2}) parameters of Zn^{2+} models with the TIP3P water model evaluated in this study

Parameter set	Model	Rmin/2	ε	κ	m_{Zn}	q_{zn}	References
NB-Merz	LJ12-6	1.1000	0.012500		65.4	+2	43
NB-Li^{HFE}	LJ12-6 (HFE set)	1.1750	0.000716		65.4	+2	11
NB-Li^{IOD}	LJ12-6 (IOD set)	1.3950	0.014917		65.4	+2	11
NB-Li^{CM}	LJ12-6 (CM set)	1.2710	0.003303		65.4	+2	11
NB-Zhang	LJ12-6	0.5152	295.5289		65.4	+2	18
NB-Macchiagodena[a]	LJ12-6	1.4561	0.012500		65.4	+2	25, 26
ECC-Nikitin	LJ12-6 (ECC)	1.1900	0.015243		65.4	+1.5	31
ECC-Duboué-Dijon	LJ12-6 (ECC)	0.9822	0.250000		65.4	+1.5	32
LJ1264	12–6-4 LJ-type	1.455	0.026628	1.623	65.4	+2	12
DU-Pang[b]	Dummy atom model (Td)	3.1000	0.000001		53.38 (4×3.0)	0 (4× + 0.5)	47
DU-Duarte[b]	Dummy atom model (Oct)	0.6814	112.734		47.39 (6×3.0)	−1 (6× + 0.5)	39
DU-Jiang[b]	Dummy atom model (Oct)	0.4895	100.7481		47.39 (6×3.0)	−1 (6x + 0.5)	40
BM-Seminario[c]	Bonded model, Seminario	1.3950	0.014917		65.4	+0.5038/+0.7351[d]	33
BM-EZAFF[c]	Bonded model, EZAFF	1.3950	0.014917		65.4	+0.5038/+0.7351[d]	36

[a]Additionally contain new force field parameters for Zn^{2+} coordinating residues.
[b]Dummy atom models additionally differ in Zn^{2+}-dummy atom bond/angle/dihedral parameters.
[c]Bonded models differ in parameters regarding Zn^{2+}-bonded residues, provided in Table S2.
[d]Retrieved from the RESP fit.

Last, the performance of the bonded model (BM) was analyzed as a reference for the simulations applied with the nonbonded models, with parameters derived from either the Seminario method or the Extended Zinc Amber Force Field (EZAFF).[36] Since the partial charges used in this model are both retrieved from the same RESP procedure,[49] these values are identical between these two sets. A python-based metal center parameter builder (MCPB)[33] was used to parameterize the divalent bonds for the bonded models: the positions of the hydrogen atoms from the large model generated by the MCPB were optimized at the B3LYP/6-31G(d) level of theory. This optimized geometry was used for the RESP fitting, applying the ChgModB scheme,[33] by restraining all charges of the backbone heavy atoms to the values from the ff14SB force field, following the approach by Peters et al.[34] that produced the most accurate results. The van der Waals radius of the Zn^{2+} ion was set to 1.395 Å, obtained from the IOD set of Li et al.[11] For the empirical model (BM-EZAFF), the bonded parameters were derived from the EZAFF. For the Seminario model (BM-Seminario), a small system containing the Zn^{2+} ion and the coordination residues and ligand was generated with the MCPB, the geometry was optimized at the B3LYP/6-31G (d) level of theory, and the bonded parameters were derived based on the method by Seminario et al.[35] For VIM-2, we optimized the small system at the TPSS/def2-TZVP level of theory prior to the application of the Seminario method, as this theory level predicted stable geometries for the bimetallic site. Additionally, a small van der Waals radius obtained from the GAFF2 force field was placed on the hydrogen of the hydroxide ion. All dihedral force constants involving a Zn^{2+} ion were set to zero, based on the strategy applied in the MCPB procedure. The relevant parameters of the benchmarked Zn^{2+} models are shown in Table 1.

2.3 | Simulation setups

Carbonic anhydrase II (CAII) in complex with the RA1 inhibitor (PDB ID: 5NXG) resolved at 1.2 Å resolution was used for the benchmarking of Zn^{2+} models in monometallic systems, while the metallo-β-lactamase VIM-2 with the ANT-431 inhibitor (PDB ID: 6HF5), resolved at 1.8 Å resolution, was applied as model system for the benchmarking of bimetallic binding sites.[50,51] During the ligand parameterization procedure, the ligands were protonated with Open Babel[52] (v. 2.3.2.) at pH 7.0, keeping a deprotonated NH^- from the sulfonamide group of the CAII inhibitor, as this has been shown to be the state the ligand binds to the Zn^{2+} ion.[12,52,53] For the bonded and van der Waals parameters of the ligand atoms, the General Amber Force Field[54] (GAFF) parameters were used. Charges of the ligand atoms for all but the bonded models were derived applying the RESP procedure based on a Merz-Singh-Kollman population analysis[55,56] at a QM-geometry optimized structure of the inhibitor, which was performed at the HF/6-31G(d) level of theory, with Gaussian09 (revision E.01).[57]

The protein-ligand complexes were prepared as follows: from all multi-resolved residues, the first occurrence was selected, and all water molecules within a sphere of 8 Å of the Zn^{2+} ion(s) were preserved. The protonation states of ionizable protein residues were predicted with the PROPKA3.0 software package[58,59] at pH 7.0, followed by a visual check. For CAII, the protonation states were ensured to match experimentally determined protonation states and hydrogen bond networks.[60] For all standard protein residues, the AMBER ff14SB[61] force field was applied. The system was solvated in a rectangular box containing TIP3P[62] water, applying a buffer region of 12 Å around the protein atoms. The system was neutralized with

Na^+ ions using parameters from Joung and Cheatham,[63] using the LEaP module of the Amber18/AmberTools18 software package.[17] The bridging hydroxide ion in VIM-2 was parameterized as described in a previous study by Marion et al.[64] with −1.3 *e* and +0.3 *e* partial charges for the oxygen and hydrogen atom, respectively. The polarizability of the hydroxide oxygen, required for the C_4 parameter in the 12-6-4 LJ-type model, was set to 2.03 $Å^3$, based on Shannon et al.,[65] while the polarizability of the hydrogen atom was set to zero, analogously to the hydrogen of water. The Zn^{2+} parameters from Table 1 were applied, and for the 12-6-4 LJ-type model, the C_4 parameters were included with ParmEd.

2.4 | Simulation procedure

An energy minimization was performed using sander from Amber18/AmberTools18 with the XMIN method (ntmin = 3), thereby gradually adjusting the box size to bring the density from 0.8 to 1.0 $g{\cdot}cm^{-3}$, in steps of 0.02 g cm^{-3}, using in-house scripts. During this minimization, a positional restraint of 20.0 kcal mol $Å^{-2}$ was applied to all protein atoms. When the target density was reached, an additional minimization was performed, applying the positional restraint solely to the binding site, defined as all residues with at least one atom in a sphere of 5 Å around the Zn^{2+} ion(s). During the heat-up procedure, the temperature was gradually increased to 300 K, while decreasing the positional restraint. The precise methodology is provided in Table S3. The Langevin thermostat[66] was applied with a collision frequency of 4.0 ps^{-1}. During the MD simulations in the *NPT* ensemble, the Berendsen barostat[67] was applied to keep the pressure at 1 bar, with a relaxation time of 1 ps and compressibility of 44.6×10^{-6} bar^{-1}. Periodic boundary conditions were applied and the SHAKE algorithm[68] was used to constrain all bonds involving a hydrogen atom at their equilibrium distance. A cut-off distance of 12 Å was used for all nonbonded interactions, while the particle mesh Ewald method[69] was applied to describe long range electrostatics. The simulations were performed with an integration step size of 1 fs. After the heat-up, a 200 ns production MD simulation was performed at 300 K in an *NPT* ensemble, saving the atomic coordinates and velocities every 10 ps. The heat-up and production simulations were conducted in three replicas, with the pmemd.cuda MD engine from Amber18.

2.5 | Trajectory analyses

From each replica, the first 100 ns were considered as equilibration, thus the last 100 ns from the three replicas were merged, and analyzed as a single trajectory of 30,000 frames. The distances and distance RMSDs between Zn^{2+} and the ligating atoms were calculated with cpptraj.[70] The coordination geometry was determined every 2 ns for each Zn^{2+} individually, using a python package FindGeo.[71] Every non-carbon heavy atom within 2.8 Å of the Zn^{2+} ion was considered as ligating atom, except the sulfur from the sulfonamide moiety in the CAII ligand.

