JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

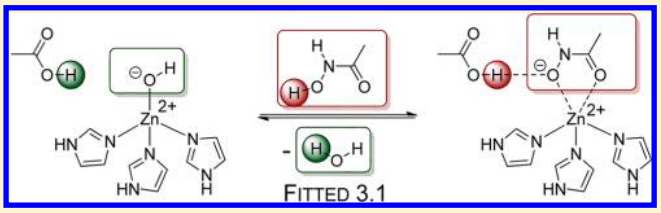
Docking Ligands into Flexible and Solvated Macromolecules. 6. Development and Application to the Docking of HDACs and other Zinc Metalloenzymes Inhibitors

Joshua Pottel, Eric Therrien, James L. Gleason, and Nicolas Moitessier*

Department of Chemistry, McGill University, 801 Sherbrooke St W, Montreal, QC, Canada H3A 0B8

Supporting Information

ABSTRACT: Metalloenzymes are ubiquitous proteins which feature one or more metal ions either directly involved in the enzymatic activity and/or structural properties (i.e., zinc fingers). Several members of this class take advantage of the Lewis acidic properties of zinc ions to carry out their various catalytic transformations including isomerization or amide cleavage. These enzymes have been validated as drug targets for a number of diseases including cancer; however, despite their pharmaceutical relevance and the availability of crystal structures, structure-based drug design methods have been poorly and indirectly parametrized for these classes of enzymes. More specifically, the metal coordination component and proton transfers of the process of drugs binding to metalloenzymes have been inadequately modeled by current docking programs, if at all. In addition, several known issues, such as coordination geometry, atomic charge variability, and a potential proton transfer from small molecules to a neighboring basic residue, have often been ignored. We report herein the development of specific functions and parameters to account for zinc−drug coordination focusing on the above-listed phenomena and their impact on docking to zinc metalloenzymes. These atom-type-dependent but atomic charge-independent functions implemented into FITTED 3.1 enable the simulation of drug binding to metalloenzymes, considering an acid−base reaction with a neighboring residue when necessary with good accuracy.



INTRODUCTION

Over the years, several zinc metalloenzymes have been validated as drug targets. Among these metalloproteins are the matrix metalloproteinases (MMPs) which represent a family of zinc endopeptidases including stromelysin-1 (MMP-3), gelatinases (e.g., MMP-2), and collagenases (e.g., MMP-1 and MMP-9), the carbonic anhydrases, the mannosidases (e.g., α-mannosidase), β-lactamase, phospholipase C, alcohol dihydrogenase, and the histone deacetylases (e.g., HDAC-8). As with several other enzymes, docking methods have been tested as tools for inhibitor design and discovery. Our interests in MMP[1−3] and α-mannosidase[4] inhibitors led us to further evaluate docking programs for the design of these classes of inhibitors.

Over the past few decades, docking methods have evolved from simple rigid body assembling tools (rigid body docking) to methods modeling flexible ligand/flexible protein complexes.[5] Our efforts in the field led to the development of a docking program, FITTED (flexibility induced through targeted evolutionary description), which accounts for ligand and protein flexibility as well as for the presence of displaceable water molecules.[6,7] In order to cover a large range of drug classes, further implementations provided FITTED with the ability to model the binding of covalent drugs.[8] In 2007, the first version of FITTED was tested for its ability to dock mannosidase inhibitors which revealed the intricacies of modeling zinc coordination.[4] Later on, we also demonstrated that scoring metalloenzyme inhibitors was poor.[9] Throughout the years, docking programs have been assessed for their ability to dock inhibitors coordinating metal ions (often zinc) of metalloenzymes: AutoDock, DOCK, GOLD, and FlexX being widely used.[10−15] Within most studies, the metal coordination has been modeled through simple electrostatic and van der Waals interactions and no treatment of the covalent nature of metal coordination was considered. In addition, the zinc ion atomic charge has often been assumed to be +2 while more advanced molecular dynamics simulations have often required finely tuned charging schemes.[10] Nondirectional nonbond interactions cannot properly model the geometrical constrains of metal coordination. In order to correct for the coordination geometry, some docking programs including GOLD,[16] PLANTS,[17,18] FlexX,[19] and FITTED are using coordination sites to guide the docking of ligands. These coordination sites are positioned on polyhedrons defining the preferred geometries of protein-bound metal ions (e.g., hexacoordinate in octahedral geometry). Although these approximations provided reasonable overall binding modes and even good enrichments in some studies,[20] scoring the metal coordination is expected to be poorly predictive and more advanced scoring is required, in particular scoring the displacement of water molecules coordinating the zinc ion in the unbound state might be

Received: September 24, 2013
Published: December 23, 2013

critical in order to identify poor zinc binding groups.[12] In an attempt to account for the covalent nature of zinc coordination, hydrogen bond-like terms have also been evaluated (e.g., GOLD[16]); however, this is merely circumventing the problem; a metal coordination is neither a covalent bond nor a hydrogen bond. Despite these limitations, docking-based virtual screening has been reasonably successful but can certainly be improved.[21]

Our recent interest in developing HDAC inhibitor hybrids[22−24] revived our curiosity in improving the ability to dock to metalloenzymes. Here we report our efforts to further develop our docking program (version: FITTED 3.1) in order to accurately model zinc coordination and more specifically to discover novel HDAC inhibitors.

THEORY AND CURRENT STATE

Metalloenzymes and Classical Molecular Mechanics. When simulating the metal−ligand coordination using classical molecular mechanics, two models have been proposed differing in the nature of the coordination bond: the bound model and the nonbound model. First, the bound model, in which the zinc−ligand bond is considered covalent, enables the coordination geometry to be considered with the *normal* bonded terms (bond, angles, torsions). In contrast, the nonbound model relies on nondirectional interactions (electrostatic, van der Waals) and is thus expected to provide less reliable geometries but eliminate the potentially overly strict constraints placed on a typical covalent interaction such as, for example, a carbon−carbon bond. Docking methods cannot easily rely on the bound model as, following the positioning and orientation of the ligand, the zinc binding group may or may not coordinate the zinc ion. If it is not coordinated, the bound energy must be ignored. In this context, covalent docking methods exist, in which both covalent and noncovalent poses can be considered and could be extended to metal coordination.[8] When either model is used, the charge transfer between the zinc ion and the coordinating groups varies as a function of the coordination geometry and of the zinc binding group nature; consequently, it remains difficult to simulate the binding process with the traditional point charges and thus, polarizable force fields have been envisioned.[25] The change in pK_a of the zinc binding group, a potential proton transfer to a neighboring basic residue and the presence of water are also to be considered for accurate docking of zinc binding molecules.

Docking to Metalloenzymes, Coordination Geometry, Proton and Charge Transfers, and Displacement of Water Molecules. Accurately docking small molecules to metalloenzymes requires consideration of the coordination geometry, charge transfer, change in pK_a, and displacement of coordinating water molecules. First, as zinc(II) has a saturated electronic configuration (d^{10}), electrostatic interaction is the major component, and the geometry is not as well-defined as with other transition metals. In fact, although zinc adopts an octahedral geometry in water through hexacoordination with water molecules, our survey of available crystal structures revealed that the coordination geometry in proteins ranges from tetrahedral to octahedral and may be highly distorted from ideality. As a result, since the coordination geometry is not stringent, simple terms such as electrostatic and van der Waals have been used in several simulation studies.[26]

Second, changes in pK_a and proton transferability significantly modify the zinc coordination energy. As an example, hydroxamic acid-based HDAC8 inhibitors bind to the zinc ion and make an additional two hydrogen bonds with neighboring residues Tyr307 and His142 (Figure 1). A close look at another twelve metalloenzymes revealed that a basic residue not

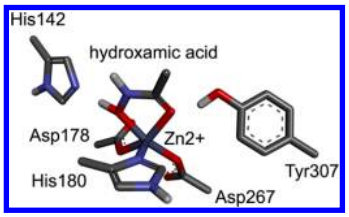


Figure 1. HDAC zinc binding site with a hydroxamic acid-containing ligand. Only the zinc binding group of the ligand is shown for clarity (pdb code: 1t67).

coordinated to the zinc ion (e.g., His 142 in HDAC8) and/or a hydrogen bond donating residue (e.g., Tyr307 in HDAC8) can be found in all of these cases. The prevalence of such residues in close proximity to the catalytic zinc may be explained by their role in the enzyme catalytic function. In fact, these residues participate in the catalytic activity by, for instance, converting $L_3Zn(OH_2)$ with L being histidine or glutamic acid to $L_3Zn(OH)$; hence converting the water molecule into a more reactive hydroxide ion (Figure 2). For example, the pK_a of a water molecule bound to the catalytic zinc ion of carbonic anhydrase is as low as 7 while it is 15.7 in bulk water.[27] These changes in pK_a sometimes result in a proton exchange (acid−base reaction). For the zinc coordination energy to be computed correctly, this property must be investigated and considered.

