A Multiscale Workflow for Modelling Ligand Complexes of Zinc Metalloproteins

Zongfan Yang[a,b], Rebecca M. Twidale[a], Silvia Gervasoni[a,c,d], Reynier Suardíaz[a,e], Charlie Colenso[a,b], Eric J. M. Lang[a], James Spencer[b] and Adrian J. Mulholland[a]*

[a]Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, BS8 1TH, U.K.

[b]School of Cellular and Molecular Medicine, University of Bristol, Bristol, BS8 1TD, U.K.

[c]Department of Pharmaceutical Sciences, University of Milan, Via Mangiagalli, 25, I-20133 Milano, Italy.

[d]Department of Physics, University of Cagliari, S.P. Monserrato-Sestu km 0.700 I-09042, Monserrato (CA), Italy (present address)

[e]Departamento de Química Física, Facultad de Química, Universidad Complutense, 28040 Madrid, Spain (present address)

KEYWORDS: Zinc, Metalloenzyme, Molecular Dynamics, DFTB3, QM/MM, Metallo-beta-lactamase

ABSTRACT: Zinc metalloproteins are ubiquitous, with protein zinc centers of structural and functional importance, involved in interactions with ligands and substrates and often of pharmacological interest. Biomolecular simulations are increasingly prominent in investigations of protein structure, dynamics, ligand interactions and catalysis, but zinc poses a particular challenge, in part because of its versatile, flexible coordination. A computational workflow generating reliable models of ligand complexes of biological zinc centers would find broad application. Here we evaluate the ability of alternative treatments, using (non-bonded) molecular mechanics (MM) and quantum mechanics/ molecular mechanics (QM/MM) at semiempirical (DFTB3) and density functional theory (DFT) levels of theory, to describe the zinc centers of ligand complexes of six metalloenzyme systems differing in coordination geometries, zinc stoichiometries (mono- and di-nuclear), and the nature of interacting groups (specifically the presence of zinc - sulfur interactions). MM molecular dynamics (MD) simulations can overfavor octahedral geometries, introducing additional water molecules to the zinc coordination shell, but this can be rectified by subsequent semiempirical (DFTB3) QM/MM MD simulations. B3LYP/MM geometry optimization further improved the accuracy of description of coordination distances, with the overall effectiveness of the approach depending upon factors including the presence of zinc – sulfur interactions that are less well described by semiempirical methods. We describe a workflow, comprising QM/MM MD using DFTB3 followed by QM/MM geometry optimization using DFT (e.g., B3LYP), that well describes our set of zinc metalloenzyme complexes and is likely to be suitable for creating accurate models of zinc protein complexes when structural information is more limited

Introduction

Zinc has an essential role in biology, with zinc sites contributing to the structural integrity, stability and catalytic activity of a wide range of proteins with highly diverse functions in both pro- and eukaryotic cells.[1] Many such proteins are actual or potential targets for pharmaceutical intervention, including by small molecule therapeutics used to treat conditions such as hypertension, cancer, infectious disease etc. Interaction of small molecule ligands with their protein targets frequently involves participation of bound zinc; in some cases, interactions involving zinc ions are essential to ligand binding and removal of zinc abolishes this.[2]

Molecular simulation methods play an increasingly prominent role in ligand and drug discovery, driven in part by hardware and software innovations and by the growing availability of high-resolution crystal structures for many biologically and/or pharmacologically important protein targets.[3,4] Their application to zinc metalloproteins is desirable, given the abundance of zinc-containing systems in the proteomes of many species, including humans. Unfortunately, many of the properties that enable zinc to play a diverse range of roles in biological systems make modelling protein zinc centers and their complexes challenging. These include the ability of zinc to coordinate different types of ligands, including N, O and S; flexibility of coordination geometry (zinc is 6-coordinated (octahedral) in aqueous solution but may be tetrahedral, 5- or, in some catalytic sites, 6-coordinated in proteins[5–8]);[9,10] the ability of coordinating water molecules to exchange with substrates or inhibitors during complex formation;[7,11] and the existence of single and multi-nuclear sites. As we have recently demonstrated,[12] treatment of zinc centers as point charges (as in for example many ligand docking methods) often leads to unrealistic coordination geometries.

A variety of treatments has been applied to zinc (and other metal) centers in proteins in molecular dynamics (MD) simulations. In the widely used AMBER MD package (https://ambermd.org/)[13–16] bonded,[17] non-bonded[18] and cationic dummy atom[19] approaches have all been implemented as molecular mechanics (MM) treatments for metal ions and their complexes. Bonded models do not allow for ligand exchange and/or changes in zinc coordination geometry, while cationic dummy atom approaches require a pre-defined zinc coordination geometry, limiting exploration of alternative ligation patterns, and are more challenging to set up. In contrast, nonbonded models, as typified by the Lennard-Jones (LJ) 12-6 model, are widely used owing to their simplicity and transferability.[20] The LJ12-6 model does, however, require user selection of the most appropriate parameter set: the IOD set, specifically designed to reproduce ion-oxygen distances, the HFE set specifically designed to reproduce hydration free energies, or the CM set designed as a compromise set for more general application.[18] More recently, the LJ12-6-4 non-bonded model, proposed and parameterized for divalent metal ions by Li and Merz,[21] includes an additional C4 term to represent the impact of ion-induced dipole interaction, and is claimed to simultaneously reproduce the hydration free energy (HFE), ion-oxygen distance (IOD) and coordination number, whereas the LJ12-6 model reproduced only one or two these experimental values in a given simulation.[21–23] Since the LJ12-6-4 model fulfils many of the performance requirements for MD simulations using a single parameter set, its ease of use in practice has led to its wide application in simulations of divalent metal cations (e.g., Zn^{2+}, Mg^{2+} and Mn^{2+}, etc.) involved in biological systems.[24–28]

Although the MM MD approach is advantageous in terms of computational efficiency, its accuracy is highly dependent on the predefined MM force fields. However, most current MM force fields do not accurately describe interactions of protein zinc centers, and force field parameter optimization may be required for a particular biological system.[29] The lack of any accurate description of polarization effects and the ability to simulate charge transfer, as well as the flexible coordination geometry of zinc, makes correct simulation of zinc ions in proteins using MM methods difficult. Quantum mechanics (QM) provides another approach to modelling interactions of zinc ions in proteins. QM methods can provide more accurate description of interactions involving zinc ions than MM methods, but at a computational cost that increases rapidly with increasing system size.[30] Accordingly, quantum mechanics/molecular mechanics (QM/MM) approaches provide a balance between calculation accuracy and efficiency. QM/MM approaches have been used in efforts to obtain more accurate descriptions of protein zinc centers in simulations. [31–33] Results can be dependent upon the appropriate partitioning of the system under investigation between the QM and MM regions.[34]

The computational requirements, and potential accuracy, can also be dependent upon the level of QM theory applied in QM/MM calculations. QM/MM calculations can be performed using semiempirical QM methods, ab initio QM or density functional theory (DFT) treatment. One of the most popular semiempirical QM methods is density functional tight binding (DFTB) approaches are derived in the framework of DFT. [35] DFTB methods have previously been used to simulate a variety of metalloprotein systems.[36–42] Higher levels of accuracy are in principle possible with ab initio QM techniques, but the computational costs of wave function optimization make ab initio QM/MM calculations impractical for routine application. DFT calculations, e.g. with hybrid functionals such as B3LYP provide a good balance between accuracy and computational cost in describing the structures of transition metal complexes[43–45] and have been widely used for zinc protein studies.[46–48] The choice of QM method then represents a compromise between accuracy and computational cost.

We have previously[49] developed a computational workflow, involving docking, MM and QM/MM simulations at two levels of QM theory, with which we have successfully reproduced crystal structures of complexes of the zinc-dependent (metallo-) beta-lactamase (MBL) IMP-1 (imipenemase-1) with thiol-based mercaptomethyl thiazolidine (MMTZ) inhibitors.[50] In this work we develop

and expand this work to test the ability of multiscale modelling to replicate crystal structures of a wider range of zinc metalloenzyme complexes varying in their zinc stoichiometry (mono- and dinuclear zinc sites), zinc coordination number (CN) or coordination geometry, and the identity of zinc ligating atoms from both protein and small molecule ligands. (Specifically, these include protein Asp, His and Cys ligands, representing Zn - N, Zn - O and Zn - S interactions, respectively; and thiolate, carboxylate and hydroxamate small molecule ligands). Our initial model system, the MBL Sfh-I in complex with an MMTZ inhibitor, features a typical tetrahedral zinc geometry with His, Asp, Cys and inhibitor thiolate ligands.[51] Subsequently we extend our investigations to a further five systems (**Figure 1**): Angiotensin-Converting Enzyme (ACE) complexed with the thiol inhibitor L-captopril (PDB: 2X8Z),[52] ACE-2 (the ACE isoform involved in viral spike protein processing during infection by SARS-CoV-2) complexed with the carboxylate inhibitor MLN-4760 (PDB: 1R4L),[53] histone deacetylase (HDAC) 2 complexed with the hydroxamate inhibitor SAHA (PDB: 4LXZ),[54] and the dinuclear IMP-1(PDB: 6JED)[55] and L1 MBLs (PDB: 7A63)[56] complexed with thioglycolic acid (TGA) and the hydrolyzed form of the penem antibiotic, faropenem, respectively. Collectively, these represent single-zinc systems with 4- and 5- coordination geometries and dinuclear systems with 4, 5 and 4, 6 coordination with and without sulfur ligands.

The results show that, while MM MD approaches can in some circumstances provide reasonable descriptions of zinc coordination distances, these usually introduce changes to zinc geometry that require rectification by semi-empirical QM/MM MD. The accuracy with which semi-empirical methods can describe protein zinc centers varies, with those involving Zn – S interactions among the most challenging. Our data suggest that a multiscale approach involving increasing levels of theory is necessary to obtain accurate models of zinc enzymes and their complexes, and identify a workflow that may be broadly applicable in such cases.

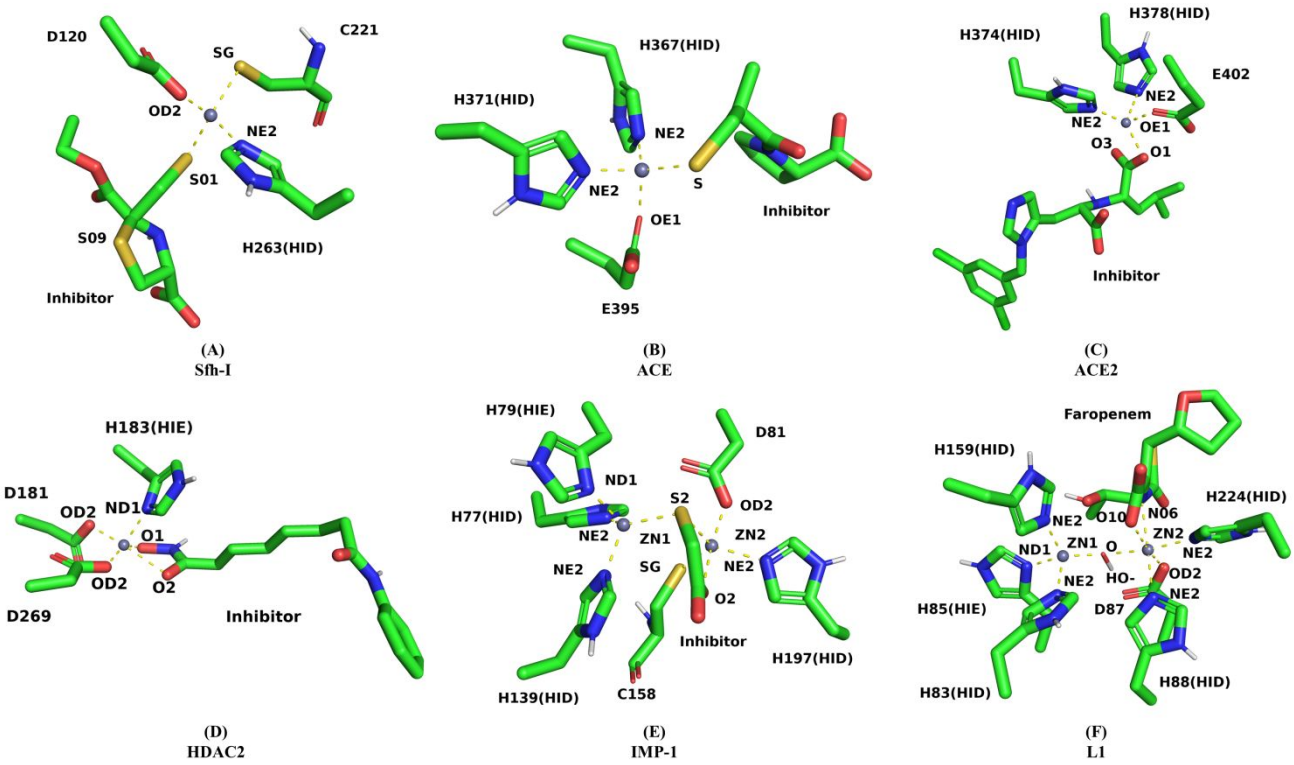


Figure 1. Zinc site geometries of the 6 protein-ligand systems included in this study. (A) Sfh-I MBL with MMTZ inhibitor L-anti-1a, coordination number (CN) = 4, PDB code: 7BJ9;[51] (B) ACE with the thiol inhibitor L-captopril, CN=4, PDB code: 2X8Z;[52] (C) ACE2 with the carboxylate inhibitor MLN-4760, CN=4, PDB code:1R4L;[53] (D) HDAC2 with the hydroxamate inhibitor SAHA, CN=5, PDB code : 4LXZ;[54] (E) IMP-1 MBL with thioglycolic acid (TGA), CN = 4,5, PDB code: 6JED;[55] (F) L1 MBL with hydrolyzed faropenem, CN=4,6, PDB code:7A63.[56] Carbon atoms are in green, nitrogen atoms are in blue, oxygen atoms are in red and sulfur atoms are in yellow, with zinc ions represented as gray balls.

Results

Our long-term goal is to develop a robust computational workflow able to generate realistic models of complexes of zinc metalloproteins. In this work we investigate the ability of different computational approaches to successfully maintain zinc centers in architectures consistent with starting crystal structure(s) for a range of representative complexes. As our aim is to develop a methodology that can be applied to suboptimal starting structures, derived from either poorer quality experimental data or models from e.g. docking experiments, and for which MM MD optimization might be necessary, we tested the inclusion of MM MD steps prior to any QM optimization.