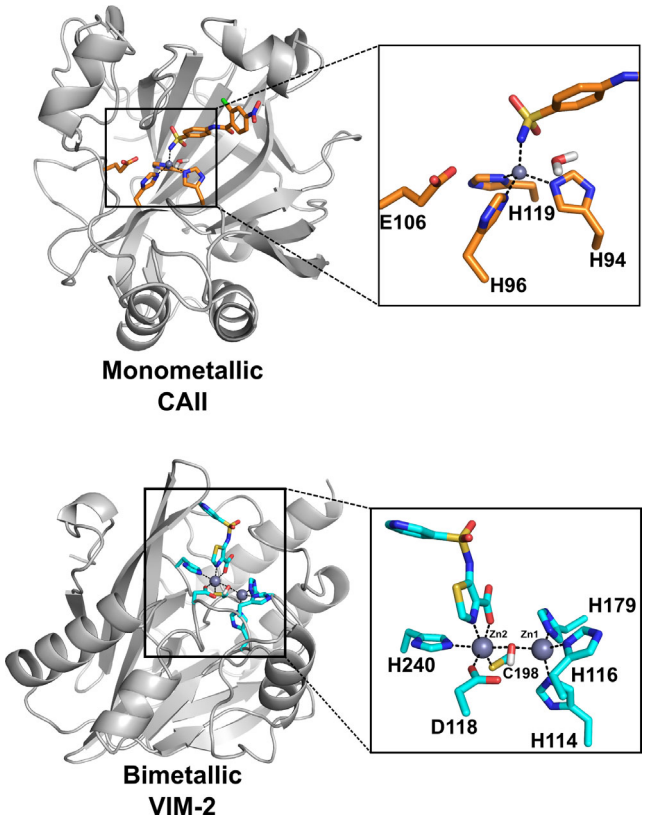

FIGURE 1 Benchmarking set used in this study. Carbonic anhydrase II (CAII) was used as model system for monometallic systems, while the β-lactamase VIM-2 was applied as model system for bimetallic systems. The proteins are shown in cartoon representations. The Zn^{2+} binding site, containing all coordinating residues and the inhibitor, as well as water molecules and protein residues that are close to the binding site and contain a heavy atom (potentially) able to coordinate the Zn^{2+} ion(s), are shown as licorice. The Zn^{2+} ions are illustrated as gray spheres. Dotted lines represent coordinate bonds

2.6 | Potential energy scan

A system containing of either a single Zn^{2+} ion or a Zn^{2+} ion with dummy atoms together with a water molecule was manually generated, followed by a geometry optimization at the MP2/cc-pVTZ level. This geometry was used for the interaction energy scanning, where the distance between the Zn^{2+} ion and the oxygen of the water molecule ranged from 1.2 to 5.0 Å with steps of 0.1 Å. For the reference calculation, the single point energies along the potential energy surface were calculated at the MP2/cc-PVTZ level of theory, applying the basis set superposition error (BSSE) correction. For the force field calculations, single point energies were calculated with AMBER, applying the Zn^{2+} parameters of the respective Zn^{2+} models, and TIP3P for the water molecule. The LJ fits for the attempt to parameterize a tetrahedral Zn^{2+} model were performed with a modified version of the FitFF suite.[72] The quantum chemical calculations were performed with Gaussion09 (revision E.01).

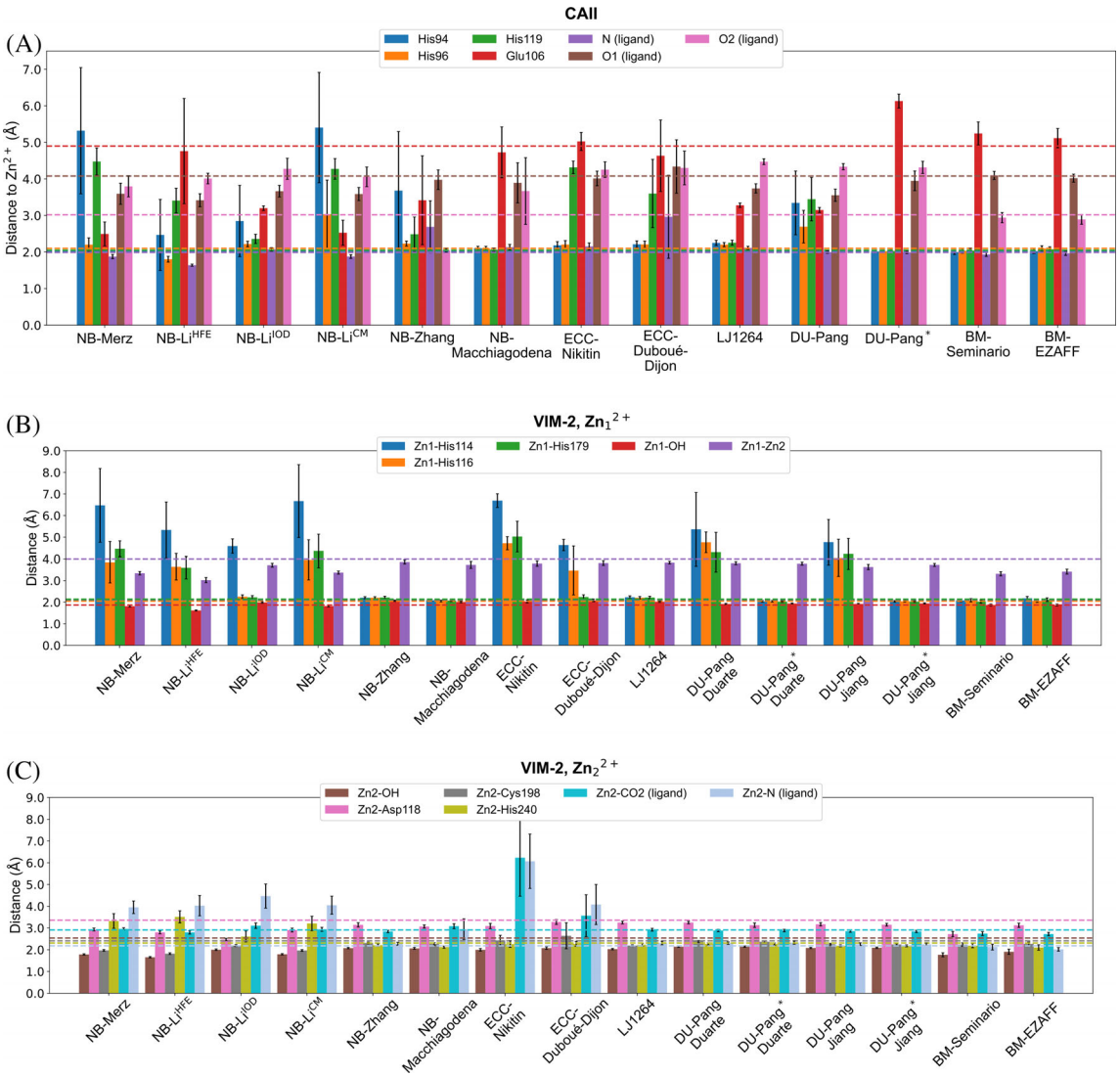

FIGURE 2 Interatomic distance between Zn^{2+} and selected binding site atoms, as a measure for binding site integrity for the Zn^{2+} ion in (A) CAII, and (B) Zn_1^{2+} and (C) Zn_2^{2+} in VIM-2. The bars represent the average distance between the Zn^{2+} ion and the ligating residue during the simulation, while the dotted lines represent the value of the respective distance in the X-ray structures. All distances to Asp or Glu are measured to the side-chain's carboxyl carbon. O1 (ligand) and O2 (ligand) represent sulfonamide oxygen atoms, N (ligand) represents the amine nitrogen of the sulfonamide moiety of the ligand. Zn2-CO_2 (ligand) and Zn2-N (ligand) represent the carboxyl carbon, and the nitrogen of the thiazole moiety from the inhibitor, respectively. Labels consisting of two dummy atom models (e.g., DU-Pang, Duarte) represent simulations where these respective models were applied for Zn_1^{2+} and Zn_2^{2+}, respectively

3 | RESULTS

3.1 | Coordination geometry in benchmarking systems

The performance of the Zn^{2+} models was analyzed in the binding sites of carbonic anhydrase II and β-lactamase VIM-2, as model systems for monometallic and bimetallic binding sites, respectively (Figure 1). CAII was used for the analysis of monometallic Zn^{2+} binding sites, since the Zn^{2+} ion in CAII is coordinated in a tetrahedral geometry, which is the dominant coordination geometry of Zn^{2+} monometallic binding sites.[8] In the CAII structure (PDB ID: 5NXG), the Zn^{2+} ion is coordinated by three histidine residues and the nitrogen of the sulfonamide moiety of the inhibitor. Due to the opposite charge between Glu106 and the Zn^{2+} ion, we hypothesized that certain Zn^{2+} models may overestimate this electrostatic interaction, and thereby falsely consider Glu106 as a coordinating residue. Therefore, the non-coordinating glutamate residue Glu106, which is close to the Zn^{2+} ion (4.0 and 5.5 Å between Zn^{2+} and glutamate oxygen atoms (OE1/OE2), respectively) was also included in the analysis. Moreover, as Glu106 plays an important role in an elaborate hydrogen-bonding network in CAII, the residue positions the binding site residues such that they can interact with the ligand, an effect which has been described in a variety of studies,[60,73,74] illustrating the importance of proper conformational sampling of this residue. Glu106 forms hydrogen bonds via a water molecule to Tyr7, and directly to the side chain

