Article

Comparison of Empirical Zn²⁺ Models in Protein–DNA Complexes

Senta Volkenandt ⬤ and Petra Imhof *⬤

Computer Chemistry Center, Friedrich-Alexander-Universität (FAU) Erlangen-Nürnberg, Nägelsbachstrasse 25, 91052 Erlangen, Germany
* Correspondence: petra.imhof@fau.de

Abstract: Zinc ions are the second most abundant ions found in humans. Their role in proteins can be merely structural but also catalytic, owing to their transition metal character. Modelling their geometric–coordination versatility by empirical force fields is, thus, a challenging task. In this work, we evaluated three popular models, specifically designed to represent zinc ions with regard to their capability of preserving structural integrity. To this end, we performed molecular dynamics simulations of two zinc-containing protein–DNA complexes, which differed in their zinc coordination, i.e., four cysteines or two cysteines and two histidines. The most flexible non-bonded 12-6-4 Lennard–Jones-type model shows a preference for six-fold coordination of the Zn^{2+}-ions in contradiction to the crystal structure. The cationic dummy atom model favours tetrahedral geometry, whereas the bonded extended zinc AMBER force field model, by construction, best preserves the initial geometry of a regular or slightly distorted tetrahedron. Our data renders the extended zinc AMBER force field the best model for structural zinc ions in a given geometry. In more complicated cases, though, more flexible models may be advantageous.

Keywords: zinc ions; protein–DNA complex; molecular dynamics simulations

1. Introduction

Metalloproteins make up a large portion of proteins in general [1] and of human proteins in particular [2]. The most abundant metal ion after iron found in humans is zinc [3,4]. It can play functional roles in enzymes such as the RNA polymerase [5]. It can also have mainly structural and stabilising roles, in particular in DNA-binding proteins. Transcription factors are an important group of DNA-binding proteins. There, the most typical zinc-binding motif is the zinc finger (ZnF), where a β-sheet with two cysteine residues is followed by an α-helix with two histidine residues, resulting in a tetrahedral C2H2-coordination of zinc [6]. Another common example of tetrahedral coordination is provided by four cysteine residues (C4), which can be found in the hormone receptor family. In such receptors, the zinc ions stabilise the secondary structure, such that protein–DNA and protein–protein contacts can be established when binding as a dimer to DNA [7,8].

Zinc is a transition metal (TM) and, thus, does not form pure covalent or ionic, i.e., non-bonded, but instead coordinate bonds, which share characteristics with both types and are thus more flexible and can break easier than covalent bonds [9]. In biological systems, zinc exists as a divalent cation Zn^{2+} with an electronic configuration of $[Ar]3d^{10}$ without its two 4s-electrons. Consequently, the bonding characteristic is complex due to the complicated shape of the d-orbitals.

In solution, a Zn^{2+}-ion is coordinated by six water molecules, whereas in proteins, it can have coordination numbers (CNs) of four, five, or six in different geometries [10,11]. The rather low energy barriers for switching the CN and geometry are ideal for catalysing reactions in enzymes [2]. Structural zinc ions are generally coordinated by four residues [11].

While a full quantum mechanical (QM) description of zinc and its immediate environment would provide the most accurate representation of all properties and effects, the

typical sizes and timescales, especially up to microseconds, required to describe biological zinc-containing systems of interest are computationally too demanding for QM methods. Therefore, it is desirable to describe zinc ions within classical force fields for molecular dynamics (MD) simulations. The difficulty, however, is the trade-off between a force field model that is flexible enough to allow changes in CN and geometry around the Zn^{2+}-ion while still maintaining structural integrity.

For that reason, as reviewed by Li et al. [9], many different models for describing all kinds of ions, including Zn^{2+}, have been developed and validated over the past few decades. Among the non-polarisable models, the most popular ones are non-bonded, bonded, and cationic dummy models. Zinc parameters of polarisable models, such as those for the AMOEBA force field [12], do exist, but will not be discussed here because they need more parameters due to their complexity and variety, as well as smaller time steps compared to unpolarised models for helping energy conservation [9].

The most intuitive treatment of metal ions is to just use the non-bonded parts of the potential energy function,

$$U_{\text{NB}}(r_{ij}) = U_{\text{vdW}}(r_{ij}) + U_{\text{Coulomb}}(r_{ij})$$
$$= \varepsilon_{ij} \left[\left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}}. \tag{1}$$

The ion only has electrostatic and van der Waals (vdW) interactions with its surroundings and possesses the formal charge of the metal ion. The model is very simple and requires only a few parameters. ε_{ij} is the potential well depth, r_{ij} is the distance between atoms i and j, and R_{min} is the distance at which the Lennard–Jones (LJ) potential is minimum. q_i and q_j are the charges of the two atoms and ε_0 and ε_r are the vacuum and relative permittivity, respectively. Interactions of the ion with ligands can be formed and broken. Using the 12-6 LJ potential for vdW interactions reproduces the ionic interaction between monovalent ions and low electronegativities (such as alkali metal ions) very well [9]. This model underestimates short-range interactions and cannot reproduce hydration free energy (HFE) and ion oxygen distance (IOD) simultaneously [13]. Macchiagodena et al. developed a new approach for parametrising the coordinating protein ligands by redefining their atomic charges as well as their LJ parameters [14,15]. This approach is aimed at reproducing mean distance distributions. Melse et al. found that the combination of the force field parameters of the coordinating residues from Macchiagodena et al. with the non-bonded CM Zn^{2+} parameters from Li et al. led to highly stable simulations of the tetrahedral Zn^{2+}-ions even on longer time scales [13,16].

To additionally mimic dipole interactions, Li and Merz proposed adding a $1/r^4$ term to the 12-6 LJ potential [17]. In the AMBER force field, it is described as follows:

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

where κ is a scaling factor and has the unit Å^{-2}. With this new 12-6-4 LJ-type potential HFE, IOD and CN can be reproduced simultaneously in water. Those parameters are available for a multitude of water models [18]. Song, Sengupta, and Merz later optimised this model (m12-6-4) further for TM ions in the protein environment, so as to reproduce the solvation free energy of the TM ion and thermodynamics of the TM ion-ligand coordination as well as ion binding to an enzyme's active site. With this approach, Song et al. calculated parameters for cobalt and nickel ions [19]. Panteva et al., on the other hand, optimised

the 12-6-4 LJ-type potential parameters of several TM ions, including Zn^{2+}, for balanced interactions with nucleic acids [20].

In general, ions are parametrised under infinite dilution by reproducing experimental properties. Certain properties (e.g., free energies) require extensive sampling, which makes a direct comparison with experimental data difficult [9]. The water model used for parametrisation also plays an important role. The same parameters can produce, e.g., different HFEs in different water models but similar structural properties [9,13]. If the parameters are determined in an aqueous solution, which is typically the case, they are consequently biased toward the CN of this environment, so tetrahedrally coordinated Zn^{2+} will likely change to a 6-coordination in the presence of other potential coordination partners such as water molecules [21].

Another strategy is to describe the zinc ion via a bonded model, as in the (extended) zinc AMBER force field (ZAFF/EZAFF), where the ligating atoms are covalently bonded to the ion [22,23]. The potential for a zinc ion now includes not only the non-bonded electrostatic and vdW terms but also environment-dependent bond, angle, and dihedral terms. In this model, the zinc ion only carries a partial charge whereas LJ parameters are directly taken from the non-bonded model. The ZAFF is an empirical model and was developed for the TIP3P water model to best represent fourfold-coordinated zinc ions in a protein environment [22]. It was later extended and methodologically refined (EZAFF) to include not only four- but also five- and six- coordinated systems [23]. EZAFF is available in the metal center parameter builder MCPB.py, which is part of the AmberTools [24]. The coordination sphere and the resulting electrostatics are most accurately described by the bonded model but a lot of non-transferable parameters are needed. This model is the most rigid one as it does not allow an exchange between ligands or metal ions and CN [21]. It is actually problematic when zinc is also coordinated by at least two adjacent water molecules as the 1-3 non-bonded interaction is usually not considered, meaning that the two water oxygen atoms do not repel each other electrostatically. On the other hand, a 1-4 non-bonded interaction is included, and the oxygen of one water molecule and the hydrogen of the adjacent water molecule strongly attract each other without experiencing vdW repulsion because water hydrogen atoms typically have zero vdW parameters. By this attraction, the angle is distorted [25].