Initial testing: complex of the MBL Sfh-I with the MMTZ inhibitor L-anti-1a.

Initial investigations were carried using the complex of the mono-zinc MBL Sfh-I (a carbapenem hydrolyzing beta-lactamase from the environmental bacterium and occasional pathogen *Serratia fonticola*[57]) with the MMTZ inhibitor L-anti-1a (PDB code: 7BJ9)[51] as a model system. This structure contains a single zinc ion in tetrahedral geometry coordinated by Asp, His and Cys residues and the inhibitor thiolate. Three different modelling methods were tested with the Sfh- I: L-anti-1a complex: MM MD (using a non-bonded model), QM/MM MD with a semi-empirical QM method (DFTB3) and QM/MM geometry optimization/ energy minimization with a DFT QM method (B3LYP).

Four different nonbonded models were applied to model the zinc center of the Sfh- I: L-anti-1a complex and tested in triplicate 100 ns MM MD simulations: the unrestrained LJ12-6 model (LJ12-6), the restrained LJ12-6 model (LJ12-6-R), the unrestrained LJ12-6-4 model (LJ12-6-4) and the restrained LJ12-6-4 model (LJ12-6-4-R). Atomic distance restraints when used were placed between the zinc ion and coordinating protein residue atoms during MD production runs in order to maintain crystallographically observed zinc coordination geometry and leave the fourth coordination site open for small molecule ligand binding. The RMSD plots (compared to the crystal structure) and representative zinc center geometries of individual models are presented in **Figures 2** and **3**, respectively.

Figure 2 shows the RMSD values of the Zn^{2+} binding site in the unrestrained LJ12-6-4 model to be consistently lower than those using the LJ12-6 model, whereas the LJ12-6 and LJ12-6-4 models have similar RMSD values when restraints are applied. Although the mean RMSD value for the MMTZ ligand is about 0.5 Å higher using the restrained LJ12-6-4 models than the restrained LJ12-6 models, the pose of the ligand in the former simulation is more stable. Taken together, the RMSD plots for the ligand and binding sites indicate that protein-ligand interactions are more stable over the duration of the simulation when the LJ12-6-4 model, rather than the LJ12-6 model, is used.

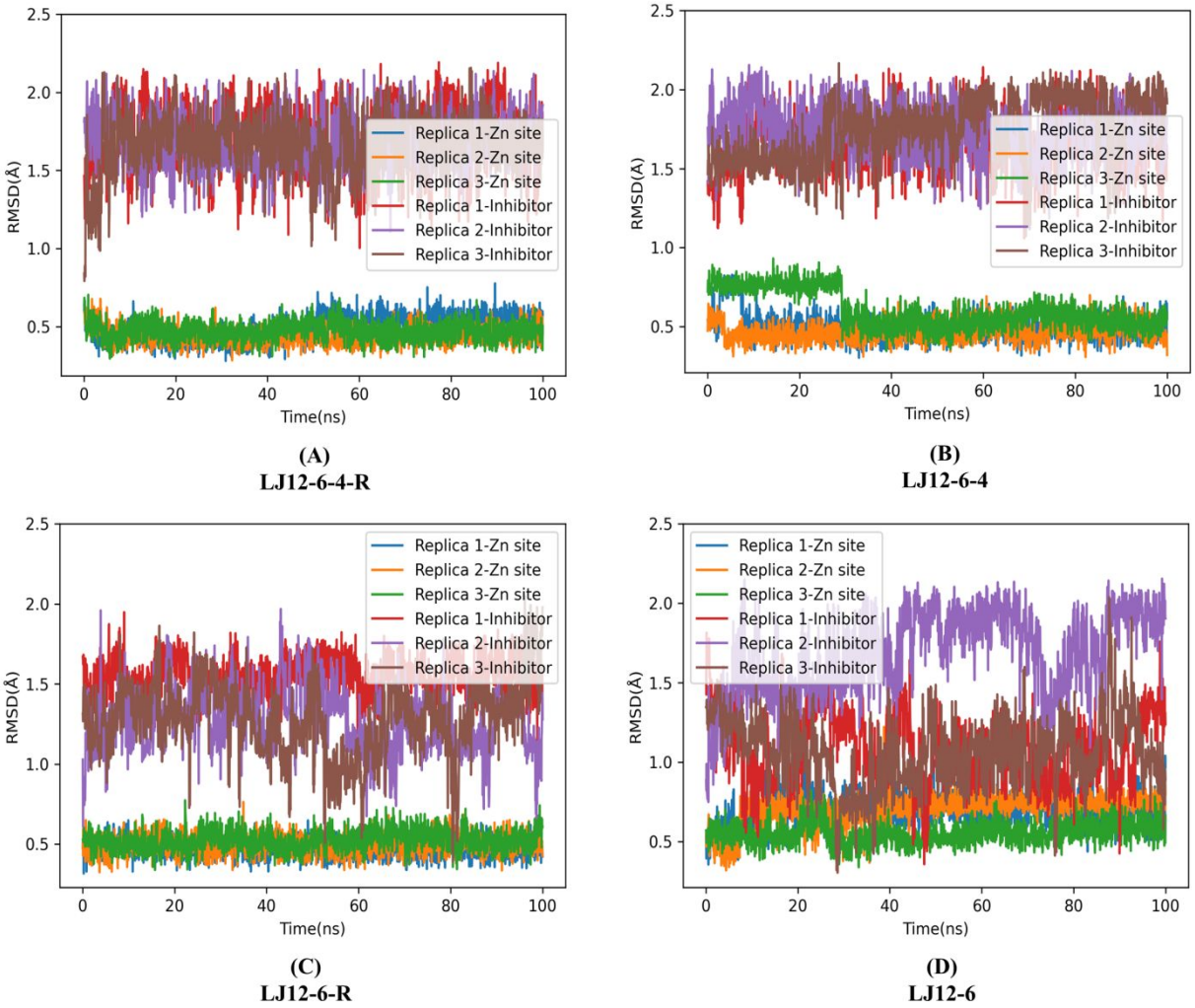


Figure 2. Time-dependence of RMSD values (compared to the crystal structure) for MM MD simulations of Sfh-I:L-anti-1a complex using non-bonded models. Three replicate simulations were performed for each model and each replicate is 100 ns. The 'Zn site' refers to zinc ions and zinc coordinating residues and 'Inhibitor' refers to the zinc bound compound.

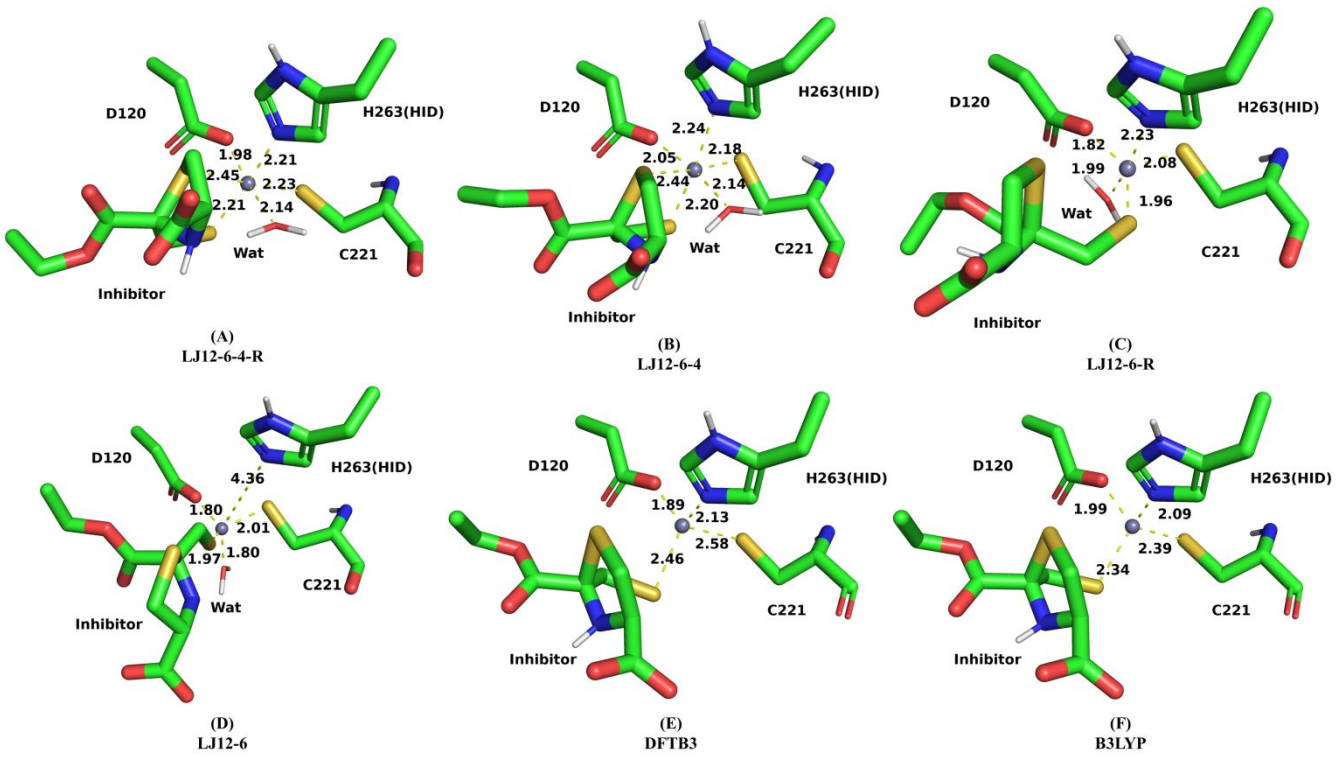


Figure 3. Zinc site geometries of Sfh-I:L-anti-1a complexes. (A-D) The representative zinc site geometry of the four non-bonded models after 100ns MM MD simulation. (E) The representative zinc geometry after 100 ps DFTB3/MM MD simulation. The extra coordination by an additional water molecule was removed. (F) The representative zinc geometry after B3LYP-D3BJ based QM/MM geometry optimization

A comparison of zinc coordination distances obtained from simulations using the four MM models is shown in **Table S1**. The performance of the two simulations using the LJ12-6-4 model is significantly better than that of those using the LJ12-6 model when considering the distances between the zinc ion and coordinating atoms. It is worth noting that during 100ns MD production runs using the unrestrained LJ12-6 model zinc coordination by the residue His263 was lost. Overall, the coordination distances predicted by the LJ12-6-4 models more closely approach the crystallographically determined values (the coordinate error of the crystal structure is 0.10 Å) and the similarity can be slightly improved with the use of distance restraints. In the majority of simulations using the restrained LJ12-6-4 model the distance restraint penalties were not triggered, demonstrating the improved ability of the LJ12-6-4 model, compared to the LJ12-6 model, to describe bond lengths. Our results suggest that the LJ12-6-4 models not only provide positional predictions for binding site residues closer to the crystal structure, but also can better reproduce coordination bond length values than the LJ12-6 model. However, when coordination geometry is considered, all four models showed a strong tendency to increase the coordination number of the Zn^{2+} ion to five or six, through the addition of an extra Zn-coordinating water molecule in the case of the LJ12-6 model, and with both a water molecule and the inhibitor thiazolidine sulfur atom forming additional bonds to Zn^{2+} in the LJ12-6-4 model. For the LJ12-6 model, application of distance restraints partially corrected this, resulting in improved performance yielding zinc coordination numbers (CN) closer to the experimental value.

Semi-empirical QM/MM MD. The data presented above demonstrate that, although the bond lengths between zinc and crystal coordinating ligands were largely consistent (~ 0.10 Å difference on average) with experimental values, the geometries were not. Accordingly, in an effort to improve the outcome, 100 ps DFTB3/MM MD was performed. As the MD trajectory and zinc site coordination were stable, the last frame of the production run of a LJ12-6-4 model was chosen as a typical snapshot after MD simulation for the following DFTB3/MM MD. The Zn^{2+} binding site (i.e., the zinc ion, the side chains of zinc coordinating residues and zinc coordinating water molecules added by MM MD) and the inhibitor were simulated at the semi-empirical level of QM theory using DFTB3 with the 3OB-3-1 parameter set,[58] while the rest of system was simulated using the Amber ff14SB forcefield.[59] The two additional coordination interactions introduced by the MM MD were not retained after the DFTB3 QM/MM dynamics simulation, reducing the

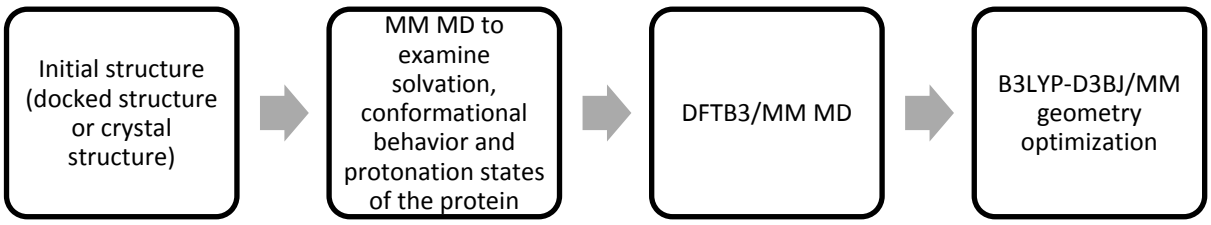
Zn^{2+} coordination number to four, which restored the structure to the crystal geometry (**Figure 2(C)**). The Zn^{2+}-ligand coordination distance was also optimized, getting closer to the value of the crystal structure (**Table S2**). Inspection of the plot of Zn^{2+}-ligand coordination distance against simulation time (**Figure S1**) demonstrated that that the major shift in zinc geometry (specifically removal of 'inappropriate' coordinating atoms) happened at the beginning of the QM/MM MD simulation, with the distance between Zn^{2+} and the additional water molecule immediately rising to above 5 Å, indicating loss of coordination, and that the system was relatively stable after 100 ps. Therefore, a simulation time scale of 100 ps appears sufficient to obtain a stable zinc coordination geometry with DFTB3/MM MD.