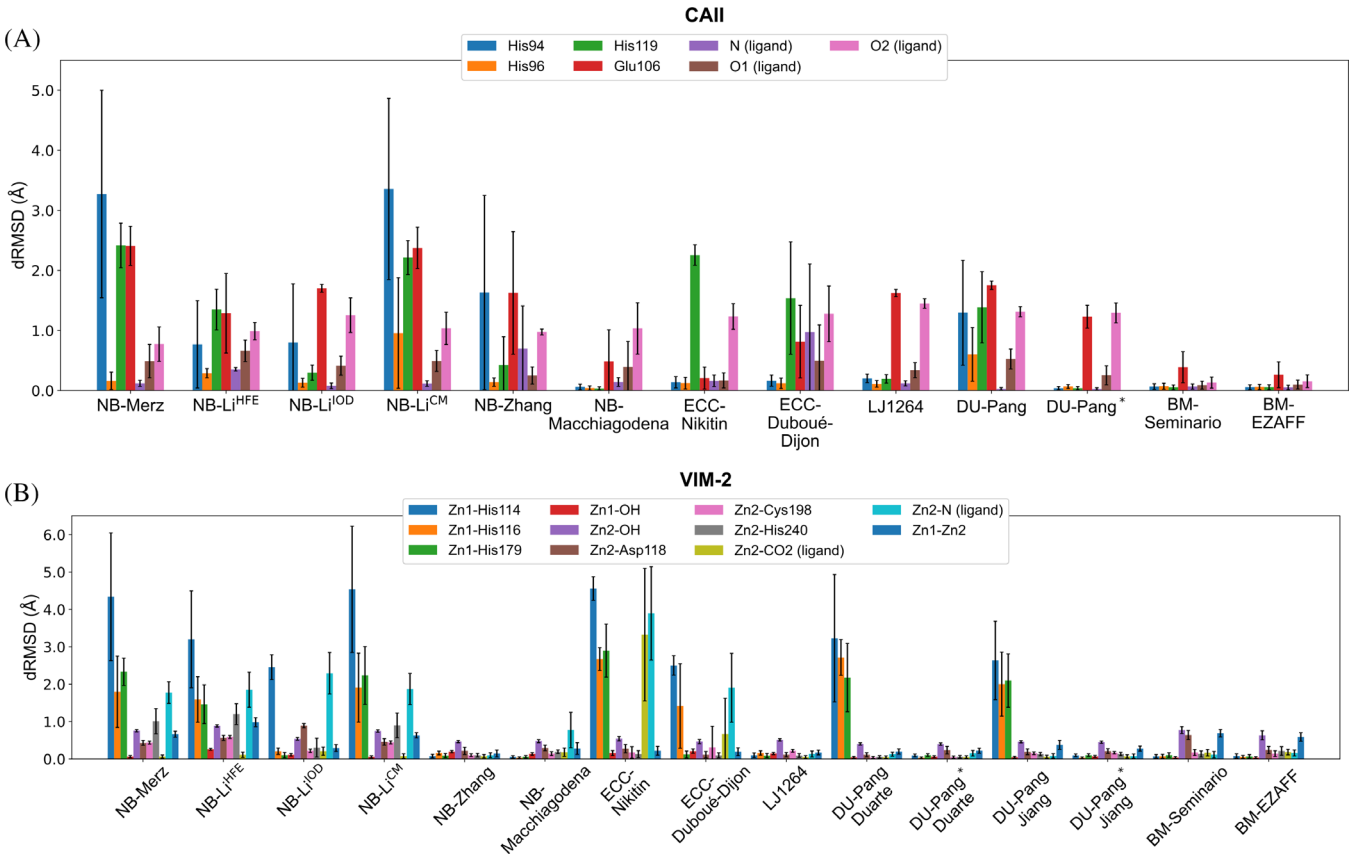

FIGURE 3 RMSD of interatomic distances (dRMSD) as a measure for the stability of the sampled Zn^{2+} binding site in (A) CAII and (B) VIM-2. dRMSDs were measured between Zn^{2+} and selected binding site atoms with respect to the average simulated interatomic distance of the respective atom pair. All distances to Asp or Glu are measured to the carboxyl carbon of the side-chain. O1 (ligand) and O2 (ligand) represent sulfonamide oxygen atoms, N (ligand) represents the amine nitrogen of the sulfonamide moiety of the ligand. Zn2-CO_2 (ligand) and Zn2-N (ligand) represent the carboxyl carbon, and the nitrogen of the thiazole moiety from the inhibitor, respectively. Labels consisting of two dummy atom models (e.g., DU-Pang, Duarte) represent simulations where these respective models were applied for Zn_1^{2+} and Zn_2^{2+}, respectively

of Thr199, which in turn interacts both via the backbone nitrogen and the side chain hydroxyl group to the sulfonamide moiety of the inhibitors.

The binding site of VIM-2 contains two Zn^{2+} ions, bridged by a hydroxide ligand (Figure 1).[75–77] The Zn_1^{2+} is coordinated in a tetrahedral geometry by three histidine residues and the bridging hydroxide, while Zn_2^{2+} adopts an octahedral coordination geometry via an aspartate, cysteine and histidine residue, the bridging hydroxide ion, as well as the thiazole nitrogen and carboxylate oxygen of the ligand. In addition to the interaction of the ligand with the Zn^{2+} ion, the ligand forms several hydrogen bonds with the protein: Arg228 forms a salt-bridge with the carboxylate group of the ligand, whereas Tyr67 forms an aromatic interaction with the pyridine moiety of the ligand.[51] The two differently coordinated Zn^{2+} ions make this an interesting model system, and allow us to probe how the models discriminate between different coordination geometries.

The studied Zn^{2+} models show clear performance differences in the benchmarking set, as well as implications of an incorrect physical description of the Zn^{2+} ion in some binding sites and models. We find that the NB-Merz, the NB-Li models, and the ECC models lead to distorted binding sites for both CAII and VIM-2, while the NB-Zhang and NB-Macchiagodena reproduce better the Zn^{2+} geometry, especially in VIM-2 (Figure 2). The octahedral dummy atom models also reproduce the conformation of the surrounding amino acids around Zn_2^{2+}, with a performance comparable to the bonded models (Figure 2C). To further analyze the stability of the Zn^{2+} binding site during the simulations, as well as the trends between the replicas, the root-mean-square-deviation was calculated over the interatomic distances (dRMSD) between the Zn^{2+} ion and a selection of atoms defining the structure of the binding site (Figure 3). The dRMSD was calculated with respect to the average distance (over all three replicas) simulated between the respective atom pair rather than to the respective distance in the X-ray structure, as this results in an indicator for the stability of Zn^{2+} ligation, that is, the average deviation from the equilibrium distance simulated. Small dRMSD values thus indicate a stable ligation of the Zn^{2+} ion, but not necessarily a correct ligation.

For CAII, the nonbonded models NB-Merz and NB-Li^{CM} have comparable average dRMSDs (averaged over the ligating atoms) of 1.38 Å and 1.51 Å, while the other nonbonded models perform better with average dRMSDs below 1.00 Å. A similar performance was

TABLE 2 Zn^{2+} coordination geometries sampled (in percentage).

Coordination geometry	NB-Merz	NB-Li^{HFE}	NB-Li^{IOD}	NB-Li^{CM}	NB-Zhang	NB-Macchia-godena	ECC-Nikitin	ECC-Duboué-Dijon	LJ-1264	DU-Pang	DU-Pang*	DU-Duarte	DU-Jiang	BM-Seminario	BM-EZAFF
	CAII geometry occupancy (%)														
Tetrahedral[a]	1.3	**99.3**	0.0	0.7	0.0	2.7	0.0	0.0	0.0	20.7	**96.7**	n/a	n/a	**80.7**	**84.0**
Octahedral	86.7	0.0	94.7	96.0	12.7	26.7	100.0	100.0	95.3	18.7	0.0	n/a	n/a	0.0	0.0
Trigonal bipyramidal	0.0	0.0	0.0	2.0	0.0	34.0	0.0	0.0	0.0	10.0	1.3	n/a	n/a	0.0	0.0
Square pyramidal	4.0	0.0	1.3	1.3	0.0	3.3	0.0	0.0	0.0	4.0	0.0	n/a	n/a	0.0	0.0
Irregular geometry	6.7	0.7	0.7	0.0	0.0	32.0	0.0	0.0	0.0	38.0	2.0	n/a	n/a	19.3	16.0
Other[b]	1.3	0.0	3.3	0.0	87.3	1.3	0.0	0.0	4.7	8.7	0.0	n/a	n/a	0.0	0.0
	Zn_1^{2+} geometry occupancy (%)														
Tetrahedral[a]	0.0	**96.0**[c]	0.0	0.0	0.0	**62.7**	0.0	0.0	0.0	4.7	**99.3**	n/a	n/a	**89.3**	**94.0**
Octahedral	95.3	0.0	100.0	96.7	100.0	0.0	98.7	100.0	100.0	22.7	0.0	n/a	n/a	0.0	0.0
Trigonal bipyramidal	0.0	1.3	0.0	0.0	0.0	26.7	1.3	0.0	0.0	40.7	0.0	n/a	n/a	0.0	0.7
Square pyramidal	4.0	0.0	0.0	2.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	n/a	n/a	0.0	0.0
Irregular geometry	0.7	2.7	0.0	0.7	0.0	3.3	0.0	0.0	0.0	27.3	0.0	n/a	n/a	0.7	4.7
Other[b]	0.0	0.0	0.0	0.0	0.0	7.3	0.0	0.0	0.0	4.7	0.7	n/a	n/a	10.0	0.7
	Zn_2^{2+} geometry occupancy (%)														
Tetrahedral	98.0	100.0	0.0	89.3	0.0	0.0	0.0	0.0	0.0	n/a	n/a	0.0	0.0	0.0	0.0
Octahedral[a]	0.0	0.0	43.3	0.0	**100.0**	**56.0**	**58.7**	**82.0**	**99.3**	n/a	n/a	**100.0**	**100.0**	**100.0**	**100.0**
Trigonal bipyramidal	0.7	0.0	0.0	2.0	0.0	4.0	31.3	1.3	0.0	n/a	n/a	0.0	0.0	0.0	0.0
Square pyramidal	0.0	0.0	0.0	0.0	0.0	32.7	9.3	4.7	0.0	n/a	n/a	0.0	0.0	0.0	0.0
Irregular geometry	1.3	0.0	18.7	8.7	0.0	6.7	0.7	6.0	0.0	n/a	n/a	0.0	0.0	0.0	0.0
Other[b]	0.0	0.0	38.0	0.0	0.0	0.7	0.0	6.0	0.7	n/a	n/a	0.0	0.0	0.0	0.0

