

# Quantum Mechanical Calculations for Binding Sites of Metalloproteins

Art E. Cho<sup>1</sup>

<sup>1</sup>Department of Biotechnology and Bioinformatics, Korea University, Jochiwon, Chungnam 339-700, Korea  
Correspondence and requests for materials should be addressed to Art E. Cho (artcho@korea.ac.kr)

Accepted 30 April 2008

## Abstract

Conventional docking methods which assume fixed charge model from force field parameters fail to predict right binding modes in a few groups of protein targets including metalloproteins. A new novel docking method with combined quantum mechanics/molecular mechanics (QM/MM) method has been applied to docking as a variable charge model and shown to exhibit improvement on the docking accuracy over fixed charge based methods. However, it has also been shown that there are a number of examples for which adoption of variable charge model fails to reproduce the native binding modes. The original implementation of QM/MM docking treated only ligands as quantum regions, which leaves metal ions present in binding sites with non-optimized charges. To address this problem, we extend the QM/MM docking method so that metal ions are included in quantum region, along with ligand atoms. This extension effectively rescales metal ion charge, but the results of docking experiment for binding mode prediction are unsatisfactory. Further analysis suggests that charge on metal ions transfers more greatly to surrounding protein atoms rather than ligand atoms, which explains the apparent over-correction of metal ion charge.

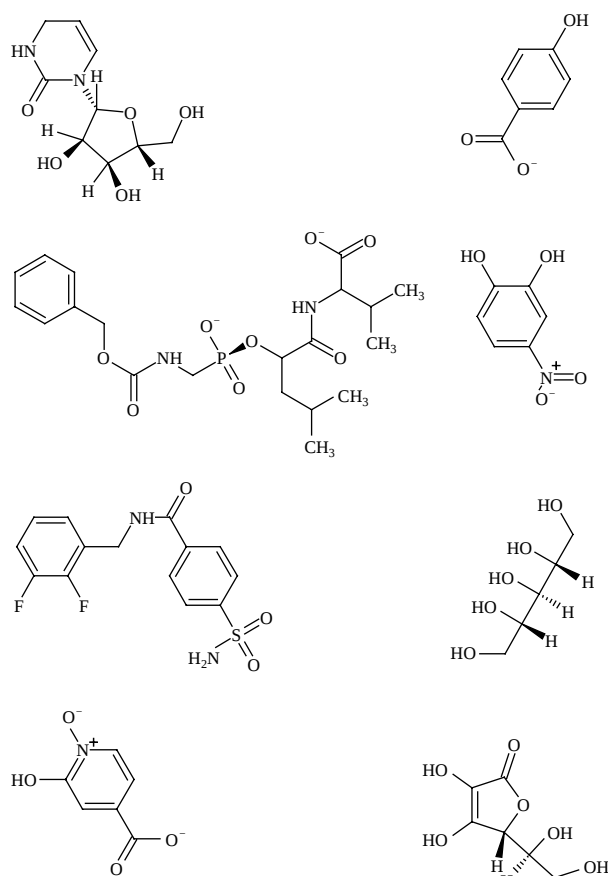
**Keywords:** QM/MM, DFT, Metalloprotein, Docking, Binding site

## Introduction

In modern day drug discovery programs, protein docking is regarded as a standard procedure, which is indispensable for reduction of development time and cost<sup>1-5</sup>. Especially in the last decade, a number of docking algorithms/programs have emerged as a rather