A hybrid version of the non-bonded and bonded model is the cationic dummy atom model (CDAM). In the CDAM, covalently bonded dummy atoms, according to the CN, are added to the ion. They usually have an equilibrium distance of 0.9 Å and carry fractional masses and charges. They all sum up to the formal mass and charge of the ion. Consequently, the charge of the ion itself can even be negative and has no physical meaning [9]. Pang introduced and validated a CDAM for Zn^{2+} in tetrahedral geometry in 1999, where the ion itself was neutral and the four dummy atoms carried a charge of $+0.5\,e$ each [26–28]. Duarte et al. later developed and Jiang et al. then refined parameters for octahedral coordination [29,30]. There, each dummy atom possesses a charge of $+0.5\,e$ and the ion consequently a charge of $-1\,e$. The CDAM does not need as many parameters as the bonded model. An exchange of ligands is possible but only works in practise if the first shell ligands are anionic, and the alternative ligands of the second sphere are neutral or positively charged [21].

The quality of different models for divalent ions, including zinc, has been studied extensively within proteins [16,21,23,25,31–33]. In this study, we focus on structure-stabilising tetrahedrally coordinated Zn^{2+}-ions in the context of protein–DNA systems as they may have additional challenges on the $>1\,\mu s$ time scale. We take the Wilms tumor protein (WT1) $ZnF2 - 4$ in complex with DNA as an example of C2H2 coordination [34] and the CXXC domain of mixed-lineage leukemia 1 (MLL1) in complex with DNA [35] as an example of C4 coordination. In those two seemingly simple proteins, the multiple zinc ions are responsible for their structural stability, which is crucial for proper DNA binding.

2. Materials and Methods

2.1. Model Setup and Zn^{2+} Models

The initial structures of WT1 (PDB ID: 5kl2; resolution: 1.69 Å) [34,36] and the MLL1 CXXC domain (PDB ID: 4nw3; resolution: 2.82 Å) [35,37], both in complex with their cognate DNA structure, were taken from the PDB [38]. Only coordinates of the protein–DNA complexes and zinc ions were used. Missing side chains were added with the AutoPSF plugin of VMD [39]. In the case of WT1, the overhanging bases of the DNA were deleted. Protonation states of histidine residues were verified by visual inspection and by using PlayMolecule ProteinPrepare [40], which utilises PROPKA 3.1. The sites of cysteines and histidines facing a Zn^{2+}-ion are all unprotonated (Table S1). Then both of the systems were prepared in the AmberTools suite with three different zinc force field models, i.e., the nonbonded 12-6-4 Lennard–Jones-type model [17], the tetrahedral cationic dummy atom model [26], and the bonded extended zinc AMBER force field model [23], which are shortly described in the following. All models were solvated in cubic boxes with TIP3P water [41] and a minimum buffer distance of 12.5 Å between the protein–DNA system and the box sides. DNA was described by the parmbsc1 parameters [42] and protein by the AMBER ff14SB [43]. The systems were neutralised by adding Na^+ ions.

2.1.1. 12-6-4 Lennard–Jones-Type Potential (12-6-4)

Compared to the normal 12-6 LJ potential [13], which is used by default in AMBER, additional parameters are needed for the C_4 term of the 12-6-4 LJ-type potential (see Equation (2)) [17]. They can be added with ParmEd as part of AmberTools [44,45].

2.1.2. Cationic Dummy Atom Model (CDAM)

For setting up the CDAM, both cysteine and histidine residues need to be in their anionic forms, CYM and HIN, respectively. Since GLU427 (index 77) of WT1 forms a hydrogen bond with HID431 (index 81), it was changed to its neutral form GLH, and the coordinates of OE1 and OE2 were switched to account for correctly protonated oxygen. The setup procedure is described in more detail online, where the corresponding parameter files for AMBER can be found [46].

2.1.3. Extended Zinc AMBER Force Field (EZAFF)

For building the bonded EZAFF model, bonds between the ions and surrounding ligands as well as their charges can be parametrised with the help of the metal center parameter builder MCPB.py [24,47]. There, a large model consisting of ions and surrounding ligands is generated. The positions of hydrogen atoms in this model were optimised with Gaussian 16 [48] using the B3LYP functional and the 6-31G(d) basis set. Based on the geometry-optimised model, partial charges were calculated via the RESP fitting procedure with the ChgModB scheme [22], where the charges of backbone heavy atoms were restrained to AMBER ff14SB values [43]. Bond and angle force constants are taken from an empirical relation between equilibrium bond/angle values and force constants.

2.2. Molecular Dynamics Simulations

The solvated systems were geometry-optimised with restraints of $50 \, \text{kcal·mol}^{-1} \cdot \text{Å}^{-2}$ on protein and DNA for 5000 steps, followed by 10,000 optimisation steps without restraints. The method of minimisation was switched from steepest descent to conjugate gradient after the first 500 and 1000 steps, respectively. After that, during a 500 ps simulation, the systems were heated to 298.15 K in an NVT ensemble with weak restraints of $50 \, \text{kcal·mol}^{-1} \cdot \text{Å}^{-2}$ on protein and DNA. For each system, three independent runs of 2 µs were performed in the NPT ensemble with Langevin dynamics at 298.15 K and 1 bar (weak pressure coupling, isotropic position scaling, pressure relaxation time of 2 ps, and collision frequency of $2 \, \text{ps}^{-1}$). We used an integration time step of 2 fs and bonds involving hydrogen atoms were constrained via SHAKE [49]. Periodic boundary conditions, a non-bonding cut-off of 10 Å and particle mesh Ewald [50,51] for treating electrostatic interactions were also applied.

To prevent fraying of the DNA termini, Watson–Crick and C1′-C1′ distance restraints with a force constant of $2\,kcal\cdot mol^{-1}\cdot Å^{-2}$ were imposed as in [52] in accordance with B-DNA geometry.

All simulations were performed on GPUs using pmemd.cuda of Amber 18 [44] or Amber 22 [45]. For each system, we conducted three individual runs of 2 µs, run1, run2, and run3, with different initial velocities. The first 100 ns were considered NPT equilibrations and were not used for analysis.

2.3. Analysis

Root mean square deviations (RMSDs) were calculated for backbone atoms with respect to the crystal structure, while root mean square fluctuations (RMSFs) only included C_α atoms for protein and P atoms for DNA. The median structure (i.e., the structure that is closest to the average structure) based on the same atoms of each replica was used as the reference.

Radial distribution functions of water around the zinc ions were calculated considering only the oxygen atoms. These were determined for chunks of 50 ns and then averaged to have the standard deviation as an error estimate.

Hydrogen bonds were defined based on geometric criteria, i.e., a maximum donor–acceptor distance of 3.2 Å and a donor–hydrogen–acceptor angle deviating from linearity by not more than 42 °.

The fraction of native contacts was calculated using the definition by Best, Hummer, and Eaton [53]

$$Q(X) = \frac{1}{N} \sum_{(i,j)} \frac{1}{1 + \exp[\beta(r_{ij}(X) - \lambda r_{ij}^0)]},$$ (3)

where X is a conformation, $r_{ij}(X)$ is the distance between heavy atoms i and j in conformation X and r_{ij}^0 is the corresponding distance in the native state conformation, i.e., the crystal structure. N is the number of all pairs of heavy atoms (i, j) belonging to different residues θ_i and θ_j such that $|\theta_i - \theta_j| > 3$ and $r_{i,j}^0 < 4.5\,Å$. As suggested by Best et al., we used a smoothing factor of $\beta = 5\,Å^{-1}$ and a fluctuation parameter of $\lambda = 1.8$.

For determining the metal coordination geometry, FindGeo was used on frames taken every 1 ns [54]. This tool finds metal ions in a given PDB file, identifies the coordinating atoms that are less than 2.8 Å away (C- and H-atoms excluded), and compares the given coordination geometry via RMSD calculation after superimposition to a library of various ideal geometries. The best-fitting geometry is then the one that is the lowest in RMSD.