Note: The experimental coordination geometry is marked with the superscript a, and models reproducing this marked in bold font.

Abbreviation: n/a, not applicable.

[a]Native coordination geometry, as observed in the X-ray structure for CAII (PDB ID: 5NXG) and VIM-2 (PDB ID: 6HF5).

[b]All coordination geometries which were visited less than 5% in all evaluated Zn^{2+} models in both CAII and VIM-2 were merged in this category. A complete list of all sampled coordination geometries for CAII and VIM-2 is provided in Tables S4 and S5, respectively.

[c]All coordinating ligands were water molecules or a hydroxide ion, and in one replica the carboxylate oxygen of the ligand.

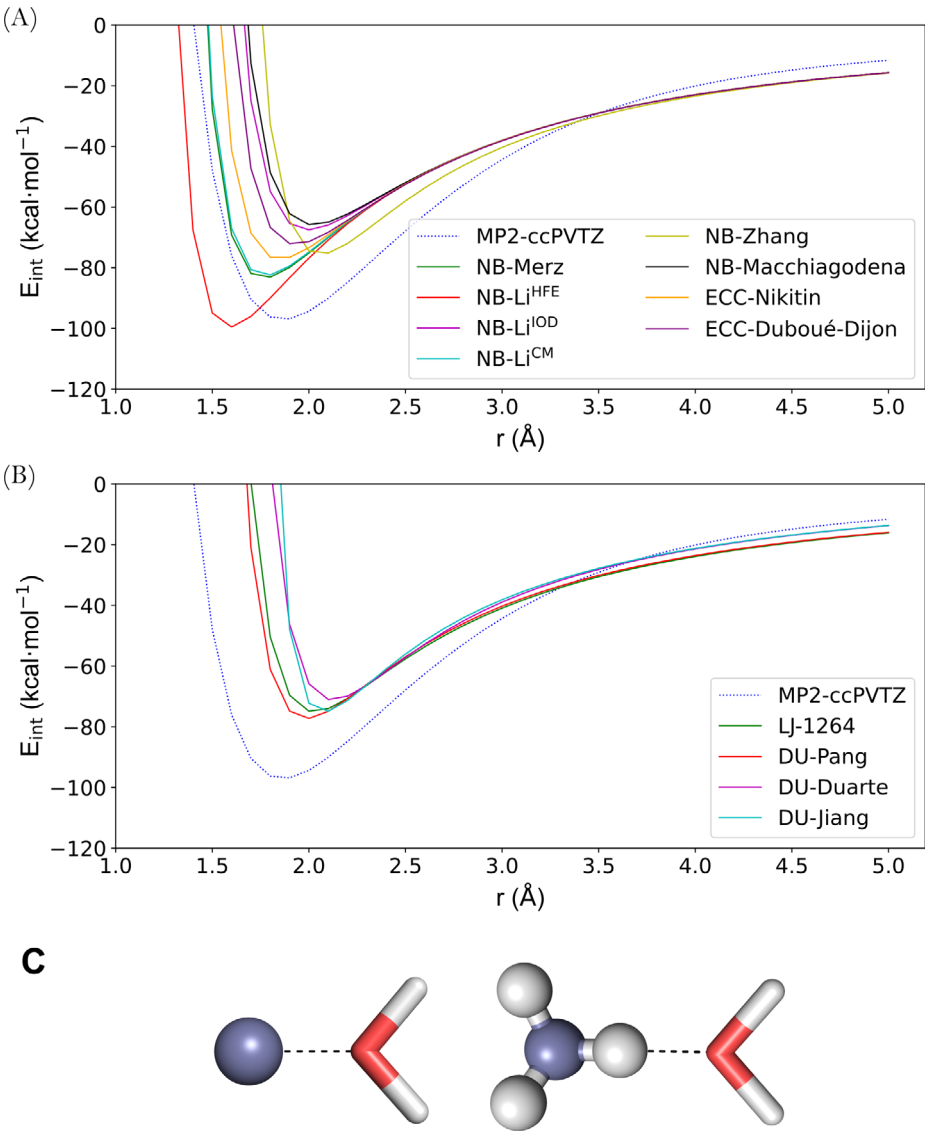

FIGURE 4 Potential energy surfaces between Zn^{2+} and a TIP3P water molecule, observed for (A) nonbonded and ECC Zn^{2+} models, as well as (B) 12-6-4 LJ-type, and (B) dummy atom models. The classical energy represents the full interaction energy, including both the Lennard–Jones and Coulomb terms. A BSSE-corrected MP2/cc-pVTZ PES scan shown as reference, in a blue, dotted line. (C) The model system used for the PES scans

observed for the tetrahedral dummy atom models with average dRMSDs ranging from 0.42 to 0.99 Å, depending on the protonation state of coordinating residues. As expected, the bonded models lead to the most stable Zn^{2+} ligation with average dRMSDs below 0.20 Å. Similar trends were observed for the VIM-2 systems, with the exception of the highly stable ligation observed for NB-Zhang and the 12-6-4 LJ-type model (average dRMSD of 0.16 Å for both models). The ECC models show high dRMSD values especially for ligations including histidine residues or the inhibitor. Simulations of VIM-2 applying NB-Pang* and DU-Duarte or DU-Jiang led to the most stable Zn^{2+} ligation, with average dRMSDs of 0.13 Å and 0.15 Å, respectively.

3.2 | Coordination geometry

We next studied how the coordination geometry is reproduced by the different Zn^{2+} models. This coordination geometry determines the exact position of the Zn^{2+} ligating atoms, which can play an important role in the stabilization of the protein structure, or introducing strain on certain bonds that can be important for catalysis. The experimental coordination geometry of the Zn^{2+} ion in CAII is tetrahedral, which is also sampled in the NB-Li^{HFE}, DU-Pang*, as well as the bonded models (Table 2). The remaining Zn^{2+} models either prefer an octahedral Zn^{2+} geometry, with 87%–100% occurrence, or sample multiple coordination geometries, including trigonal bipyramidal and square pyramidal geometries. Similar trends can also be observed for the bimetallic VIM-2 system (Table 2). The octahedral coordination geometry for Zn_1^{2+} is clearly preferred by most nonbonded models, while a tetrahedral coordination geometry is observed in the X-ray structure of VIM-2. The experimental coordination geometry for Zn_2^{2+} is octahedral based on to the X-ray structure (PDB ID: 6HF5), which is correctly reproduced by the NB-Zhang, NB-Machhiagodena (to 56%), the ECC models (to 59% and 82%), and 12-6-4 LJ-type model, in addition to the dummy atom and bonded models (Table 2).