reliable tool for prediction of binders to protein targets and their binding modes. Recent survey of widely-used programs shows that these programs can reproduce a set of experimentally determined structures within 2Å of RMSD up to 80% of the time. To achieve such accuracy, contemporary docking programs exploit various empirical scoring schemes as well as energy based ones. For the electrostatic energy portion of scoring, most of current docking methods utilize force field based fixed electric charges for both protein and ligand atoms<sup>2,6-9</sup>. In an earlier work, we tested the variable charge model for docking by applying the combined quantum mechanics/molecular mechanics (QM/MM) methods to correct the charges on ligand atoms in the binding sites. This work has been proven to be quite successful in predicting the correct binding modes of ligand-protein complexes<sup>10</sup>. It was shown that the correction of the atomic charges on ligand atoms according to the polarization of the binding site environment can improve docking accuracy. With a more extensive test on a larger set of samples, however, it was found that there still remain some cases in which even the new method, which we will call QM/MM docking henceforth, fails to improve the results of conventional docking practice. An analysis to understand the failure of QM/MM docking revealed that there are certain classes of proteins for which it failed more often than others in redocking experiment. One of them was a group of proteins which contain metal ions in the binding sites, namely, metalloprotein. Careful analysis of these specific cases led us to conclude that charges on the metal ions determined by the force field parameter are just not appropriate for docking<sup>11</sup>.

Metalloproteins play important roles in many different biological processes<sup>12</sup>. Docking studies involving metalloproteins are notoriously difficult since the ligand interactions with transition metals can be treated appropriately only at the quantum mechanical level whereas the current docking methods rarely adopt quantum mechanical calculations<sup>13-15</sup>. Attempts have been made to improve docking accuracy on metalloproteins, especially zinc complexes, by deviating from force field based atomic charges<sup>16,17</sup>. The authors of these works emphasize on practicality of the methods and therefore attempt to improve the docking accuracy and binding affinity prediction by reparametrization of the metal ion force field. In another work, Stern-



**Figure 1.** Structures of ligands used for test. From top, in PDB ID: 1CTT, 6TMN, 1G52, 3PCJ, 2BUR, 1NO3, 2XIM, and 1XID.

berg *et al.* used fluctuating atomic charge model of force field, which is parametrized by semi-empirical quantum chemical method to study zinc complexes<sup>18</sup>.

Armed with the knowledge obtained from the analysis of previous results, we test the idea of docking of metalloproteins with rescaled metal ion charge. The idea is a direct extension of the previous algorithm: to include metal ions inside the binding pocket in addition to ligand atoms as quantum region for QM/MM calculations. Inclusion of metal ions in the quantum region will certainly alter the charge on them, and we proceed to docking with the altered charge to see if such alteration benefits the final docking results.

In this article we describe the following work. We start out by slightly modifying the methodology of the original QM/MM docking. Our new calculations simply add metal ions in the quantum region. The method is tested on 8 metalloprotein complex structures by redocking the native ligands. For this test set

**Table 1.** RMSD in Å of top scoring poses from native poses for Glide SP 4.0, QM/MM docking (QM Dock), and metal-included-QM/MM docking (M-QM Dock)

| PDB ID | Glide SP 4.0 |      | QM Dock |      | M-QM Dock |      |
|--------|--------------|------|---------|------|-----------|------|
|        | Ecvdw        | RMSD | Ecvdw   | RMSD | Ecvdw     | RMSD |
| 1ctt   | -33.5        | 4.85 | -50.9   | 0.59 | -53.7     | 0.74 |
| 6tmn   | -48.9        | 8.17 | -89.7   | 0.35 | -65.0     | 2.18 |
| 1g52   | -44.8        | 4.54 | -48.9   | 3.79 | -56.6     | 5.28 |
| 3pcj   | -48.6        | 0.33 | -58.6   | 0.35 | -39.6     | 0.31 |
| 2bur   | -32.2        | 4.14 | -31.4   | 4.05 | -22.6     | 4.09 |
| 1no3   | -23.3        | 3.89 | -26.6   | 3.54 | -27.1     | 0.27 |
| 2xim   | -34.9        | 4.25 | -36.5   | 4.39 | -25.8     | 2.45 |
| 1xid   | -26.0        | 4.07 | -32.7   | 4.48 | -30.3     | 2.86 |

