JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

AutoDock4$_{Zn}$: An Improved AutoDock Force Field for Small-Molecule Docking to Zinc Metalloproteins

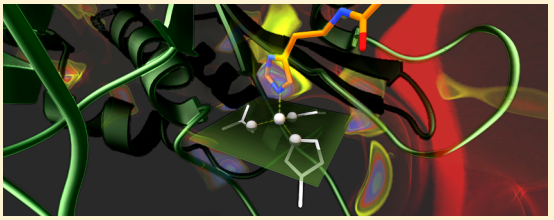
Diogo Santos-Martins,[†,‡,¶] Stefano Forli,[†,¶] Maria João Ramos,[‡] and Arthur J. Olson*[,†]

[†]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California 92037, United States

[‡]REQUIMTE, Departamento de Química e Bioquímica, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre s/n, 4169-007, Porto, Portugal

Supporting Information

ABSTRACT: Zinc is present in a wide variety of proteins and is important in the metabolism of most organisms. Zinc metalloenzymes are therapeutically relevant targets in diseases such as cancer, heart disease, bacterial infection, and Alzheimer's disease. In most cases a drug molecule targeting such enzymes establishes an interaction that coordinates with the zinc ion. Thus, accurate prediction of the interaction of ligands with zinc is an important aspect of computational docking and virtual screening against zinc containing proteins. We have extended the AutoDock force field to include a specialized potential describing the interactions of zinc-coordinating ligands. This potential describes both the energetic and geometric components of the interaction. The new force field, named AutoDock4$_{Zn}$, was calibrated on a data set of 292 crystal complexes containing zinc. Redocking experiments show that the force field provides significant improvement in performance in both free energy of binding estimation as well as in root-mean-square deviation from the crystal structure pose. The new force field has been implemented in AutoDock without modification to the source code.



INTRODUCTION

Zinc is present in numerous biological structures and is found in virtually all aspects of metabolism across multiple species.[1] It can play a structural role as in zinc finger proteins, the most prevalent proteins in eukaryotic genomes,[2] and is present in all enzyme classes,[3] usually in the form of coordinated Zinc(II) or Zn^{2+} ion. Zinc metalloenzymes are therapeutically relevant targets in many diseases, like heart disease,[4] cancer,[5,6] bacterial infections,[7] and Alzheimer's.[8,9] In most cases, a drug molecule establishes coordination bonds with the zinc ion[10] present in the protein; thus, an accurate description of this interaction is crucial for drug design.

To properly model the zinc coordination interactions, two issues should be addressed: the coordination geometry and the interaction strength.

Most force fields describe metal coordination using descriptions derived from the original Stote and Karplus nonbonded model,[11] where the interaction is described using Lennard-Jones and Coulomb potentials. This description relies on assignment of partial charges,[11−13] and thus accuracy becomes strongly dependent on the choice of charge model. Also, an electrostatic model based on the filled valence orbital of the Zn^{2+} ion fails to explain the prevalence of histidine and cysteine over the more electronegative carboxylate groups of glutamate and aspartate as the most frequent zinc coordinating residues.[14−17]

Moreover, some high potency inhibitors coordinate Zn^{2+} via uncharged nitrogens with electron lone pairs, such as those found in sulfonamides[18] and imidazoles[6] (see Figure 13), that seem to interact more strongly than negatively charged nitro groups.[7] Recently, DFT calculations were used to calibrate a nondirectional zinc coordination force field independent of atomic partial charges.[19]

Polarization and charge transfer models[20−22] could provide a more accurate description, although their computational complexity makes them unsuitable for dockings, which typically involve a large number of energy estimations over the course of the calculation.

The coordination geometry issue is addressed differently by bonded and nonbonded models. It has been recently demonstrated that zinc exhibits a strong preference for the tetrahedral geometry, with some of the previously observed variability in coordination spheres being artifactual.[17] Bonded models, such as the Zinc AMBER Force Field,[23] describe the tetrahedral coordination with harmonic potentials and angle terms for an explicit bond that provides directionality. Due to the requirement of the explicit bonds where ligands coordinate with zinc, bonded models are not suitable for docking calculations.

In nonbonded models, few force fields provide directional potentials. Two examples that do are the cationic dummy atom model[5] and the scoring function implemented in FlexX.[24] However, in this latter case, while improving coordination geometry accuracy, no improvements in binding energy prediction were reported.

Received: April 3, 2014
Published: June 15, 2014

   dx.doi.org/10.1021/ci500209e | J. Chem. Inf. Model. 2014, 54, 2371−2379

To be suitable for docking, and virtual screens in particular, modeling the interaction with zinc must provide a description of the geometry that is computationally efficient, and good accuracy in the estimation of the interaction strength.

In this paper, we report the development of a directional, charge-independent model for zinc-coordination force field for AutoDock4 which provides higher accuracy than the standard force field. Interactions are modeled independently for different atom types, providing specific potentials for each one.

Over the years, AutoDock and its force field were modified by us and others to improve scoring of low affinity ligands[25] or to obtain a scoring function tailored to specific targets, like kinases.[26] Other methods added new features, like receptor flexibility models,[27,28] flexible macrocycle docking,[29] and docking with waters.[30] The highly customizable architecture of the program allows implementation of substantial modifications relatively easily and without requiring source code changes.

METHODS

To implement the new zinc-coordination model, first we identified a data set of high-quality complexes for which experimental affinity values had been determined. The data set analysis enabled determination of the parameters for geometrical terms that were then calibrated to fit within the AutoDock force field. The new force field, named AutoDock4$_{Zn}$, was then cross-validated on the data set.

Data Set Creation. To design the new force field, a suitable set of zinc metalloprotein−ligand complexes was defined. The ligands cover a wide range of structure diversity and binding affinity, thus providing an optimal calibration set for generic applicability of the force field for drug design.

In order to build the data set, the Binding MOAD[31] was filtered using the following criteria: (a) presence of at least one zinc ion; (b) experimentally determined inhibition (K_i) or dissociation (K_d) constants; (c) no alternate conformation or missing atoms for the ligand, and (d) no alternate side chain conformations in receptor residues within 5 Å from any ligand atom. This filtering led to a set of 510 complexes, which were downloaded from the Protein Data Bank.[32]

These complexes were then analyzed to isolate and characterize the zinc coordination geometry within the receptor and its interaction with ligands. Each complex was classified accordingly to the number of receptor (r) and ligand (l) atoms within coordination distance (≤2.8 Å for sulfur atoms, ≤2.5 Å for all others) from the zinc ion,[24] and denoted as Zn$_{r,l}$.