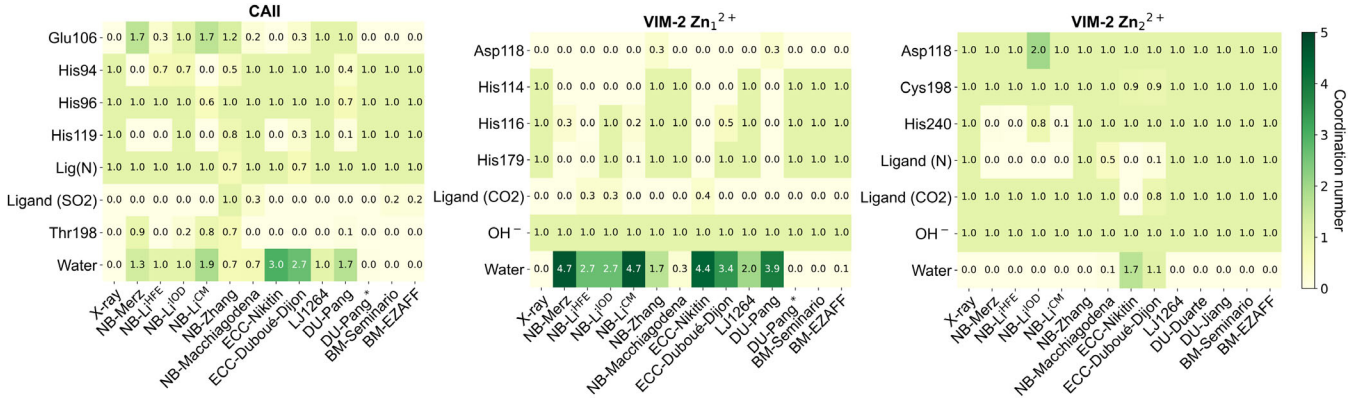

FIGURE 5 Contribution of ligating residues to the coordination of Zn^{2+} during simulations with the corresponding Zn^{2+} models. The values represent a normalized coordination number for each ligating residue. A residue is defined as ligating if the ligating atom of this residue is present within the coordination sphere (here: 2.8 Å radius) around the Zn^{2+} ion. The heat maps are colored from light to dark green, ranging from small to large coordination numbers

Finally, the potential energy surface (PES) scan between Zn^{2+} and water generated by the evaluated models further illustrate the large differences between these models (Figure 4). For example, only NB-Li^{HFE} reached the energy minimum calculated by the MP2/cc-pVTZ reference scan, but this minimum was reached at a much shorter atomic separation (1.6 Å and 1.9 Å, respectively). We note that the PES for NB-Merz and NB-Li^{CM} are almost identical, whereas the other nonbonded models show a large variation, both in R_{min} and well depth.

4 | DISCUSSION

We have benchmarked here a large set of available Zn^{2+} models, which can be subdivided in five model types: classical nonbonded models (NB), electronic continuum correction models (ECC), the 12-6-4 LJ-type model (LJ1264), dummy atom models (DU) and bonded models (BM). All these models except NB-Macchiagodena, DU-Pang, and the bonded models were parameterized for hexa-coordinated Zn^{2+}, either by design (as for the dummy atom models) or as an effect of the parameterization environment: bulk water in which Zn^{2+} adopts a hexa-coordination.[78,79] Since little is known about the performance of the nonbonded models in protein environment, especially ligand binding sites, the performance and preferred coordination geometry of the nonbonded models was evaluated here for both the tetrahedral and octahedral Zn^{2+} ions, aiming to identify suitable simulation conditions for MD-based docking simulations.

4.1 | Nonbonded models

Proper sampling of the metal ion coordination geometry is highly important, as it defines both structural and thermodynamical properties of the metal-protein interactions. However, this property is often neglected.[80,81] We found that the coordination geometry is least accurately reproduced by the nonbonded models (Table 2), often predicting incorrect binding modes for both CAII and VIM-2 (Figures 2 and 5). This observation is consistent with findings from a study of Mg^{2+} models by Zuo and Liu.[81] Although the NB-Li^{HFE} model reproduces the tetrahedral coordination geometry for the tetrahedral Zn^{2+} ions (CAII, and Zn_1^{2+} in VIM-2, Table 2), it predicted incorrect ligation: in all replicas, His119 was replaced by a ligating water molecule and the ligating role of His94 was taken over by Glu106 in one of the replicate simulations. This effect was also observed for NB-Merz, NB-Li^{CM} and NB-Zhang, for which most coordination bonds with the histidine residues were broken and replaced by water molecules, Glu106, or the sulfonamide oxygen atoms of the ligand (Figure 5). Simulations of VIM-2 applying these nonbonded models, with the exception of NB-Zhang, further showed that all coordination bonds between the Zn^{2+} ions and the histidine residues, as well as the ligand's thiazole nitrogen were broken and replaced by oxygen atoms, mostly from water molecules (Figures 5 and 6). In simulations applying NB-Zhang, additional water molecules were attracted in order to fill an octahedral coordination, but the interactions with the histidine residues were maintained (Figures 2, 5, and 6). Our data reveals a strong preference of NB-Merz, NB-Li^{HFE}, NB-Li^{CM}, and to a lesser extend the ECC models and NB-Zhang, towards hard bases as ligating atoms (as in water). The only exception is the interaction between Zn_2^{2+} and Cys198 in VIM-2, which remained intact in all Zn^{2+} models, despite the rather soft base character of cysteine. However, the interaction between Cys198 and Zn_2^{2+} may be prevailed by the coulomb term, as this residue was modeled in the deprotonated state that could favor the observed stable interaction mode.

Both NB-Merz and NB-Li^{CM} models show similar dRMSD measures and preferred coordinating residues (Figures 2, 3, and 5), and interaction energy profile (Figure 4). Furthermore, both models preferred an octahedral coordination for CAII and Zn_1^{2+} in VIM-2, while they simulated a tetrahedral coordination geometry for Zn_2^{2+} (Table 2). The instable simulation trajectories and incorrect Zn^{2+} coordination sampled by these models result mainly from unstable binding of the

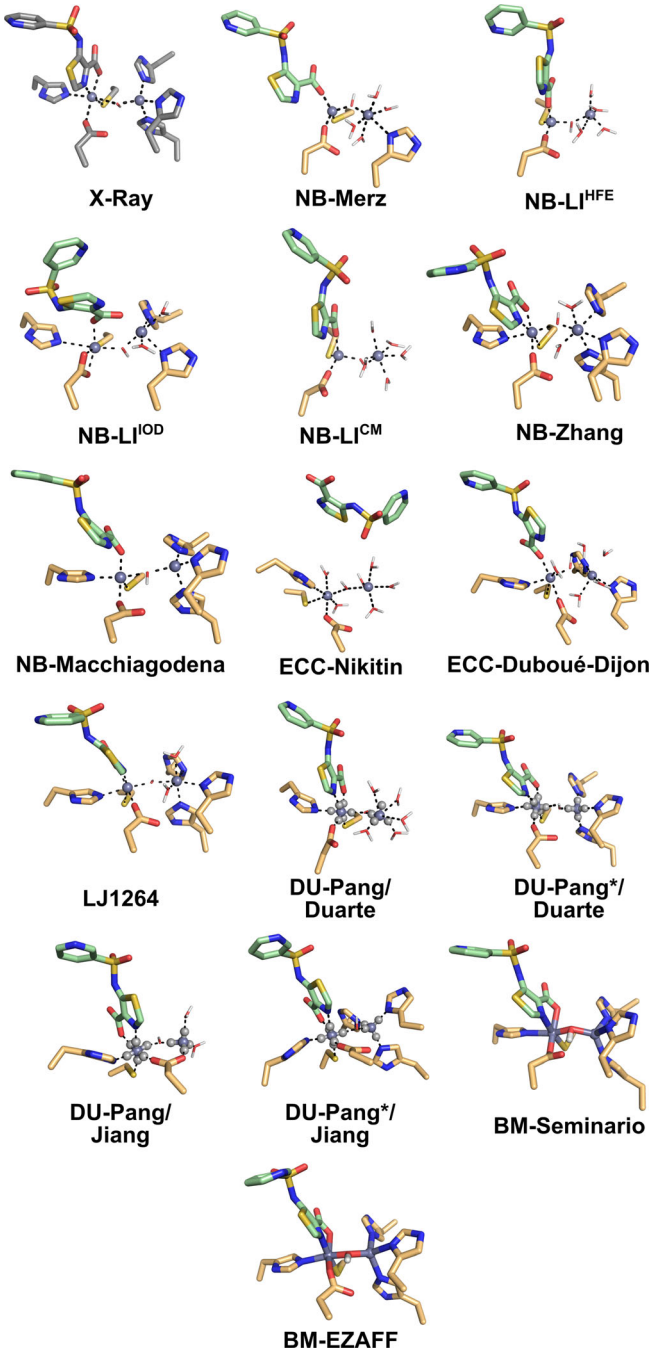

FIGURE 6 Sampled binding sites of VIM-2 by the evaluated Zn^{2+} models. Carbon atoms of the reference X-ray structures are shown in gray, and the ligand and coordinating residues in the sampled structures are shown in respective green and light orange. Zn^{2+} ions are represented as gray spheres, and the dummy atoms as small light gray spheres. Dotted lines represent the coordination geometry

Zn^{2+} ions, as the Zn^{2+} ions move up to 3 Å away from their crystallized position during the simulations (Figure S1). NB-Li^{IOD}, which was optimized to reproduce ion-oxygen distances, thus properly describe the Zn^{2+}-His interactions for His96 and His119 in CAII and His116 and His179 in VIM-2, in contrast to the other 12-6 LJ nonbonded models as discussed above. However, this nonbonded model does not predict the interaction between the metal and His94 in CAII (in one replica) and His114 in VIM-2 (in all replicas), as well as the thiazole nitrogen of the ligand in VIM-2 (Figures 5 and 6). We also observe that an octahedral coordination geometry was sampled with the NB-Li^{IOD} model for all evaluated Zn^{2+} ions (Table 2), suggesting that this model does not reproduce tetrahedral binding geometries.