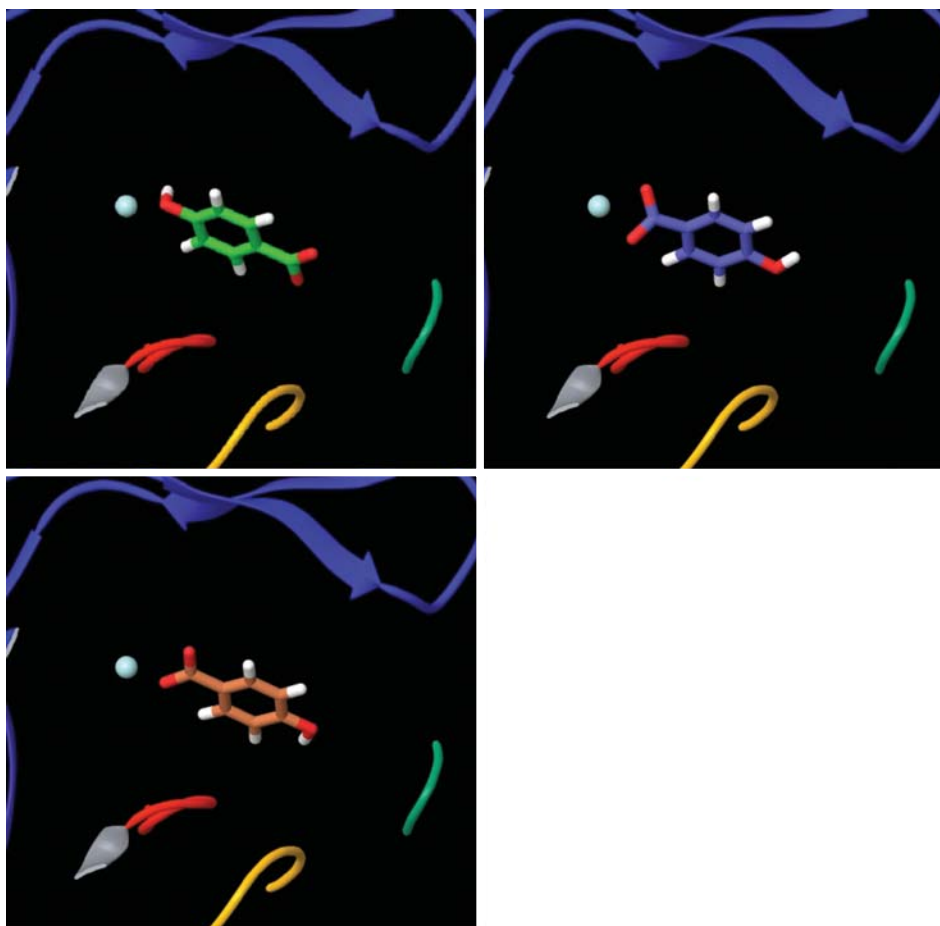
■ Zn, ■ Fe, ■ Mg, ■ Mn

of proteins, Glide 4.0 failed 88% of times, and previous QM/MM docking algorithm failed 63% of times. The new extension brings some improvements but there still exist discrepancies. We analyze the results and attempt to explain the reason for failure. Finally, we make suggestions for future directions in the conclusion.