A specific treatment was used to analyze the coordination geometry of carboxylic acids. Carboxylic acids from aspartate or glutamate side chains have been described to coordinate zinc mainly with bidentate, monodentate, syn or anti modes. However, it has been demonstrated that carboxylate groups can adopt any coordination geometry ranging between mono- and bidentate,[33,34] which is poorly described by a discrete classification scheme. To address this issue, carboxylic acid groups on receptors were always considered as monovalent and represented by a weighted average of the position of the two coordinating oxygens (see Figure 6); the method used to calculate the weighted average for carboxylic acids is described in the Supporting Information. The distribution of different coordination geometries is summarized in Table 1. The most represented coordination geometry in our data set is the tetrahedral one (Zn$_{3,1}$, Zn$_{4,0}$), that was indeed found to be the most common in biological systems.[17] Other geometries, like five- (Zn$_{4,1}$, Zn$_{3,2}$) and six-coordinated (Zn$_{3,3}$, Zn$_{4,2}$), were also found, but were much rarer.

Table 1. Number of Zn$_{r,l}$ Classes for Each Zinc Ion Found in the Initial Data Set[a]

	l			
r	0	1	2	3
0	14	3	1	0
1	56	1	0	0
2	72	2	5	0
3	57	**244**	**43**	**3**
4	214	**12**	**15**	0
5	8	0	0	0

[a]Complexes with at least one the classes in bold were selected for the final data set.

Unoccupied coordination geometry locations are usually engaged by a of the resides coordinating the metal modulates association energies of the ligands.[35] To our knowledge, no experimental structures have been reported where ligands interacted with zinc through the mediation of a water molecule. In fact, dockings performed with the hydrated ligand protocol[30] consistently predicted the displacement of the waters. Therefore, zinc was considered always desolvated, while desolvation energy was implicitly accounted for during the calibration of the force field.

Complexes where ligands were not directly involved in zinc coordination (i.e., $l = 0$) were discarded. This included also Zn$_{4,0}$ cases, where zinc plays a structural role helping protein folding,[14,15] coordinating four cysteine side chains. Some of the Zn$_{4,0}$ cases were misclassified as Zn$_{2,0}$ because the zinc ion bridges two monomeric units that were split during the analysis process.

Geometries where receptor atoms were not involved, or only partially involved in zinc interactions ($0 \le r \le 2$) were also discarded upon visual inspection. In particular, the Zn$_{0,0}$ class contains complexes where Zn is used as an aid in crystallization and has no biological significance, surrounding the protein structures often in large number and at toxic concentrations.[17] Finally, 5 complexes involving serine protease inhibitors from Zn$_{2,2}$ class, were discarded because zinc is known to be recruited transiently as coinhibitor only, and it is not consistently present in the binding site.[36]

This left four coordination classes, Zn$_{3,1}$, Zn$_{3,2}$, Zn$_{3,3}$, Zn$_{4,1}$, and Zn$_{4,2}$, resulting in a calibration set of 292 unique complexes. A summary of ligand properties in the set is shown in Figure 1.

For complexes where the tetrahedral coordination geometry is possible (Zn$_{3,x}$), we analyzed the distribution of the ligand atoms coordinating zinc, using the AutoDock atom types: NA (nitrogen HB acceptor), N (nitrogen non-HB acceptor), OA (oxygen HB acceptor), and SA (sulfur HB acceptor). The ideal zinc tetrahedral geometry was calculated with respect to the averaged position of receptor atoms. The tetrahedral plane was defined as the plane calculated between average coordinating receptor atoms and the zinc atom (Figure 7). QM optimizatons performed on few representative cases confirmed high quality of experimental tetrahedral geometries (see the Supporting Information).

Then, we measured the deviation of ligand atoms from the ideal position in the tetrahedral geometry, defined as the angle between the vector Zn-ligand atom and the tetrahedral plane.