4.2 | Fitted force field parameters for zinc-binding residues

The NB-Macchiagodena model performed rather well in terms of simulating correct Zn^{2+}-ligating atom distances: these interatomic distances only differ on average 0.17 Å and 0.23 Å with respect to the X-ray structures, for CAII and VIM-2, respectively (Figure 2). Moreover, this model performs well in the choice of ligating atoms, where interactions with all histidine residues remained intact due to the use of modified force field parameters for the Zn^{2+}-coordinating residues (Figure 5). Furthermore, no coordination with charged residues was observed, in contrast to the other nonbonded models. However, multiple coordination geometries were observed during the simulation of both CAII and VIM-2 (Table 2), which could result from an incorrect description of the interaction between Zn^{2+} and residues other than the "special" coordinating residues, including the ligand. For example, the interaction between Zn_1^{2+} and the thiazole nitrogen in VIM-2 is transient in the simulation, which could explain the sampling of a square pyramidal coordination geometry (Table 2). Moreover, Zn_2^{2+} often attracted an additional water molecule in addition to the hydroxide ion and histidine residues, resulting in a five-coordinated square pyramidal geometry. In CAII, alternating coordination with the sulfonamide oxygen atoms of the ligand, and surrounding water molecules resulted in a variety of different sampled coordination geometries (Table 2). These observations indicate that NB-Macchiagodena performs best when solely protein residues coordinate the Zn^{2+} ion, for which modified force field parameters are available (currently: His, Cys, Asp, and Glu). In an attempt to reduce this dependency on these coordinating residues, we combined the force field parameters of the coordinating residues from NB-Macchiagodena with the Zn^{2+} parameters from NB-Li^{CM}. Surprisingly, this led to highly stable simulations of the tetrahedral zinc ions (i.e., CAII and Zn_1^{2+} of VIM-2; Figure S3 and Table S6). Zn_2^{2+} of VIM-2 was additionally observed in a tetrahedral coordination, resulting in the dissociation of the interaction between the ion and the ligand (Figure S3).

4.3 | Effect of electronic polarization and ion-induced dipole effects

The ECC models rely on charge scaling to account for missing electronic polarization, and thereby the overestimation of interactions between charged groups, an approach which has been discussed repeatedly.[27–29] While the charge of the metal ion is generally scaled by a factor of 0.75 in the charge scaling ECC approach, various

strategies to scale the protein residues can be found in literature.[29,30,46] The ECC models benchmarked here contain scaled charges on side chains of coordinating residues, resulting in a net charge of the residue scaled by factor 0.75, as described in the Materials and Methods section. We observed that the ECC models perform slightly better in maintaining Zn^{2+}-His coordination in our benchmarking compared to NB-Merz and the NB-Li models. Both ECC models maintain the ligation with His94 and His96 in CAII, as well as His240 and His179 (the latter only for ECC-Duboué-Dijon) in VIM-2, but breaking the interactions with the remaining histidine residues. ECC-Nikitin is the only model in our benchmarking set breaking the interaction between Zn_2^{2+} and the carboxylate oxygen of the inhibitor in VIM-2, thereby completely losing Zn^{2+} coordination by the inhibitor (Figures 5, 6 and S2). During these simulations, Zn_1^{2+} was completely ligated by water and the hydroxide ion, leading to a large displacement (>4 Å) of this ion from its crystallized position (Figure S1). Moreover, a preference towards an octahedral coordination geometry was observed, in which additional water molecules were attracted to fill the remaining ligation sites. In contrast, Del Frate et al.[46] were able to simulate tetrahedral Zn^{2+} in CAII (without inhibitor) with an ECC model by specifically parameterize coordinating histidine residues to avoid water displacement, illustrating the sensitivity of this method on the applied protein parameters.

Besides the Lennard–Jones parameters on Zn^{2+}, we further show that alternative scaling approaches can significantly affect the description of the metal site (Figure S4 and Table S7). For example, simulations with scaled charges for all protein residues completely lost binding of both Zn^{2+} ions in VIM-2, especially when combined with ECC-Nikitin parameters. Simulations with only slightly modified charges of coordinating residues to restore neutrality (in line with Reference 30) performed similarly as the ECC models with scaled side chains. Thus, these benchmarking results illustrate the sensitivity of the ECC models on LJ parameters and charge scaling approaches, indicating that there is still room to further optimize this method for simulations of protein-ligand complexes.

The inclusion of ion-induced dipole effects via the 12-6-4 LJ-type model leads to a significantly improved description of interactions between Zn^{2+} and particularly soft/borderline bases compared to the nonbonded models applying standard force field parameters for coordinating residues (Figures 2 and 5). All interatomic distances between Zn^{2+} and ligating nitrogen atoms simulated by the 12–6-4 LJ-type model deviate <0.20 Å from the X-ray structure. However, we observed that the 12-6-4 LJ-type model almost solely models the Zn^{2+} ion in an octahedral coordination geometry (Table 2). This can be problematic, for example when MD simulations are used to refine docked ligand poses, as Zn^{2+} often adopts a tetrahedral coordination in protein structures.[8] The octahedral coordination geometry in the tetrahedral binding sites CAII, and Zn_1^{2+} in VIM-2, was obtained by ligation of additional water molecules in the simulations applying the 12-6-4 LJ-type potential (Figure 5), an effect which was also observed in simulations with other metal ions.[82] Furthermore, our simulations of CAII applying the 12-6-4 LJ-type model overshoot the interaction between Zn^{2+} and the charged Glu106, as illustrated by the too short distance between them (3.28 Å vs. 4.90 Å observed in the X-ray structure; Figure 2). Finally, the 12-6-4 LJ-type model additionally improved the sampling of the ligand binding pose in VIM-2 (Figures 6 and S2): the heavy-atom RMSD of the ligand is 1.84 Å in the simulations applying the 12-6-4 LJ-type potential, using the crystal structure as reference, while the ligand RMSD of the other NB-Li models are higher than 2.80 Å.

4.4 | Dummy atom models and the influence of protonation states

The octahedral Zn^{2+} dummy atom models, DU-Duarte and DU-Jiang, the latter being a re-parameterization of DU-Duarte based on revised solvation free energies,[40] both contain six dummy atoms.[39,40] Both the DU-Duarte and DU-Jiang models perform well for the octahedral Zn_2^{2+} in VIM-2, applying the correct ligating residues (Figure 5), and adopt an octahedral geometry throughout the entire simulation (Table 2). The simulation of the tetrahedral binding sites can be improved substantially by the DU-Pang* dummy atom model. This model sampled the tetrahedral binding sites almost exclusively in their correct coordination geometry: 96.7% and 99.3% for CAII, and Zn_1^{2+} in VIM-2, respectively (Table 2). However, this result could only be reached applying the (artificial) deprotonated histidine residues and modified protonation states of the other coordinating residues, as also used by Pang.[38] The performance was drastically reduced once the histidine residues were modeled in their neutral protonation states as in the other Zn^{2+} models. In this "DU-Pang" set, no stable coordination geometry was observed, and only 20.7% and 4.7% of the simulation showed a tetrahedral coordination for CAII and Zn_1^{2+} in VIM-2, respectively. Furthermore, following Pang's rules to determine which residues should be protonated, the required protonation state of Asp118 in VIM-2 remains ambiguous. Asp118 should be deprotonated according to Pang's rules because of its interaction with Zn_2^{2+}, yet it must be protonated because of its hydrogen-bonding with the hydroxide ion. Therefore, we performed the simulations applying the DU-Pang* model with both protonated and deprotonated Asp118: the simulation containing a deprotonated Asp118 resulted in a highly distorted binding site, thereby losing ligand binding (data not shown), while the simulation with a protonated Asp118 stayed rather close to the experimentally observed structure. Thus, the simulation with the protonated Asp118 represents the DU-Pang* model in this study. This protonated Asp118 did however not form a hydrogen bond with the hydroxide ion during the simulation, which was the rationalization to protonate this residue. In contrast, Asp118 formed a hydrogen bond with solvent molecules and thereby losing its coordination of the Zn^{2+} ion (Figures 2 and 5).

The above results show that Zn^{2+} ions adopting an octahedral coordination geometry can be simulated with an octahedral dummy atom model, but the description of the tetrahedral Zn^{2+} remains challenging. Therefore, we decided to probe if a re-parameterization of DU-Pang could lead to a dummy atom model without the requirements of deprotonated coordinating residues. The standard parameterization procedure of nonbonded- and dummy atom models consists of a fitting or scanning procedure of LJ parameters to

reproduce hydration free energies and ion-oxygen distances in bulk water. Zn^{2+} however adopts an octahedral coordination geometry in bulk water, making this approach not suitable for a tetrahedral dummy atom model. Therefore, we attempted to parameterize a tetrahedral dummy atom model by fitting the LJ parameters to reproduce the MP2 interaction energy profile between a dummy atom model and TIP3P water. We fitted both a LJ 12-6 and a 12-6-4 LJ-type model, but none of these models showed convincing performance in MD simulations (Figure S5). Thus, a proper description of the interaction energy profile does not necessarily reflect stable metal ligation and coordination, indicating that more sophisticated methods may be required to parameterize a tetrahedral dummy atom model.