Some of the above-mentioned basic residues also participate in drug binding through interactions other than usual nonbonded interactions. Previous computational studies have shown that a proton transfer from the drug to the basic residue or change in protonation state of this residue by buffer proton uptake can occur (Figure 3). Approximately 10 years ago, Cross and co-workers reported a change in the hydroxamic acid pK_a of over 3 pK_a units when bound to TACE with a concomitant increase in the pK_a of the neighboring Glu406 of nearly 2 pK_a units.[28] These changes in acidity led the hydroxamic ligand to be more acidic than Glu406 and hence induced a proton shift. The protonation of Glu406 was also proposed when an acetate group was bound to zinc. Earlier this year, this protonation of neighboring Glu when an acetate is bound to zinc was demonstrated experimentally in MMP12,[29] while the proton transfer to a remote glutamic acid was proposed in carbonic anhydrases.[30] A similar proton transfer was initially proposed for HDAC8[31] but was recently found to be disfavored by nearly 4.0 kcal/mol (Figure 3c).[32] When docking ligands to metalloenzymes, docking programs should therefore include routines to transfer protons when required. To our knowledge, this has never been done. Another approach is to dock the drugs in their anionic form to the properly protonated binding site. Once more, docking of anions has been carried out (e.g., sulfonamidate ions to carbonic anhydrases) although the protonation state of the neighboring residues was not set to consider the proton transfer.[33,34] When more than one anion can be formed, all the tautomers have to be prepared and docked separately.[34] The approach was not desirable since this technique can be tedious and bias the result.

As shown in Figures 3 and 4, one water molecule (or hydroxide ion) is coordinating the catalytic zinc ions (distance lower than 2.5 Å) as observed in several crystal structures such

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

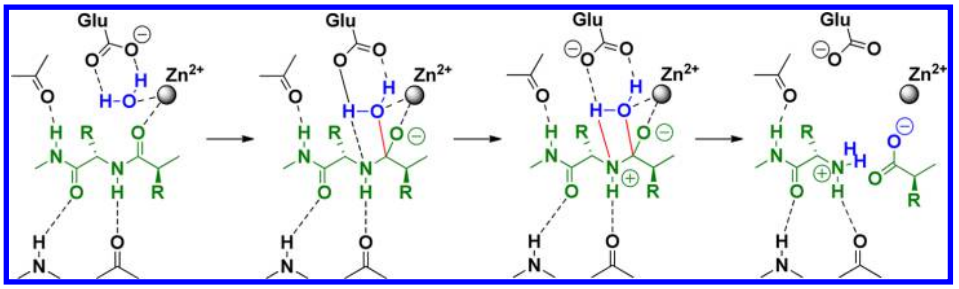


Figure 2. Catalytic process of matrix metalloproteinases (substrate in green, water in blue, and enzyme in black).

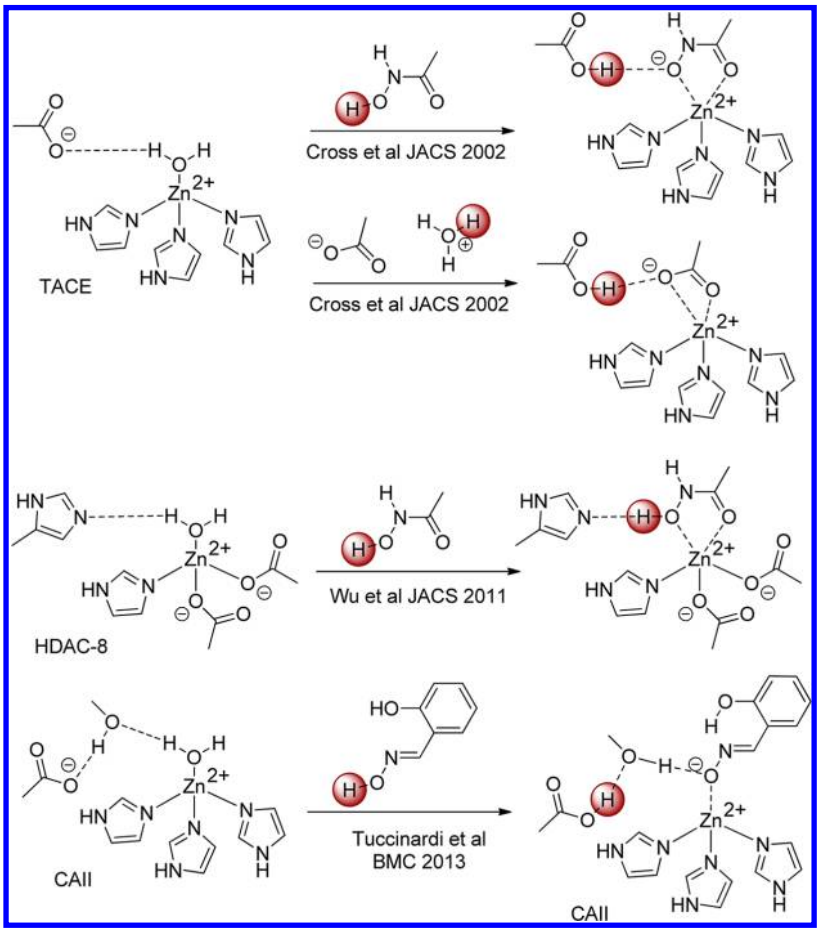


Figure 3. Zinc coordination and proton transfer.

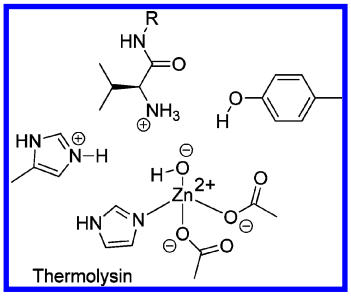


Figure 4. Ligand interacting with water molecule which coordinates to zinc (thermolysin, 8tln).

as MMPs (e.g., 1xuc) and thermolysin (e.g., 3tln) while a second one may appear at greater distances (greater than 3 Å) somewhat solvating the first one (e.g., ACE, 1j38). Even enzymes in which zinc coordinates two hydroxyl groups of carbohydrates, such as α-mannosidase, have only one water molecule coordinating zinc when unbound (α-mannosidase, 3bub). Displacement of this water molecule is necessary for ligand binding. In order to properly score the ligand−metal coordination energy, the water−metal energy should be subtracted. Thus, the water coordination energy should be also known.

Among the other major factors that docking programs should account for (implicitly or explicitly) is charge transfer. When zinc is coordinated to, for example, 3 histidine residues, its actual charge is not +2 but rather closer to +1 or lower. In addition, the atomic charge also significantly changes when an additional ligand (e.g., a drug) is bound to a free coordination site. For example, Merz and co-workers have developed force field parameters for zinc-containing enzymes in which the zinc atomic charge varies from 0.43 to 0.92 depending on both the ligand (such as water or hydroxide) and coordinated protein

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

residues. In addition, large variations of charge transfer were observed between a neutral (water, charge transfer of 0.17) and negatively charged ligands (hydroxide, 0.41).[35] Consequently, when docking a library of small molecules, a distinction should be made when neutral or charged molecules are considered. Classical molecular mechanics cannot account for these effects, as atomic charges are fixed unless specific charges are developed for each system. In order to account for this charge transfer, polarizable force fields can be used.[25] However their implementation in docking programs can be challenging due to the computational power and time needed to dynamically modify atomic properties.

IMPLEMENTATION

FITTED. Before the major developments reported herein, FITTED 3.0 was handling metal coordination in a very specific way. First, the metal coordination geometry was not considered when energy was computed, but FITTED guided the docking of zinc binding groups toward the free metal coordination sites. In a nutshell, FITTED relies on a hybrid matching algorithm/genetic algorithm to dock small molecules to proteins. More specifically, the matching algorithm implemented in the very early versions of FITTED relies on interaction sites identified within the binding site including the automatically defined free metal coordination sites. These use of interaction sites located at the free metal coordination sites implicitely account for coordination geometry and will still be used with the current implementations although with a number of additional functionalities.

DFT Studies and Testing Set. We first planned to develop a function that will evaluate the zinc coordination potential energy. This function will be independent of the currently used hydrogen bond, electrostatic or van der Waals terms. To do so, we turned our attention to quantum mechanical (QM) methods. In this area, density functional theory (DFT) studies have been reported although most include data computed solely with optimized structures and not with structures away from their ideal coordination in order to evaluate the coordination energy surface. In addition, in some of these studies, the role played by the basic neighboring residues was not considered.[36] An exhaustive survey of the PDB led us to collect 121 structures of metalloenzymes bound to various ligands and with most having a resolution lower than 2.5 Å. The ligands cover a number of zinc binding groups (e.g., *o*-amidoaniline, terminal sulfonamides, hydroxamic acids, thiols), and even include ligands bearing two potential zinc binding groups such as captopril cocrystallized with the angiotensin converting enzyme (pdb code: 2x8z, Figure 5) or a sulfanyl butanoic acid derivative (3i1u). These latter systems will enable us to test whether the proper zinc binding group can be identified by our docking method if a small molecule features more than one. Some of these ligands do not coordinate to zinc such as the hydrolysis product Val-Trp bound to thermolysin (pdb code: 3tmn, Figure 5). These structures were assembled by family (e.g., MMP1, MMP3, MMP8, α-mannosidase, thermolysin, HDAC8, see the Supporting Information for a complete set).