B3LYP/MM optimization. Although DFTB3/MM MD treatment substantially improved the quality of the model for the Sfh-I:L-anti-1a complex, discrepancies remained with respect to the experimental crystal structure, specifically regarding the coordination distances for Zn-S interactions involving both the Cys221 ligand and the inhibitor thiolate. Hence a QM/MM energy minimization using higher level DFT theory was performed to further optimize the geometry of the zinc site. The hybrid functional B3LYP, widely used in QM/MM studies of enzymes including zinc-containing proteins[43,47,48,60,61], and the 6-31G(d) basis set were selected for the DFT calculations. Two parameters were tested in the DFT QM/MM geometry optimization: the diffuse function and the empirical dispersion correction. The B3LYP/MM optimization was carried out with and without diffuse functions for heavy atoms, and with and without the D3BJ (D3 version of Grimme's dispersion with Becke-Johnson damping) empirical dispersion correction,[62,63] to test their effects. The results of these comparisons are shown in **Table S2**.

Use of diffuse functions improved the accuracy of Zn^{2+} - ligand coordination distance predictions, particularly for electron-rich sulfur atoms. The absence of empirical dispersion corrections leads to longer Zn^{2+} - ligand bond lengths, reducing the accuracy with which the experimental geometry is reproduced. Accordingly, the results presented here are those obtained using B3LYP/MM optimization with diffuse functions for heavy atoms and GD3BJ empirical dispersion. Compared to the DFTB3-optimized starting structure, we found that this resulted in further improvement in prediction of Zn^{2+}-ligand coordination distances, especially those between the zinc ion and ligating sulfur atoms. Ultimately the values for DFT-refined Zn^{2+} coordination distances were almost identical to those observed in the crystal structure (the total absolute deviation of all zinc ligating distances was 0.05 Å). The geometry of the zinc center in the DFT - based QM/MM optimized structure is shown in **Figure 3**. We found that the atomic distance between the zinc ion and its ligands only changes slightly in the zinc site during the B3LYP geometry optimization (**Figure S2**). We also found no significant difference in accuracy between a fully converged B3LYP-D3BJ/MM geometry optimization, and one progressed over 250 steps. As B3LYP/MM geometry minimization normally takes around 500-750 steps to converge, use of the 250th step as the end point for geometry refinement is thereby much more computationally efficient. A comparison of the Zn^{2+} site geometry predicted from this simulation and the crystal structure is shown in **Figure 4 (A).** In additional experiments, we also attempted application of the B3LYP/MM geometry optimization process directly after the 100ns MM MD simulation, instead of starting from the DFTB3-optimized structure. The results showed that the QM/MM optimization process cannot remove the additional coordinating water molecules to the zinc site introduced by the MM treatment. Instead, B3LYP/MM optimization led to the stable incorporation of additional water molecules into the Sfh-I zinc site, increasing the discrepancy between the modelled zinc site and that observed in the crystal structure.

The results of simulations of the Sfh-I:L-anti-1a complex suggested the workflow shown in **Scheme 1** as an approach capable of accurately describing this tetrahedrally coordinated zinc center. Non-bonded LJ12-6-4 MM MD simulation was first used to examine the general motion of the complex (e.g., examine solvation, conformational behavior and protonation states) and establish the stability of the system over a relatively long timescale. DFTB3 QM/MM MD is then applied to better describe the zinc site, and to rectify changes in coordination introduced by MM treatment. Subsequent B3LYP-D3BJ /MM optimization can be used to further refine the geometry of the zinc site (e.g., atomic distances, particularly for Zn – S interactions).

Scheme 1. Schematic of the workflow for modelling zinc-containing protein-ligand complexes.



Workflow validation: Inhibitor complexes of ACE and ACE2

To assess the broader applicability of this approach, the pipeline described above was then tested with two additional tetrahedral zinc systems (complexes of angiotensin converting enzyme (ACE) and its ACE2 isoform with the inhibitors L-captopril and MLN-4760). Compared to Sfh-I, above, the protein zinc sites lack a Cys ligand, and the inhibitors are coordinated via thiolate and carboxylate groups respectively, so representing a wider range of enzyme zinc centers. In the L-captopril complex[52] the ACE zinc ion is coordinated by three protein ligands (His367, His371 and Glu399) and the captopril thiolate (**Figure 1(B)**). In the ACE2:MLN-4760 complex[53] (a structure determined at lower resolution (3.00 Å) enabling assessment of the performance of the modelling workflow with a lower-quality starting structure) the zinc ion is coordinated by the equivalent protein ligands (His374, His378, Glu402) and the inhibitor carboxylate (**Figure 1(C)**). Based upon results obtained for the Sfh-I:D-anti-1a complex (above) the LJ12-6 models were discarded for the MM MD simulations and the LJ12-6-4 and LJ12-6-4-R models only were used. RMSD plots (compared to the crystal structures) and representative geometries for the ACE and ACE2 zinc centers at different stages of the modelling workflow are presented in **Figures S3 - S6**.

For both the ACE and ACE2 models, the RMSD values of residues within the zinc binding site were consistently lower than 1 Å, suggesting that stable geometry is maintained across the 100 ns of the simulation. (Notably, although the simulations overall remained stable the inhibitor RMSD value of the bound ACE-2 inhibitor in one replicate using the LJ12-6-4 model increased suddenly from ~ 1 Å to 2.5 Å at around 35ns of the simulation and remained at ~ 2 Å for a further 30 ns before reverting to baseline). This restoration of inhibitor binding geometry, and the disappearance of other occasional RMSD peaks, suggested that the MD treatment using LJ12-6-4 models was able to handle incorrect ligand poses and form a reasonable binding geometry. Subsequent DFTB3/MM MD simulations were carried out starting from the last frame of the 100ns MM MD simulation, and were followed by a 250 step DFT-based QM/MM geometry optimization using the B3LYP functional with GD3BJ dispersion correction (B3LYP-D3BJ) and the 6-31+G(d) basis set. In the case of ACE, the DFTB3 MM MD step was trialed starting from the last frame of the unrestrained MM MD simulations using the LJ12-6-4 model, with negligible differences between the starting or end points. For ACE2 DFTB3 QM/MM MD began from the last frame of both the restrained and unrestrained MM MD simulation. Zn^{2+} coordination geometries from the various treatments are reported in **Table S3**. (Note that both oxygen atoms of the ACE2:MLN-4760 carboxylate were included to monitor the performance of the model on Zn^{2+}-carboxylate interactions).

In general, the QM/MM optimized structures were close to the crystal structures and, although the coordination distances were not perfectly predicted, the Zn^{2+} binding geometry in the crystal structure was successfully restored. Alignment of the ACE and ACE2 Zn^{2+} sites to the respective crystal structures is shown in **Figure 4(B) and (C)**. The results showed that the workflow developed using Sfh-I worked well for these additional systems with tetrahedral zinc binding centers that do not contain Cys, with both thiolate (ACE) and non-thiolate (ACE2) small-molecule ligands.

For the ACE2:MLN-4760 complex, **Table S3** also includes the interaction between the zinc ion and the inhibitor O3 atom, even though the crystallographically observed distance (2.62 Å) is beyond the 2.5 Å boundary considered to be the limit for a coordination bond. This is due to our wish to investigate whether the MM and QM/MM models can correctly handle this weaker interaction. The results showed that, during MM simulations using the LJ12-6-4 series models, the Zn^{2+} – O3 distance reduced to form an additional coordination bond not present in the crystal structure, while the QM/MM simulation correctly handles this interaction. The modelled ACE2 structure after QM/MM refinement showed an unexpectedly large (0.3 Å) deviation from the experimentally observed distance for the interaction between Zn and His378-NE2. As this is a low-resolution (3 Å) structure, the ACE2 zinc center was inspected using the CheckMyMetal server[64,65]. The results indicate that the experimentally determined value (2.31 Å) for this Zn^{2+} - N distance is likely to be an outlier, given that Zn^{2+} - N distances in protein zinc sites are mainly distributed in the range: ~1.9 - 2.2 Å, and that the gRMSD value (defined as the RMSD of the observed ligand-metal-ligand angles compared to their idealized values) (23.3°), is also an outlier. These findings, when considered together with the low resolution of this structure, may explain why the QM/MM optimized values in Table S1 lie relatively far from the experimental figures.

Extension to a 5-coordinate system: HDAC2 complex with SAHA

After testing the ability of this approach to model 4-coordinated Zn^{2+} centers, we then tested it on a Zn^{2+} center with 5-coordination. We selected the crystal structure of HDAC2 in complex with the hydroxamate inhibitor suberanilohydroxamic acid (SAHA, also known as Vorinostat and used clinically for treatment of cutaneous T cell lymphoma) solved at a resolution of 1.85 Å (PDB code: 4LXZ)[54] as a model system. In this structure the HDAC2 zinc ion is coordinated by three amino acids (Asp181, His183 and Asp269) and by two oxygen atoms of the SAHA hydroxamate head group (**Figure 1(D)**).

MM MD. The performance of MM MD simulations using nonbonded models was investigated first. Four models were tested: the LJ12-6 and LJ12-6-4 models, each with and without restraints. Three replicate simulations (100ns/replicate) were performed in each case, RMSD values, compared to the starting crystal structure, for the active site and bound SAHA are shown in **Figure S7**. Although RMSD values of the zinc binding site were consistent (~0.4 Å) across the LJ12-6 models and the LJ12-6-4 restrained model in general. the unrestrained LJ12-6-4 model yielded higher RMSD values than the other models.

However, compared to the relative stability of the zinc-coordinating residues, the RMSD values of the SAHA ligand show much greater fluctuation, with bound SAHA adopting a similar pose at the end of each simulation that is distinct from that observed in the crystal structure. Comparison of zinc coordination distances shown in **Table S4** indicates that the overall performance of the four models with respect to the zinc site is similar, although the accuracy with which coordination distances are predicted varies. In general, the LJ12-6-4 models yield Zn-O coordination distances that are more realistic, and more consistent with the crystallographically determined values, than those obtained from the LJ12-6 models. In contrast to the outcome from our earlier calculations with Sfh-I (above) application of distance restraints did not further improve the performance of either model, possibly indicating that the non-bonded MM models handle zinc binding sites that include only Zn-O and Zn-N interactions more easily than those that also include a Zn-S interaction. However, in all cases the zinc coordination number at the end of the simulation had increased to 6 as a result of a move to bi-, rather than monodentate coordination by Asp 181.

DFTB3/MM MD. 100ps of DFTB3/MM MD was then performed starting from the last frame of the 100 ns MM MD production run with the restrained LJ12-6-4 model, with the HDAC2 zinc site and SAHA ligand included in the QM region. After DFTB3 QM/MM MD simulation, zinc coordination by Asp 181 shifted from bi- to mono-dentate, returning the Zn coordination number to five, as observed in the crystal structure. However, we also found occasionally that zinc was in a tetrahedral geometry after DFTB3 QM/MM simulation, with the detachment of the SAHA O2 atom.

B3LYP/MM optimization. We then ran a B3LYP-D3BJ /MM geometry optimization after the DFTB3 treatment to further optimize the zinc-ligand coordination distance. The 6-31G(d) basis set was first applied, and the diffuse function subsequently added. The results of this further round of QM/MM optimization are shown in **Table S5**. In this case, incorporation of the diffuse function had no effect on the accuracy with which Zn^{2+} coordination distances were predicted. The results clearly show that in this case B3LYP-D3BJ/MM yielded a better result for Zn^{2+} coordination distance and the overall geometry closely resembles the crystal structure (**Figure 4(D)**). Moreover, as detailed for simulations of Sfh-I (above) we observed only slight improvement in prediction accuracy for structures that had undergone 542 steps (converged) of B3LYP-D3BJ/MM geometry optimization, compared to those that had only undergone 250 steps. This increases our confidence that 250 steps of B3LYP QM/MM geometry optimization should be sufficient to build accurate models of complexes of zinc metalloproteins.

Application to di-zinc systems: complexes of the IMP-1 and L1 metallo-beta-lactamases.

In addition to the wide range of zinc metalloproteins with mono-zinc centers, several enzyme classes of mechanistic and pharmacological and/or biotechnological interest possess dinuclear zinc centres.[66,67] Accordingly, we also investigated the application of the combined MM and QM/MM MD approach to model complexes of di-zinc metalloenzymes, specifically metallo-beta-lactamases (MBLs), enzymes that confer resistance to a broad range of beta-lactam antibiotics upon producer bacteria.[68,69] Two MBL model complexes were selected: the complex of IMP-1 with the small thiol thioglycolic acid (TGA, resolution 1.57 Å, PDB code: 6JED) [55] and the complex of the L1 MBL from *Stenotrophomonas maltophilia* with the hydrolyzed form of the penem antibiotic faropenem (resolution 1.57 Å, PDB code: 7A63). [56] Both complex structures contain two zinc ions that are close together in space (3.55 Å and 3.57 Å respectively for the IMP-1 and L1 complexes). In the IMP1:TGA complex one zinc ion (Zn1) is tetrahedrally coordinated by three histidine residues (His77, His79, His139) and the thiolate of TGA (**Figure 1(E)**), while the second (Zn2) is 5-coordinated by three amino acids (Asp81, Cys158 and His197) and the thiolate and carboxylate oxygen atom of TGA. In the L1:faropenem complex the equivalent zinc ions are respectively tetrahedrally coordinated by three histidine residues (His83, His85 and His159) and a water molecule (which bridges the two zinc ions); and octahedrally coordinated by three amino acids (Asp87, His 88, His224), the carboxylate oxygen and ring nitrogen of hydrolyzed faropenem, and the zinc-bridging water. The two systems thus resemble one another in some aspects (sharing a tetrahedrally coordinated zinc ion in a tri-histidine center and with a bridging, non-protein ligand); but differ in others (with the second zinc ion in IMP-1 5-, as opposed to 6-coordinated and with IMP-1 containing a Cys ligand).

Workflow validation for the di-zinc system using the IMP-1 complex

MM MD. MM MD simulations on the IMP-1:TGA complex were carried out as above, using both the LJ12-6 and LJ12-6-4 models with and without restraints. Of note, the GAFF2 (Ver 2.11) forcefield was used to parameterize the inhibitor. Three replicate simulations were run for each model; RMSD values for the zinc centers and bound inhibitor, compared to the crystal structure, are presented in **Figure S9**.