Finally, we investigated if a nonbonded model and a dummy atom model can be combined in the bimetallic system VIM-2. To this end, we simulated Zn_1^{2+} with NB-Macchiagodena, and Zn_2^{2+} with DU-Jiang, that is, the most promising models in both categories. This resulted in highly stable simulations, where both ions were simulated in the correct coordination geometry with a highly stable ligation (Figure S3C and Table S6).

4.5 | Restraining metal-coordination via explicit bonds

The reference simulations applying bonded models reproduced the Zn^{2+} sites rather accurately, which was expected because of the system-specific parameterization and application of explicit bonds between the Zn^{2+} ions and ligating atoms. Correct usage of the ligating residues, proper reproduction of the coordination geometry and highly stable simulation trajectories were observed. It is however surprising that simulations of VIM-2 applying the dummy atom models DU-Duarte and DU-Jiang were even more stable than the bonded models, and the simulated bimetallic binding site geometries were closer to the experimental structure than for both bonded models, while not having the disadvantage of fixed ligating residues. Especially the distance between the two Zn^{2+} ions (which were connected via explicit bonds in the bonded model via the hydroxide ion) was simulated closer to the experimental value by these dummy atom models (Figure 2). This is a result of the MCPB approach, in which BM-Seminario is parameterized based on a QM-geometry optimized structure, during which the distance between the Zn^{2+} ions reduced to 3.67 Å, with respect to 4.00 Å in the X-ray structure (which lies within the experimental uncertainty of the experimental structure). The distance between the two Zn^{2+} ions during the BM-Seminario simulations was even smaller (3.32 Å). Finally, we barely observed any performance differences between BM-Seminario and BM-EZAFF, indicating that the empirical derivation of the bonded parameters works well, both for monometallic and bimetallic binding sites.

5 | CONCLUSIONS

Caution should be taken when applying nonbonded Zn^{2+} models in biomolecular simulations, in particular in challenging environments such as ligand binding sites. Our benchmarking of a large variety of Zn^{2+} models show that interactions between Zn^{2+} and non-charged ligating atoms are often underestimated, such as the imidazole nitrogen in histidine residues. These ligating atoms are replaced by water molecules or charged residues during the simulation. The performance can be improved by the application of special AMBER coordinating residues (NB-Macchiagodena), or by the introduction of the ion-induced dipole interactions via the 12-6-4 LJ-type potential. However, we also show that the 12-6-4 LJ-type model strongly prefers an octahedral coordination geometry and overestimates interactions with charged atoms, although this was found to have only a minor effect on the binding site conformations sampled in this study. The NB-Macchiagodena model showed a strong improvement concerning the description of protein residues with the Zn^{2+} ion, but interactions with a ligand showed partially perturbed coordination geometries. However, combination of the NB-Macchiagodena coordinating residues and the Zn^{2+} parameters of NB-Li^{CM} led to highly stable coordination geometries in the simulated systems, indicating that there is potential to further improve this model to overcome its current limitations. The ECC models on the other hand only slightly improved the description of Zn^{2+} ligation with non-charged residues, and were shown to be sensitive on both Zn^{2+} parameters and the applied charge scaling approach, leaving space for further optimization of these type of models for biomolecular (here: protein) simulations.

We also found that the dummy atom models establish good alternatives to the nonbonded models, from which DU-Pang* reproduces a tetrahedral coordination geometry rather well, although the performance was also found to strongly depend on the protonation state of the metal surrounding. Since many Zn^{2+} ions in proteins are found in tetrahedral coordination geometries, the development of a tetrahedral dummy atom model parameterized specifically in a protein environment would be highly useful.

Our benchmarking calculations suggest that the octahedral dummy atom models DU-Duarte and DU-Jiang are good choices for Zn^{2+} ions coordinated in an octahedral geometry because of their solid performance in this study regarding the reproduction of the environment of the metal. The difference in performance between DU-Duarte and DU-Jiang is small, whereas the latter showed the best performance in this current benchmarking set. In the case of the bimetallic VIM-2 binding site, these octahedral dummy atom models perform even slightly better than the bonded models studied here. If the application allows for bonded models, the EZAFF bonded model may suffice. BM-Seminario and BM-EZAFF perform almost identical in our benchmarking set, while BM-EZAFF has the advantage of being independent of QM-calculations for its parameterization, except for the charge fitting. In summary, our benchmarking data illustrates the importance of a suitable simulation setup for Zn^{2+}-containing ligand binding sites, revealing preferences and trends of the evaluated Zn^{2+} models. The data can support the design of suitable modeling setups and guide further development of biomolecular Zn^{2+} models.

ACKNOWLEDGMENTS

The authors thank the Deutsche Forschungsgemeinschaft for financial support via SFB 749, project C08 and CIPSM. VRIK is supported by the Knut and Alice Wallenberg (KAW) foundation, and the Swedish National Infrastructure for Computing (SNIC 2021/6-143), partially funded by the Swedish Research Council through grant agreement no. 2016-07213. We thank Prof. Dr. Antoine Marion for useful insights and discussions. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Okke Melse https://orcid.org/0000-0002-0021-3466