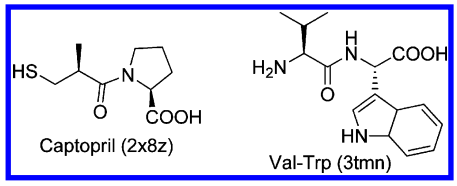


Figure 5. Selected ligands cocrystallized with metalloenzymes.

In order to investigate the energetics of the coordination, 19 representative structures out of the 121 were selected for DFT studies. These 19 systems were selected to cover as much of a diverse set of zinc coordination spheres (i.e., coordinating ligand and protein residues) as possible. The key residues and ligands were truncated (the main chain atoms were removed) leading to systems such as the one shown in Figure 1. Then hydrogen atoms were added and optimized using GAMESS-US (B3LYP/6-31G*). For the following studies, we kept all the protein atoms frozen as was previously reported.[37] We are aware that DFT functionals are not optimal for these metal coordination energies. Despite the large use of B3LYP in the area, recently reported work by Friedman and co-workers demonstrated that on significantly smaller and highly polar systems (e.g., H_3C-S- :::: Zn^{2+}), B3LYP overestimates the interaction energy by about 5% at the equilibrium distance and by even more at longer distance due to an overestimation of the polarization energy.[38] However as mentioned by Friedman and co-workers, the presence of four or five coordinating ligands in our truncated systems should reduce the polarization error hence the overall error of B3LYP. As the coordination number will not change whether the noncoordinating neighboring residue is removed or not, the error associated to this polarization effect will be kept to a minimum. In addition, the size of our systems (four or five coordinating ligands, 40−90 atoms) precludes the use of MP2 leaving DFT as an acceptable alternative.

Computing Zinc Coordination Energy. Scripts were developed to move the ligand away from the zinc atom in 0.20 Å increments, and single point energy was measured for each configuration. In order to uncouple the effect of the zinc coordination from the effect of neighboring residues, the same trajectory was alternatively computed with and without these residues; the bound state/unbound state energy difference must be attributed to the zinc coordination energy only if the residue is not present. As mentioned previously and shown in Figure 3, the neighboring residues may have significant effects on the binding affinities of zinc binding groups (ZBG). In the case of HDAC8-hydroxamic acid interactions, the overall potential energy drops by as much as 45 kcal/mol when His142 is kept in the truncated system (Figure 6). When this additional residue is ignored, the gain in energy of the hydroxamic acid coordination is reduced to about 20 kcal/mol for a difference of approximately 25 kcal/mol. In practice this histidine residue participates to a strong hydrogen bond with the hydroxamic acid increasing the polarity of the O−H bond; hence the coordination energy of the oxygen to zinc. In turn, this coordination increases the acidity of this hydrogen hence the strength of the hydrogen bond. This demonstrates that the two effects (hydrogen bond and zinc coordination) are acting in concert.

Computing Proton Transfer to a Neighboring Residue. In addition to this zinc coordination, cases where proton transfer is expected were evaluated. Hydroxamic acids are neutral at physiological pH. However, as mentioned above, when approaching the coordination sphere of the zinc ions, the pK_a decreases and the proton may transfer to a neighboring (not zinc coordinated) glutamate or histidine residue or even to a remote glutamic acid as observed in carbonic anhydrase.[30] To probe this effect, the potential energy of hydroxamate and

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

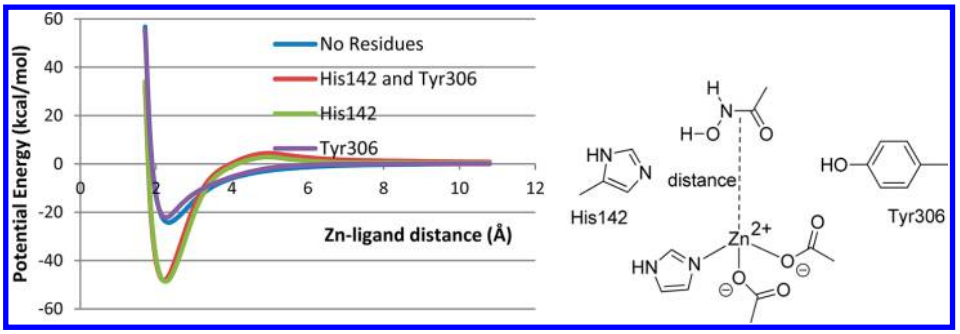


Figure 6. Zinc coordination energy for HDAC8 (pdb: 1t67). The shoulder between 4 and 6 Å observed when His142 was kept is due to light steric clashes upon ligand removal.

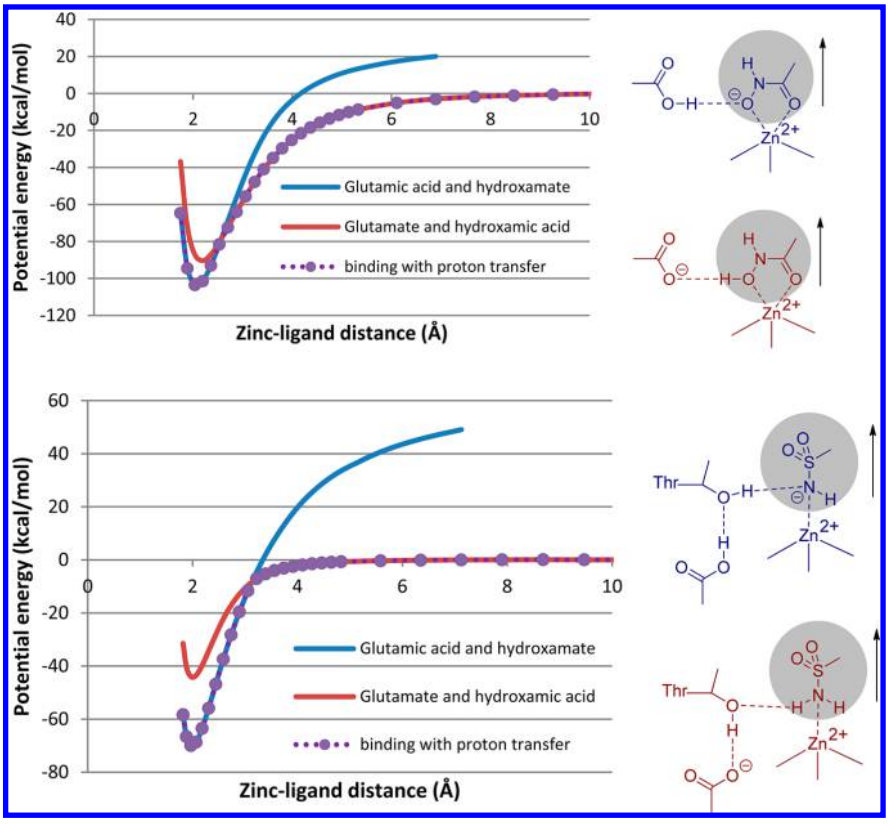


Figure 7. Effect of transferring a proton from ligand (hydroxamic acid or sulfonamide) to residue (glutamate with or without a threonine relay): (a) 1kbc, human neutrophil collagenase (MMP-8); (b) 3s71, carbonic anhydrase. The blue curve shows the favorable situation when the ligand is coordinated, and the red shows the favorable situation at further distances. The purple curve is the combination of the two that is required for docking.

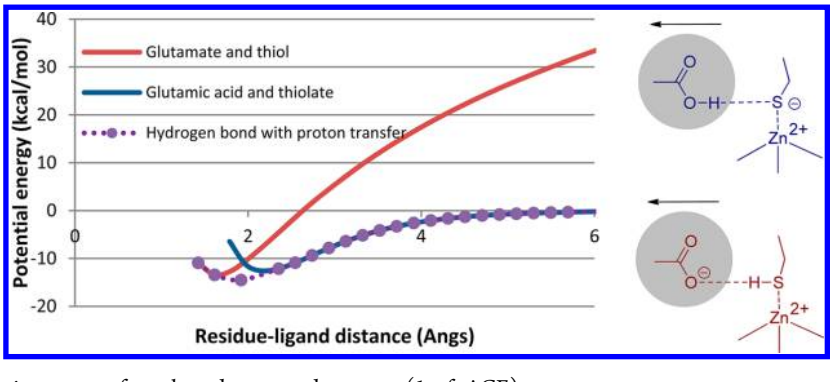


Figure 8. Computed interaction energy for selected truncated systems (1uzf, ACE).