DNA parameters were calculated with Curves+ and Canal [55].

All analyses were performed using self-written Python scripts based on MDTraj [56]. For analysis, all three runs of 1.9 µs were merged.

3. Results

Throughout this section, the Zn^{2+}-ions are named after their occurrence in the structures, i.e., ZN1 and ZN2 for MLL1, and ZN1, ZN2, and ZN3 for WT1, respectively (cf. Figure 1).

3.1. Protein Structure

In Figure 1, representative median structures show that the 12-6-4 model deviates the most from the initial crystal structure for both proteins whereas the other two models are in very good agreement with the crystal structures of both proteins, MLL1 and WT1.

RMSD and RMSF values were calculated to probe the overall stability and flexibility of the systems.

In the case of MLL1, high protein RMSD values (Figure 2) are partially due to more flexible behaviour in the C-terminal end, as can be seen from the RMSF in Figure 3. Such behaviour cannot be attributed to one specific model but can rather occur in all of them. Another reason for large RMSD values is the change in secondary structure, especially from sheet to coil or even to helix (Figure 4). With regard to the secondary structure, all of the

models look rather comparable with 12-6-4 being slightly worse. A closer look at the time series of the secondary structures (Figure S1) reveals that in none of the three runs in the 12-6-4 model the beta-sheet, formed by residues 1153–1154 and 1198–1199 (indices 3–4 and 48–49), reforms after being lost at approximately 0.2, 0.75, and 1 μs in the respective runs. A similar behaviour can be observed in only one run of the CDAM model. In the EZAFF model, the sheet structure is also lost in one run at approximately 1 μs,but reforms again (Figure 2). All coordinating residues behave similarly across all models in terms of RMSF (Figure 3).
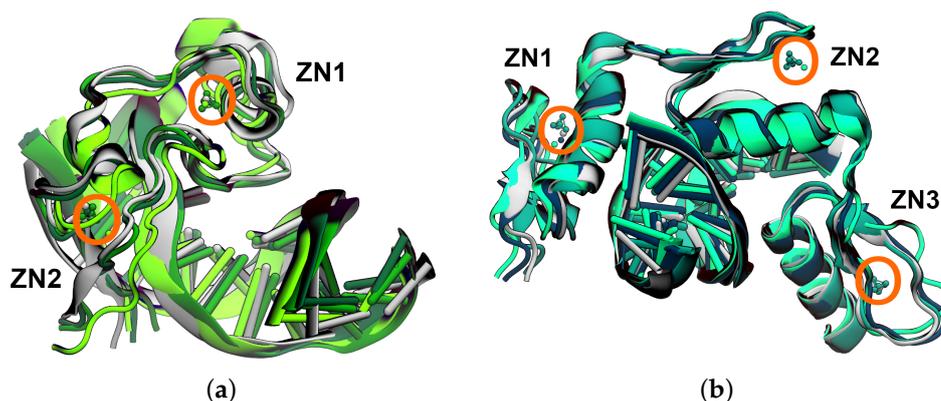
Figure 1. Representative median structures (coloured) of the simulations with the three ion models aligned (protein and DNA backbone) to the respective crystal structure (grey) of (**a**) MLL1 (PDB ID: 4nw3 [35,37]) and (**b**) WT1 (PDB ID: 5kl2 [34,36]). The colours represent (from light to dark) the non-bonded 12-6-4 model, the CDAM, and the bonded EZAFF model. Zn^{2+}-ions are highlighted with orange circles. For a detailed view of the coordination geometry of the Zn^{2+}-ions, see Figure 6. For median structures of the last 100 ns of all runs, see Figures S10 and S11 for MLL1 and WT1, respectively.

For WT1, the increased RMSD is due to multiple and different higher flexible residues. This can best be seen for the 12-6-4 model, e.g., around the cysteines coordinating ZN3 (for run3) and around the cysteines coordinating ZN1 (for run1) in Figure 3. It is also noteworthy that the sheet regions around the ZN1- and ZN3-coordinating cysteines are smaller than in the crystal structure and the other two models, and that a helix can form between the ZN1-coordinating cysteines (Figures 4 and S1). For the CDAM, a slight upward trend in protein RMSD is visible in Figure 2. The ZN3-coordinating residues fluctuate most when compared to the others and, thus, seem to be the most critical (Figure 3). In the case of the bonded EZAFF model and run3, the helices to which the ZN2- and ZN3-coordinating histidines belong become shorter after ~500 ns and ~1 μs, respectively (Figure S1).

The number of hydrogen bonds within the protein is rather unrevealing with median values of 19, 20, and 20 for the 12-6-4, CDAM, and EZAFF models of MLL1, respectively (for distributions see Figure S2). The same is true for WT1 with 35, 34, and 36 median numbers of hydrogen bonds.