Inspection of **Figure S9** shows that in all cases bound TGA remains close to the crystallographically observed binding pose throughout the duration of the simulation, with RMSD values remaining stable and below 0.5 Å throughout the simulation. However, when the unrestrained LJ12-6 model was used RMSD values for the binding site residues increased dramatically, although these remained consistently low (~0.7 Å) when restraints were applied or the LJ12-6-4 model was used. Visual inspection of the simulation trajectory identified that in the case of the unrestrained LJ12-6 model zinc coordination was disrupted by replacement of coordinating histidine residues with water molecules. Compared to the LJ12-6 models, the zinc coordination distances obtained from simulations using the

LJ12-6-4 models were overall closer to the crystallographic values (**Table S6**). Accuracy also improved when restraints were applied, although the effect was far less pronounced than in the case of the LJ12-6 model. The distance between two zinc ions was however better predicted by the LJ12-6 series models, whereas this increased by ~0.4 Å when the LJ12-6-4 model was used. Overall, and consistent with our findings for the mono-zinc systems, we consider the LJ12-6-4 model to also be the first choice for this di-zinc system. However, the coordination number for each zinc ion increased to 6, as a result of the introduction of additional water molecules into the coordination shells of both zinc ions after LJ12-6-4 MM MD simulations.

DFTB3/MM MD. As with our previous approaches, 100 ps of QM/MM MD using DFTB3 was then performed starting from the last frame of the MM production run. The QM region included the inhibitor, the zinc ions and the side chains of their coordinating residues, and the zinc-bound water molecules. The additional water molecules were successfully removed by this treatment, restoring the total Zn^{2+} coordination number to nine. However, with the exception of the interaction with His197, the Zn^{2+}-ligand distances for Zn2 increased by 0.1 ~ 0.2 Å, and approached the coordination boundary of 2.5 Å. This was particularly the case for Zn-S interactions involving both Cys221 and the TGA thiolate.

B3LYP/MM optimization. BL3YP-D3BJ based QM/MM geometry optimization was carried out using the last frame from the 100ps DFTB3/MM MD. In the interests of computational efficiency, the 6-31G(d) basis set was applied first, and the diffuse function was added subsequently. The results of the DFTB3/MM MD calculation and DFT/MM optimization are shown in **Table S7**. After DFT refinement the Zn^{2+} coordination distances were very close to their values in the crystal structure. However, the incorporation of the diffuse function did not show a large improvement on the accuracy of predictions for Zn^{2+} coordination distances. We also found that there was no improvement in accuracy between structures that had met the convergence criteria (449 steps) for DFT/MM geometry optimization and structures that had only undergone 250 steps. Snapshots of the active site of the IMP-1:TGA complex at different stages of the simulation workflow are shown in **Figure S10** and alignment of Zn^{2+} site geometries obtained from simulations and the starting crystal structure, is shown in **Figure 4(E).**

Workflow implementation for L1 complex with hydrolyzed faropenem

After demonstrating successful treatment of the IMP-1:TGA complex, the pipeline was tested against the complex of the L1 metallo-beta-lactamase with the hydrolysis product of the penem antibiotic faropenem. As observed for IMP-1, MM MD simulations using the LJ12-6-4 model yield similar results regardless of the inclusion of restraints: the RMSD plots showed that the zinc center remained stable across the simulations while the geometry visualization showed that in each case both zinc ions were both octahedrally coordinated after 100 ns MM MD simulation (**Figures S11, S12**). MD treatment leads to coordination of both zinc ions by the faropenem C6 carboxylate group, and moves coordination of Zn2 by Asp87 from mono- to bidentate. In addition, the hydroxyl group connected to the faropenem C1 carbon moved to coordinate Zn1. As the above modes of coordination generated by simulations using MM non-bonded models are clearly unrealistic (compared to the crystal structure), DFTB3/MM MD was carried out from the last snapshot of a restrained LJ12-6-4 simulation to address this artificial Zn^{2+} coordination. The distorted zinc geometry was successfully restored to that observed in the crystal structure by this subsequent 100 ps DFTB3/MM MD step. The structure obtained from DFTB3 treatment was then further refined to optimize zinc coordination distances with the B3LYP-D3BJ based QM/MM minimization approach (**Figure S12, Table S**). The 6-31G(d) basis set was first applied and the diffuse function was then added. In this case, incorporation of the diffuse function slightly improves the prediction accuracy of the Zn^{2+} coordination distances but significantly increased the computational cost. The alignment of the Zn^{2+} site geometry predicted from these simulations to the crystal structure is shown in **Figure 4(F)**.

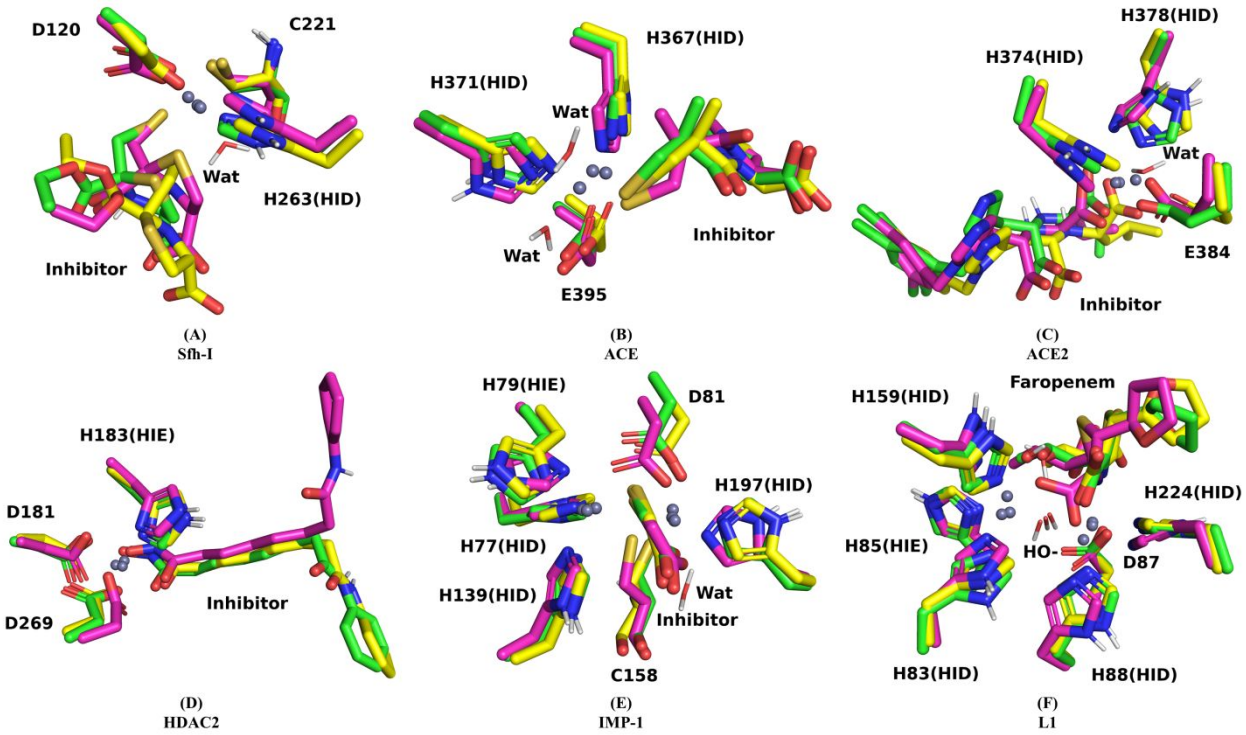


Figure 4. Zinc binding sites of the 6 protein-ligand systems included in the study after simulations at different levels of theory. Carbon atoms are in green/yellow/purple depending on the model: yellow represents the crystal structure; purple represents a typical structure after 100 ns MM simulation with LJ12-6-4 models; green represents a representative structure after DFTB3/MM MD and B3LYP-D3BJ based QM/MM geometry optimization. Nitro atoms are in blue, oxygen atoms are in red and sulfur atoms are in dark yellow, zinc ions are represented as gray balls.

DFTB3 calculations directly from crystal structures.

Comparison of the results obtained from the various systems identified that when DFTB3/MM treatment was used to rectify distortions (e.g. additional zinc-coordinating interactions, such as by water molecules, introduced during MM MD simulations) this often resulted in preferential production of 4-coordinate geometry for individual zinc ions. To investigate whether DFTB3 can maintain appropriate zinc coordination when applied to a more accurate starting structure, we ran 2ns DFTB3/MM MD simulations directly from the crystal structures of each system using the same QM regions as in the QM/MM calculations described above. The outcomes of these simulations are shown in **Figure 5** and **S13**. The results showed that some structures (the ACE, ACE2 and L1 complexes) were well described by direct application of the DFTB3 QM/MM approach, but that this was not the case for the Sfh- I, HDAC2 and IMP-1 complexes. For the Sfh-I:L-anti-1a complex, the inhibitor showed a relatively high RMSD value compared to the ligands in other simulations. The interaction of the inhibitor thiolate with Zn^{2+} was lost, with Zn^{2+} coordination substituted by a water molecule. This may reflect the relatively high importance of the interaction between the inhibitor thiolate and zinc ion to the affinity of this system, compared to the much less extensive interactions made with the remainder of the active site[51], and is also consistent with the relatively poor performance of our previous DFTB3-based simulations in describing zinc-sulfur interactions.

For the HDAC2:SAHA inhibitor system, although the SAHA binding pose remained similar to that observed in the crystal structure, the SAHA O2 atom sometimes lost its attachment to the zinc ion (with the Zn^{2+} – O2 distance increasing to > 2.5 Å) after DFTB3 QM/MM simulation, with the zinc ion adopting a tetrahedral geometry. This was the same outcome as observed on DFTB3 optimization after MM MD simulations of this system (see above). The situation was however improved when DFTB3 treatment was applied using a larger QM region, that included not only the zinc ion, the side chains of coordinating residues and the inhibitor, but also the side chains of residues that are not in the zinc site but may interact with the inhibitor. In the case of HDAC2, when Tyr297 and His135 were included in an enlarged QM region coordination between SAHA O2 and the zinc ion was retained. However, analysis of the distances between SAHA and interacting residues suggested that these were not affected by changes to the size of the QM region. In the case of IMP-1, the major inaccuracy when using DFTB3 treatment is the incorrect calculation of Zn^{2+}-inhibitor coordination distances, with both Zn^{2+} - S and Zn^{2+} - O coordination distances predicted to be greater than 2.5 Å. As in the case of Sfh-I, above, this may reflect the involvement of sulfur atoms in interactions with zinc by both the protein and small-molecule ligand.

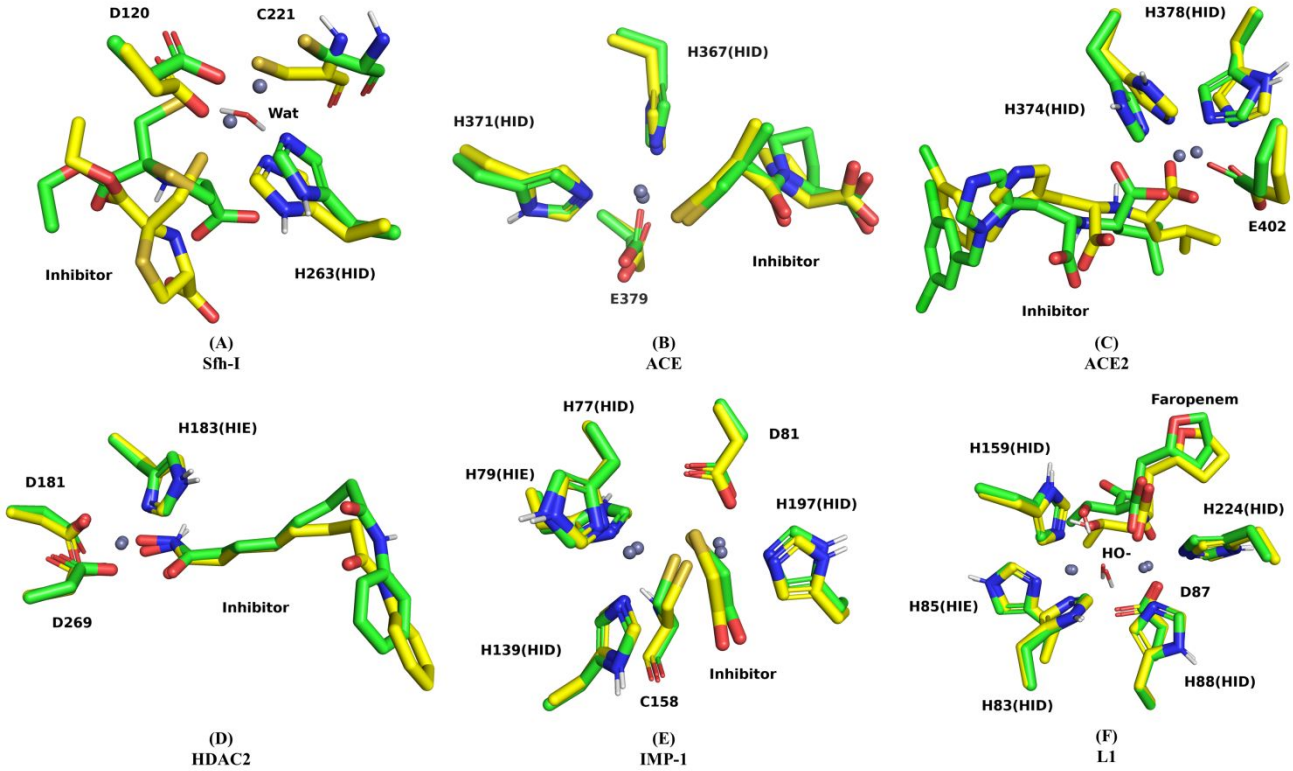


Figure 5. Zinc binding sites of the 6 protein-ligand systems included in the study after DFTB3/MM MD starting directly from the crystal structure. Carbon atoms are in green and yellow depending on the model: green represents a representative zinc site structure after 2ns DFTB3/MM MD starting directly from the crystal structure, yellow represents the crystal structure. Nitro atoms are in blue, oxygen atoms are in red and sulfur atoms are in dark yellow, zinc ions are represented as gray balls.