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

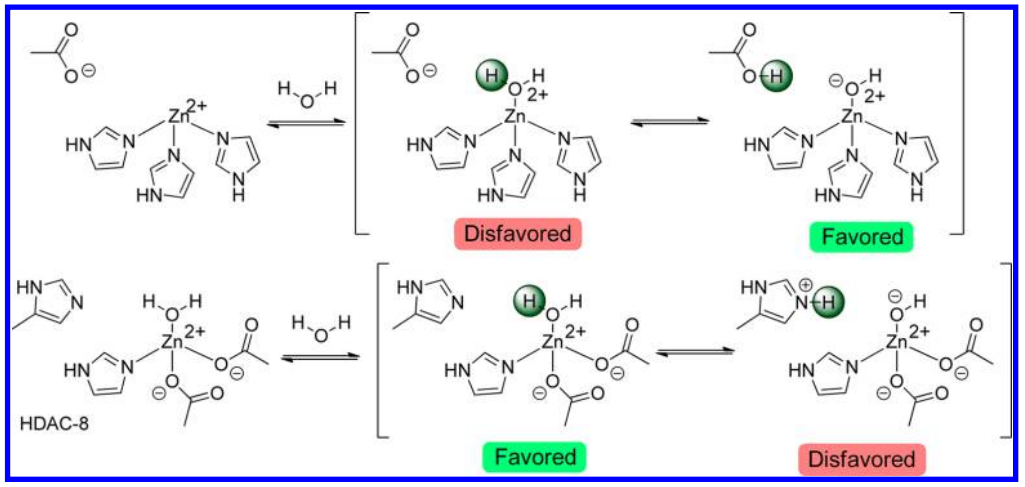


Figure 9. Zinc coordination and proton transfer with water molecule.

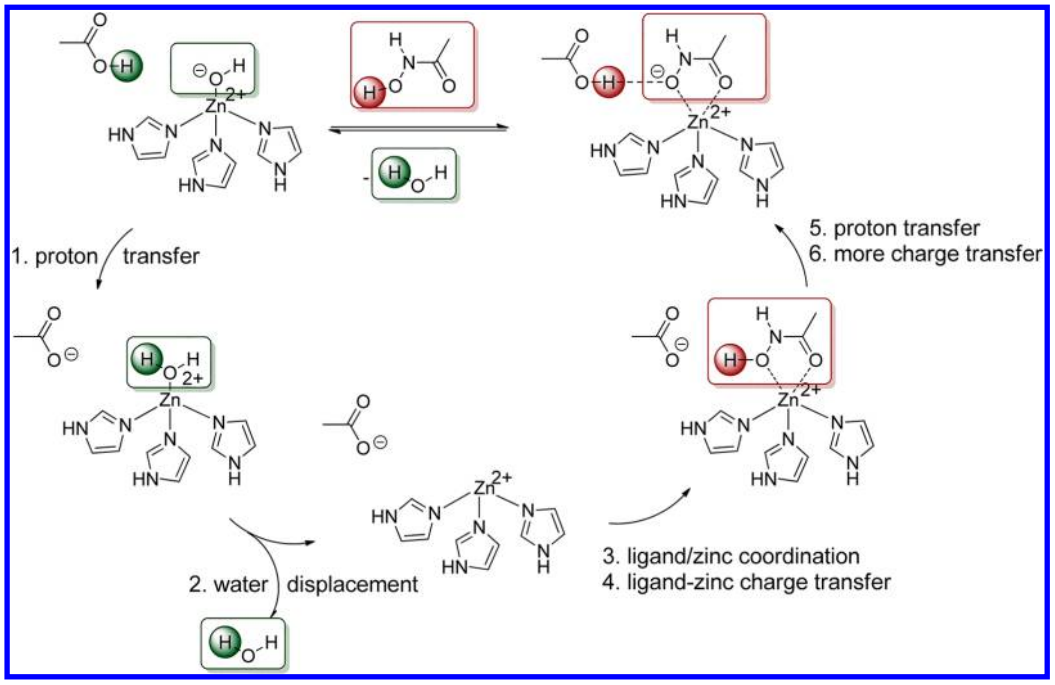


Figure 10. Binding process.

hydroxamic acid approaching the zinc coordination sphere next to a glutamic acid or a glutamate respectively was first computed (Figure 7). Our calculations on the representative systems revealed that at short ligand-zinc distances, the proton of hydroxamic acids, sulfonamides, and thiols is likely transferred to the glutamate (via a threonine in the case of carbonic anhydrase). The same calculations were performed when a histidine was the basic residue. In contrast to the glutamate containing systems, we found that the hydroxamic acid proton does not transfer in HDAC8 in which the zinc is neutralized (i.e., coordinated with two aspartates and one histidine). These observations, which are in agreement with the reports from Cross et al.[28] and Wu et al.,[32] validated our approach in which the crystallographic Cartesian coordinates were used without further optimization.

In order to uncouple zinc coordination and proton transfer for implementation into FITTED, investigations were carried out on diversely truncated systems. In contrast to the ligand−zinc coordination energy which was computed with or without neighboring residues by moving the ligand away, the hydrogen bond strength was evaluated (QM methods mentioned above) with the ligand bound to zinc in its equilibrium position while moving the neighboring residue away from the ligand along the hydrogen bond coordinate (Figure 8).

If we combine the two potential energy curves once more for the 19 representative systems, we see that the distances at which hydrogen bonds occur are solely controlled by the ionized ligands and glutamic acid residues and the neutral ligand/glutamate energies can be ignored (Figure 8). The point at which the two curves cross is assumed to be the energy barrier for the proton transfer to occur. However, this is difficult to model with molecular mechanics and, thus, a flatter-minimum curve was implemented.

A close look at crystal structures also showed that the residue for which no transfer was observed can still form hydrogen bonds that are shorter and stronger than usual. Among the examples is Tyr307 (Figure 1) in HDAC8. An approach similar to that used to develop zinc coordination parameters has been

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

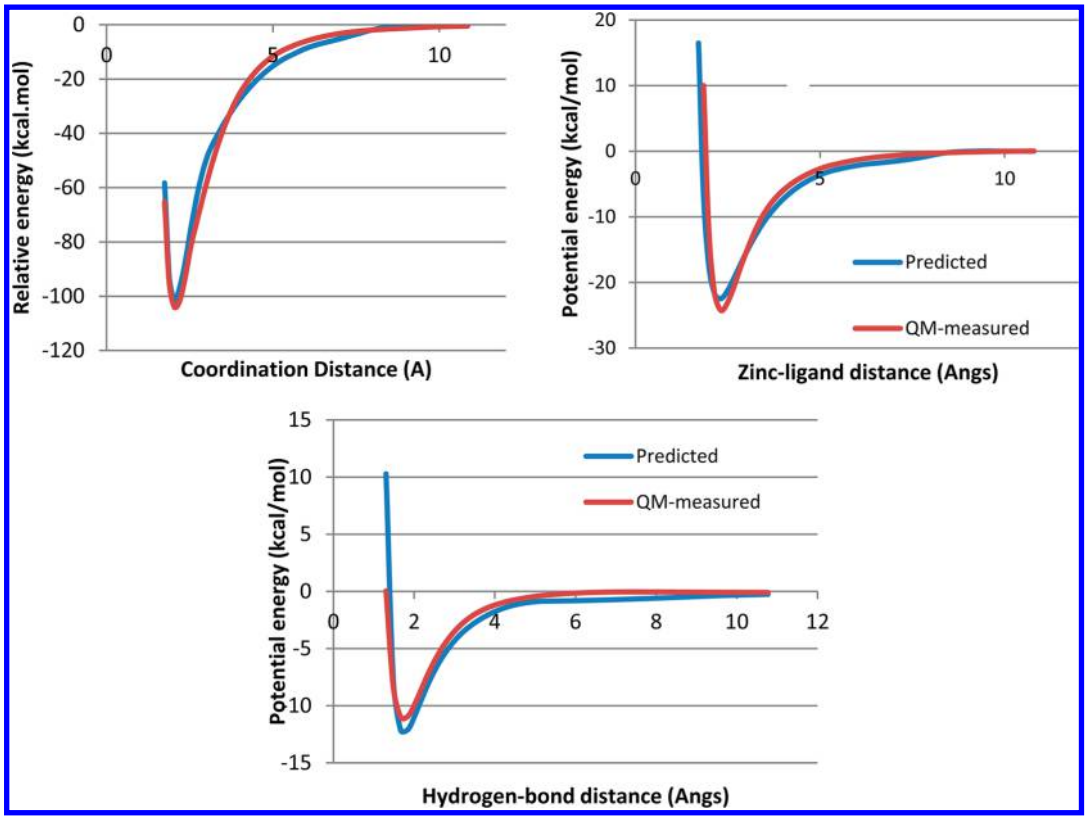


Figure 11. FITTED-derived energy curves using a LJ 6−3 equation vs DFT-derived energy curves for (1) zinc−ligand interaction energy 1kbc; (2) zinc−ligand interaction energy HDAC8 (1t67); (3) hydrogen bond energy (1kbc).

applied to develop the corresponding parameters. In a nutshell, the neighboring basic residues were moved away by 0.20 Å, and the corresponding energy was recorded.