3.2. DNA Structure

Intra- and inter-base pair parameters, and axis DNA parameters were calculated to check for potential long-range effects of different ion models (Figure S3, Figure S4, and Figure S5, respectively). For all DNA parameters, the models behave identically within the error range for MLL1 and WT1, which indicates that the specific ion model does not influence the DNA structure. The WT1 simulations also agree quite well with the crystal structure except for the axis bend. MLL1, on the other hand, shows more deviations between the crystal structure and simulations (see, for example, shear and stretch in Figure S3).
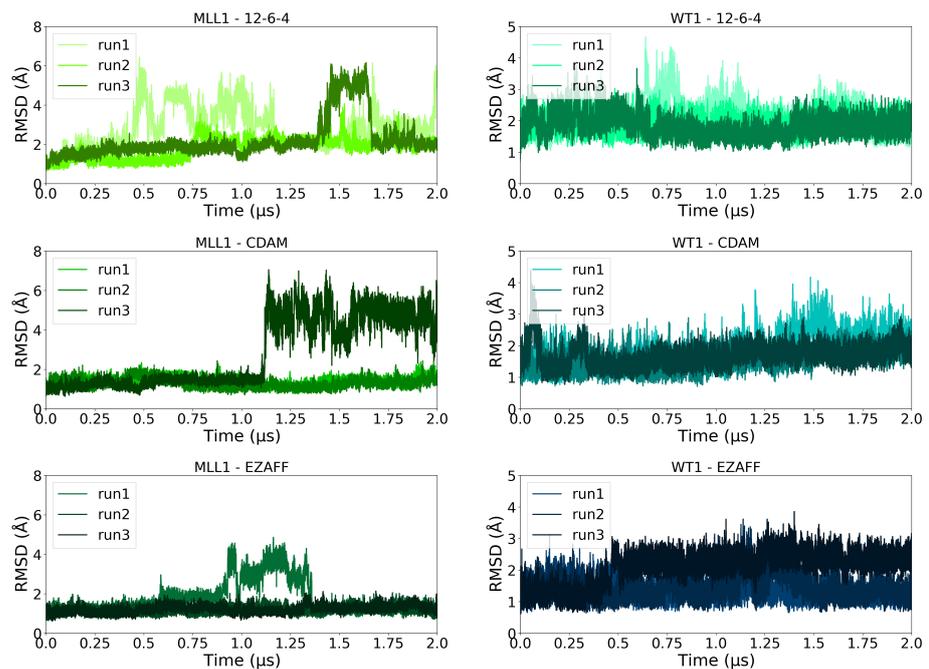
Figure 2. RMSD of protein backbone with respect to the crystal structure for the three different zinc models. Note that the scale of the vertical axis is different for MLL1 and WT1. For RMSD plots of all individual Zn^{2+}-ions and their ligating residues, see Figure S9.
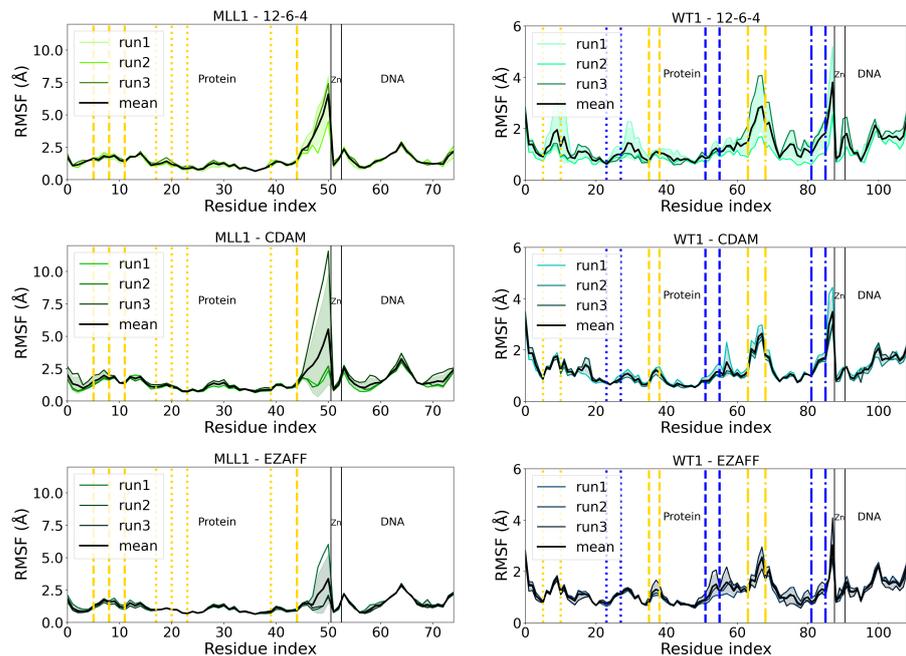
Figure 3. RMSF of C_α (protein, residue indices 0–50 and 0–87 for MLL1 and WT1, respectively), Zn^{2+} (residue indices 51–52 and 88–90 for MLL1 and WT1, respectively) and P (DNA, residue indices 53–74 and 91–108 for MLL1 and WT1, respectively) atoms after aligning to the median structure of the same atoms for MLL1 and WT1, respectively, simulated with the different zinc models. The standard deviation of the mean is shown in the transparent shade. Coordinating cysteines are shown in yellow and histidines in blue. The residues coordinating ZN1, ZN2, and ZN3 are indicated as dotted, dashed, and dash-dotted lines, respectively. Black vertical lines are added for visually separating protein, Zn^{2+}-ions, and DNA. For residue names of coordinating residues see Table S1.
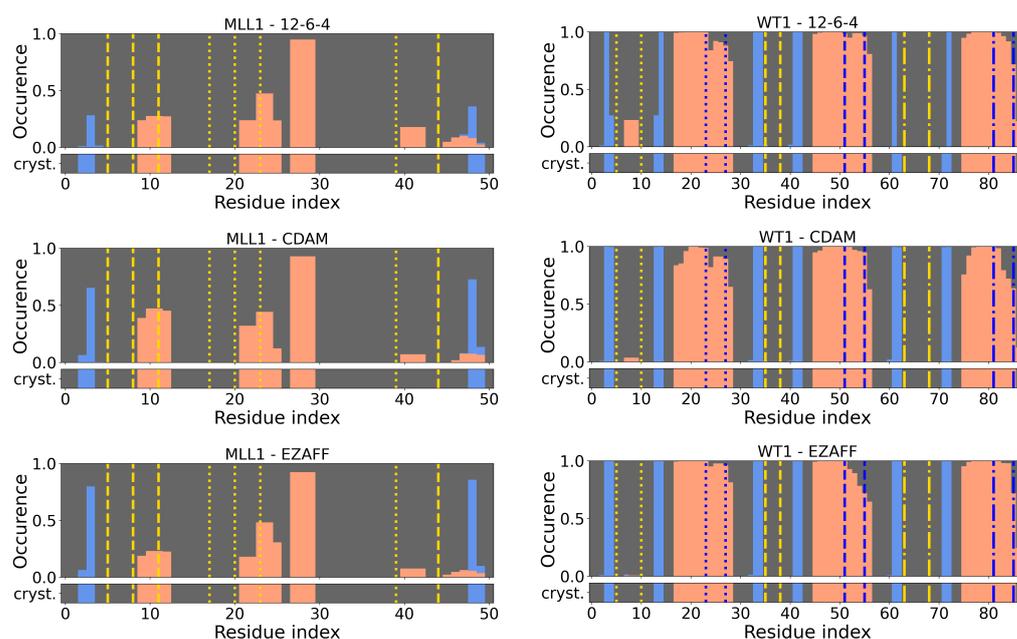
Figure 4. Percentage of secondary structure elements for all protein residues; helices are shown in light red, sheets in light blue, and everything else in grey. Coordinating cysteines are shown in yellow and histidines in blue. The residues coordinating ZN1, ZN2, and ZN3 are indicated as dotted, dashed, and dash-dotted lines, respectively. The secondary structure of the corresponding crystal structure is also shown in each figure. For residue names of coordinating residues, see Table S1.

3.3. Protein–DNA Interactions

For proper DNA recognition, the protein and the DNA need to be intact on their own, as well as their interactions. The number of hydrogen bonds, as one metric for such interactions, differs only slightly between the different models for both systems. The median number is 11 for all models of MLL1 and WT1 apart from the CDAM of WT1 where it is 10 (distribution see Figure S6).

The fraction of native contacts, on the other hand, provides further insights into the behaviour of the different ion models. Again, the 12-6-4 model is the worst among the Zn^{2+}-models tested in this study. The biggest penalty contribution is from within the proteins rather than from the protein–DNA interactions, as shown by the median fraction of native contacts for the entire system, within the protein, and between the protein and DNA in Figure 5a. The CDAM and EZAFF models perform similarly, with CDAM being slightly better in terms of maintaining native contacts between protein and DNA, and the EZAFF model performing better within protein. The distribution for MLL1 in all Zn^{2+}-models is generally much broader than for WT1 because Q declines over time (Figure S7). For WT1, the median of the whole system is above 0.9 for all models tested and, thus, very good. We also see that the model does not have a significant long-range influence on the protein–DNA interactions for WT1. The bimodal distribution for the EZAFF model is due to a change in secondary structure in one of the runs (Figures S1 and S7), which gives rise to a different set of contacts. The CDAM and EZAFF models can be considered equally good in this context.

3.4. Zn^{2+} Coordination

Looking at the individual Zn^{2+}-ions and their coordination spheres in more detail reveals the reason for the poorer protein structure quality for both proteins in the 12-6-4 model (as can be seen in RMSD values, secondary structure, and native contact analysis). All Zn^{2+}-ions, in both systems, change their CN from four to predominantly six in octahedral geometry (ZN2 in MLL1 also has some amount of CN = 5 in a trigonal bipyramid, which is due to one run, data not shown). A detailed breakdown of the

observed geometries during the simulations can be found in Table 1. While CDAM enforces the tetrahedral geometry throughout all simulations and Zn^{2+}-ions, the situation for the EZAFF model is more nuanced. This can best be seen for ZN2 of MLL1, whose initial crystal geometry is a distorted tetrahedron instead of a regular one. Consequently, observed geometries are more diverse and not purely tetrahedral, though with a CN of four.
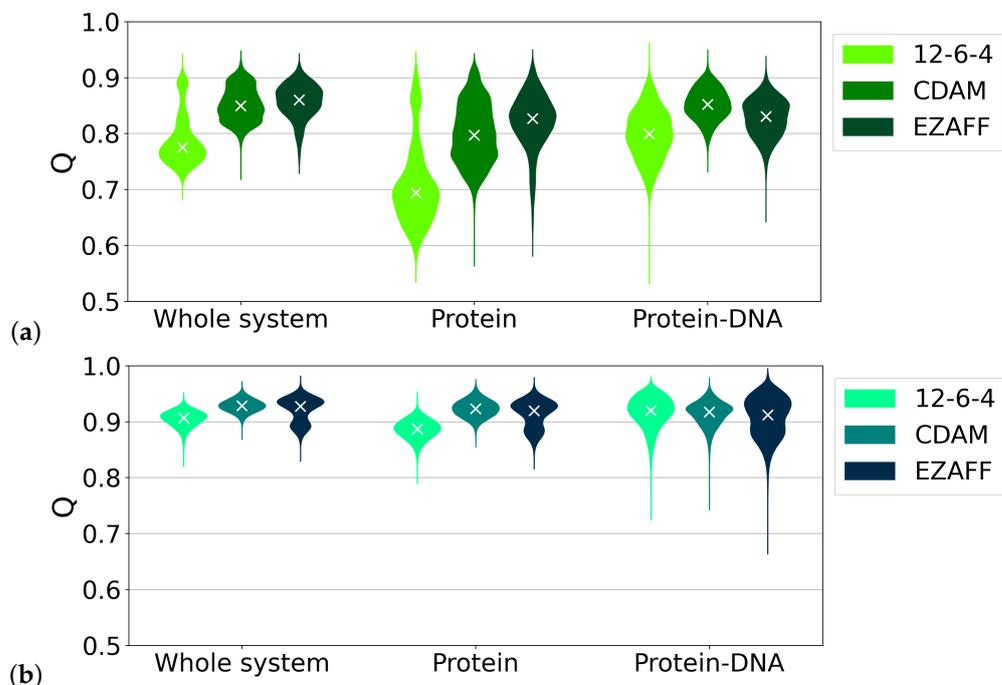
Figure 5. Fraction of native contacts, *Q*, of (**a**) MLL1 and (**b**) WT1. White crosses mark the median fraction of native contacts.