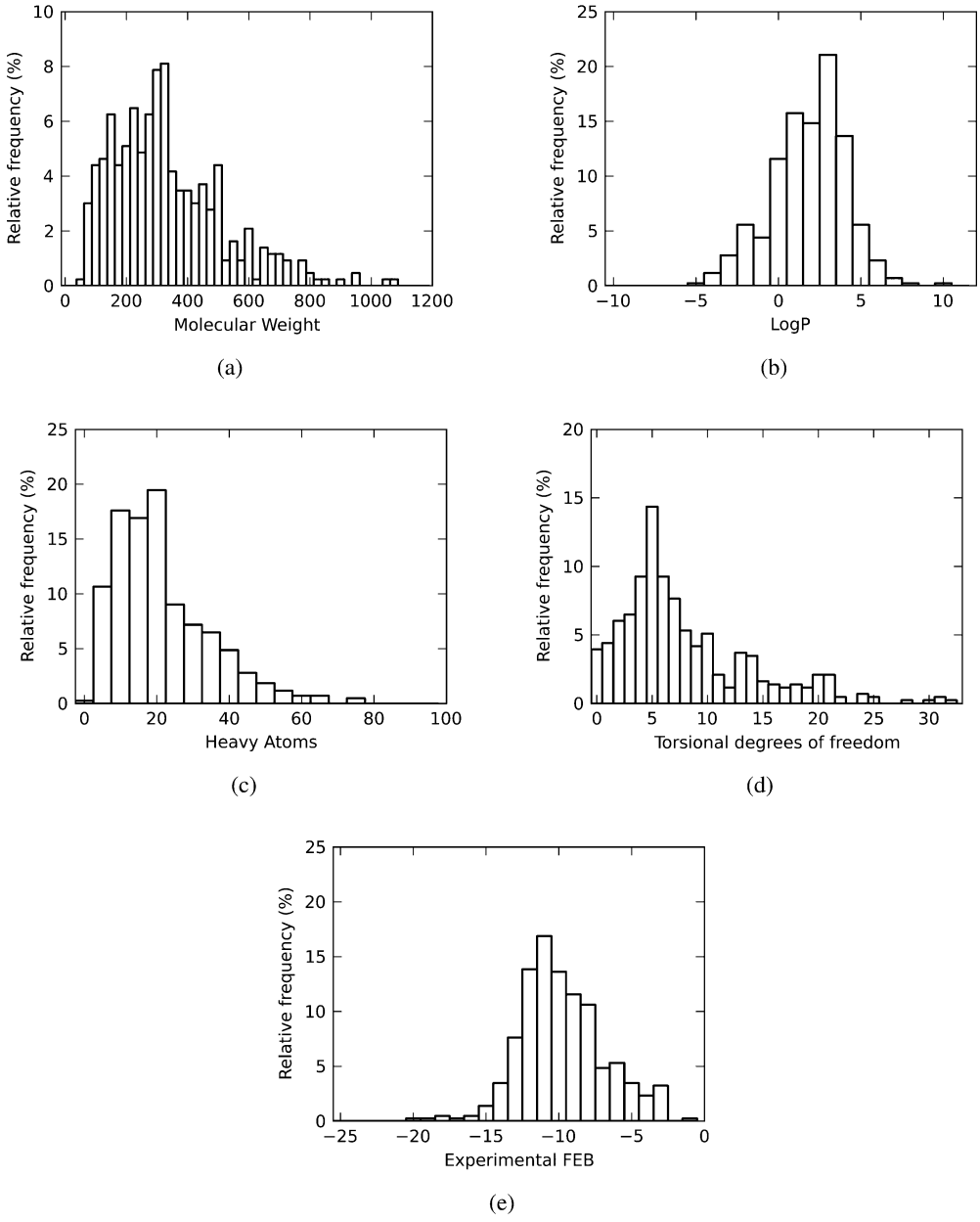


Figure 1. Summary of the distributions of ligand properties in the final data set: molecular weight (*a*), LogP (*b*), number of heavy atoms (*c*), torsional degrees of freedom (*d*), experimental free energy of binding (*e*).

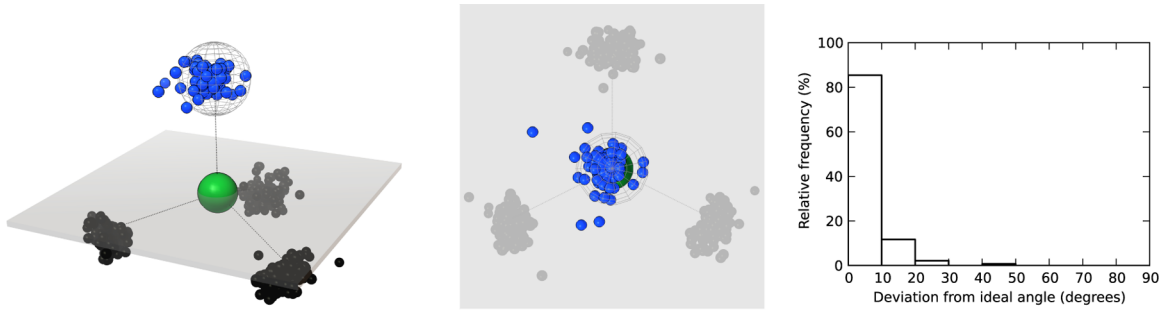


Figure 2. Distribution of 137 NA atom types coordinating zinc: (a) perspective projection; (b) top view; (c) angle histogram. Atoms are shown as spheres: receptor atoms (black), zinc (green), NA atoms (blue). Tetrahedral geometries are colored in gray; tetrahedral plane is shown as semitransparent polygon; pseudoatom location is shown as wireframe sphere.

dx.doi.org/10.1021/ci500209e | J. Chem. Inf. Model. 2014, 54, 2371−2379

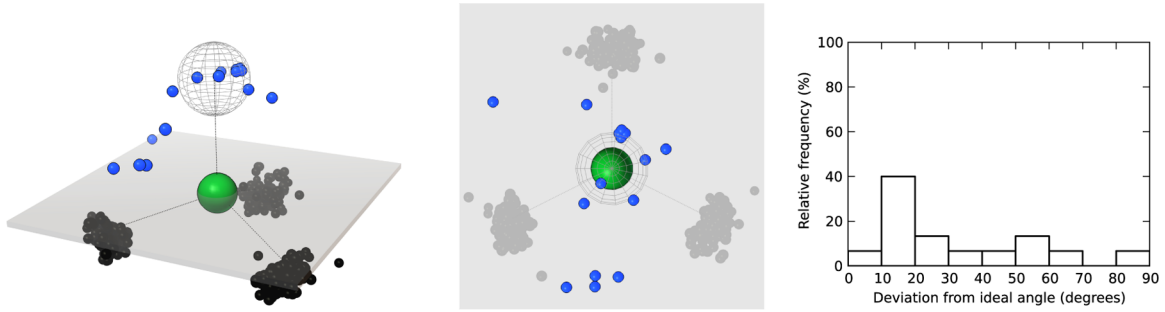


Figure 3. Distribution of 15 N atom types coordinating zinc: (a) perspective projection; (b) top view; (c) angle histogram. Atoms are shown as spheres: receptor atoms (black), zinc (green), N atoms (blue). Tetrahedral geometries are colored in gray; tetrahedral plane is shown as semitransparent polygon; pseudoatom location is shown as wireframe sphere.

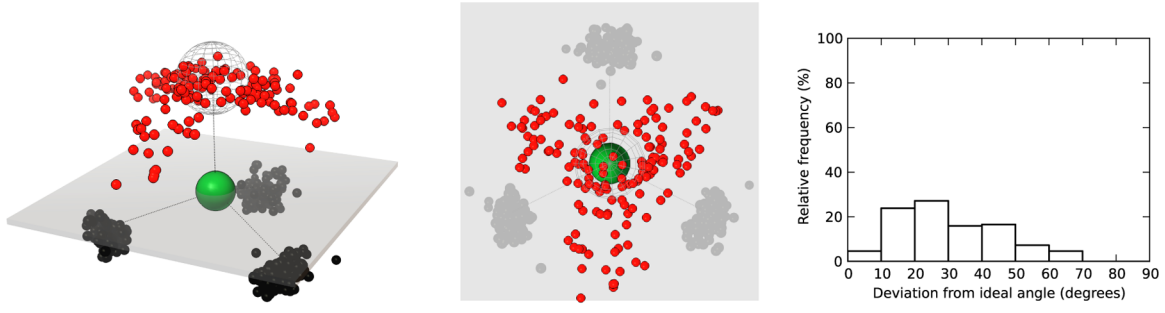


Figure 4. Distribution of 151 OA atom types coordinating zinc: (a) perspective projection; (b) top view; (c) angle histogram. Atoms are shown as spheres: receptor atoms (black), zinc (green), OA atoms (red). Tetrahedral geometries are colored in gray; tetrahedral plane is shown as semitransparent polygon; pseudoatom location is shown as wireframe sphere.