Computing Water Coordination Energy. As discussed above, optimal binding energy can only be computed if the energy associated with the displacement of a water molecule is considered. For this purpose, the above systems were used in which the ligand was replaced by a water molecule and the position of the latter was optimized through DFT energy optimization. In some cases, the acidifying effect of the zinc ion led to a spontaneous proton transfer from the water molecule to a glutamate and contrarily, when many aspartic acids are coordinated to zinc, the proton spontaneously transferred from the glutamic acid to the hydroxide ion (Figure 9). In order to compute the water coordination energy, the difference in energy between the lowest in energy bound states (either water−glutamate or hydroxide−glutamic acid, Figure 9) and the unbound state was computed. For optimal transferability of these calculations, the set included systems with either Glu (alone or via Thr) or His as a basic residue and either His/His/His, His/His/GA, His/GA/GA, His/His/GA/GA or His/GA/GA/GA as residues in the zinc coordination sphere, where GA represents Glu or Asp.

Energy Function Parametrization. As discussed above and shown in Figure 10, the binding of small molecules to zinc-containing metalloenzymes is often not a simple bimolecular binding.

In order to properly model the zinc coordination (no. 3 in Figure 10), we proposed to develop a novel energy function term for computing the zinc−ligand interaction. This function should be independent of the charge eliminating the charge transfer effect (nos. 4 and 6 in Figure 10). From the collected DFT data, the energy well depth (ε) and zero energy point (σ) were determined for each system and used to derive a Lennard−Jones-like potential equation. The derived equations predict the energy differences with good accuracy (eqs 1−3).

$$\frac{A}{r^x} - \frac{B}{r^y} - \frac{C}{r^2} \tag{1}$$

$$\frac{A}{(r-0.25)^x} - \frac{B}{(r-0.25)^y} \tag{2}$$

$$A = 4\varepsilon^*\sigma^6 \quad \text{and} \quad B = 4\varepsilon^*\sigma^3 \tag{3}$$

This approach was applied to both zinc coordination and the strong hydrogen bonds described above. The terms (x, y) = (6, 3), (8, 4), and (10,6) were assessed with A and B being products of the energy well depth and zero points energies as defined in Lennard-Jones' theory,[39] C being a constant term to be trained and r being the distance between the coordinated atom and zinc. The third term was initially introduced to account for electrostatic interactions that may vary from one system to another. However, we found that upon fitting the curves for each system using the 6−3 relationship, removing this third term was possible without significantly affecting the fit of the DFT and MM curves. Unexpectedly, a shift of 0.25 Å of the position of the minimum of the energy well was observed in the majority of cases. All these observations led us to implement function 2 into FITTED with the parameters given as the Supporting Information. Examples of the predicted energy curves are given in Figure 11.

In addition to the new zinc-coordination energy equation, upon investigating the special hydrogen bond between the bound ligand and the neighboring basic residue (no. 5 in Figure

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

10), we determined that a new equation should be implemented for this type of interactions as well. We were able to model this as a 6−3 relationship as well, but with different parameters. The depth of the energy-well upon distancing the basic residue, as shown in Figure 8, is far beyond the expected stabilization obtained from currently modeled hydrogen bonds.

Implementation. At this stage, the newly implemented FITTED energy function can now properly evaluate the energy of such zinc-ligand systems. Then a routine identifying whether the proton transfer should be carried out based on distances and chemical nature of the coordinating functional group has been implemented. First, a routine identifying acidic zinc binding groups (e.g., hydroxamic acid, terminal sulfonamide, and thiols) was implemented into SMART, a program of the FITTED suite used to prepare the small molecules prior to the actual docking. This information is then output in the ligand file. Within FITTED itself, a routine was implemented that can read this information from the ligand file and a switching function was introduced which recognizes whether the zinc binding group (e.g., hydroxamic acid) is close enough to zinc to be ionized or not. If it is ionized, the energy function is applied to the system on the left in Figure 12a, or if neutral, it is applied

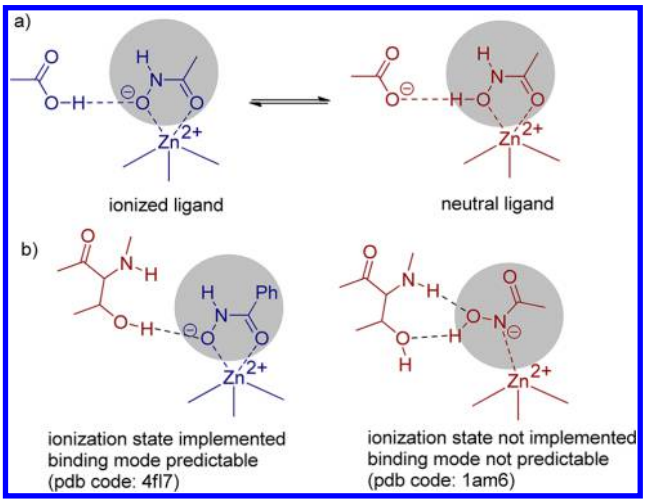


Figure 12. Limitations in modeling proton transfer.

to the system on the right. In practice, hydrogens on both the functional group and on the basic residue are present and a list of interactions for each is prepared. The newly implemented routine identifies which of the two lists should be selected. As a major drawback of this approach, any acidic proton that has not been implemented to be recognized by the software will not be considered as candidates for the proton transfer. For example, the unusual binding mode observed with a small, aliphatic hydroxamic acid (Figure 12b) within the binding site of carbonic anhydrase II[40] can not be predicted with the current implementations.

Since one of the major features of FITTED is the option to displace water molecule, no additional modifications were necessary for displacing the coordinating water in the unbound state. However, scoring the displacement of water molecule was modified to incorporate the water/zinc coordination energy which is highly dependent on the zinc coordination sphere. In parallel to this explicit displaceable water molecule method, we also assessed the use of a more implicit approach. For this purpose, scaling factors were applied to the zinc coordination energies to implicitly account for the displacement of the water molecules coordinating the zinc ion.

RESULTS AND DISCUSSION

Validation—Pose Prediction. With these implementations in hand, a first set of validation experiments was carried out. For this purpose, self-docking experiments using the 121 complexes selected as a testing set were performed. All these docking experiments were carried out 10 times to ensure that the result was statistically significant. In order to validate the novel energy function and implementations, the accuracy of FITTED, using either the traditional electrostatic/van der Waals energy, the previously implemented hydrogen bond-like term or the current version with an explicit or implicit water molecule coordinating zinc, was assessed. With the developed docking procedure, the metal coordination geometry was not explicitly evaluated but as mentioned above, the conformation search algorithm oriented zinc binding groups into the free coordination sites.

At this stage, we expected a significant improvement of the zinc coordination geometry that will be therefore more accurately scored. In practice, we observed that the novel implementation significantly improved the positioning of the zinc binding group (Figure 13 top). As a result, the overall pose prediction is significantly enhanced (Figure 13 bottom) with improvement as large as 10% (RMSD lower than 2 Å) and even 13% if the best scoring pose of the 10 runs was used for each of the 121 systems. More unexpectedly, the implicit water displacement energy treatment turned out to lead to more accurate predictions. However, a closer look at the docked poses revealed that some of the docked molecules favor an interaction with the coordinated water over zinc coordination. In a large fraction of our set, this water is displaced, and this alternative binding not observed. Obviously, the implicit water mode precludes this alternative (i.e., the zinc binding group can only bind to zinc) and appears as more accurate. Thus, we believe that this apparent improvement is due to the simpler potential energy surface while the presence of an explicit water molecule still models more closely the ligand binding process.

We next looked more specifically at HDAC/inhibitor complexes (Table 1). While the previous implementations have overall success of 43% (electrostatic, LJ 12−6 treatment of metal coordination) and 40% (electrostatic, LJ 12−10), the novel implementation predicted the pose correctly 71% of the time. 3c0z (HDAC7) and 1t67 and 1t69 (HDAC8) are three examples of significant pose prediction improvement. While the previous two implementations predicted poses with RMSDs of about 4 Å in all 10 runs with 3c0z, the novel implementation enabled the prediction of poses within 2.0 Å of the observed poses 8 of the 10 runs. The latest implementations also allowed a more precise ligand placement as shown in Figure 14. While the RMSD for the whole ligand was good (1.35 and 1.56 Å for the previous two implementations), it was improved with the implementations reported herein (1.09 Å). This improvement was even more pronounced in the ZBG; while the ligand atoms in the coordination sphere were above 2.0 Å with the previous implementations, it was as low as 0.51 Å with the current version of the program. Although this did not significantly affect the overall binding mode, it indicates the scoring of the zinc coordination and proton transfer are more accurate and should be reflected in more accurate scoring in virtual screening experiments.