Discussion

The diversity and versatility of zinc coordinating interactions in zinc metalloproteins and their complexes with small molecule ligands makes accurate modelling of these systems an undertaking that is frequently challenging. In this work our aim was to develop a computational approach able to accurately describe ligand complexes of 6 diverse zinc metalloprotein systems, representing a range of coordination chemistries and geometries. Whilst we here focus on complexes of known crystal structure against which the accuracy of the various approaches can be assessed, in many cases crystal structures of zinc metalloprotein complexes with known or putative small molecule ligands will not be available. Hence we sought to develop a pipeline that is sufficiently robust to deal with more approximate starting structures. For this reason, our evaluation included MM MD methods (as might be applied to optimize structures of putative complexes generated from e.g. docking experiments) as well as QM/MM MD at both semi-empirical (DFTB3) and higher (B3LYP-D3BJ based DFT) levels of theory, with each approach representing increasing levels of accuracy in treatment of zinc and its interactions, at the expense of increasing demands upon computational resources.

For the 6 tested systems, non-bonded MM MD approaches using the unrestrained LJ12-6-4 and restrained (both LJ12-6-R and LJ12-6-4-R) models yielded RMSD values of whole protein backbone atoms, compared to the starting crystal structures, for the 6 systems that were consistent across the 4 tested non-bonded models. This suggests that the differences between the LJ12-6 and LJ12-6-4 models mainly occur at the zinc center. In general, the LJ12-6-4 nonbonded model outperformed the LJ12-6 models in terms of accuracy of Zn^{2+}-ligand coordination distance (smaller overall deviations from crystallographic values), consistency of simulation (lower standard error and more stable RMSD plots) and lower dependence on restraints. However, the LJ12-6-4 models show a much greater tendency than the LJ12-6 models to increase the zinc ion coordination number to six, usually by means of introduction of additional water molecules to the zinc site, resulting in octahedral coordination geometries (e.g., Sfh-I, ACE, ACE2, IMP-1, etc.). This difference likely reflects differences between the Zn^{2+} sphere radius used in the various models. Specifically, the R_{min} value for zinc (1.276Å) for the LJ12-6 CM parameter set is much smaller than that for the 12-6-4 parameter set (1.454 Å). Smaller sphere sizes (R_{min} values) reduce the space available for additional water molecules to "crowd" into the zinc site. Despite inclusion in the LJ12-6-4 model of a C4 term to represent dipoles, its performance still reflects design and validation based on

aqueous Zn^{2+} complexes, in which the preference for octahedral geometry is well established, compared to a relatively earlier stage of development for describing the behavior of zinc in a protein environment where a wider range of geometries are evident, with tetrahedral particularly common. Fortunately, however, our work here demonstrates that additional zinc interactions introduced in MM MD simulations using the LJ12-6-4 model are normally water molecules, and that these can usually be removed with subsequent QM/MM MD approaches. With this caveat, the LJ12-6-4 model is a good choice for modelling small molecule complexes of zinc containing proteins, being easy to be set up, delivering reproducible simulations and overall predicting reasonable values for zinc coordination distances, over a longer timescale than is possible with QM/MM MD approaches and at reasonable computational cost.

The use of distance restraints was essential for obtaining reasonable results with LJ12-6 models, as without these some ligand or protein side-chain interactions with the zinc could be lost (e.g., in the above IMP-1:TGA complex). Although the application of distance restraints did not affect the LJ12-6-4 model as much as the LJ12-6 model, these did improve the consistency between individual simulations in our sets of three replicates and contributed during the initialization of the system (energy minimization and equilibration). Although small molecule ligands were not restrained during either the equilibration or production MD processes, these did require restraining during energy minimization to retain or obtain a reasonable starting pose. Selection of appropriate restraints then represents an important determinant of the overall success of MM MD steps, as inappropriate choices may lead to artificial interactions resulting in unrealistic descriptions of zinc binding or detachment of the small molecule from the zinc ion(s); and requires input based on prior knowledge of common interaction patterns between zinc and typical zinc-binding groups (e.g. thiolates, carboxylates, hydroxamates).

An additional factor influencing the quality of the results obtained with MM MD simulations is the ligand parameters. GAFF and GAFF2 from the AMBER package are general forcefields for parameterizing small molecule ligands; in the work presented here GAFF generally performed well. (Difficulties with implementation in LJ12-6-4 models precluded more extensive use of GAFF2). Of the model systems investigated here, limitations to ligand parameterization may be apparent in MM MD simulations of Sfh-I, where the ligand pose is poorly replicated, possibly due to the presence of the thiazolidine ring system and the additional thiolate sulfur atom; and HDAC2, where the aromatic ring at the end of the SAHA alkyl tail is oriented oppositely to its position in the crystal structure. In both cases the accuracy with which the ligand position was modelled was improved by subsequent DFT/MM optimization. Nevertheless, the GAFF approach retains the major advantages of speed and ease of use, making it a good choice for fast model building or ligand screening tasks with acceptable levels of accuracy.

The DFTB3 and B3LYP-D3BJ methods were the two QM approaches investigated here. In our simulations, the DFTB3 approach showed a preference for formation of a 4-fold (tetrahedral) zinc coordination geometry, and was an effective tool for removing additional water molecules introduced during MM MD with LJ12-6-4 models. However, when both sulfur atoms and water molecules added during MM MD simulations are present in the zinc binding site, with DFTB3 the Zn - S contact is occasionally lost and zinc instead interacts with water molecules. In addition, the coordination distances for Zn - S interactions were often badly described with DFTB3 treatment (i.e., the Zn - S coordination distances were often close to 2.5 Å, substantially longer than the typical experimental values of 2.15 – 2.35 Å). Taken together, these findings suggest that the DFTB3 approach may be less effective in handling Zn - S interactions. Although DFTB3/MM MD simulations were run for 100ps, additional water molecules introduced in preceding MM MD simulations departed from the zinc center at the very beginning (i.e., first 10ps) of the run, suggesting that shorter timescales of 50ps or even 20ps may be sufficient for this purpose. Limiting the duration of DFTB3/MM MD simulations may also reduce the possible impact on other Zn^{2+} - ligand interactions, leaving final refinement of the zinc center for subsequent B3LYP QM/MM optimization.

Our results demonstrate that removal of additional zinc-coordinating water molecules requires DFTB3/MM MD simulations. In comparison, B3LYP/MM optimization had only limited capability to alter the geometry of the zinc site, and in most cases was unable to remove artificial zinc coordinating water molecules (resulting in 6-coordinate zinc geometry) introduced during MM MD simulations. Computational expense precluded B3LYP/MM MD simulations. Our data show that B3LYP/MM optimization should be carried out from an appropriate starting geometry, and not one that contains significant distortions, such as those that may be introduced during preceding MM MD steps.

We found the GD3BJ dispersion correction to be important for accurate description of zinc coordination distances in simulations using B3LYP, with inclusion of the diffuse function important in systems containing sulfur atoms. In addition, only minor differences were observed when comparing structures obtained from a fully converged B3LYP-D3BJ simulation and after 250 steps of B3LYP-D3BJ treatment. Given the difficulty in reaching convergence when using DFT QM/MM optimization (which with our hardware infrastructure normally required around 750 steps and approximately 10 days of calculations), we then consider a 250-step optimization likely to be sufficient to obtain model geometries of acceptable levels of accuracy. Furthermore, the computational efficiency of the B3LYP-D3BJBJ/6-31+G(d) treatment was greatly improved when the structure was first optimized with B3LYP-D3BJ/6-31G*. A computationally efficient B3LYP-D3BJ-based QM/MM optimization approach could then be 250 steps of B3LYP-D3BJ/6-31G(d) first, followed by a further 250 steps with diffuse functions included. Of the three methods tested, the B3LYP-D3BJ approach then provided the most accurate descriptions of the systems under test, but required

previous geometry optimization of the starting structure by e.g. DFTB3 treatment.

Although MM MD may provide a robust and computationally efficient approach to obtaining a proper complex system for simulation (e.g., protonation state, stable conformation of the complex), especially when docked structures are used as initial structures, our results indicate that incorporation of an initial MM MD step is not essential when simulating ligand complexes of zinc containing proteins from crystal structures. In such cases reasonable descriptions of metal centers can be obtained by using QM/MM methods directly. In particular, although the semi-empirical DFTB3 method is less effective than DFT in predicting coordination bond lengths, and can present particular difficulties in dealing with sulfur atoms, DFTB3 consumes less computational resources than B3LYP-D3BJ, making possible QM/MM MD simulations on the nanosecond timescale. Moreover, DFTB3 QM/MM MD simulations have some ability to restore distorted zinc geometries, although caution should be taken when sulfur atoms are involved.

Conclusions

The extensive set of simulations described here for a range of zinc metalloprotein systems allows some general conclusions to be drawn regarding approaches that may be suitable to generate realistic models of zinc metalloprotein complexes. MM MD simulations using non-bonded models failed to maintain experimental zinc coordination and (as expected) showed strong tendencies towards octahedral geometries, due most likely to the treatment of zinc ions. DFTB3 MM/MD treatment retained experimental coordination when crystal structures were used as starting models, and was often able to remove water molecules added by prior MM MD treatment, but was less well able to describe zinc centers with coordinating waters exposed to bulk solvent and, in particular, zinc-sulfur interactions. B3LYP-D3BJ/MM geometry optimization well describes zinc centers and can restore distorted coordination distances close to experimental values, but does not result in large-scale changes to geometry and so requires an appropriate initial structure. Our data suggest that inclusion of empirical dispersion corrections in these simulations is beneficial, while use of diffuse functions in B3LYP-D3BJ/MM geometry optimization may improve the quality of the final model but does not always do so, and imposes a dramatic increase in computational cost. We conclude that a computational pipeline involving DFTB3 QM/MM MD simulations at the semi-empirical DFTB3 level of theory, followed by B3LYP-D3BJ/MM geometry optimization, should be sufficient to generate models of protein zinc centers, with preceding classical MM MD not essential unless there are requirements for significant optimization of the starting model, observation of the motions of the system over longer time scales, or examination of protonation states. In such cases LJ12-6-4 models provide a more accurate description of zinc than LJ12-6 models in terms of coordination distance, and the unrestrained LJ12-6 model is not recommended. While unrestrained simulations could be used to examine the potential movement of zinc binding residues, restrained simulations are more appropriate to simulate the dynamics of the zinc site while as far as possible retaining crystallographically observed geometry. Overall, the pipeline of MM MD, DFTB3/MM MD and B3LYP-D3BJ/MM geometry optimization provided good results in our tests. We suggest that this constitutes a robust and versatile approach suitable for modelling diverse zinc metalloproteins and their complexes.

Methods & Materials

Molecular Mechanics (MM) dynamics simulations. The PDB file for the Sfh-I inhibitor complex (PDB code: 7BJ9)[51] was obtained from Dr Philip Hinchliffe prior to release. Crystal structures of other protein-ligand complexes were obtained from the PDB database (https://www.rcsb.org/).[70,71] Crystallographic water molecules within 10 Å of the ZN ion were retained, other in the PDB files were removed. PDB headers and all lines other than 'ATOM', 'HEATM', 'TER' and 'END', were deleted. The protonation state of the protein was determined by PROPKA 3.0[72,73] through the PDB2PQR[74] server. Packages including tLEaP, Antechamber and ParmEd of AmberTools20[15] were used to parameterize the protein-ligand system. Hydrogens were added to the protein using tLEaP. Antechamber was used to generate parameterization files for the ligands: inhibitors were described by the general AMBER force field (GAFF)[75] and AM1-BCC charge method unless otherwise specified. The complex was solvated in a 12 Å water box using tLEaP. The Amber ff14SB forcefield and SPC/E water model were used to parameterize the system of protein and water molecules. After balancing the charge of the system using Na^+ and Cl^- counter ions, the topology and the coordinate files of a typical nonbonded LJ12-6 model of the system were generated (the LJ12-6 CM parameter set). The topology file editor ParmEd was then called to add C4 terms to the LJ12-6 topology file and coordinate file generating corresponding files for the LJ 12-6-4 model. After system preparation, the simulation used the pmemd engine of Amber18.[13]

The system first went through minimization of the hydrogen atoms followed by minimization of the water molecules, then minimization of side chains and finally of the whole protein-ligand complex. After the minimization, the system was slowly heated to 298K over 200 ps. After that, a 2-nanosecond equilibration process was carried out to optimize the system configuration. Distance restraints were implemented in the above steps to keep the zinc binding site stable. Positional restraints were applied to the protein backbone atoms and the zinc ion. The strength of positional restraints gradually decreased (starting from 25 kcal/mol) as the equilibrium progressed and these were totally removed in the last equilibration step (1 ns). Distance restraints were applied to help maintain the zinc coordination geometry. The values of the lower (r2) and upper (r3) bounds were respectively set to be ~0.15 Å from the crystallographically observed values. The constants rk2 and rk3 were set to be 50.0 kcal/mol. MD simulations

(production runs) with or without (depending on model type) restraints were launched when the equilibration was completed. The non-bond cut off distance was set to 10 Å and the system temperature was set to 298K and regulated by Langevin dynamics with a collision frequency of 2. The Berendsen barostat was applied to control the system pressure at 1 atm. The SHAKE algorithm was applied to allow a larger time step. The simulation ran for 100 ns with a time step of 2 femtoseconds under the constant pressure periodic condition. MD trajectories were analyzed and RMSD values calculated by Cpptraj, a trajectory processing package included in AmberTools20. Records of the system status during MD simulations were analyzed by the Python script 'mdout_analyzer' of AmberTools19.[16] MD trajectories were visualized by VMD (Version 1.9.4)[76] and snapshot analysis (e.g., distance measurement between atoms) carried out using PyMoL (https://pymol.org/)[77]. 2.5 Å was set as the boundary distance for zinc coordination analysis.