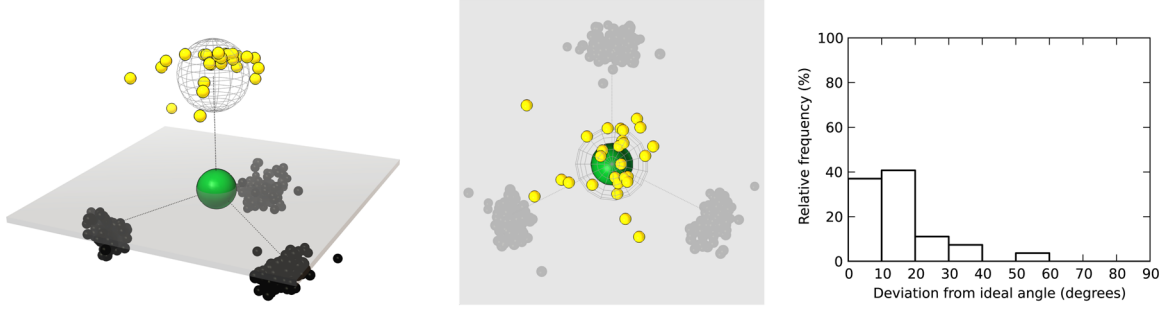


Figure 5. Distribution of 27 SA atom types coordinating zinc: (a) perspective projection; (b) top view; (c) angle histogram. Atoms are shown as spheres: receptor atoms (black), zinc (green), SA atoms (yellow). Tetrahedral geometries are colored in gray; tetrahedral plane is shown as semitransparent polygon; pseudoatom location is shown as wireframe sphere.

In Figures 2−5 are shown the tridimensional scattering coordinating ligand atoms with respect to the tetrahedral zinc (a and b) and their angle deviations (c). More details about the alignment method and analysis are reported in the Supporting Information.

The deviation analysis showed that nitrogen HB acceptor (NA) is consistently found very close to the ideal position (>80% within ≤10°, Figure 2). On the other hand, the placement of nitrogen non-HB acceptor (N) and oxygen (OA) and sulfur (SA) is less well-defined, appearing to be dependent solely on the accessibility of the zinc atom in the receptor (Figures 3−5).

New Force Field. The standard AutoDock force field supports several ligand−metal interactions.[37] Similar to all other pairwise interactions in the force field, the interaction between ligand atoms and metals contained in the receptor is described mainly by van der Waals (ΔH_{vdW}) and Coulomb electrostatic (ΔH_{elec}) terms and, to a smaller degree, by the desolvation term (ΔG_{desolv}). This approach has several limitations. First, the van der Waals equilibrium distances for the atoms involved in zinc coordination are significantly larger than the coordination distances[13,19] (i.e., for nitrogen, the vdW equilibrium distance is 2.49 Å, compared to coordination distance of 2.0 Å). Second, due to the lack of a specialized terms for the metal coordination, directionality is not

dx.doi.org/10.1021/ci500209e | J. Chem. Inf. Model. 2014, 54, 2371−2379

accounted for. Finally, while the electrostatic term is very effective in describing interactions involving partial charges, it makes the energy function highly sensitive to strongly charged groups, such as metals with formal charges. Also, in the Gasteiger[38] charge model used in AutoDock,[37] oxygen atoms are systematically assigned a more negative charge than nitrogen and sulfur, thus resulting in the preferred candidates for chelating positively charged metal. While this approach is accurate enough for magnesium ion interactions, it is not sufficient to properly describe zinc coordination preferences.

From our data set analysis, we found that the coordination of zinc requires a specialized treatment, so we modified the standard AutoDock force field. The standard force field includes the following terms (eq 1):[37]

$$\mathrm{FEB} = W_{\mathrm{vdw}}(\mathrm{vdW}) + W_{\mathrm{hb}}(\mathrm{Hbond}) + W_{\mathrm{elec}}(\mathrm{elec}) + W_{\mathrm{sol}}(\mathrm{desolv}) + W_{\mathrm{tor}}(\mathrm{torsDoF}) \quad (1)$$

where the free energy of binding (FEB) is calculated as a sum of van der Waals (vdW), hydrogen bond (Hbond), Coulomb electrostatic (Elec), desolvation (desolv) and ligand torsional entropy (torsDoF); each term is weighted by a specific value (W_{term}) estimated using a linear regression model.[37] To extend the force field, we first disabled the electrostatic potential for zinc by setting its partial charge to zero. Then, the pairwise interactions of each atom types involved in zinc coordination was defined as a new potential energy term. For N, OA, and SA atom types, spherical potentials $V_{\mathrm{Zn,N}}$, $V_{\mathrm{Zn,OA}}$, and $V_{\mathrm{Zn,SA}}$ were defined to reflect the known coordination distances, by adapting the van der Waals potential in the AutoDock force field (eq 2):

$$V_{ij} = \varepsilon_{ij}\left[\left(\frac{r_{ij}}{r}\right)^{12} - 2\left(\frac{r_{ij}}{r}\right)^{6}\right] \quad (2)$$

The pairwise equilibrium distance r_{ij} between zinc and N, OA, and SA atom types was set to 2.0, 2.1, and 2.25 Å, respectively, and independent ε well-depth values were estimated. Spherical potentials are particularly suitable for accurately reproducing hydroxamate coordination geometries.[39,40]

For the NA type a new directional tetrahedral potential $V_{\mathrm{TZ,NA}}$ was defined and the interaction with zinc was split in two separate components. The repulsive component is mediated by the zinc atom, while the attractive component is mediated by a new pseudoatom TZ that has been added to the standard force field table.[41] The pseudoatom interacts only with NA, therefore no interaction is defined with any other atom type. The pseudoatom is added in the receptor structure for all complexes where the tetrahedral coordination geometry is present, i.e., all $\mathrm{Zn}_{3,x}$ classes, where only three receptor atoms are coordinating zinc. The pseudoatom is placed at the unoccupied vertex of the tetrahedral geometry, located at the optimal coordination distance for nitrogen (r_{ij} = 2.0 Å) (Figure 7a), and an attractive 12−6 potential with a corresponding ε is defined (Figure 7b). Finally, the zinc−hydrogen pairwise interaction was eliminated to prevent clashes that would interfere with the proper interaction between groups like sulfonamide $-NH_2$, or hydroxyl, with zinc. This allows ligands to establish the proper coordination interaction independent of the orientation of the hydrogen with respect to the heavy atom.

Therefore, the following potential was added to eq 1:

$$\mathrm{ZincCoord} = V_{\mathrm{TZ,NA}} + V_{\mathrm{Zn,N}} + V_{\mathrm{Zn,OA}} + V_{\mathrm{Zn,SA}} \quad (3)$$

and the FEB becomes the linear combination of the five standard AutoDock terms plus the new zinc coordination pairwise potential.

All modifications to the AutoDock force field were made by adapting the force field table and parameter files, without source code modifications. The details of the implementation are described in the Supporting Information.

The ε values for eq 3 were then calibrated independently from each other and from the other terms in eq 1.

Calibration Protocol. The new force field was calibrated with an iterative least-squares scheme. Initial attempts to calibrate combined terms from eqs 1 and 3 led to performance degradation in nonzinc complexes. Optimization of different term combination were tried, and best results were obtained by optimizing only terms in eq 3, while keeping the standard terms (eq 1) unmodified.