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

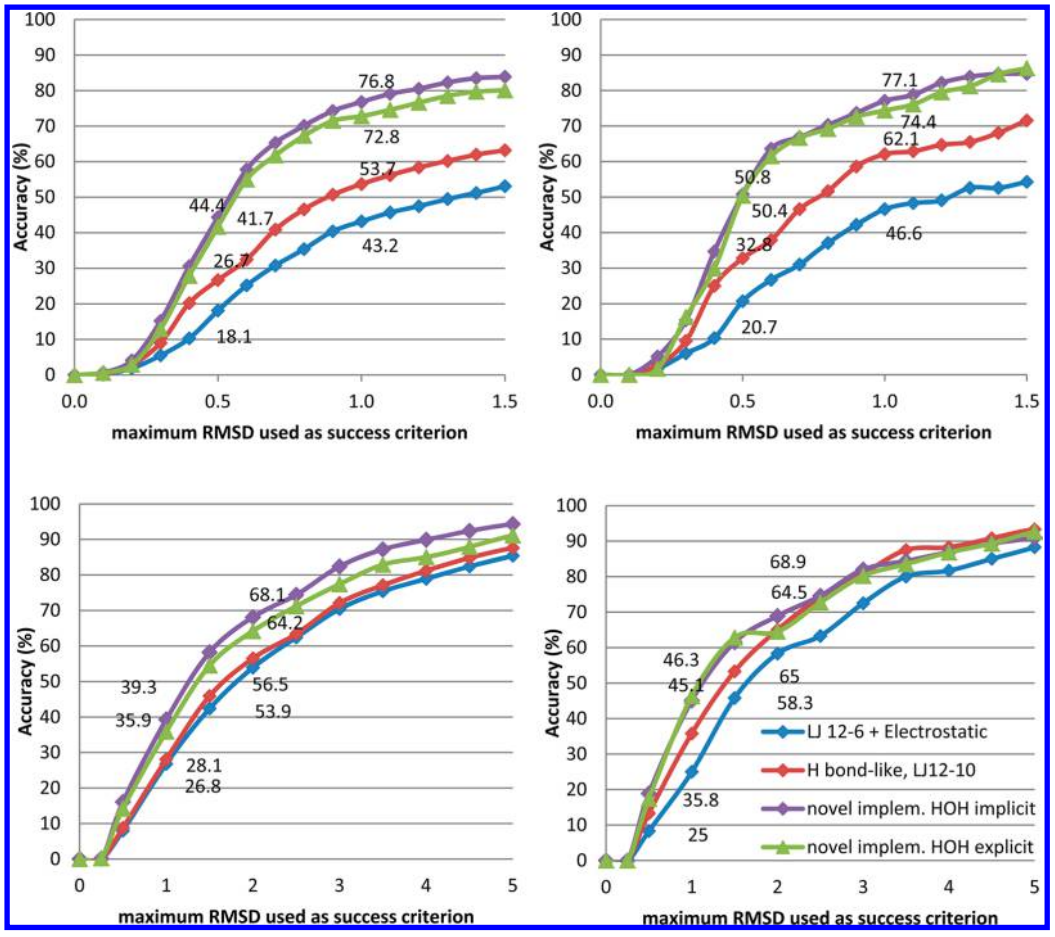


Figure 13. Pose prediction accuracy. (top panel) Accuracy of the zinc coordination geometry (the RMSD of only the zinc binding group is computed) average over 10 runs (left), best-scoring of the 10 runs (right). (bottom panel) Accuracy of the pose prediction average over 10 runs (left), best-scoring of the 10 runs (right).

Validation—Virtual Screening. At this stage, these novel implementations were applied to the screening of potential metalloenzyme inhibitors. For this purpose, known actives and decoys should be collected. The DUD-e (http://dude.docking.org/targets) set includes inhibitors and decoys for five metalloenzymes which were used herein. In order to test our entire set of programs, the tautomers from the sets of ligands and decoys from the DUD-e sets were identified and removed and 100 unique ligands and 5000 unique decoys were selected and hydrogen atoms added. All these steps were done using routines of our FORECASTER platform as described in the Experimental Section. Table 2 summarizes the accuracy of these screens. While DOCK provided AU-ROC values ranging from 0.71 to 0.80 with an average of 0.744, FITTED provided AU-ROC above 0.79 (explicit water molecule) for 4 of the 5 targets used in this validation study with an improved average of 0.831 with an explicit water molecule bound to zinc and 0.821 with implicit water. Surprisingly HDAC2 provided a lower AU-ROC of 0.67 (0.63 with implicit water).

Further investigation of the somewhat disappointing result with HDAC2 revealed that the set of actives includes several macrocyclic molecules, as exemplified by ChEMBL424189 (trapoxin B, Figure 15), as well as several inhibitors with bulky and/or branched cap groups that did not fit into the binding site of the enzyme cocrystallized with a much smaller ligand (3max, Figure 15). The N-terminal L1 loop in HDAC2, which is on the surface lining the opening to the active site, is several amino acids longer than the corresponding loop in HDAC8 (Figure 16). The result is that HDAC2 possesses a longer access tube and a more restricted surface that may limit the docking of larger or conformationally restricted cap groups. In addition, it is established in some HDACs, including HDAC8, that the surface is malleable and thus may accommodate larger ligands,[41] but this motion is not currently modeled in FITTED. In addition, we have found that among the actives included in the DUD-e validation set, some are most likely not simple competitive inhibitors. For example, trapoxin B is a known irreversible covalent inhibitor[42] and that ChEMBL402341, a dithiolethione-containing inhibitor releases H_2S over time (Figure 15).[43] Thus the actual active inhibitors might be species other than those in the database, although this could not explain that the AU-ROC observed for HDAC2 was significantly lower than for the other four investigated proteins.

Earlier this year, another two crystal structures of HDAC2 (4lxz and 4ly1) have been reported. In order to assess the postulated role of the protein conformation on the poor AU-ROC observed, the screening was carried out with structure 4lxz. First, we observed that, in contrast to 3max, all the compounds were successfully docked into 4lxz, which features a larger binding site. Second, an excellent AU-ROC was measured (4lxz: 0.87 vs 3max: 0.68). This remarkable improvement demonstrated the impact of the selection of the protein structure on the accuracy of a screen. If this structure

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

Table 1. Docking Accuracy on HDACs[a]

	implementation	average RMSD[b]	lowest RMSD[b]	success rate[c]
HDAC2—3max	elec./vdW	0.39	0.35	100%
	H-bond-like	0.39	0.36	100%
	new implem.	0.34	0.30	100%
HDAC4—2vqj	elec./vdW	4.24	3.49	0%
	H-bond-like	3.08	2.24	0%
	new implem.	4.06	2.98	0%
HDAC7—3c0z	elec./vdW	4.40	4.00	0%
	H-bond-like	4.21	4.00	0%
	new implem.	2.07	1.82	80%
HDAC7—3c10	elec./vdW	1.07	0.85	90%
	H-bond-like	1.62	1.03	80%
	new implem.	0.92	0.81	100%
HDAC8—1t67	elec./vdW	2.06	1.56	50%
	H-bond-like	3.19	1.35	50%
	new implem.	2.3	1.09	50%
HDAC8—1t69	elec./vdW	2.72	1.69	40%
	H-bond-like	3.43	2.34	0%
	new implem.	2.15	1.65	90%
HDAC8—1w22	elec./vdW	3.26	0.71	60%
	H-bond-like	1.74	1.09	80%
	new implem.	4.27	0.78	60%
HDAC8—3f07	elec./vdW	3.33	2.68	0%
	H-bond-like	2.79	2.26	0%
	new implem.	2.38	1.46	10%

[a]The three implementations are compared for their accuracy in predicting the ligand binding modes (implicit water displacement energy mode). [b]Over 10 runs. [c]Percentage of runs with RMSD < 2 Å (out of 10 runs).

was considered, the average AU-ROC over the five metalloenzymes rose to 0.869.