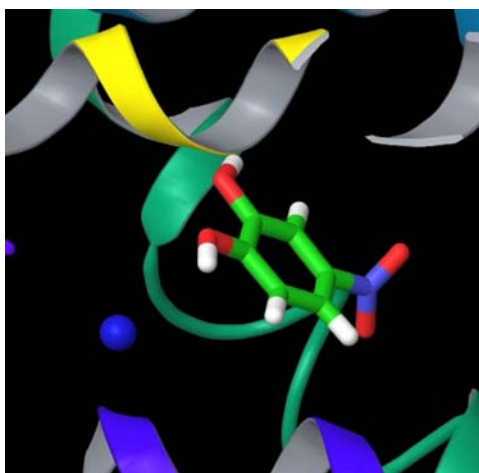
## Results and Discussion

To test our method, we select 8 metalloproteins, which contain Zn, Fe, Mg, and Mn. For this set of complex structures, Glide 4.0 SP mode failed to predict the binding mode within 2.0Å RMSD of the native pose as the top scoring pose in 7 of them. All of the examples were downloaded from protein data bank (PDB) and prepared according to the procedure described in the methods section. The ligands for these complex structures are depicted in Figure 1. Table 1 shows the result of our test. Metal-included QM/MM docking protocol (M-QM Dock) produced moderately improved poses in 2 cases (2xim and 1xid) and greatly improved one in just one case (1no3). In another case (1ctt), the new protocol regenerated already-improved results of original QM/MM docking protocol (QM Dock in Table 1) and maintained the good results of Glide SP and QM/MM docking in the case of 3pcj. Overall, the improvement over original QM/MM docking was minimal. Rescaling of metal ion charge seems to have done little to improve on docking with ligand atom charge modified with quantum level calculations in metalloproteins. To understand the apparent contradiction to our initial expectation, we analyzed the following cases.

We first turned our attention to the case of protocatechuate 3,4-dioxygenase with PDB id 2BUR. As in Figure 2, both Glide 4.0 SP and QM Dock gave as



**Figure 2.** Redocking of 2BUR. Crystallographic structure (top left), Glide SP prediction (top right), and QM Dock prediction (bottom).



**Figure 3.** Crystallographic structure of 1NO3.

the top scoring pose one that has flipped orientation: the carboxylate group pointing to iron center as opposed to hydroxyl group pointing to the same direction as in the crystallographic complex structure. The

reason why Glide SP and QM Dock both predicted a wrong orientation is that the atomic charge of iron was set to  $+3e$  from force field parameter and therefore the carboxylate group, with its strong negative charge is attracted towards the iron center. With the inclusion of iron atom in the quantum region for QM/MM calculation, the atomic charge on iron was reduced significantly to  $+1.98e$ . However, at the same time, the charge value on oxygen atom of hydroxyl group turned more positive. This means that, with our setting of quantum region in M-QM Dock calculation, some of the positive charge was transferred over to the ligand diluting negativity on nearby oxygen atom. With this new set of atomic charge values, we were still led to prediction of a wrong orientation in which the carboxylate group pointing towards the iron center. On the other hand, the case of soybean lipoxygenase-3 (PDB id: 1NO3) tips us with a hint of how a further modification should be done. As in Figure 2, the ligand in this example shows similar configuration as in the previous example: the oxygen atoms in nitric dioxide group are negatively charged. With regular force field based docking, these atoms were predicted

to be placed near the iron center. Although with iron in the quantum region QM/MM calculation altered the charge configuration of the ligand, the charge transfer from the iron atom is more evenly distributed throughout the whole ligand than the previous example due to the geometry of the native pose. This fact compounded with the reduction of atomic charge on iron certainly helped M-QM Dock to produce the right binding mode prediction. What we can deduce from these two examples is that though reduction of atomic charge on iron was important, charge configuration on ligand should not be altered greatly.

## Conclusion

Our analysis shows although that the inclusion of metal ions in quantum region rescales the charge on them, charge configuration on ligands should not be affected greatly for subsequent docking to work properly. With further inspection of the binding sites of these examples, we conclude that surrounding protein atoms of metal ions must also be included in quantum region during QM/MM calculations. There usually exist a few histidine residues coordinated with metal ions in the binding sites and one can easily suspect that in reality charge transfer happens between these residues and metal ions. If that is the case, inclusion of these residues in quantum region will enable QM/MM calculations to accurately describe the charge transfer within binding sites of the corresponding proteins. With further studies along this line, one can identify the right application of quantum mechanical calculations for modeling of metalloprotein binding sites.

## Methods

### Docking Method

We employed the Glide 4.0 program<sup>19</sup> in SP (standard precision) mode for our study. The docking algorithm in Glide utilizes a hierarchical search protocol. Selection of the final ligand pose for the regular Glide is done with Glide score, which is an extension of an empirically based Chem-Score function of Eldridge *et al.*<sup>20</sup>. For our QM/MM docking protocol, however, final ligand selection is primarily determined by the total Coulomb-van der Waals energy (Ecvdw), with the Coulomb energy screened by a distance-dependent dielectric constant, since we have in effect emphasized Coulombic energy portion of the scoring function by recalculating the atomic charges. The effectiveness of this choice of scoring function was