The calibration protocol consisted of the following steps: (a) crystallographic structures of the ligands were minimized with the current version of the force field, using Solis-Wet local search implemented in AutoDock;[37] (b) unweighted terms were calculated from minimized structures; (c) a regression model was built; (d) weights from the new regression model were used in the next minimization step. The protocol iterated through steps c and d five times to achieve convergence; stable weight values were achieved after the first two iterations.

Initial calibration results and cross-validation tests showed that no statistical significance could be achieved for the $V_{\mathrm{Zn,N}}$ term. This is likely due to insufficient experimental data for the N atom type. Therefore, standard force field term for this interaction (i.e., van der Waals) was restored, while keeping the correct equilibrium radius (2.0 Å) identified in the analysis. Then the calibration was repeated omitting the $V_{\mathrm{Zn,N}}$ term from eq 3. Final force field weights were selected from the last iteration, with a residual standard error was 2.804 kcal/mol. Final coefficients and extended analysis of the iterative calibration are described in the Supporting Information.

RESULTS AND DISCUSSION

Predictive capabilities of the regression model were assessed with 5-fold cross-validation. The data set was divided in five bins containing an approximately uniform distribution of ligand atom types coordinating zinc, then redocking calculations were performed. Cross-validation docking results are summarized in Table 2. Details on docking preparation and RMSD calculations are available in the Supporting Information. Reproducing proper metal-coordination geometries and accurate energy estimations are notoriously difficult, especially for zinc.[42] The performance of $\mathrm{AutoDock4}_{\mathrm{Zn}}$ was evaluated accordingly to three different criteria: FEB estimation error, ligand pose RMSD calculated on all heavy atoms, and deviation from ideal zinc coordination geometry. Overall, the new $\mathrm{AutoDock4}_{\mathrm{Zn}}$ force field performed significantly better than both standard AutoDock and Vina force fields. These force fields provide roughly the same prediction errors in FEB

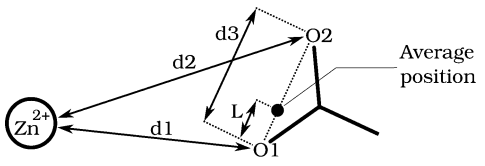
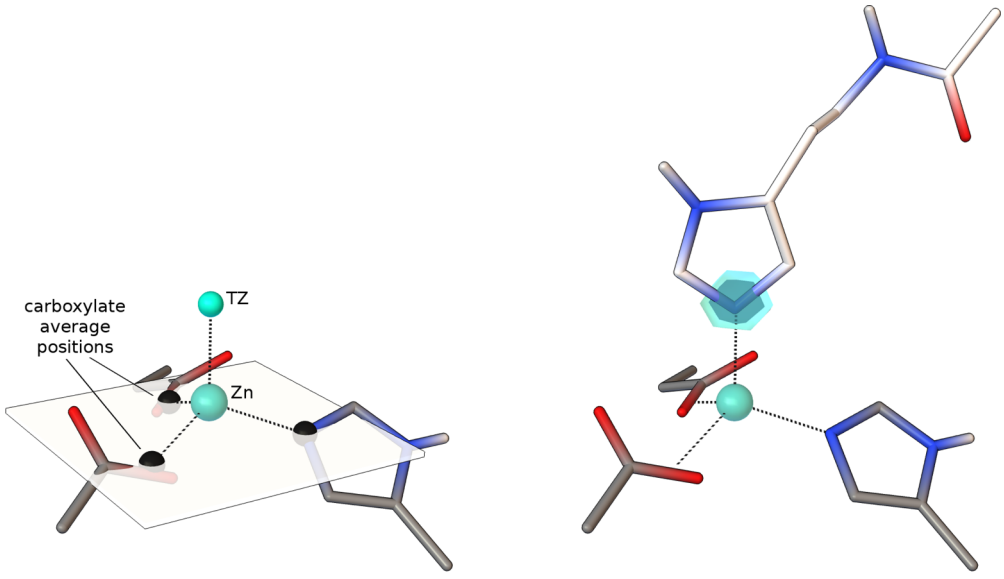


Figure 6. Definition of carboxyl group average atom. Details about the methods are reported in the Supporting Information.

dx.doi.org/10.1021/ci500209e | J. Chem. Inf. Model. 2014, 54, 2371−2379



(a) (b)

Figure 7. Tetrahedral zinc geometry. (a) Ligand and receptor atoms are shown as sticks colored by atom type. The tetrahedral plane defined by three receptor atoms (black spheres) is determined. The TZ pseudoatom is located at unoccupied corner of the ideal tetrahedral geometry. Coordination geometry is calculated on weighted average oxygen positions from carboxylic side chains. (b) The potential for atom type NA (nitrogen acceptor) is shown as iso-contour surfaces (cyan).

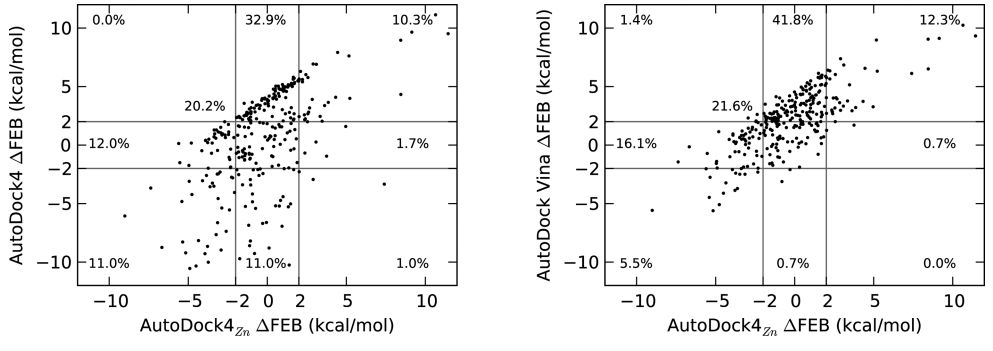
Table 2. Cross-validation of Docking Performances and FEB Estimation Accuracy

	FEB error (kcal/mol)			RMSD (Å)		$RMSD_{Zn}$ (Å)	
	<1.0	<2.0	<3.0	<2.0	<2.5	<1.0	<1.5
$AutoDock4_{Zn}$	32%	64%	81%	45%	51%	75%	80%
AutoDock4	18%	34%	53%	36%	42%	33%	46%
Vina	20%	38%	64%	45%	52%	37%	52%

estimations, while $AutoDock4_{Zn}$ consistently improved success rate (+50% with <1.0 kcal/mol) (Figure 8). The new force field also improves RMSD accuracy over the standard AutoDock force field in pose prediction accuracy (RMSD), producing results comparable with Vina (Figure 9). Not surprisingly, a remarkable improvement was achieved in reproducing the proper zinc-coordination geometry ($RMSD_{Zn}$), where $AutoDock4_{Zn}$ outperforms the two other force fields by a large amount (+127% success rate <1 Å, Figure 10).