In order to further demonstrate that these novel implementations significantly increased the accuracy of our docking program, the same screens were carried out with the other two metal coordination computation methods (i.e., electrostatic and van der Waals as well as electrostatic and hydrogen bond). Overall the enrichment in active compounds in the top of the ranked list was found to be very poor. When compared to the average AU-ROC of 0.869 obtained with the new version of the program, turning off these new functionalities led to significantly lower average AU-ROC's of 0.639 (elec−vdW) and 0.510 (elec−HB). This drop is most likely due to the very poor prediction of the key hydrogen bond network associated to the zinc coordination. Our latest implementations consider more aspects of the binding and more specifically the proper ionization states of the proteins and ligands and include energy terms for zinc coordination and interactions with the neighboring residues. In fact, we would likely observe better predictions using the electrostatic and/or hydrogen bond terms if properly protonated states (e.g., ionized ligands and protonated neighboring residues) were used. With HDAC2 and HDAC8, for which no proton transfer occurs, the AU-ROC observed with the electrostatic/van der Waals implementation is higher than with the other three

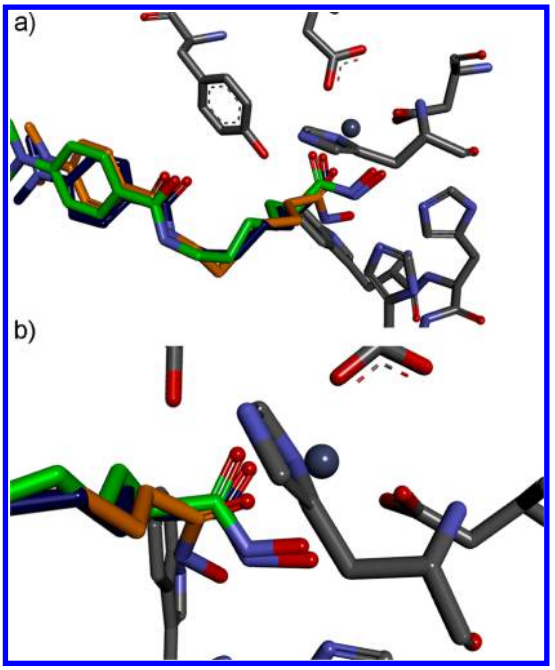


Figure 14. Predicted poses for 1t67. HB-mode: orange carbons, new implementations. Dark blue carbons vs crystal structure (green carbons): (a) ligand binding mode; (b) zinc coordination.

enzymes for which proton transfer should occur. However, the unusually short hydrogen bonds observed in these systems could not be predicted using the standard van der Waals (Lennard-Jones 12−6) or hydrogen bond (Lennard-Jones 12−10) terms, and unless these zinc coordination and hydrogen bond models using LJ 12−6 or LJ 12−10 are revised, they can not be as predictive as the novel approach presented herein.

CONCLUSION

In conclusion, our docking program FITTED has been modified to account for zinc coordination of ligands and other related processes such as unusually strong hydrogen bonds and proton exchange with neighboring residues. These implementations significantly improved the pose prediction accuracy over the more traditional electrostatic/Lennard-Jones treatment or even the hydrogen bond-like treatment previously implemented. Based on the success of this technique, it is likely that this approach could be applied to other transition metals if desired. A close look at the results showed that docking to HDACs was also more accurate and that identification of actives in screening is possible with this current version of the program.

EXPERIMENTAL SECTION

DFT Calculations. All the quantum mechanical calculations were performed using DFT, more specifically the B3LYP functional (restricted Hartree−Fock) and a custom basis set combining 6-311G* for the zinc atom and 6-31+G* for all other atoms (based on ref 24). All calculations were performed in vacuum. The B3LYP calculations were performed using GAMESS-US v.Aug2011−64bit. The crystal structures from the PDB were truncated to include the zinc atom, the residues coordinated to zinc (also truncated), and the residues interacting with the ligand. The ligands were also truncated to include only the zinc binding group (ZBG). Hydrogen atoms were added to the ligand only where they could interact with

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

Table 2. Area under the Receiver Operating Curve

enzyme	PDB code	DOCK[a]	FITTED[b] HOH explicit	FITTED HOH implicit	FITTED elec−vdW[b]	FITTED Hbond[b]
ACE	3bkl	0.72	0.79	0.78	0.60	0.41
CA	1bcd	0.73	0.93	0.94	0.60	0.56
HDAC2	3max	0.77	0.68	0.63	0.64	0.59
	4lxz	nd[c]	0.87	0.87	0.76	0.71
HDAC8	3f07	0.80	0.87	0.86	0.71	0.55
MMP13	830c	0.71	0.89	0.89	0.66	0.44
average		0.744	0.831 (0.869)	0.821 (0.869)	0.641 (0.665)	0.509 (0.534)

[a]Data provided at http://dude.docking.org/targets. [b]With the ligands prepared as with the new implementation (neutral ligands) or preionized. [c]Not determined. [d]Average with HDAC2 represented by 3max, average with 4lxz in brackets.

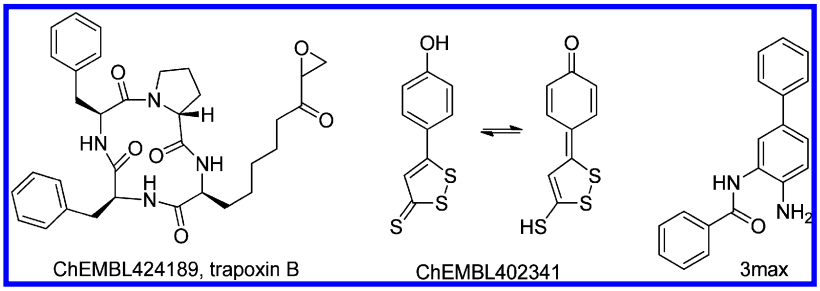


Figure 15. HDAC2 inhibitors.

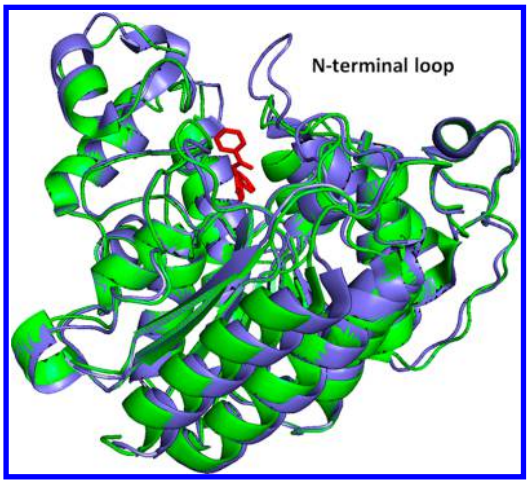


Figure 16. Comparison of HDAC2 (3max, blue) and HDAC8 (2v5x, green). The additional length of the *N*-terminal loop increases the length of the access tube. The ligand shown in red is the *N*-(2-aminophenyl)-2-benzamide from the HDAC2 structure (3max).

the neighboring residue in the active site (i.e., where a proton shift could occur).

Force Field Parameters. The first step was to freeze all atoms in the system other than the manually added hydrogen atoms which were optimized using the DFT methods mentioned previously. This was deemed to be the equilibrium state (crystal structure + optimized hydrogen atoms). Next, the system was broken down and the neighboring residues were omitted. The ligand was moved systematically away from the zinc along a designated vector by steps of 0.2 Å in order to obtain the energy profile (single point energies). Similarly, the neighboring residue was moved systematically along the hydrogen bond vector in order to obtain the hydrogen bond energy profile. These two sets of calculations were performed with the hydrogen atom either on the ligand or on the neighboring residue in order to establish which configuration would be optimal depending on the conditions (i.e., distance of the ligand from the zinc atom); either one configuration was always preferential (no proton transfer) or the two profiles had to be combined (proton transfer). Special considerations were made for the complex systems (carbonic anhydrase) where the hydrogen transfer occurred via a threonine to a glutamic acid.

The generated structures along the vectors were used to compute the FITTED energy and to derive force field parameters for optimal curve matching.

Docking with FITTED. The subversion 3475 of the FORECASTER platform including all the necessary programs used in this work (FITTED, CONVERT, SMART, SELECT, PREPARE, PROCESS) has been used for this study.

Construction of the Testing Sets. The sets of decoys and ligands were downloaded from the DUD-e Web site (http://dude.docking.org/targets) and further processed as follows. First the hydrogens were removed from all the molecules in order to make the tautomers as similar as possible using a routine of FORECASTER. Then, the molecules were clustered by similarity using SELECT, and 100 ligands and 5000 decoys with the largest diversity were selected. Hydrogens were added using CONVERT, and all these ligands prepared for docking using SMART.

Preparation of the Protein Files. PREPARE and PROCESS were applied using the specific keyword identifying metalloenzymes (`Macromolecule metalloprotein`) and other parameters to the default.

Docking with FITTED. Each docking is a set of three runs starting from a different seed. The different implementations were identified by a specific keyword (`Macromolecule protein/metalloprotein_HB/metalloprotein`). For the pose prediction tests, each set of three runs was carried out 10 times. For the screening, three runs for each small molecule were also carried out.