Table 1. Observed coordination geometries in MLL1 and WT1 (percentage).

Coordination Geometry (CN)	12-6-4	CDAM	EZAFF
MLL1 ZN1, Regular Tetrahedron (4)			
Tetrahedron (4)	0.82	100.00	99.16
Trigonal bipyramid with a vacancy (axial) (4)	–	–	0.25
Trigonal bipyramid (5)	13.26	–	–
Square pyramid (5)	0.11	–	–
Trigonal prism with a vacancy (5)	0.02	–	–
Octahedron (6)	78.18	–	–
Pentagonal bipyramid with a vacancy (equatorial) (6)	7.30	–	–
Irregular geometry	0.32	–	0.60
MLL1 ZN2, Distorted Tetrahedron (4)			
Tetrahedron (4)	0.26	100.00	55.18
Trigonal bipyramid with a vacancy (axial) (4)	–	–	28.14
Square pyramid with a vacancy (equatorial) (4)	–	–	15.30
Trigonal bipyramid with a vacancy (equatorial) (4)	–	–	1.39
Trigonal bipyramid (5)	13.54	–	–
Square pyramid (5)	0.02	–	–
Octahedron (6)	82.12	–	–
Pentagonal bipyramid with a vacancy (equatorial) (6)	4.02	–	–
Irregular geometry	0.04	–	–

Table 1. *Cont.*

Coordination Geometry (CN)	12-6-4	CDAM	EZAFF
WT1 ZN1, Regular Tetrahedron (4)			
Tetrahedron (4)	–	100.00	99.12
Trigonal bipyramid with a vacancy (axial) (4)	–	–	0.88
Square pyramid (5)	0.02	–	–
Octahedron (6)	96.79	–	–
Pentagonal bipyramid with a vacancy (equatorial) (6)	3.17	–	–
Octahedron, face monocapped with a vacancy (capped face) (6)	0.02	–	–
WT1 ZN2, Regular Tetrahedron (4)			
Tetrahedron (4)	–	99.98	99.96
Trigonal bipyramid with a vacancy (axial) (4)	–	0.02	0.04
Octahedron (6)	99.12	–	–
Pentagonal bipyramid with a vacancy (equatorial) (6)	0.86	–	–
Octahedron, face-monocapped with a vacancy (capped face) (6)	0.02	–	–
WT1 ZN3, Regular Tetrahedron (4)			
Tetrahedron (4)	–	99.98	96.81
Trigonal bipyramid with a vacancy (axial) (4)	–	0.02	3.19
Octahedron (6)	92.58	–	–
Pentagonal bipyramid with a vacancy (equatorial) (6)	7.42	–	–

Representative structures for each Zn^{2+}-ion and model are depicted in Figure 6, along with the corresponding crystal structure geometry. It can be observed that EZAFF closely resembles the original geometry due to parameter construction. The Zn^{2+}-ions in the 12-6-4 model are coordinated by two water molecules, resulting in octahedral coordination. In the CDAM, the structure is also comparable to the crystal but one can see that the distance between the coordinating atoms and Zn^{2+} has decreased. The average distances are visualised in Figure 7. For MLL1, almost all distances simulated by all models are shorter than in the crystal structure. The models that are closest to the crystal structure in terms of distances are EZAFF and 12-6-4. For WT1, the Zn^{2+}–S_γ distance is, on average, systematically underestimated by all models, while the 12-6-4 model overestimates the Zn^{2+}–$N_{\varepsilon2}$ distance and CDAM and EZAFF are comparable. Still, EZAFF produces the most consistent representation of distances.

In an ideal tetrahedron, the atom–Zn^{2+}–atom angles for all ligating atoms should be $\approx 109.5°$. Figure S8 shows the distributions for all angles in all models as well as the respective angles in the crystal structures. Again, EZAFF is in the best agreement with the crystal structure due to the model construction, and CDAM benefits from a regular tetrahedral structure.

The RDFs of water oxygen atoms around Zn^{2+}-ions in Figure 8 show that water molecules move into the coordination shell in all simulations using the 12-6-4 model. This is not the case for the other two models. However, due to the lower distance to the coordinating ligands, slightly more water can move closer to the Zn^{2+}-ions for the CDAM than for the EZAFF model.