The use of specific potentials for describing the interaction is the main factor responsible for such an improvement, as shown by redocking experiments.

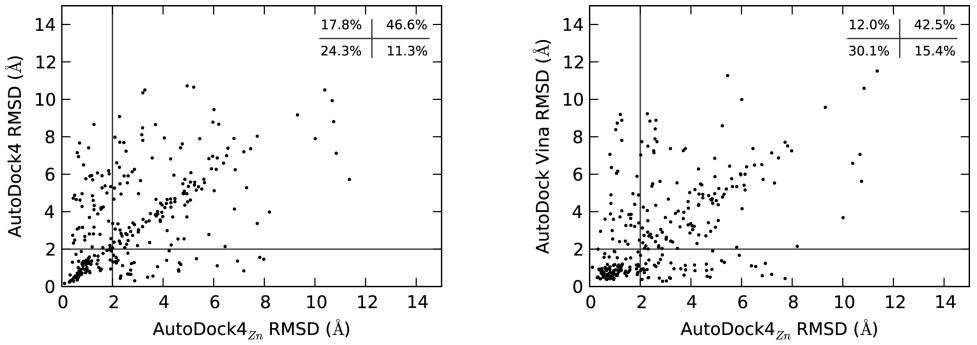
Examples. In some cases, where sulfur is directly involved in coordinating zinc, neither AutoDock4 nor Vina force fields were able to establish the proper interactions between the ligand and the zinc ion. A key example of this is provided by reredocking results of a potent inhibitor of neutral endopeptidase (NEP)[43] in the crystallographic complex with the PDB ID 1r1j (Figure 11). Results are summarized in Table 3. Both AutoDock4 and Vina predicted zinc to be coordinated by the carboxylate group, resulting in a misalignment of the ligand with respect to the receptor. The $AutoDock4_{Zn}$, on the other hand, predicted the proper coordination by sulfanyl and carbonyl groups and provided more accurate FEB estimation. Similar results were found when redocking a potent aryl sulfonamide TACE inhibitor in the crystallographic complex with PDB ID 1oi0 (Figure 12). The increase in docking pose prediction and accuracy in the coordination geometry indentification resulted in a more precise FEB estimation



(a) (b)

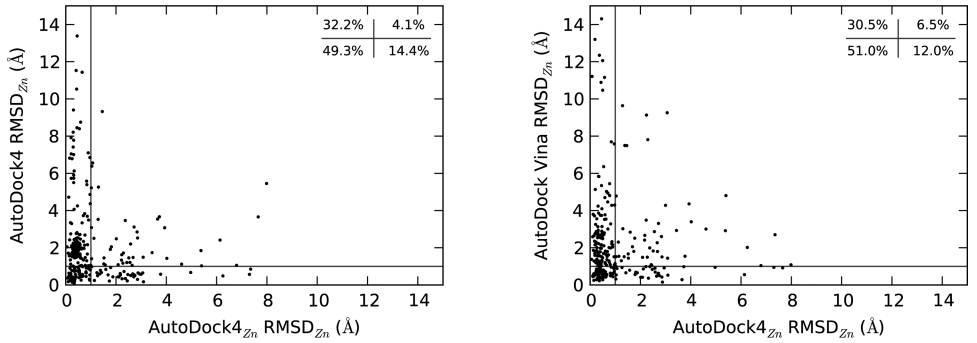
Figure 8. Comparison of FEB prediction errors of the new force field with (a) standard AutoDock4 force field and (b) AutoDock Vina.

dx.doi.org/10.1021/ci500209e | J. Chem. Inf. Model. 2014, 54, 2371−2379



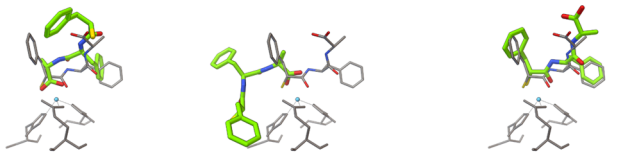
(a) (b)

Figure 9. Comparison of RMSD error of the new force field with (a) standard AutoDock4 force field and (b) AutoDock Vina.



(a) (b)

Figure 10. Comparison of RMSD error on zinc coordination geometry of the new force field with (a) standard AutoDock4 force field and (b) AutoDock Vina.



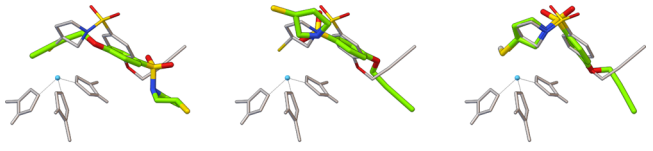
(a) AutoDock4 (4.85 Å) (b) AutoDock Vina (9.19 Å) (c) AutoDock4$_{Zn}$ (1.21 Å)

Figure 11. Comparinson of redocking accuracy with 1r1j using (a) standard AutoDock4, (b) AutoDock Vina, and (c) AutoDock4$_{Zn}$ force fields (RMSD are shown in parentheses). Zinc-coordinating residues and experimental ligand pose are shown as thin gray sticks; docked poses are shown as green thick sticks. Hydrogens are not shown for sake of clarity.

Table 3. Docking Performances and FEB Estimation Accuracy on NEP (1r1j)

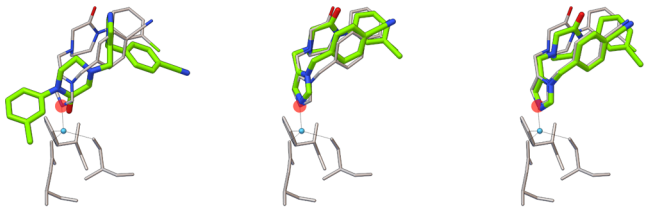
	FEB error (kcal/mol)	RMSD (Å)	RMSD$_{Zn}$ (Å)
AutoDock4$_{Zn}$	+0.74	1.21	1.04
AutoDock4	+1.76	4.85	6.38
Vina	+4.68	9.19	4.78

(Table 4). It must also be noted also that no performance degredation was found when docking compounds without ligand-zinc interactions using the new force field. Improvements provided by AutoDock4$_{Zn}$ makes it suitable for virtual screening campaigns involving zinc.