Otherwise, default parameters implemented in FITTED have been used.

dx.doi.org/10.1021/ci400550m | J. Chem. Inf. Model. 2014, 54, 254−265

■ ASSOCIATED CONTENT

Supporting Information

Set of 121 crystal structures used for validation. This material is available free of charge via the Internet at http://pubs.acs.org. The program is available on request from the authors (www.fitted.ca).

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: nicolas.moitessier@mcgill.ca.

Notes

The authors declare the following competing financial interest(s): FITTED is distributed by Molecular Forecaster Inc., cofounded by E.T. and N.M. (FITTED is free to academic groups).

■ ACKNOWLEDGMENTS

We thank FQRNT (Equipe program) for financial support. Calcul Québec and Compute Canada are acknowledged for generous CPU allocations.

■ REFERENCES

(1) Hanessian, S.; Moitessier, N.; Therrien, E. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 873.
(2) Hanessian, S.; MacKay, D. B.; Moitessier, N. *J. Med. Chem.* **2001**, *44*, 3074.
(3) Hanessian, S.; Moitessier, N.; Cantin, L. D. *Tetrahedron* **2001**, *57*, 6885.
(4) Englebienne, P.; Fiaux, H.; Kuntz, D. A.; Corbeil, C. R.; Gerber-Lemaire, S.; Rose, D. R.; Moitessier, N. *Proteins: Struct. Funct. Bioinf.* **2007**, *69*, 160.
(5) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. *Br. J. Pharmacol.* **2008**, *153*, S7.
(6) Corbeil, C. R.; Englebienne, P.; Moitessier, N. *J. Chem. Inf. Model.* **2007**, *47*, 435.
(7) Corbeil, C. R.; Moitessier, N. *J. Chem. Inf. Model.* **2009**, *49*, 997.
(8) De Cesco, S.; Deslandes, S.; Therrien, E.; Levan, D.; Cueto, M.; Schmidt, R.; Cantin, L. D.; Mittermaier, A.; Juillerat-Jeanneret, L.; Moitessier, N. *J. Med. Chem.* **2012**, *55*, 6306.
(9) Englebienne, P.; Moitessier, N. *J. Chem. Inf. Model.* **2009**, *49*, 1568.
(10) Marcial, B. L.; Sousa, S. F.; Barbosa, I. L.; Dos Santos, H. F.; Ramos, M. J. *J. Phys. Chem. B* **2012**, *116*, 13644.
(11) Norris, R.; Casey, F.; FitzGerald, R. J.; Shields, D.; Mooney, C. *Food Chem.* **2012**, *133*, 1349.
(12) Marques, S. M.; Tuccinardi, T.; Nuti, E.; Santamaria, S.; André, V.; Rossello, A.; Martinelli, A.; Santos, M. A. *J. Med. Chem.* **2011**, *54*, 8289.
(13) Omanakuttan, A.; Nambiar, J.; Harris, R. M.; Bose, C.; Pandurangan, N.; Varghese, R. K.; Kumar, G. B.; Tainer, J. A.; Banerji, A.; Perry, J. J. P.; Nair, B. G. *Mol. Pharmacol.* **2012**, *82*, 614.
(14) Feng, J.; Jin, K.; Zhu, H.; Zhang, X.; Zhang, L.; Liu, J.; Xu, W. *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5863.
(15) Tuccinardi, T.; Bertini, S.; Granchi, C.; Ortore, G.; Macchia, M.; Minutolo, F.; Martinelli, A.; Supuran, C. T. *Bioorg. Med. Chem.* **2013**, *21*, 1511.
(16) Cole, J. C.; Nissink, J. W. M.; Taylor, R. In *Virtual Screening in Drug Discovery*; Shoichet, B.; Alvarez, J., Eds.; Taylor & Francis CRC Press: Boca Raton, FL, 2005.
(17) Korb, O.; Stutzle, T.; Exner, T. E. *Lect. Notes Comput. Sci.* **2006**, *4150*, 247.
(18) Korb, O.; Stutzle, T.; Exner, T. E. *J. Chem. Inf. Model.* **2009**, *49*, 84.
(19) Seebeck, B.; Reulecke, I.; Kamper, A.; Rarey, M. *Proteins Struct. Funct. Bioinf.* **2008**, *71*, 1237.
(20) Irwin, J. J.; Raushel, F. M.; Shoichet, B. K. *Biochemistry* **2005**, *44*, 12316.
(21) Norris, R.; Casey, F.; FitzGerald, R. J.; Shields, D.; Mooney, C. *Food Chem.* **2012**, *133*, 1349.
(22) Tavera-Mendoza, L. E.; Quach, T. D.; Dabbas, B.; Hudon, J.; Liao, X.; Palijan, A.; Gleason, J. L.; White, J. H. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 8250.
(23) Lamblin, M.; Dabbas, B.; Spingarn, R.; Mendoza-Sanchez, R.; Wang, T. T.; An, B. S.; Huang, D. C.; Kremer, R.; White, J. H.; Gleason, J. L. *Bioorg. Med. Chem.* **2010**, *18*, 4119.
(24) Fischer, J.; Wang, T. T.; Kaldre, D.; Rochel, N.; Moras, D.; White, J. H.; Gleason, J. L. *Chem. Biol.* **2012**, *19*, 963.
(25) Zhang, J.; Yang, W.; Piquemal, J. P.; Ren, P. *J. Chem. Theory Comput.* **2012**, *8*, 1314.
(26) Kalyaanamoorthy, S.; Chen, Y. P. P. *J. Chem. Inf. Model.* **2012**, *52*, 589.
(27) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry*, 5th ed.; WH Freeman: New York, 2002.
(28) Cross, J. B.; Duca, J. S.; Kaminski, J. J.; Madison, V. S. *J. Am. Chem. Soc.* **2002**, *124*, 11004.
(29) Czarny, B.; Stura, E. A.; Devel, L.; Vera, L.; Cassar-Lajeunesse, E.; Beau, F.; Calderone, V.; Fragai, M.; Luchinat, C.; Dive, V. *J. Med. Chem.* **2013**, *56*, 1149−1159.
(30) Tuccinardi, T.; Bertini, S.; Granchi, C.; Ortore, G.; Macchia, M.; Minutolo, F.; Martinelli, A.; Supuran, C. T. *Bioorg. Med. Chem.* **2013**, *21*, 1511.
(31) Vanommeslaeghe, K.; Van Alsenoy, C.; De Proft, F.; Martins, J. C.; Tourwé, D.; Geerlings, P. *Org. Biomol. Chem.* **2003**, *1*, 2951.
(32) Wu, R.; Lu, Z.; Cao, Z.; Zhang, Y. *J. Am. Chem. Soc.* **2011**, *133*, 6110.
(33) Akdemir, A.; Güzel-Akdemir, Ö.; Scozzafava, A.; Capasso, C.; Supuran, C. T. *Bioorg. Med. Chem.* **2013**, *21*, 5228.
(34) Gavernet, L.; Gonzalez Funes, J. L.; Palestro, P. H.; Bruno Blanch, L. E.; Estiu, G. L.; Maresca, A.; Barrios, I.; Supuran, C. T. *Bioorg. Med. Chem.* **2013**, *21*, 1410.
(35) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M. *J. Chem. Theory Comput.* **2010**, *6*, 2935.
(36) Wang, D.; Helquist, P.; Wiest, O. *J. Org. Chem.* **2007**, *72*, 5446.
(37) Vanommeslaeghe, K.; De Proft, F.; Loverix, S.; Tourwé, D.; Geerlings, P. *Bioorg. Med. Chem.* **2005**, *13*, 3987.
(38) Ahlstrand, E.; Spångberg, D.; Hermansson, K.; Friedman, R. *Int. J. Quantum Chem.* **2013**, *113*, 2554−2562.
(39) Lennard-Jones, J. E. *Proc. Phys. Soc.* **1931**, *43*, 461.
(40) Scolnick, L. R.; Clements, A. M.; Liao, J.; Crenshaw, L.; Hellberg, M.; May, J.; Dean, T. R.; Christianson, D. W. *J. Am. Chem. Soc.* **1997**, *119*, 850.
(41) Somoza, J. R.; Skene, R. J.; Katz, B. A.; Mol, C.; Ho, J. D.; Jennings, A. J.; Luong, C.; Arvai, A.; Buggy, J. J.; Chi, E.; Tang, J.; Sang, B. C.; Verner, E.; Wynands, R.; Leahy, E. M.; Dougan, D. R.; Snell, G.; Navre, M.; Knuth, M. W.; Swanson, R. V.; McRee, D. E.; Tari, L. W. *Structure* **2004**, *12*, 1325.
(42) Kijima, M.; Yoshida, M.; Sugita, K.; Horinouchi, S.; Beppu, T. *J. Biol. Chem.* **1993**, *268*, 22429.
(43) Rossoni, G.; Sparatore, A.; Tazzari, V.; Manfredi, B.; Soldato, P. D.; Berti, F. *Br. J. Pharmacol.* **2008**, *153*, 100.