This behaviour of the zinc ion coordination is also reflected in the RMSD time series calculated for the Zn^{2+}-ions and their ligating residues (Figure S9). Additionally, we see the effect of change in the secondary structure, resulting in changed dihedral angles but not necessarily altered coordination geometry.
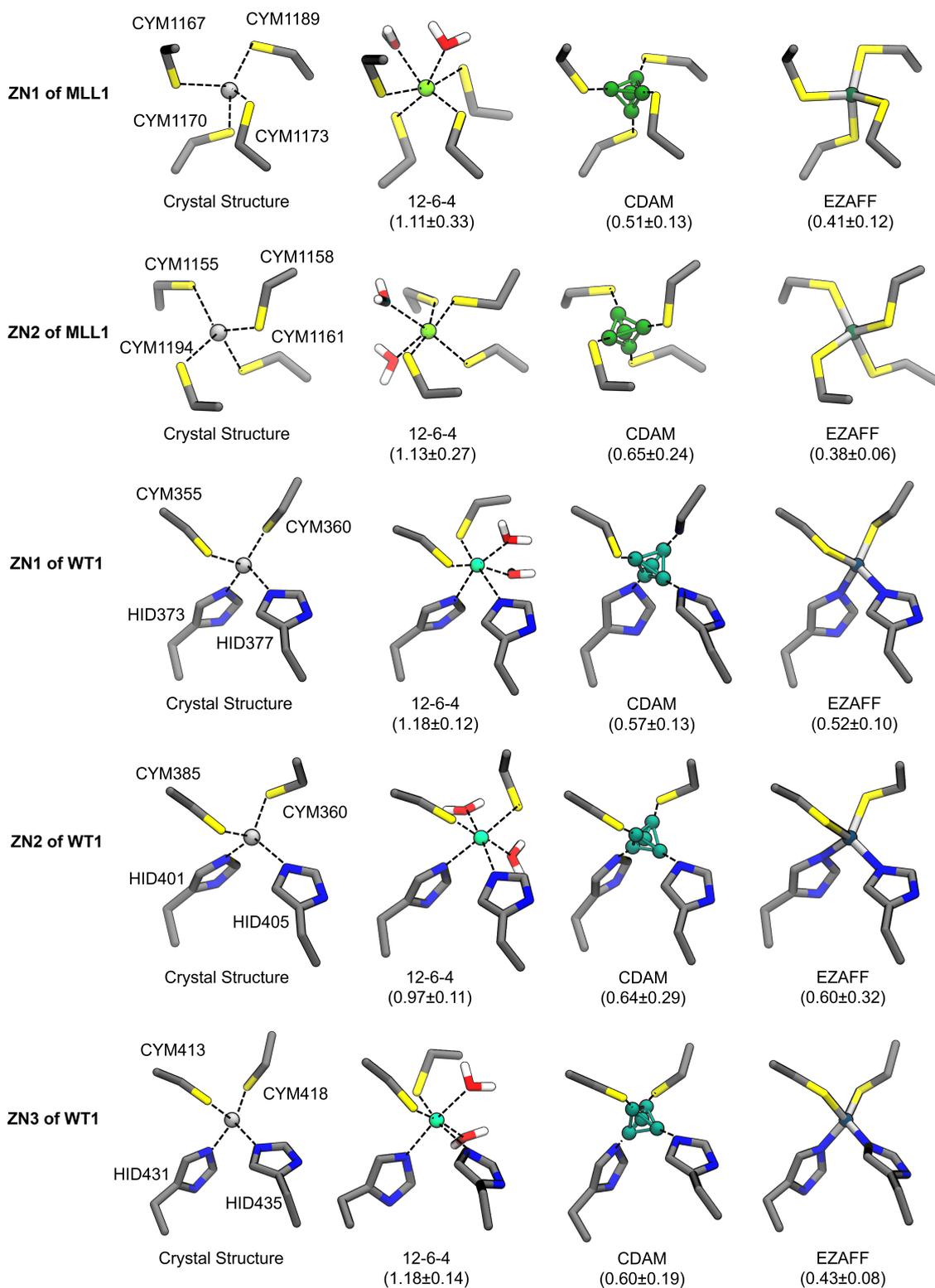
Figure 6. Coordination geometries of all Zn^{2+}-ions in MLL1 and WT1, respectively, as observed in the crystal structures and MD simulations with the different ion models (shown as representative median structures). Average RMSD values with respect to the crystal structure in Å are shown in parentheses. Hydrogen atoms are not shown for clarity. For corresponding residue indices, see Table S1.
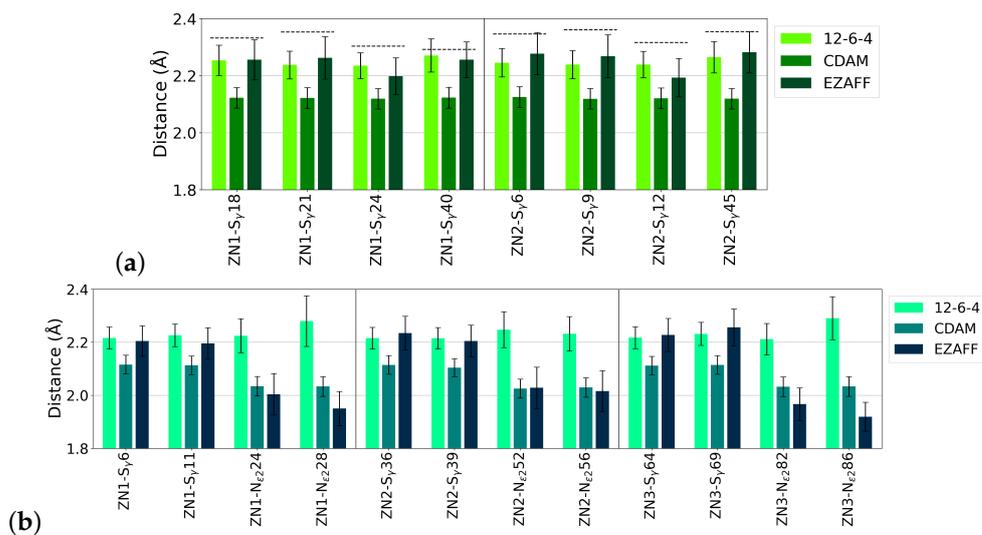
Figure 7. Distance between the Zn^{2+}-ions and their coordinating atoms (residue indices given; for corresponding residue names, see Table S1) in (**a**) MLL1 and (**b**) WT1. The error is given as the standard deviation. Crystal structure values are indicated as black dashed lines.
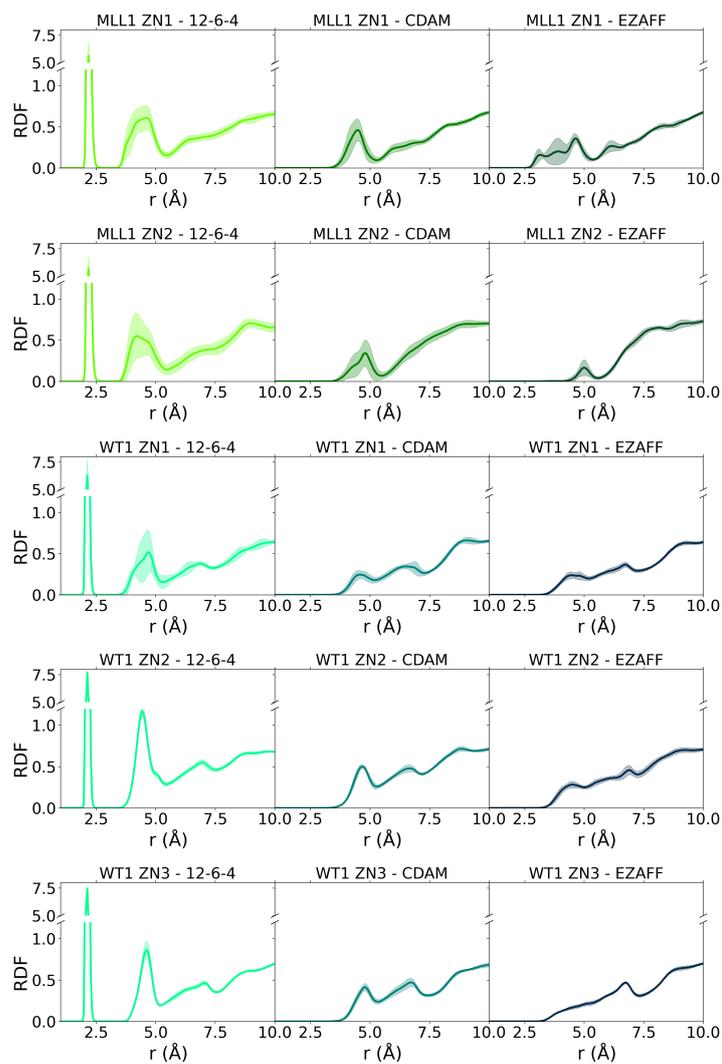
Figure 8. Radial distribution function of oxygen atoms of water molecules around the different Zn^{2+}-ions in MLL1 and WT1, respectively, for all ion models. Note the broken RDF axis.

4. Discussion

Our results show that the Zn^{2+}-model used, i.e., the 12-6-4 LJ-type non-bonded model, the cationic dummy atom model, or the bonded EZAFF model, has an impact on the Zn^{2+} coordination geometry and, thus, on the structural stability of at least the immediate environment of the Zn^{2+}-ions. We observe deviations from the initial crystal structure to different extents, which is clearly dependent on the flexibility of the model.

In the 12-6-4 LJ non-bonded model, the CN changes to six in mainly octahedral geometry as two water molecules enter the first coordination shell of all Zn^{2+}-ions. This destroys the geometry around the Zn^{2+}-ions, the secondary structure is altered, and some of the native contacts, mainly in protein, break. Since the model has been parametrised in solution, where octahedral coordination is dominant, and not in the protein environment, different geometries that can be observed in a protein, such as the tetrahedral coordination of the two systems studied in this work, cannot be adequately represented. This model is only a good choice if the CN changes to six in octahedral geometry or a flexible ligand exchange is desired (or expected), as was the case for the investigation of the divalent cation binding protein PsaA by MacDermott-Opeskin et al. [33].

The CDAM by Pang maintains a tetrahedral coordination geometry reliably but underestimates Zn^{2+}–S distances by approximately 0.2 Å as already pointed out by the author in the original paper [26]. These too-short distances cause slight changes in the geometry around the Zn^{2+}-ion, which may propagate through the protein. The slowly increasing RMSD and the slowly decreasing fraction of the native contacts within WT1 are indicative of such a Zn^{2+}–S bond contraction affecting the entire protein, at least for longer time scales.