(a) AutoDock4 (7.98 Å) (b) AutoDock Vina (2.34 Å) (c) AutoDock4$_{Zn}$ (2.10 Å)

Figure 12. Comparison of redocking accuracy of 2oi0 using (a) standard AutoDock4, (b) AutoDock Vina, and (c) AutoDock4$_{Zn}$ force fields (RMSD are shown in parentheses). Zinc-coordinating residues and experimental ligand pose are shown as thin gray sticks; zinc is cyan; docked poses are shown as green thick sticks. Hydrogens are not shown for sake of clarity.



(a) AutoDock4 (7.62 Å) (b) AutoDock Vina (0.5 Å) (c) AutoDock4$_{Zn}$ (0.68 Å)

Figure 13. Comparison of redocking accuracy of 1s63 using (a) standard AutoDock4, (b) AutoDock Vina, and (c) AutoDock4$_{Zn}$ force fields (RMSD are shown in parentheses). Zinc-coordinating residues and experimental ligand pose are shown as thin gray sticks; zinc is cyan; docked poses are shown as green thick sticks; the location and the optimal radius of the TZ pseudoatom potential is shown as semitransparent sphere (red). Hydrogens are not shown for sake of clarity.

dx.doi.org/10.1021/ci500209e | J. Chem. Inf. Model. 2014, 54, 2371−2379

Table 4. Docking Performances and FEB Estimation Accuracy on TACE (2oi0)

	FEB error (kcal/mol)	RMSD (Å)	$RMSD_{Zn}$ (Å)
$AutoDock4_{Zn}$	+0.33	2.10	0.65
AutoDock4	+1.56	7.98	11.43
Vina	+2.50	2.34	5.03

Table 5. Docking Performances and FEB Estimation Accuracy on Farnesyltransferase (1s63)

	FEB error (kcal/mol)	RMSD (Å)	$RMSD_{Zn}$ (Å)
$AutoDock4_{Zn}$	+0.07	0.68	0.34
AutoDock4	+3.52	7.62	7.78
Vina	+2.93	0.50	0.54

CONCLUSIONS

We extended the standard AutoDock force field to include a specialized potential describing interactions of zinc-coordinating ligands. The potential has a description for both energetic and geometric components of the interaction

The new force field, named $AutoDock4_{Zn}$, was calibrated on 292 complexes containing zinc using an iterative linear regression model. Redocking experiments showed that the force field provides a considerable improvement in performance when compared to both standard AutoDock4 and Vina force fields. Improvements are particularly relevant in the accuracy of FEB estimations, as well as in reproducing the proper coordination geometry. In fact, $AutoDock4_{Zn}$ provides a significant advantage when docking zinc-coordinating ligands, as shown in the examples described.

Due to the fully configurable nature of AutoDock4, the new potential was implemented by modifying the standard force field tables and few Python helper scripts available at http://autodock.scripps.edu/.

Moreover, the potential itself does not add any overhead to the docking calculation. There is no increase in the search complexity nor in the computational power requirement, therefore docking speed is completely unaffected. Also, performance on dockings not involving zinc-coordination are unchanged. Therefore, accuracy increase and lack of computational overhead make the $AutoDock4_{Zn}$ suitable for virtual screening campaigns, particularly when coordination of zinc is important.

ASSOCIATED CONTENT

Supporting Information

Procedure to determine the carboxylate average position; definition of TZ pseudoatom potential; data set and linear regression analysis; new force field implementation; docking preparation and analysis. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: olson@scripps.edu.

Author Contributions

¶D.S.-M. and S.F. contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

All figures were generated using Python Molecular Viewer (PMV) version 1.5.6.[44] The authors thank the Molecular Graphics Laboratory members for insightful discussions and technical help. D.S.-M. is supported by grant SFRH/BD/84922/2012 from Fundação para a Ciência e Tecnologia (FCT) for financial support, and M.J.R. thanks FCT for project grant EXCL/QEQ-COM/0394/2012. A.J.O. is supported by grant R01-GM069832-01 from the National Institutes of Health. This is manuscript 27023 from The Scripps Research Institute.

REFERENCES

(1) Vallee, B. L.; Auld, D. S. Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry* **1990**, *29*, 5647−5659.

(2) Laity, J. H.; Lee, B. M.; Wright, P. E. Zinc finger proteins: new insights into structural and functional diversity. *Curr. Opin. Struct. Biol.* **2001**, *11*, 39−46.

(3) Sousa, S. F.; Lopes, A. B.; Fernandes, P. A.; Ramos, M. J. The Zinc proteome: a tale of stability and functionality. *Dalton Trans.* **2009**, 7946−7956.

(4) Nawarskas, J.; Rajan, V.; Frishman, W. Vasopeptidase inhibitors, neutral endopeptidase inhibitors, and dual inhibitors of angiotensin-converting enzyme and neutral endopeptidase. *Heart Dis.* **2001**, *3*, 378−385.

(5) Pang, Y.-P. Novel zinc protein molecular dynamics simulations: Steps toward antiangiogenesis for cancer treatment. *Molec. Model. Annual* **1999**, *5*, 196−202.

(6) Reid, T. S.; Long, S. B.; Beese, L. S. Crystallographic analysis reveals that anticancer clinical candidate L-778,123 inhibits protein farnesyltransferase and geranylgeranyltransferase-I by different binding modes. *Biochemistry* **2004**, *43*, 9000−9008.

(7) Docquier, J.-D.; Benvenuti, M.; Calderone, V.; Stoczko, M.; Menciassi, N.; Rossolini, G. M.; Mangani, S. High-resolution crystal structure of the subclass B3 metallo-β-lactamase BJP-1: rational basis for substrate specificity and interaction with sulfonamides. *Antimicrob. Agents Chemother.* **2010**, *54*, 4343−4351.

(8) Schilling, S.; et al. Glutaminyl cyclase inhibition attenuates pyroglutamate Aβ and Alzheimer's disease−like pathology. *Nat. Med.* **2008**, *14*, 1106−1111.

(9) Jalkute, C. B.; Barage, S. H.; Dhanavade, M. J.; Sonawane, K. D. Molecular dynamics simulation and molecular docking studies of angiotensin converting enzyme with inhibitor lisinopril and amyloid beta peptide. *Protein J.* **2013**, *32*, 356−364.

(10) Jacobsen, F. E.; Lewis, J. A.; Cohen, S. M. The design of inhibitors for medicinally relevant metalloproteins. *ChemMedChem.* **2007**, *2*, 152−171.

(11) Stote, R. H.; Karplus, M. Zinc binding in proteins and solution: a simple but accurate nonbonded representation. *Proteins: Struct., Funct., Bioinf.* **1995**, *23*, 12−31.

(12) Wu, R.; Lu, Z.; Cao, Z.; Zhang, Y. A transferable nonbonded pairwise force field to model zinc interactions in metalloproteins. *J. Chem. Theory Comput.* **2011**, *7*, 433−443.