QM/MM dynamics and geometry optimization. QM/MM calculations were carried out at two levels of QM theory: DFTB3 and DFT (B3LYP). In both DFTB3 and B3LYP QM/MM calculations, the QM region was defined as the inhibitor, zinc ions, any water molecule coordinated to the zinc ions, and the side chains of the zinc coordinating residues; while the rest of system was defined as the MM region. Link hydrogen atoms were automatically added to the system by Amber's QM/MM engine Sander. QM/MM MD was carried out from the last snapshot of preceding MM MD simulations to restore the coordination geometry of Zn^{++}, or directly from crystal structures. Calculations were performed using AMBER 18 which has built-in DFTB3 code and can handle both the QM and MM portions of the calculation. The QM region was modelled by DFTB3 theory, and the MM part was modelled using Amber forcefield ff14SB (the same profile as the MM production run). For the QM region, the nonbond cutoff distance was set to 8 Å and the SHAKE algorithm was used. QM/MM geometry optimization using DFT theory for the QM region was performed based on the result of DFTB3/MM MD or the last frame of MM MD simulation (depending on the system). The B3LYP/MM optimization was performed with AMBER 18 via its external QM program interface: Gaussian 16[78] was called to initiate single point calculations for the QM region and the MM portion was handled by AMBER 18 using the Amber ff14SB forcefield[59]. In the DFT optimization a QM cut-off distance of 8 Å was set. For the QM portion, we used B3LYP hybrid functionals with the 6-31G(d) or 6-31+G(d) basis sets. The D3 version of Grimme's dispersion correction with Becke-Johnson damping[62,63] was also applied, which is important for predicting accurate protein structures. In all QM/MM calculations, the particle mesh Ewald (PME) method was implemented to calculate long-range QM-QM and QM-MM electrostatic interactions. Covalent C-C bonds at the boundary of the QM and MM region were treated by adding hydrogen link-atoms, which were automatically placed by AMBER. SCF convergence thresholds was set to be SCF=(Conver=8). The outputs from QM/MM calculations were analyzed by Cpptraj in AmberTools20[16] and visualized using VMD (Version 1.9.4) and PyMoL (https://pymol.org/)[77]. As above, 2.5 Å was set as the boundary distance for zinc coordination analysis.

ASSOCIATED CONTENT

Supporting Information. Plots of zinc coordination distances during QM/MM simulations of Sfh-I:L-anti-1a complex, Figures showing RMSD plots and representative geometries for other simulations, Tables of coordination distances. This material is available free of charge via the Internet at https://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Professor Adrian J Mulholland, Email: Adrian J. Mulholland@bristol.ac.uk

Author Contributions

All authors designed the research. ZY conducted and analyzed the molecular simulations with guidance from RMT, CKC, SG, RS and AJM. ZY, JS and AJM wrote the manuscript, with all authors giving approval to the final version.

Funding Sources

We thank the China Scholarship Council and the U.K. Engineering and Physical Sciences Research Council (EPSRC) for funding studentships to ZY and RMT, respectively. This work was supported by the U.K. Biotechnology and Biological Sciences Research Council (BBSRC)-funded SouthWest Biosciences Doctoral Training Partnership (training grant reference BB/J014400/1). AJM thanks EPSRC and BBSRC for funding (grant numbers EP/M022609/1, EP/M013219/1, BB/M000354/1 and BB/L01386X/1). RS thanks RSC for Research Funds (Grants R19-3409 and R20-6912) and MCIN/AEI/10.13039/501100011033 (Grant PID2020-113147GA-I00). This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (NIH) to JS under Award Number R01AI100560. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

DATA AVAILABILITY STATEMENT

Underlying data are provided as Supporting Information accompanying this paper. The computational data of this work are available at the University of Bristol data repository, data.bris, at {DOI to be added}.

ACKNOWLEDGEMENTS

This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol - http://www.bris.ac.uk/acrc/. We thank Pengfei Li (Loyola University) for advice regarding MD parameters and Natalie Fey (School of Chemistry, University of Bristol) for her careful reading of the manuscript.

ABBREVIATIONS

DFT, density functional theory; MBL, metallo-beta-lactamase; MD, molecular dynamics; MM, molecular mechanics; MMTZ,

thiol-based mercaptomethyl thiazolidine; QM, quantum mechanics; TGA, thioglycolic acid.

REFERENCES

(1) Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. Zinc through the Three Domains of Life. *J. Proteome Res.* **2006**, *5* (11), 3173–3178. https://doi.org/10.1021/pr0603699.

(2) Rasia, R. M.; Vila, A. J. Exploring the Role and the Binding Affinity of a Second Zinc Equivalent in B. Cereus Metallo-β-Lactamase. *Biochemistry* **2002**, *41* (6), 1853–1860. https://doi.org/10.1021/bi010933n.

(3) Amaro, R. E.; Mulholland, A. J. Multiscale Methods in Drug Design Bridge Chemical and Biological Complexity in the Search for Cures. *Nat. Rev. Chem.* **2018**, *2* (4). https://doi.org/10.1038/s41570-018-0148.

(4) Huggins, D. J.; Biggin, P. C.; Dämgen, M. A.; Essex, J. W.; Harris, S. A.; Henchman, R. H.; Khalid, S.; Kuzmanic, A.; Laughton, C. A.; Michel, J.; Mulholland, A. J.; Rosta, E.; Sansom, M. S. P.; van der Kamp, M. W. Biomolecular Simulations: From Dynamics and Mechanisms to Computational Assays of Biological Activity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2019**, *9* (3), 1–23. https://doi.org/10.1002/wcms.1393.

(5) Dudev, T.; Lin, Y. lin; Dudev, M.; Lim, C. First-Second Shell Interactions in Metal Binding Sites in Proteins: A PDB Survey and DFT/CDM Calculations. *J. Am. Chem. Soc.* **2003**, *125* (10), 3168–3180. https://doi.org/10.1021/ja0209722.

(6) Zhang, J.; Yang, W.; Piquemal, J. P.; Ren, P. Modeling Structural Coordination and Ligand Binding in Zinc Proteins with a Polarizable Potential. *J. Chem. Theory Comput.* **2012**, *8* (4), 1314–1324. https://doi.org/10.1021/ct200812y.

(7) Dudev, T.; Lim, C. Tetrahedral vs Octahedral Zinc Complexes with Ligands of Biological Interest: A DFT/CDM Study. *J. Am. Chem. Soc.* **2000**, *122* (45), 11146–11153. https://doi.org/10.1021/ja0010296.

(8) Laitaoja, M.; Valjakka, J.; Jänis, J. Zinc Coordination Spheres in Protein Structures. *Inorg. Chem.* **2013**, *52* (19), 10983–10991. https://doi.org/10.1021/ic401072d.

(9) Dudev, T.; Lim, C. Principles Governing Mg, Ca, and Zn Binding and Selectivity in Proteins. *Chem. Rev.* **2003**, *103* (3), 773–787. https://doi.org/10.1021/cr020467n.

(10) Dudev, T.; Lim, C. Competition among Metal Ions for Protein Binding Sites: Determinants of Metal Ion Selectivity in Proteins. *Chem. Rev.* **2014**, *114* (1), 538–556. https://doi.org/10.1021/cr4004665.

(11) Daniel, A. G.; Farrell, N. P. The Dynamics of Zinc Sites in Proteins: Electronic Basis for Coordination Sphere Expansion at Structural Sites. *Metallomics* **2014**, *6* (12), 2230–2241. https://doi.org/10.1039/c4mt00213j.

(12) Gervasoni, S.; Spencer, J.; Hinchliffe, P.; Pedretti, A.; Vairoletti, F.; Mahler, G.; Mulholland, A. J. A Multiscale Approach to Predict the Binding Mode of Metallo Beta-Lactamase Inhibitors. *Proteins Struct. Funct. Bioinforma.* **2021**. https://doi.org/10.1002/prot.26227.

(13) Case DA, I. Y. B-S, Brozell SR, Cerutti DS, Cheatham TE III, Cruzeiro WD V., Darden TA,; Duke RE, Gilson MK, Gohlke H, Goetz AW, Greene D, Harris R, Homeyer N, Huang Y,; Izadi S, A. K., Kurtzman T, Lee TS, LeGrand S, Li P, Lin C, Liu J, Luchko T, Luo R, M, Y. Dm. K. P. Amber 2018. University of California, San Francisco: 2018.

(14) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, *26* (16), 1668–1688. https://doi.org/10.1002/jcc.20290.

(15) D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, G. Giambasu, T. Giese, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, D. M. Y. and P. A. K. AMBER 2019. University of California, San Francisco:2019.

(16) D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R Harris, S. Izadi, K. Kasava- jhala, A. Kovalenko, R. Krasny, T. Kurtzman, T., D. M. Y. and P. A. K. AMBER 2020. University of California, San Francisco: 2020.

(17) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. Structural Survey of Zinc-Containing Proteins and Development of the Zinc AMBER Force Field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6* (9), 2935–2947. https://doi.org/10.1021/ct1002626.

(18) 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**. https://doi.org/10.1021/ct400146w.

(19) Pang, Y. P. Successful Molecular Dynamics Simulation of Two Zinc Complexes Bridged by a Hydroxide in Phosphotriesterase Using the Cationic Dummy Atom Method. *Proteins Struct. Funct. Genet.* **2001**, *45* (3), 183–189. https://doi.org/10.1002/prot.1138.

(20) Li, P.; Merz, K. M. Metal Ion Modeling Using Classical Mechanics. *Chem. Rev.* **2017**, *117* (3), 1564–1686. https://doi.org/10.1021/acs.chemrev.6b00440.

(21) Li, P.; Merz, K. M. Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**. https://doi.org/10.1021/ct400751u.

(22) Li, P.; Song, L. F.; Merz, K. M. Systematic Parameterization of Monovalent Ions Employing the Nonbonded Model. *J. Chem. Theory Comput.* **2015**. https://doi.org/10.1021/ct500918t.

(23) Li, P.; Song, L. F.; Merz, K. M. Parameterization of Highly Charged Metal Ions Using the 12-6-4 LJType Nonbonded Model in Explicit Water. *J. Phys. Chem. B* **2015**, *119*, 29. https://doi.org/10.1021/jp505875v.

(24) Marek, M.; Shaik, T. B.; Heimburg, T.; Chakrabarti, A.; Lancelot, J.; Ramos-Morales, E.; Da Veiga, C.; Kalinin, D.; Melesina, J.; Robaa, D.; Schmidtkunz, K.; Suzuki, T.; Holl, R.; Ennifar, E.; Pierce, R. J.; Jung, M.; Sippl, W.; Romier, C. Characterization of Histone Deacetylase 8 (HDAC8) Selective Inhibition Reveals Specific Active Site Structural and Functional Determinants. *J. Med. Chem.* **2018**, *61* (22), 10000–10016. https://doi.org/10.1021/acs.jmedchem.8b01087.

(25) 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* (3), 1913–1923. https://doi.org/10.1021/acs.jctc.9b01180.

(26) Panteva, M. T.; Giambaşu, G. M.; York, D. M. Comparison of Structural, Thermodynamic, Kinetic and Mass Transport Properties of Mg2+ Ion Models Commonly Used in Biomolecular Simulations. *J. Comput. Chem.* **2015**, *36* (13), 970–982. https://doi.org/10.1002/jcc.23881.

(27) Chen, H.; Giese, T. J.; Golden, B. L.; York, D. M. Divalent Metal Ion Activation of a Guanine General Base in the Hammerhead Ribozyme: Insights from Molecular Simulations. *Biochemistry* **2017**, *56* (24), 2985–2994. https://doi.org/10.1021/acs.biochem.6b01192.

(28) Pu, Z.; Zhao, M.; Zhang, Y.; Sun, W.; Bao, Y. Dynamic Description of the Catalytic Cycle of Malate Enzyme: Stereoselective Recognition of Substrate, Chemical Reaction, and Ligand Release. *J. Phys. Chem. B* **2018**, *122* (51), 12241–12250. https://doi.org/10.1021/acs.jpcb.8b05135.

(29) Xu, M.; Zhu, T.; Zhang, J. Z. H. Molecular Dynamics Simulation of Zinc Ion in Water with an Ab Initio Based Neural Network Potential. *J. Phys. Chem. A* **2019**, *123* (30), 6587–6595. https://doi.org/10.1021/acs.jpca.9b04087.

(30) Xu, M.; He, X.; Zhu, T.; Zhang, J. Z. H. A Fragment Quantum Mechanical Method for Metalloproteins. *J. Chem. Theory Comput.* **2019**, No. Mm. https://doi.org/10.1021/acs.jctc.8b00966.

(31) Duarte, F.; Amrein, A.; Caroline, S.; Kamerlin, L. Modeling Catalytic Promiscuity in the Alkaline Phosphatase Superfamily. *Phys. Chem. Chem. Phys* **2013**, *15*, 11160. https://doi.org/10.1039/c3cp51179k.

(32) Xu, D.; Cui, Q.; Guo, H. Quantum Mechanical/Molecular Mechanical Studies of Zinc Hydrolases. *Int. Rev. Phys. Chem.* **2014**, *33* (1), 1–41.

https://doi.org/10.1080/0144235X.2014.889378.

(33) Ahmadi, S.; Barrios Herrera, L.; Chehelamirani, M.; Hostaš, J.; Jalife, S.; Salahub, D. R. Multiscale Modeling of Enzymes: QM-Cluster, QM/MM, and QM/MM/MD: A Tutorial Review. *Int. J. Quantum Chem.* **2018**, *118* (9), 1–34. https://doi.org/10.1002/qua.25558.

(34) Kulik, H. J.; Zhang, J.; Klinman, J. P.; Martínez, T. J. How Large Should the QM Region Be in QM/MM Calculations? The Case of Catechol O-Methyltransferase. *J. Phys. Chem. B* **2016**, *120* (44), 11381–11394. https://doi.org/10.1021/acs.jpcb.6b07814.

(35) Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116* (9), 5301–5337. https://doi.org/10.1021/acs.chemrev.5b00584.

(36) Wu, S.; Xu, D.; Guo, H. QM/MM Studies of Monozinc β-Lactamase CphA Suggest That the Crystal Structure of an Enzyme-Intermediate Complex Represents a Minor Pathway. *J. Am. Chem. Soc.* **2010**, *132* (51), 17986–17988. https://doi.org/10.1021/ja104241g.

(37) Wang, X.; Wu, S.; Xu, D.; Xie, D.; Guo, H. Inhibitor and Substrate Binding by Angiotensin-Converting Enzyme: Quantum Mechanical/Molecular Mechanical Molecular Dynamics Studies. *J. Chem. Inf. Model.* **2011**, *51* (5), 1074–1082. https://doi.org/10.1021/ci200083f.

(38) Wang, C.; Guo, H. Inhibitor Binding by Metallo-β-Lactamase IMP-1 from Pseudomonas Aeruginosa: Quantum Mechanical/Molecular Mechanical Simulations. *J. Phys. Chem. B* **2007**, *111* (33), 9986–9992. https://doi.org/10.1021/jp073864g.