The initial crystal geometry is best preserved by the bonded EZAFF model since it is parametrised for this input geometry. Therefore, one has to be sure that the initial structure is of good quality as the one used for WT1. The crystal structure of MLL1, on the other hand, has a lower resolution. For this protein, we observed a decreasing fraction of native contacts (much worse compared to WT1), lower distances between Zn^{2+} and coordinating atoms than in the crystal structure, and deviations in DNA parameters. While the resulting structure may be a better representation of the system in solution, it is also conceivable that the constraints imposed by the bonded model force the system into a wrong Zn^{2+}-ion coordination and geometry, and the remainder of the system responds by structural adaptions that better accommodate the forced, and possibly wrong, geometry. Such structures may need refinement, such as with the PDB-REDO procedure [57], prior to MD simulation or the CDAM may be the better choice in such cases as it allows for more flexibility of the coordinating ligands in the context of tetrahedral coordination.

In all models investigated in this work, the coordination geometry, once established, is stable, even if it is a wrong geometry. Thus, contrary to CDAM and EZAFF, the 12-6-4 model should not be used for tetrahedrally coordinated zinc that is accessible by water since it could lead to octahedral coordination.

We could not observe long-range effects of the zinc models on DNA, even though the protein secondary structure was altered. This overall stability of the investigated systems suggests the structural role of the zinc ions to be rather local, i.e., only relevant for the domains in which they are located. In more complex systems with dimerisation domains that contain Zn^{2+}-ions, as is found in many DNA binding domains of hormone receptors [7,8], the effect of an altered zinc coordination and geometry will likely have more severe effects on the overall structure.

It is always important to validate the chosen model by comparing quantities from MD simulations, as shown here, to those of the corresponding X-ray crystallographic or the NMR structure as already pointed out in other studies [16,21,32]. Care must be taken if the crystal structure only has low resolution. In such cases, it is possible to first use geometry restraints based on the AMBER FF for crystallographic refinement with Phenix, a commonly used program for crystal structure refinement [58].

Our study confirms that there is no single perfect ion model that outperforms all others, even for describing structural–tetrahedrally coordinated Zn^{2+}-ions. If the Zn^{2+} environment is challenging, e.g., a ligand-binding site where a change in CN could be important, the exact coordination is not known, or none of the available parameter sets produce adequate results, new molecular mechanics parameters (hybrid bonded/non-bonded model parameters as in [31]), more sophisticated models, or (semi-empirical) QM calculations may be needed (as shown by Yu et al. [23]).

5. Conclusions

We investigated the influence of three popular zinc models, i.e., the non-bonded 12-6-4 LJ-type model, the cationic dummy atom model, and the bonded EZAFF model, on the protein–DNA complexes of MLL1 and WT1. All Zn^{2+}-ions in these proteins are tetrahedrally coordinated and guarantee structural integrity. Our results indicate that the 12-6-4 model should not be used for this geometry if water is nearby. CDAM, on the other hand, yields a regular tetrahedral geometry, while EZAFF can additionally maintain more distorted geometries. Thus, EZAFF should only be used if the geometry in the crystal structure is correct. Consequently, the choice of an appropriate model for a given structure depends on the quality of the structure and the role of the zinc ion and should, therefore, be carefully validated for the given system.

Abbreviations

The following abbreviations are used in this manuscript:

CDAM   cationic dummy atom model
EZAFF   extended zinc AMBER force field
MD   molecular dynamics
CN   coordination number

References

1. Lu, Y.; Yeung, N.; Sieracki, N.; Marshall, N.M. Design of Functional Metalloproteins. *Nature* **2009**, *460*, 855–862. [CrossRef] [PubMed]
2. Vahrenkamp, H. Why does nature use zinc-a personal view. *Dalton Trans.* **2007**, *42*, 4751–4759. [CrossRef]
3. Gropper, S.S.; Smith, J.L.; Carr, T.P. *Advanced Nutrition and Human Metabolism*; Wadsworth: Belmont, CA, USA, 2021.
4. Cherasse, Y.; Urade, Y. Dietary Zinc Acts as a Sleep Modulator. *Int. J. Mol. Sci.* **2017**, *18*, 2334. [CrossRef] [PubMed]
5. Markov, D.; Naryshkina, T.; Mustaev, A.; Severinov, K. A zinc-binding site in the largest subunit of DNA-dependent RNA polymerase is involved in enzyme assembly. *Genes Dev.* **1999**, *13*, 2439–2448. [CrossRef]
6. Rich, A.M.; Bombarda, E.; Schenk, A.D.; Lee, P.E.; Cox, E.H.; Spuches, A.M.; Hudson, L.D.; Kieffer, B.; Wilcox, D.E. Thermodynamics of Zn2+ Binding to Cys2His2 and Cys2HisCys Zinc Fingers and a Cys4 Transcription Factor Site. *J. Am. Chem. Soc.* **2012**, *134*, 10405–10418. [CrossRef]
7. Luisi, B.F.; Xu, W.X.; Otwinowski, Z.; Freedman, L.P.; Yamamoto, K.R.; Sigler, P.B. Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **1991**, *352*, 497–505. [CrossRef]
8. Schöne, S.; Jurk, M.; Helabad, M.B.; Dror, I.; Lebars, I.; Kieffer, B.; Imhof, P.; Rohs, R.; Vingron, M.; Thomas-Chollier, M.; et al. Sequences flanking the core-binding site modulate glucocorticoid receptor structure and activity. *Nat. Commun.* **2016**, *7*, 12621. [CrossRef] [PubMed]
9. Li, P.; Merz, K.M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117*, 1564–1686. [CrossRef] [PubMed]
10. Lipscomb, W.N.; Sträter, N. Recent Advances in Zinc Enzymology. *Chem. Rev.* **1996**, *96*, 2375–2434. [CrossRef]
11. Laitaoja, M.; Valjakka, J.; Jänis, J. Zinc Coordination Spheres in Protein Structures. *Inorg. Chem.* **2013**, *52*, 10983–10991. [CrossRef] [PubMed]
12. Wu, J.C.; Piquemal, J.P.; Chaudret, R.; Reinhardt, P.; Ren, P. Polarizable Molecular Dynamics Simulation of Zn(II) in Water Using the AMOEBA Force Field. *J. Chem. Theory Comput.* **2010**, *6*, 2059–2070. [CrossRef] [PubMed]
13. Li, P.; Roberts, B.P.; Chakravorty, D.K.; Merz, K.M. Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733–2748. [CrossRef] [PubMed]
14. Macchiagodena, M.; Pagliai, M.; Andreini, C.; Rosato, A.; Procacci, P. Upgrading and Validation of the AMBER Force Field for Histidine and Cysteine Zinc(II)-Binding Residues in Sites with Four Protein Ligands. *J. Chem. Inf. Model.* **2019**, *59*, 3803–3816. [CrossRef]
15. Macchiagodena, M.; Pagliai, M.; Andreini, C.; Rosato, A.; Procacci, P. Upgraded AMBER Force Field for Zinc-Binding Residues and Ligands for Predicting Structural Properties and Binding Affinities in Zinc-Proteins. *ACS Omega* **2020**, *5*, 15301–15310. [CrossRef]
16. Melse, O.; Antes, I.; Kaila, V.R.I.; Zacharias, M. Benchmarking biomolecular force field-based Zn²⁺ for mono- and bimetallic ligand binding sites. *J. Comput. Chem.* **2022**, *44*, 912–926. [CrossRef] [PubMed]
17. Li, P.; Merz, K.M. Taking into Account the Ion-induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10*, 289–297. [CrossRef] [PubMed]
18. Li, Z.; Song, L.F.; Li, P.; Merz, K.M. Systematic Parametrization of Divalent Metal Ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB Water Models. *J. Chem. Theory Comput.* **2020**, *16*, 4429–4442. [CrossRef] [PubMed]
19. Song, L.F.; Sengupta, A.; Merz, K.M. Thermodynamics of Transition Metal Ion Binding to Proteins. *J. Am. Chem. Soc.* **2020**, *142*, 6365–6374. [CrossRef]
20. Panteva, M.T.; Giambaşu, G.M.; York, D.M. Force Field for Mg²⁺, Mn²⁺, Zn²⁺, and Cd²⁺ Ions That Have Balanced Interactions with Nucleic Acids. *J. Phys. Chem. B* **2015**, *119*, 15460–15470. [CrossRef]
21. Sousa, S.F.; Fernandes, P.A.; Ramos, M.J. Molecular Dynamics Simulations: Difficulties, Solutions and Strategies for Treating Metalloenzymes. In *Challenges and Advances in Computational Chemistry and Physics*; Springer Netherlands: Dordrecht, The Netherlands, 2010; pp. 299–330. [CrossRef]
22. Peters, M.B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M.N.; Merz, K.M. Structural Survey of Zinc Containing Proteins and the Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6*, 2935–2947. [CrossRef]
23. Yu, Z.; Li, P.; Merz, K.M. Extended Zinc AMBER Force Field (EZAFF). *J. Chem. Theory Comput.* **2018**, *14*, 242–254. [CrossRef] [PubMed]
24. Li, P.; Merz, K.M.J. MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* **2016**, *56*, 599–604. [CrossRef] [PubMed]
25. Hu, L.; Ryde, U. Comparison of Methods to Obtain Force-Field Parameters for Metal Sites. *J. Chem. Theory Comput.* **2011**, *7*, 2452–2463. [CrossRef] [PubMed]