demonstrated in the previous work<sup>21</sup>. The ligand structures and the corresponding receptor protein structures were prepared manually using utilities provided in the Schrödinger Glide suite. We started all the test cases from the raw PDB files. During this preparation, protein-ligand complexes were minimized up to 0.3 RMSD from X-ray crystal structures. We calculated the root mean square distance (RMSD) of the resulting ligand configurations to these structures to determine the accuracy of binding mode prediction. After we generated ligand poses, we discarded those that were within 0.6 Å of RMSD values to any previously accepted poses, a process we call “clustering”. For Glide SP mode, we kept 5,000 poses for each ligand docking from initial generation for refinement. After the refinement we kept 400 poses for minimization using grids during which a maximum of 100 steps were imposed. Finally, we scored 10 poses and ranked them after minimizations.

### QM/MM Docking

In the earlier work of QM/MM docking, a docking protocol coined as “Survival of the Fittest”<sup>10</sup> was implemented. In that work, an initial docking with regular force field was performed to produce a set of poses that were fed into QM/MM one-point energy calculations regarding only ligands as quantum mechanical region, which in turn were fitted to generate a new set of atomic charges based on density functional theory (DFT)<sup>22</sup> quantum mechanical calculations. Using this new set of charges, a new generation of docking runs was performed and in the end the best scoring pose was selected. The key idea in this protocol was that at a pose that is close to the native structure, quantum mechanical calculation will produce atomic charges that give rise to lower Coulombic energy values at native pose. For a validation study, one performs QM/MM calculation at the native pose of each complex to adjust the atomic charges on ligand atoms. This calculation would generate a set of charges in a given binding site environment that is theoretically the best possible one for docking. For QM/MM calculations, we used QSite<sup>19</sup> version 4 of Schrödinger’s 2006 suite, which combines IMPACT 4.0<sup>19</sup> and Jaguar 6.5<sup>19</sup>. All our calculations for QM part were done with DFT. 6-31G\* basis set of Pople and co-workers<sup>23-25</sup>, and hybrid DFT functional B3LYP<sup>26-29</sup>, which has been shown to yield excellent results for atomization energies and transition states in a wide range of chemical systems, were used.

### Charge Fitting

To calculate atomic charges after single point energy QM/MM calculations, we used ESP (electrostatic

potential) fitting, which is done by first calculating the molecular electrostatic potential (MEP) and fitting atomic charges to it. For the frozen orbital method<sup>30</sup>, which is the theory behind QSite program, ESP atomic charge fitting becomes problematic for atoms near the QM/MM boundaries because of the extra electron planted in the frozen orbital. For our definition of quantum region, this problem is not an issue since there is no explicit bond defined between QM and MM regions.

## Acknowledgements

We would like to thank Schrödinger, Inc. for providing its software. This work was partially supported by startup fund from Korea University (AEC).