(39) Wang, Y. T.; Cheng, T. L. Refined Models of New Delhi Metallo-Beta-Lactamase-1 with Inhibitors: An QM/MM Modeling Study. *J. Biomol. Struct. Dyn.* **2016**, *34* (10), 2214–2223. https://doi.org/10.1080/07391102.2015.1110834.

(40) Xu, D.; Guo, H.; Cui, Q. Antibiotic Binding to Dizinc β-Lactamase L1 from Stenotrophomonas Maltophilia: SCC-DFTB/CHARMM and DFT Studies. *J. Phys. Chem. A* **2007**, *111* (26), 5630–5636. https://doi.org/10.1021/jp068746s.

(41) Hou, G.; Cui, Q. QM/MM Analysis Suggests That Alkaline Phosphatase (AP) and Nucleotide Pyrophosphatase/Phosphodiesterase Slightly Tighten the Transition State for Phosphate Diester Hydrolysis Relative to Solution: Implication for Catalytic Promiscuity in the AP Superfamil. *J. Am. Chem. Soc.* **2012**, *134* (1), 229–246. https://doi.org/10.1021/ja205226d.

(42) Hou, G.; Cui, Q. Stabilization of Different Types of Transition States in a Single Enzyme Active Site: QM/MM Analysis of Enzymes in the Alkaline Phosphatase Superfamily. *J. Am. Chem. Soc.* **2013**, *135* (28), 10457–10469. https://doi.org/10.1021/ja403293d.

(43) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. General Performance of Density Functionals. *J. Phys. Chem. A* **2007**, *111* (42), 10439–10452. https://doi.org/10.1021/jp0734474.

(44) Douglas-Gallardo, O. A.; Shepherd, I.; Bennie, S. J.; Ranaghan, K. E.; Mulholland, A. J.; Vöhringer-Martinez, E. Electronic Structure Benchmark Calculations of CO2 Fixing Elementary Chemical Steps in RuBisCO Using the Projector-Based Embedding Approach. *J. Comput. Chem.* **2020**, *41* (24), 2151–2157. https://doi.org/10.1002/jcc.26380.

(45) Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H. J. High-Accuracy Computation of Reaction Barriers in Enzymes. *Angew. Chemie - Int. Ed.* **2006**, *45* (41), 6856–6859. https://doi.org/10.1002/anie.200602711.

(46) Zheng, M.; Xu, D. New Delhi Metallo-β-Lactamase I: Substrate Binding and Catalytic Mechanism. *J. Phys. Chem. B* **2013**, *117* (39), 11596–11607. https://doi.org/10.1021/jp4065906.

(47) Wang, D.; Helquist, P.; Wiest, O. Zinc Binding in HDAC Inhibitors: A DFT Study. *J. Org. Chem.* **2007**, *72* (14), 5446–5449. https://doi.org/10.1021/jo070739s.

(48) Hinchliffe, P.; Yang, Q. E.; Portal, E.; Young, T.; Li, H.; Tooke, C. L.; Carvalho, M. J.; Paterson, N. G.; Brem, J.; Niumsup, P. R.; Tansawai, U.; Lei, L.; Li, M.; Shen, Z.; Wang, Y.; Schofield, C. J.; Mulholland, A. J.; Shen, J.; Fey, N.; Walsh, T. R.; Spencer, J. Insights into the Mechanistic Basis of Plasmid-Mediated Colistin Resistance from Crystal Structures of the Catalytic Domain of MCR-1. *Sci. Rep.* **2017**, *7* (January), 1–10. https://doi.org/10.1038/srep39392.

(49) Silvia Gervasoni, James Spencer, Philip Hinchliffe, Alessandro Pedretti, Franco Vairoletti, Graciela Mahler, A. J. M. A Multiscale Approach to Predict the Binding Mode of Metallo Beta-Lactamase Inhibitors. *Proteins* **2021**. https://doi.org/10.1002/prot.26227.

(50) Rossi, M. A.; Martinez, V.; Hinchliffe, P.; Mojica, M. F.; Castillo, V.; Moreno, D. M.; Smith, R.; Spellberg, B.; Drusano, G. L.; Banchio, C.; Bonomo, R. A.; Spencer, J.; Vila, A. J.; Mahler, G. 2-Mercaptomethyl-Thiazolidines Use Conserved Aromatic-S Interactions to Achieve Broad-Range Inhibition of Metallo-β-Lactamases. *Chem. Sci.* **2021**, *12* (8), 2898–2908. https://doi.org/10.1039/d0sc05172a.

(51) Hinchliffe, P.; Moreno, D. M.; Rossi, M.-A.; Mojica, M. F.; Martinez, V.; Villamil, V.; Spellberg, B.; Drusano, G. L.; Banchio, C.; Mahler, G.; Bonomo, R. A.; Vila, A. J.; Spencer, J. 2-Mercaptomethyl Thiazolidines (MMTZs) Inhibit All Metallo-β-Lactamase Classes by Maintaining a Conserved Binding Mode. *ACS Infect. Dis.* **2021**. https://doi.org/10.1021/acsinfecdis.1c00194.

(52) Akif, M.; Georgiadis, D.; Mahajan, A.; Dive, V.; Sturrock, E. D.; Isaac, R. E.; Acharya, K. R. High-Resolution Crystal Structures of Drosophila Melanogaster Angiotensin-Converting Enzyme in Complex with Novel Inhibitors and Antihypertensive Drugs. *J. Mol. Biol.* **2010**, *400* (3), 502–517. https://doi.org/10.1016/j.jmb.2010.05.024.

(53) Towler, P.; Staker, B.; Prasad, S. G.; Menon, S.; Tang, J.; Parsons, T.; Ryan, D.; Fisher, M.; Williams, D.; Dales, N. A.; Patane, M. A.; Pantoliano, M. W. ACE2 X-Ray Structures Reveal a Large Hinge-Bending Motion Important for Inhibitor Binding and Catalysis. *J. Biol. Chem.* **2004**, *279* (17), 17996–18007. https://doi.org/10.1074/jbc.M311191200.

(54) Lauffer, B. E. L.; Mintzer, R.; Fong, R.; Mukund, S.; Tam, C.; Zilberleyb, I.; Flicke, B.; Ritscher, A.; Fedorowicz, G.; Vallero, R.; Ortwine, D. F.; Gunzner, J.; Modrusan, Z.; Neumann, L.; Koth, C. M.; Kaminker, J. S.; Heise, C. E.; Steiner, P. Histone Deacetylase (HDAC) Inhibitor Kinetic Rate Constants Correlate with Cellular Histone Acetylation but Not Transcription and Cell Viability. *J. Biol. Chem.* **2013**, *288* (37), 26926–26943. https://doi.org/10.1074/jbc.M113.490706.

(55) Jun-ichi Wachino, a Reo Kanechi, a Erina Nishino, a Marie Mochizuki, a Wanchun Jin, a Kouji Kimura, a Hiromasa Kurosaki, B.; A, Y. A. 4-Amino-2-Sulfanylbenzoic Acid as a Potent Subclass B3 Metallo-β-Lactamase-Specific Inhibitor Applicable for Distinguishing Metallo-β-Lactamase Subclasses. **2019**, No. June, 1–7.

(56) Lucic, A.; Hinchliffe, P.; Malla, T. R.; Tooke, C. L.; Brem, J.; Calvopiña, K.; Lohans, C. T.; Rabe, P.; McDonough, M. A.; Armistead, T.; Orville, A. M.; Spencer, J.; Schofield, C. J. Faropenem Reacts with Serine and Metallo-β-Lactamases to Give Multiple Products. *Eur. J. Med. Chem.* **2021**, *215*. https://doi.org/10.1016/j.ejmech.2021.113257.

(57) Saavedra, M. J.; Peixe, L.; Sousa, J. C.; Henriques, I.; Alves, A.; Correia, A. Sfh-I, a Subclass B2 Metallo-β-Lactamase from a Serratia Fonticola Environmental Isolate. *Antimicrob. Agents Chemother.* **2003**, *47* (7), 2330–2333. https://doi.org/10.1128/AAC.47.7.2330-2333.2003.

(58) Lu, X.; Gaus, M.; Elstner, M.; Cui, Q. Parametrization of DFTB3/3OB for Magnesium and Zinc for Chemical and Biological Applications. *J. Phys. Chem. B* **2015**, *119* (3), 1062–1082. https://doi.org/10.1021/jp506557r.

(59) 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*, 50.

https://doi.org/10.1021/acs.jctc.5b00255.

(60) Lythell, E.; Suardíaz, R.; Hinchliffe, P.; Hanpaibool, C.; Visitsatthawong, S.; Oliveira, A. S. F.; Lang, E. J. M.; Surawatanawong, P.; Lee, V. S.; Rungrotmongkol, T.; Fey, N.; Spencer, J.; Mulholland, A. J. Resistance to the "Last Resort" Antibiotic Colistin: A Single-Zinc Mechanism for Phosphointermediate Formation in MCR Enzymes. *Chem. Commun.* **2020**, *56* (50), 6874–6877. https://doi.org/10.1039/d0cc02520h.

(61) Lonsdale, R.; Harvey, J. N.; Mulholland, A. J. Effects of Dispersion in Density Functional Based Quantum Mechanical/Molecular Mechanical Calculations on Cytochrome P450 Catalyzed Reactions. *J. Chem. Theory Comput.* **2012**, *8* (11), 4637–4645. https://doi.org/10.1021/ct300329h.

(62) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132* (15). https://doi.org/10.1063/1.3382344.

(63) Stefan Grimme, S. E. and L. G. Effect of the Damping Function in Dispersion Corrected Density Functional Theory. *J. Comput. Chem.* **2011**, *32*, 1456–1465. https://doi.org/10.1002/jcc.21759.

(64) Zheng, H.; Cooper, D. R.; Porebski, P. J.; Shabalin, I. G.; Handing, K. B.; Minor, W. CheckMyMetal: A Macromolecular Metal-Binding Validation Tool. *Acta Crystallogr. Sect. D Struct. Biol.* **2017**, *73*, 223–233. https://doi.org/10.1107/S2059798317001061.

(65) Zheng, H.; Chordia, M. D.; Cooper, D. R.; Chruszcz, M.; Müller, P.; Sheldrick, G. M.; Minor, W. Validation of Metal-Binding Sites in Macromolecular Structures with the CheckMyMetal Web Server. *Nat. Protoc.* **2014**, *9* (1), 156–170. https://doi.org/10.1038/nprot.2013.172.

(66) Wilcox, D. E. Binuclear Metallohydrolases. *Chem. Rev.* **1996**, *96* (7), 2435–2458. https://doi.org/10.1021/cr950043b.

(67) Schenk, G.; Mitić, N.; Gahan, L. R.; Ollis, D. L.; McGeary, R. P.; Guddat, L. W. Binuclear Metallohydrolases: Complex Mechanistic Strategies for a Simple Chemical Reaction. *Acc. Chem. Res.* **2012**, *45* (9), 1593–1603. https://doi.org/10.1021/ar300067g.

(68) Tooke, C. L.; Hinchliffe, P.; Bragginton, E. C.; Colenso, C. K.; Hirvonen, V. H. A.; Takebayashi, Y.; Spencer, J. β-Lactamases and β-Lactamase Inhibitors in the 21st Century. *J. Mol. Biol.* **2019**, *431* (18), 3472–3500. https://doi.org/https://doi.org/10.1016/j.jmb.2019.04.002.

(69) E., B. S.; M., L. D.; C., H. D.; W., H. W. Metallo-β-Lactamases: Structure, Function, Epidemiology, Treatment Options, and the Development Pipeline. *Antimicrob. Agents Chemother.* **2021**, *64* (10), e00397-20. https://doi.org/10.1128/AAC.00397-20.

(70) Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T.N.Bhat, Helge Weissig, I. N. S. and P. E. B. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.

(71) Burley, S. K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G. V.; Christie, C. H.; Dalenberg, K.; Di Costanzo, L.; Duarte, J. M.; Dutta, S.; Feng, Z.; Ganesan, S.; Goodsell, D. S.; Ghosh, S.; Green, R. K.; Guranovic, V.; Guzenko, D.; Hudson, B. P.; Lawson, C. L.; Liang, Y.; Lowe, R.; Namkoong, H.; Peisach, E.; Persikova, I.; Randle, C.; Rose, A.; Rose, Y.; Sali, A.; Segura, J.; Sekharan, M.; Shao, C.; Tao, Y. P.; Voigt, M.; Westbrook, J. D.; Young, J. Y.; Zardecki, C.; Zhuravleva, M. RCSB Protein Data Bank: Powerful New Tools for Exploring 3D Structures of Biological Macromolecules for Basic and Applied Research and Education in Fundamental Biology, Biomedicine, Biotechnology, Bioengineering and Energy Sciences. *Nucleic Acids Res.* **2021**, *49* (1), D437–D451. https://doi.org/10.1093/nar/gkaa1038.

(72) Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of p K a Values. *J. Chem. Theory Comput.* **2011**, *7* (7), 2284–2295. https://doi.org/10.1021/ct200133y.

(73) Olsson, M. H. M.; SØndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p K a Predictions. *J. Chem. Theory Comput.* **2011**, *7* (2), 525–537. https://doi.org/10.1021/ct100578z.

(74) Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32* (WEB SERVER ISS.), 665–667. https://doi.org/10.1093/nar/gkh381.