26. Pang, Y.P. Novel Zinc Protein Molecular Dynamics Simulations: Steps toward Antiangiogenesis for Cancer Treatment. *J. Mol. Model.* **1999**, *5*, 196–202. [CrossRef]

27. Pang, Y.P.; Xu, K.; Yazal, J.E.; Prendergas, F.G. Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach. *Protein Sci. Publ. Protein Soc.* **2000**, *9*, 1857–1865.

28. Pang, Y.P. Successful molecular dynamics simulation of two zinc complexes bridged by a hydroxide in phosphotriesterase using the cationic dummy atom method. *Proteins* **2001**, *45*, 183–189. [CrossRef]

29. Duarte, F.; Bauer, P.; Barrozo, A.; Amrein, B.A.; Purg, M.; Aqvist, J.; Kamerlin, S.C.L. Force Field Independent Metal Parameters Using a Nonbonded Dummy Model. *J. Phys. Chem. B* **2014**, *118*, 4351–4362. [CrossRef]

30. Jiang, Y.; Zhang, H.; Tan, T. Rational Design of Methodology-Independent Metal Parameters Using a Nonbonded Dummy Model. *J. Chem. Theory Comput.* **2016**, *12*, 3250–3260. [CrossRef]

31. Tomić, A.; Horvat, G.; Ramek, M.; Agić, D.; Brkić, H.; Tomić, S. New Zinc Ion Parameters Suitable for Classical MD Simulations of Zinc Metallopeptidases. *J. Chem. Inf. Model.* **2019**, *59*, 3437–3453. [CrossRef]

32. Jardin, C.; Chaves, G.; Musset, B. Assessing Structural Determinants of Zn^{2+} Binding to Human HV1 via Multiple MD Simulations. *Biophys. J.* **2020**, *118*, 1221–1233. [CrossRef]

33. MacDermott-Opeskin, H.; McDevitt, C.A.; O'Mara, M.L. Comparing Nonbonded Metal Ion Models in the Divalent Cation Binding Protein PsaA. *J. Chem. Theory Comput.* **2020**, *16*, 1913–1923. [CrossRef] [PubMed]

34. Hashimoto, H.; Zhang, X.; Zheng, Y.; Wilson, G.G.; Cheng, X. Denys-Drash syndrome associated WT1 glutamine 369 mutants have altered sequence-preferences and altered responses to epigenetic modifications. *Nucleic Acids Res.* **2016**, *44*, 10165–10176. [CrossRef]

35. Xu, C.; Liu, K.; Lei, M.; Yang, A.; Li, Y.; Hughes, T.R.; Min, J. DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure* **2018**, *26*, 85–95.e3. [CrossRef] [PubMed]

36. Hashimoto, H.; Cheng, X. Wilms Tumor Protein (WT1) ZnF2-4 in Complex with DNA. 2016. Available online: https://ftp.wwpdb.org/pub/pdb/validation_reports/kl/5kl2/5kl2_full_validation.pdf.gz (accessed on 28 January 2023). [CrossRef]

37. Bian, C.; Tempel, W.; Chao, X.; Walker, J.; Bountra, C.; Weigelt, J.; Arrowsmith, C.; Edwards, A.; Min, J. Crystal Structure of MLL CXXC Domain in Complex with a CpG DNA. 2014. Available online: https://ftp.wwpdb.org/pub/pdb/validation_reports/nw/4nw3/4nw3_full_validation.pdf.gz (accessed on 28 January 2023). [CrossRef]

38. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]

39. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38. [CrossRef]

40. Martínez-Rosell, G.; Giorgino, T.; de Fabritiis, G. PlayMolecule ProteinPrepare: A Web Application for Protein Preparation for Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2017**, *57*, 1511–1516. [CrossRef] [PubMed]

41. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]

42. Ivani, I.; Dans, P.D.; Noy, A.; Pérez, A.; Faustino, I.; Hospital, A.; Walther, J.; Andrio, P.; Go ni, R.; Balaceanu, A.; et al. Parmbsc1: A refined force field for DNA simulations. *Nat. Methods* **2016**, *13*, 55–58. [CrossRef]

43. Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. [CrossRef] [PubMed]

44. 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.; et al. *AMBER 2018*; University of California: San Francisco, CA, USA, 2018.

45. Case, D.A.; Aktulga, H.M.; Belfon, K.; Ben-Shalom, I.Y.; Berryman, J.T.; Brozell, S.R.; Cerutti, D.S.; Cheatham, T.E., III.; Cisneros, G.A.; et al. *AMBER 2022*; University of California: San Francisco, CA, USA, 2022.

46. Pang, Y.P. Zinc Protein Simulations Using the Cationic Dummy Atom (CaDA) Method. Available online: https://www.mayo.edu/research/labs/computer-aided-molecular-design/projects/zinc-protein-simulations-using-cationic-dummy-atom-cada-approach (accessed on 28 January 2023).

47. Li, P.; Merz, K.M. Building Bonded Model for A Ligand Binding Metalloprotein with MCPB.py. Availabe online: http://ambermd.org/tutorials/advanced/tutorial20/mcpbpy.htm (accessed on 28 January 2023).

48. Frisch, M.E.; Trucks, G.; Schlegel, H.B.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H.; et al. *Gaussian 16*; Gaussian, Inc.: Wallingford, CT, USA, 2016.

49. Ryckaert, J.P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341. [CrossRef]

50. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]

51. Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [CrossRef]

52. Zacharias, M. Atomic Resolution Insight into Sac7d Protein Binding to DNA and Associated Global Changes by Molecular Dynamics Simulations. *Angew. Chem. (Int. Ed. Engl.)* **2019**, *58*, 5967–5972. [CrossRef] [PubMed]

53. Best, R.B.; Hummer, G.; Eaton, W.A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17874–17879. [CrossRef] [PubMed]

54. Andreini, C.; Cavallaro, G.; Lorenzini, S. FindGeo: A tool for determining metal coordination geometry. *Bioinformatics* **2012**, *28*, 1658–1660. [CrossRef]
55. Lavery, R.; Moakher, M.; Maddocks, J.H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917–5929. [CrossRef] [PubMed]
56. McGibbon, R.T.; Beauchamp, K.A.; Harrigan, M.P.; Klein, C.; Swails, J.M.; Hernández, C.X.; Schwantes, C.R.; Wang, L.P.; Lane, T.J.; Pande, V.S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528–1532. [CrossRef]
57. Joosten, R.P.; Salzemann, J.; Bloch, V.; Stockinger, H.; Berglund, A.C.; Blanchet, C.; Bongcam-Rudloff, E.; Combet, C.; Da Costa, A.L.; Deleage, G.; et al. PDB_REDO: Automated re-refinement of X-ray structure models in the PDB. *J. Appl. Crystallogr.* **2009**, *42*, 376–384. [CrossRef] [PubMed]
58. Moriarty, N.W.; Janowski, P.A.; Swails, J.M.; Nguyen, H.; Richardson, J.S.; Case, D.A.; Adams, P.D. Improved chemistry restraints for crystallographic refinement by integrating the Amber force field into Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* **2020**, *76*, 51–62. [CrossRef]