## References

- Halperin, I., Ma, B.Y., Wolfson, H. & Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins-Structure Function and Genetics* **47**, 409-443 (2002).
- Gschwend, D.A., Good, A.C. & Kuntz, I.D. Molecular docking towards drug discovery. *J. Mol. Recognit.* **9**, 175-186 (1996).
- Abagyan, R. & Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **5**, 375-382 (2001).
- Kuntz, I.D. Structure-based strategies for drug design and discovery. *Science* **257**, 1078-1082 (1992).
- Drews, J. Drug discovery: A historical perspective. *Science* **287**, 1960-1964 (2000).
- Brooijmans, N. & Kuntz, I.D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomolec. Struct.* **32**, 335-373 (2003).
- Cho, A.E. *et al.* The MPSim-Dock hierarchical docking algorithm: application to the eight trypsin inhibitor cocrystals. *J. Comput. Chem.* **26**, 48-71 (2005).
- Kramer, B., Metz, G., Rarey, M. & Lengauer, T. Ligand docking and screening with FlexX. *Med. Chem. Res.* **9**, 463-478 (1999).
- Verdonk, M.L., Cole, J.C., Hartshorn, M.J., Murray, C.W. & Taylor, R.D. Improved protein-ligand docking using GOLD. *Proteins-Structure Function and Genetics* **52**, 609-623 (2003).
- Cho, A.E., Guallar, V., Berne, B.J. & Friesner, R. Importance of accurate charges in molecular docking: Quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **26**, 915-931 (2005).
- Cho, A.E. Effect of quantum mechanical charges in binding sites of metalloproteins. *BioChip J.* **1**, 70-75 (2007).
- Lipscomb, W.N. & Strater, N. Recent advances in zinc enzymology. *Chem. Rev.* **96**, 2375-2433 (1996).
- Deerfield, D.W., Carter, C.W. & Pedersen, L.G. Models for protein-zinc ion binding sites. II. The catalytic sites. *Int. J. Quantum Chem.* **83**, 150-165 (2001).
- Dudev, T. & Lim, C. Metal binding in proteins: The effect of the dielectric medium. *J. Phys. Chem. B* **104**, 3692-3694 (2000).
- Rulisek, L. & Havlas, Z. Theoretical studies of metal ion selectivity. 1. DFT calculations of interaction energies of amino acid side chains with selected transition metal ions ( $\text{Co}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Cd}^{2+}$ , and  $\text{Hg}^{2+}$ ). *J. Am. Chem. Soc.* **122**, 10428-10439 (2000).
- Hu, X. & Shelver, W.H. Docking studies of matrix metalloproteinase inhibitors: zinc parameter optimization to improve the binding free energy prediction. *J. Mol. Graph.* **22**, 115-126 (2003).
- Khandelwal, A. *et al.* A combination of docking, QM/MM methods, and MD simulation for binding affinity estimation of metalloprotein ligands. *J. Med. Chem.* **48**, 5437-5447 (2005).
- Strynadka, N.C.J. *et al.* Molecular docking programs successfully predict the binding of a beta-lactamase inhibitory protein to TEM-1 beta-lactamase. *Nat. Struct. Mol. Biol.* **3**, 233-239 (1996).
- Friesner, R.A. *et al.* Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739-1749 (2004).
- Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V. & Mee, R.P. Empirical scoring functions. 1. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **11**, 425-445 (1997).
- Cho, A.E., Guallar, V., Berne, B.J. & Friesner, R. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *J. Comput. Chem.* **26**, 915-931 (2005).
- Kohn, W. Thoughts about density functional theory in 1998. *J. Comput. Chem.* **20**, 1-1 (1999).
- Rassolov, V.A., Pople, J.A., Ratner, M.A. & Windus, T.L. 6-31G\* basis set for atoms K through Zn. *J. Chem. Phys.* **109**, 1223-1229 (1998).
- Rassolov, V.A., Ratner, M.A., Pople, J.A., Redfern, P.C. & Curtiss, L.A. 6-31G\*basis set for third-row atoms. *J. Comput. Chem.* **22**, 976-984 (2001).
- Tran, J.M., Rassolov, V.A., Ratner, M.A. & Pople, J.A. Diffuse functions in the 6-31G basis set for atoms K through Zn. *J. Am. Chem. Soc.* **217**, U443-U443 (1999).
- Becke, A.D. A new mixing of hartree-fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372-1377 (1993).
- Kim, K. & Jordan, K.D. Comparison of density-functional and Mp2 calculations on the water monomer and dimer. *J. Phys. Chem. A* **98**, 10089-10094 (1994).
- Perdew, J.P., Emzerhof, M. & Burke, K. Rationale for mixing exact exchange with density functional ap-

- proximations. *J. Chem. Phys.* **105**, 9982-9985 (1996).
29. Stephens, P.J., Devlin, F.J., Chabalowski, C.F. & Frisch, M.J. Ab-initio calculation of vibrational absorption and circular-dichroism spectra using density-functional force-fields. *J. Phys. Chem. A* **98**, 11623-11627 (1994).
30. Philipp, D.M. & Friesner, R.A. Mixed ab initio QM/MM modeling using frozen orbitals and tests with alanine dipeptide and tetrapeptide. *J. Comput. Chem.* **20**, 1468-1494 (1999).