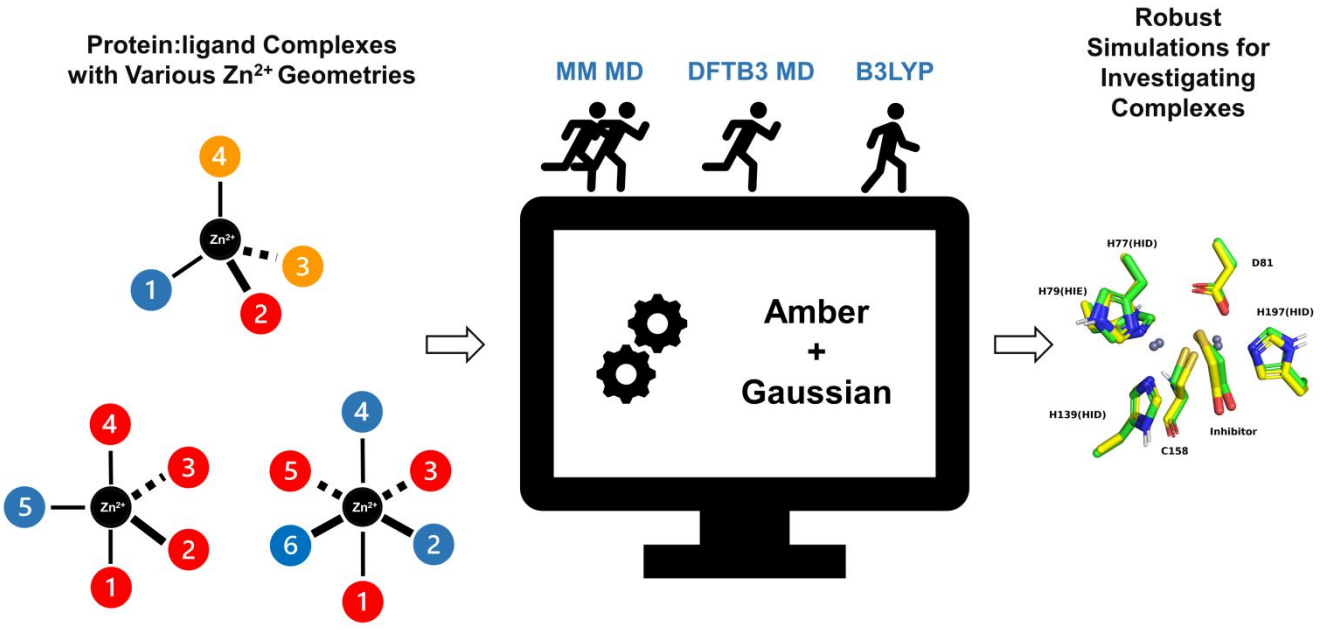
(75) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174. https://doi.org/10.1002/jcc.20035.

(76) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(77) Schrodinger LLC. The PyMOL Molecular Graphics System, Version 2.3. 2019.

(78) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian 16, Revision A.03. 2016.

SYNOPSIS TOC



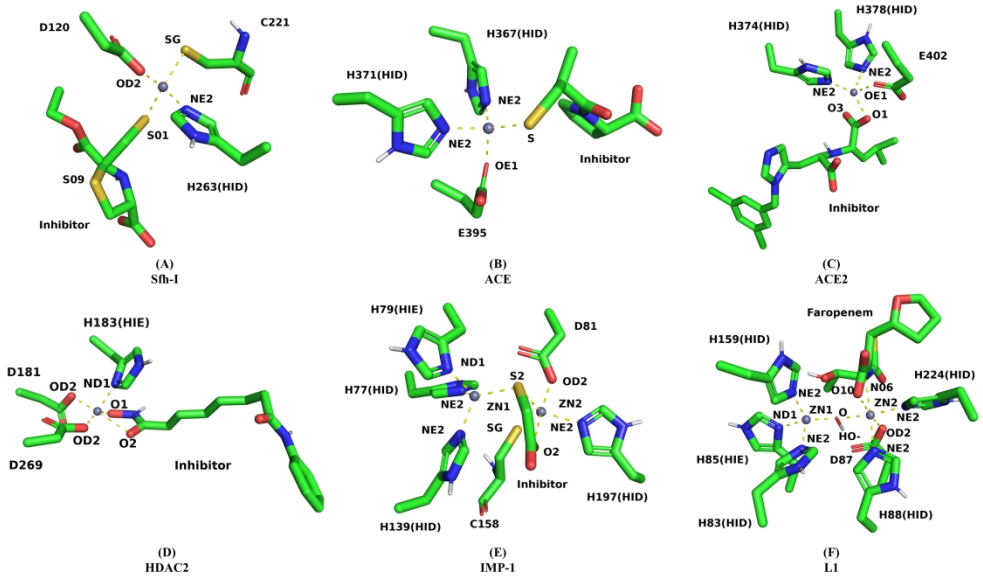


Figure 1. Zinc site geometries of the 6 protein-ligand systems included in this study. (A) Sfh-I MBL with MMTZ inhibitor L-anti-1a, co-ordination number (CN) = 4, PDB code: 7BJ9;[51] (B) ACE with the thiol inhibitor L-captopril, CN=4, PDB code: 2X8Z;[52] (C) ACE2 with the carboxylate inhibitor MLN-4760, CN=4, PDB code:1R4L;[53] (D) HDAC2 with the hydroxamate inhibitor SAHA, CN=5, PDB code : 4LXZ;[54] (E) IMP-1 MBL with thioglycolic acid (TGA), CN = 4,5, PDB code: 6JED;[55] (F) L1 MBL with hydrolyzed faropenem, CN=4,6, PDB code:7A63.[56] Carbon atoms are in green, nitrogen atoms are in blue, oxygen atoms are in red and sulfur atoms are in yellow, with zinc ions represented as gray balls

330x190mm (300 x 300 DPI)

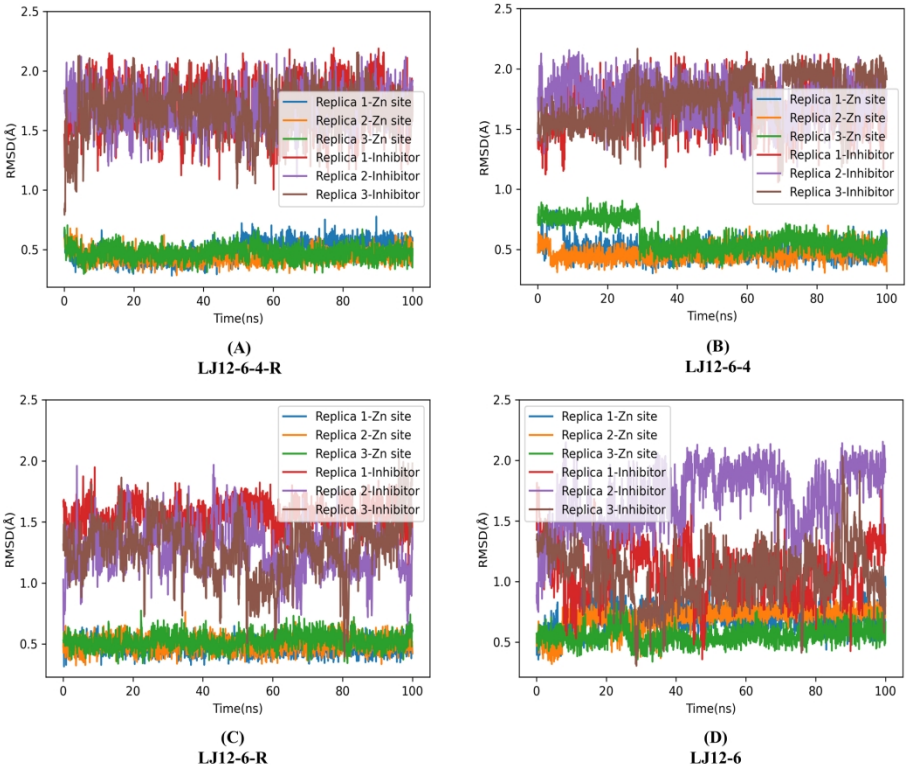


Figure 2.Time-dependence of RMSD values (compared to the crystal structure) for MM MD simulations of Sfh-I:L-anti-1a complex using non-bonded models. Three replicate simulations were performed for each model and each replicate is 100 ns. The 'Zn site' refers to zinc ions and zinc coordinating residues and 'Inhibitor' refers to the zinc bound compound

206x167mm (300 x 300 DPI)

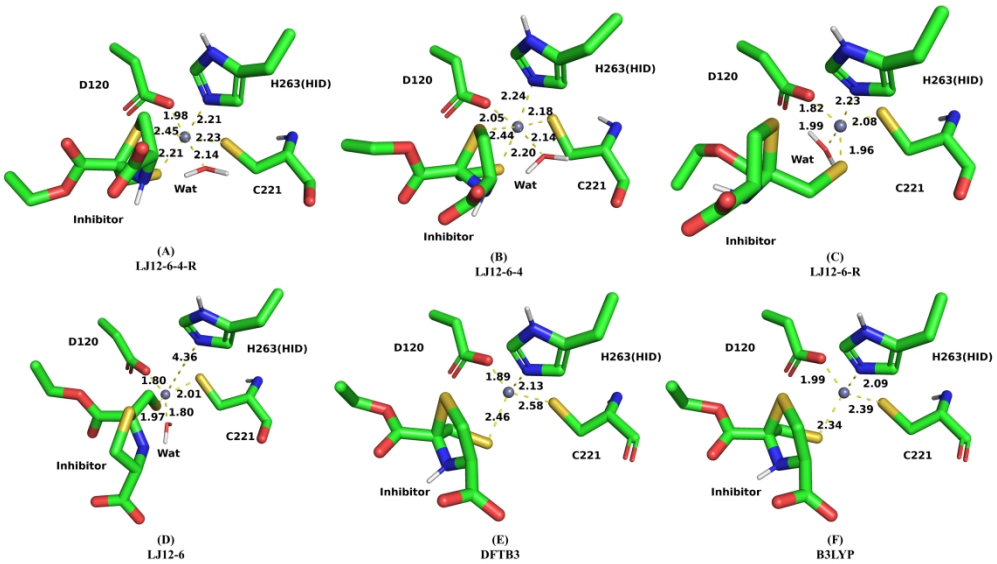


Figure 3. Zinc site geometries of Sfh-I:L-anti-1a complexes. (A-D) The representative zinc site geometry of the four non-bonded models after 100ns MM MD simulation. (E) The representative zinc geometry after 100 ps DFTB3/MM MD simulation. The extra coor-dination by an additional water molecule was removed. (F) The representative zinc geometry after B3LYP-D3BJ based QM/MM ge-ometry optimization

338x190mm (300 x 300 DPI)

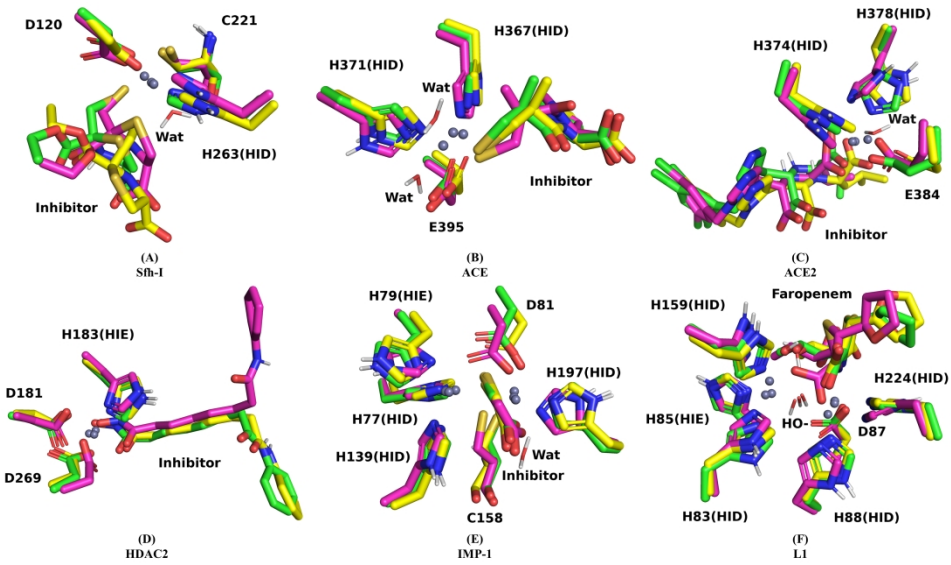


Figure 4. Zinc binding sites of the 6 protein-ligand systems included in the study after simulations at different levels of theory. Carbon atoms are in green/yellow/purple depending on the model: yellow represents the crystal structure; purple represents a typical structure after 100 ns MM simulation with LJ12-6-4 models; green represents a representative structure after DFTB3/MM MD and B3LYP-D3BJ based QM/MM geometry optimization. Nitro atoms are in blue, oxygen atoms are in red and sulfur atoms are in dark yellow, zinc ions are represented as gray balls.

338x190mm (300 x 300 DPI)

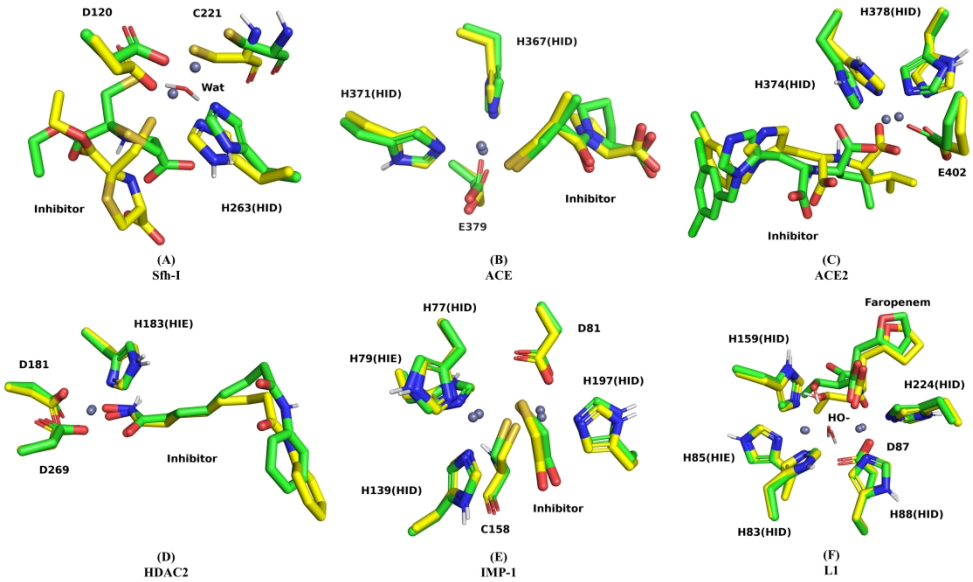


Figure 5. Zinc binding sites of the 6 protein-ligand systems included in the study after DFTB3/MM MD starting directly from the crystal structure. Carbon atoms are in green and yellow depending on the model: green represents a representative zinc site structure after 2ns DFTB3/MM MD starting directly from the crystal structure, yellow represents the crystal structure. Nitro at-oms are in blue, oxygen atoms are in red and sulfur atoms are in dark yellow, zinc ions are represented as gray balls.

324x190mm (300 x 300 DPI)