REFERENCES

[1] L. Riccardi, V. Genna, M. De Vivo, *Nat. Rev. Chem.* **2018**, *2*(7), 100.
[2] T. N. Parac-Vogt, A. Erxleben, G. Schenk, R. Prabhakar, *Front. Chem.* **2019**, *7*, 599.
[3] P. Li, K. M. Merz, *Chem. Rev.* **2017**, *117*(3), 1564.
[4] U. T. Bornscheuer, *Philos. Trans. R. Soc. A* **2018**, *376*(2110), 20170063.
[5] S. M. Ireland, A. C. R. Martin, *Database* **2019**, *2019*.
[6] M. Dudev, J. Wang, T. Dudev, C. Lim, *J. Phys. Chem. B* **2006**, *110*(4), 1889.
[7] G. Kuppuraj, M. Dudev, C. Lim, *J. Phys. Chem. B* **2009**, *113*(9), 2952.
[8] M. Laitaoja, J. Valjakka, J. Jänis, *Inorg. Chem.* **2013**, *52*(19), 10983.
[9] H. Vahrenkamp, *Dalton Trans.* **2007**, *42*, 4751.
[10] P. Śledź, A. Caflisch, *Curr. Opin. Struct. Biol.* **2018**, *48*, 93.
[11] P. Li, B. P. Roberts, D. K. Chakravorty, K. M. Merz, *J. Chem. Theory Comput.* **2013**, *9*(6), 2733.
[12] P. Li, K. M. Merz, *J. Chem. Theory Comput.* **2014**, *10*(1), 289.
[13] P. Li, L. F. Song, K. M. Merz, *J. Chem. Theory Comput.* **2015**, *11*(4), 1645.
[14] M. T. Panteva, G. M. Giambaşu, D. M. York, *J. Phys. Chem. B* **2015**, *119*(50), 15460.
[15] A. Sengupta, A. Seitz, K. M. Merz, *J. Am. Chem. Soc.* **2018**, *140*(45), 15166.
[16] L. F. Song, A. Sengupta, K. M. Merz, *J. Am. Chem. Soc.* **2020**, *142*(13), 6365.
[17] Case, D. A., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, T. E., III, Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Ghoreishi, D., Gilson, M. K., Gohlke, H., Goetz, A. W., Greene, D., Harris, R., Homeyer, N., Huang, Y., 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.; AMBER. University of California, San Francisco, **2018**.
[18] Y. Zhang, Y. Jiang, J. Peng, H. Zhang, *J. Chem. Inf. Model.* **2021**, *61*(8), 4031.
[19] C. S. Babu, C. Lim, *J. Phys. Chem. A* **2006**, *110*(2), 691.
[20] V. H. Man, X. Wu, X. He, X.-Q. Xie, B. R. Brooks, J. Wang, *J. Chem. Theory Comput.* **2021**, *17*(2), 1086.
[21] J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson, T. Head-Gordon, *J. Phys. Chem. B* **2010**, *114*(8), 2549.
[22] J. Zhang, W. Yang, J.-P. Piquemal, P. Ren, *J. Chem. Theory Comput.* **2012**, *8*(4), 1314.
[23] H. Yu, T. W. Whitfield, E. Harder, G. Lamoureux, I. Vorobyov, V. M. Anisimov, A. D. MacKerell, B. Roux, *J. Chem. Theory Comput.* **2010**, *6*(3), 774.
[24] D. V. Sakharov, C. Lim, *J. Am. Chem. Soc.* **2005**, *127*(13), 4921.
[25] M. Macchiagodena, M. Pagliai, C. Andreini, A. Rosato, P. Procacci, *J. Chem. Inf. Model.* **2019**, *59*(9), 3803.
[26] M. Macchiagodena, M. Pagliai, C. Andreini, A. Rosato, P. Procacci, *ACS Omega* **2020**, *5*(25), 15301.
[27] I. V. Leontyev, A. A. Stuchebrukhov, *J. Chem. Theory Comput.* **2010**, *6*(5), 1498.
[28] E. Duboué-Dijon, M. Javanainen, P. Delcroix, P. Jungwirth, H. Martinez-Seara, *J. Chem. Phys.* **2020**, *153*(5), 050901.
[29] I. Leontyev, A. Stuchebrukhov, *Phys. Chem. Chem. Phys.* **2011**, *13*(7), 2613.
[30] M. Kohagen, M. Lepšík, P. Jungwirth, *J. Phys. Chem. Lett.* **2014**, *5*(22), 3964.
[31] A. Nikitin, G. Del Frate, *J. Comput. Chem.* **2019**, *40*(28), 2464.
[32] E. Duboué-Dijon, P. E. Mason, H. E. Fischer, P. Jungwirth, *J. Phys. Chem. B* **2018**, *122*(13), 3296.
[33] P. Li, K. M. Merz, *J. Chem. Inf. Model.* **2016**, *56*(4), 599.
[34] M. B. Peters, Y. Yang, B. Wang, L. Füsti-Molnár, M. N. Weaver, K. M. Merz, *J. Chem. Theory Comput.* **2010**, *6*(9), 2935.
[35] J. M. Seminario, *Int. J. Quantum Chem* **1996**, *60*(7), 1271.
[36] Z. Yu, P. Li, K. M. Merz, *J. Chem. Theory Comput.* **2018**, *14*(1), 242.
[37] J. Åqvist, A. Warshel, *J. Am. Chem. Soc.* **1990**, *112*(8), 2860.
[38] Y.-P. Pang, *J. Mol. Model.* **1999**, *5*(10), 196.
[39] F. Duarte, P. Bauer, A. Barrozo, B. A. Amrein, M. Purg, J. Åqvist, S. C. L. Kamerlin, *J. Phys. Chem. B* **2014**, *118*(16), 4351.
[40] Y. Jiang, H. Zhang, T. Tan, *J. Chem. Theory Comput.* **2016**, *12*(7), 3250.
[41] M. D. Tissandier, K. A. Cowen, W. Y. Feng, E. Gundlach, M. H. Cohen, A. D. Earhart, J. V. Coe, T. R. Tuttle, *J. Phys. Chem. A* **1998**, *102*(40), 7787.
[42] I. Antes, *Proteins: Struct. Funct. Bioinform* **2010**, *78*(5), 1084.
[43] K. M. Merz, *J. Am. Chem. Soc.* **1991**, *113*(2), 406.
[44] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, P. Kollman, *J. Comput. Chem.* **2003**, *24*(16), 1999.
[45] M. C. Lee, Y. Duan, *Proteins: Struct. Funct. Bioinform* **2004**, *55*(3), 620.
[46] G. Del Frate, A. Nikitin, *ChemistrySelect* **2018**, *3*(43), 12367.
[47] Y.-P. Pang, *Proteins: Struct. Funct. Bioinform* **2001**, *45*(3), 183.
[48] J. G. Park, P. C. Sill, E. F. Makiyi, A. T. Garcia-Sosa, C. B. Millard, J. J. Schmidt, Y.-P. Pang, *Biorg. Med. Chem* **2006**, *14*(2), 395.
[49] C. I. Bayly, P. Cieplak, W. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*(40), 10269.
[50] A. Pecina, J. Brynda, L. Vrzal, R. Gnanasekaran, M. Hořejší, S. M. Eyrilmez, J. Řezáč, M. Lepšík, P. Řezáčová, P. Hobza, P. Majer, V. Veverka, J. Fanfrlík, *ChemPhysChem* **2018**, *19*(7), 873.
[51] S. Leiris, A. Coelho, J. Castandet, M. Bayet, C. Lozano, J. Bougnon, J. Bousquet, M. Everett, M. Lemonnier, N. Sprynski, M. Zalacain, T. D. Pallin, M. C. Cramp, N. Jennings, G. Raphy, M. W. Jones, R. Pattipati, B. Shankar, R. Sivasubrahmanyam, A. K. Soodhagani, R. R. Juventhala, N. Pottabathini, S. Pothukanuri, M. Benvenuti, C. Pozzi, S. Mangani, F. De Luca, G. Cerboni, J.-D. Docquier, D. T. Davies, *ACS Infect. Dis* **2019**, *5*(1), 131.
[52] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J. ChemInform* **2011**, *3*(1), 33.
[53] A. Pecina, M. Lepšík, J. Řezáč, J. Brynda, P. Mader, P. Řezáčová, P. Hobza, J. Fanfrlík, *J. Phys. Chem. B* **2013**, *117*(50), 16096.
[54] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*(9), 1157.

[55] B. H. Besler, K. M. Merz, P. A. Kollman, *J. Comput. Chem.* **1990**, *11*(4), 431.
[56] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*(2), 129.
[57] 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, J. A. J., 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, J. M., 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. and Fox, D. J., Gaussian 09, Revision E.01; Gaussian, Inc.: Wallingford CT, **2016**.
[58] M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*(2), 525.
[59] C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski, J. H. Jensen, *J. Chem. Theory Comput.* **2011**, *7*(7), 2284.
[60] A. Kovalevsky, M. Aggarwal, H. Velazquez, M. J. Cuneo, M. P. Blakeley, K. L. Weiss, J. C. Smith, S. Z. Fisher, R. McKenna, *Structure* **2018**, *26*(3), 383.
[61] J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser, C. Simmerling, *J. Chem. Theory Comput.* **2015**, *11*(8), 3696.
[62] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* **1983**, *79*(2), 926.
[63] I. S. Joung, T. E. Cheatham, *J. Phys. Chem. B* **2008**, *112*(30), 9020.
[64] A. Marion, M. Groll, D. H. Scharf, K. Scherlach, M. Glaser, H. Sievers, M. Schuster, C. Hertweck, A. A. Brakhage, I. Antes, E. M. Huber, *ACS Chem. Biol.* **2017**, *12*(7), 1874.
[65] R. D. Shannon, R. X. Fischer, *Phys. Rev. B* **2006**, *73*(23), 235111.
[66] R. J. Loncharich, B. R. Brooks, R. W. Pastor, *Biopolymers* **1992**, *32*(5), 523.
[67] H. J. Berendsen, J. v. Postma, W. F. Van Gunsteren, A. Di Nola, J. Haak, *J. Chem. Phys.* **1984**, *81*(8), 3684.
[68] J.-P. Ryckaert, G. Ciccotti, H. J. Berendsen, *J. Comput. Phys.* **1977**, *23*(3), 327.
[69] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*(19), 8577.
[70] D. R. Roe, T. E. J. Cheatham, *Chem. Theory Comput.* **2013**, *9*(7), 3084.
[71] C. Andreini, G. Cavallaro, S. Lorenzini, *Bioinformatics* **2012**, *28*(12), 1658.
[72] A. Marion, J. Góra, O. Kracker, T. Fröhr, R. Latajka, N. Sewald, I. Antes, *J. Chem. Inf. Model.* **2018**, *58*(1), 90.
[73] Z. Liang, Y. Xue, G. Behravan, B. H. Jonsson, S. Lindskog, *Eur. J. Biochem.* **1993**, *211*(3), 821.
[74] V. M. Krishnamurthy, G. K. Kaufman, A. R. Urbach, I. Gitlin, K. L. Gudiksen, D. B. Weibel, G. M. Whitesides, *Chem. Rev.* **2008**, *108*(3), 946.
[75] F. R. Salsbury, M. W. Crowder, S. F. Kingsmore, J. J. A. Huntley, *J. Mol. Model.* **2008**, *15*(2), 133.
[76] M. Aitha, A. R. Marts, A. Bergstrom, A. J. Møller, L. Moritz, L. Turner, J. C. Nix, R. A. Bonomo, R. C. Page, D. L. Tierney, M. W. Crowder, *Biochemistry* **2014**, *53*(46), 7321.
[77] S. Skagseth, S. Akhter, M. H. Paulsen, Z. Muhammad, S. Lauksund, Ø. Samuelsen, H.-K. S. Leiros, A. Bayer, *Eur. J. Med. Chem.* **2017**, *135*, 159.
[78] E. S. Marcos, J. M. Martínez, R. R. Pappalardo, *J. Chem. Phys.* **1996**, *105*(14), 5968.
[79] D. H. Powell, P. M. N. Gullidge, G. W. Neilson, M. C. Bellissent-Funel, *Mol. Phys.* **1990**, *71*(5), 1107.
[80] M. T. Panteva, G. M. Giambaşu, D. M. York, *J. Comput. Chem.* **2015**, *36*(13), 970.
[81] Z. Zuo, J. Liu, *J. Chem. Inf. Model.* **2019**, *59*(1), 399.
[82] H. MacDermott-Opeskin, C. A. McDevitt, M. L. J. O'Mara, *J. Chem. Theory Comput.* **2020**, *16*(3), 1913.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: O. Melse, I. Antes, V. R. I. Kaila, M. Zacharias, *J. Comput. Chem.* **2023**, *44*(8), 912. https://doi.org/10.1002/jcc.27052