(13) Hu, X.; Shelver, W. H. Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. *J. Molec. Graph. Modell.* **2003**, *22*, 115−126.

(14) Andreini, C.; Bertini, I. A bioinformatics view of zinc enzymes. *J. Inorg. Biochem.* **2012**, *111*, 150−156.

(15) Patel, K.; Kumar, A.; Durani, S. Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures. *Biochim. Biophy. Acta* **2007**, *1774*, 1247−1253.

(16) Alberts, I. L.; Nadassy, K.; Wodak, S. J. Analysis of zinc binding sites in protein crystal structures. *Protein Sci.* **1998**, *7*, 1700−1716.

(17) Laitaoja, M.; Valjakka, J.; Jänis, J. Zinc Coordination Spheres in Protein Structures. *Inorg. Chem.* **2013**, *52*, 10983−10991.

(18) Håkansson, K.; Liljas, A. The structure of a complex between carbonic anhydrase II and a new inhibitor, trifluoromethane sulphonamide. *FEBS Lett.* **1994**, *350*, 319−322.
(19) Pottel, J.; Therrien, E.; Gleason, J. L.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 6. Development and Application to the Docking of HDACs and other Zinc Metalloenzymes Inhibitors. *J. Chem. Inf. Model.* **2014**, *54*, 254−265.
(20) Zhu, T.; Xiao, X.; Ji, C.; Zhang, J. Z. A New Quantum Calibrated Force Field for Zinc−Protein Complex. *J. Chem. Theory Comput.* **2013**, *9*, 1788−1798.
(21) Sakharov, D. V.; Lim, C. Zn protein simulations including charge transfer and local polarization effects. *J. Am. Chem. Soc.* **2005**, *127*, 4921−4929.
(22) 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*, 1314−1324.
(23) Peters, M. B.; Yang, Y.; Wang, B.; Füsti-Molnár, L.; Weaver, M. N.; Merz, K. M., Jr Structural survey of zinc-containing proteins and development of the zinc AMBER force field (ZAFF). *J. Chem. Theory Comput.* **2010**, *6*, 2935−2947.
(24) Seebeck, B.; Reulecke, I.; Kämper, A.; Rarey, M. Modeling of metal interaction geometries for protein−ligand docking. *Proteins: Struct., Funct., Bioinf.* **2008**, *71*, 1237−1254.
(25) Cincilla, G.; Vidal, D.; Pons, M. An improved scoring function for suboptimal polar ligand complexes. *J. Comput.-Aided Molec. Des.* **2009**, *23*, 143−152.
(26) Buzko, O. V.; Bishop, A. C.; Shokat, K. M. Modified AutoDock for accurate docking of protein kinase inhibitors. *J. Comput.-Aided Molec. Des.* **2002**, *16*, 113−127.
(27) Österberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 34−40.
(28) Cosconati, S.; Marinelli, L.; Di Leva, F. S.; La Pietra, V.; De Simone, A.; Mancini, F.; Andrisano, V.; Novellino, E.; Goodsell, D. S.; Olson, A. J. Protein flexibility in virtual screening: the BACE-1 case study. *J. Chem. Inf. Model.* **2012**, *52*, 2697−2704.
(29) Forli, S.; Botta, M. Lennard-Jones potential and dummy atom settings to overcome the AutoDock limitation in treating flexible ring systems. *J. Chem. Inf. Model.* **2007**, *47*, 1481−1492.
(30) Forli, S.; Olson, A. J. A force field with discrete displaceable waters and desolvation entropy for hydrated ligand docking. *J. Med. Chem.* **2012**, *55*, 623−638.
(31) Hu, L.; Benson, M. L.; Smith, R. D.; Lerner, M. G.; Carlson, H. A. Binding MOAD (Mother Of All Databases). *Proteins* **2005**, *60*, 333−340.
(32) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F., Jr; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *Arch. Biochem. Biophys.* **1978**, *185*, 584−591.
(33) Harding, M. M. Geometry of metal-ligand interactions in proteins. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2001**, *57*, 401−411.
(34) Ryde, U. Carboxylate binding modes in zinc proteins: a theoretical study. *Biophys. J.* **1999**, *77*, 2777−2787.
(35) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. The carboxylate shift in zinc enzymes: A computational study. *J. Am. Chem. Soc.* **2007**, *129*, 1378−1385.
(36) Katz, B. A.; Luong, C. Recruiting Zn^{2+} to Mediate Potent, Specific Inhibition of Serine Proteases. *J. Molec. Biol.* **1999**, *292*, 669−684.
(37) Huey, R.; Morris, G. M.; Olson, A. J.; Goodsell, D. S. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145−1152.
(38) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity — a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(39) Wu, R.; Lu, Z.; Cao, Z.; Zhang, Y. Zinc chelation with hydroxamate in histone deacetylases modulated by water access to the linker binding channel. *J. Am. Chem. Soc.* **2011**, *133*, 6110−6113.
(40) Chiu, Y.-H.; Gabriel, G. J.; Canary, J. W. Ternary ligand-zinc-hydroxamate complexes. *Inorg. Chem.* **2005**, *44*, 40−44.
(41) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J.Comput. Chem.* **2009**, *30*, 2785−2791.
(42) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Struct., Funct., Bioinf.* **2004**, *57*, 225−242.
(43) Oefner, C.; Roques, B. P.; Fournie-Zaluski, M.-C.; Dale, G. E. Structural analysis of neprilysin with various specific and potent inhibitors. *Acta Crystallogr. Sect. D: Biol. Crystallogr.* **2004**, *60*, 392−396.
(44) Sanner, M. F. Python: a programming language for software integration and development. *J. Mol. Graph Model.* **1999**, *17*, 57−61.
(45) Svensson, M.; Humbel, S.; Froese, R. D.; Matsubara, T.; Sieber, S.; Morokuma, K. ONIOM: A multilayered integrated MO+ MM method for geometry optimizations and single point energy predictions. A test for Diels-Alder reactions and Pt (P (t-Bu) 3) 2+ H2 oxidative addition. *J. Phys. Chem.* **1996**, *100*, 19357−19363.
(46) Dapprich, S.; Komáromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. *J. Molec. Struct.: THEOCHEM* **1999**, *461*, 1−21.
(47) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G* basis set for atoms K through Zn. *J. Chem. Phys.* **1998**, *109*, 1223−1229.
(48) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712−725.
(49) Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(50) O' Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 1−14.
(51) Huey, R.; Morris, G. M. *Using AutoDock 4 with AutoDockTools: A Tutorial*; Scripps Research Institute, 2008.
(52) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Molec. Biol.* **1999**, *285*, 1735−1747.