Rank-ordering protein-ligand binding affinity by a quantum mechanics/molecular mechanics/Poisson-Boltzmann-surface area model

Cite as: J. Chem. Phys. **126**, 026101 (2007); https://doi.org/10.1063/1.2423029
Submitted: 22 September 2006 . Accepted: 22 November 2006 . Published Online: 09 January 2007

Mingliang Wang, and Chung F. Wong

Rank-ordering protein-ligand binding affinity by a quantum mechanics/molecular mechanics/Poisson-Boltzmann-surface area model

Mingliang Wang[a] and Chung F. Wong[b]
Department of Chemistry and Biochemistry, University of Missouri-Saint Louis, One University Boulevard, Saint Louis, Missouri 63121

(Received 22 September 2006; accepted 22 November 2006; published online 9 January 2007)

[DOI: 10.1063/1.2423029]

We implemented the quantum mechanics/molecular mechanics/Poisson-Boltzmann-surface area (QM/MM/PBSA) model by integrating the linear-scaling quantum mechanical code SIESTA (Ref. 1) with the PBSA model in UHBD.[2] The implementation was similar to our earlier work[3] in integrating PWSCF[4] with UHBD except that we also added a MM feature to treat part of the system, such as the protein in this work, classically rather than quantum mechanically. In solving the Kohn-Sham equation for the solvated complex, we added the MM field of the protein and the solvent reaction fields (SRFs) generated by the protein and the ligand to the Kohn-Sham Hamiltonian. For the solvated ligand, only the SRF from the ligand was added. In solving the PB equation, we used the quantum mechanical charge density directly. The MM treatment utilized the CHARMM (Ref. 5) force field. The coupled Kohn-Sham and Poisson-Boltzmann equations were solved self-consistently until the energy converged. We then used the results in the following expression to calculate the binding energy between a ligand and a protein within the fixed-conformation approximation:

$$\int \rho_{\text{ligand}}^{\text{complex}}(\mathbf{r})\phi_{\text{protein}}^{\text{complex}}d^3\mathbf{r} + \int \rho_{\text{ligand}}^{\text{complex}}(\mathbf{r})\phi_{\text{ligand}}^{\text{complex}}d^3\mathbf{r} - \int \rho_{\text{ligand}}^{\text{solvent}}(\mathbf{r})\phi_{\text{ligand}}^{\text{solvent}}d^3\mathbf{r} + \Delta G_{\text{protein}}^{\text{desolv}} + \langle \Psi_{\text{ligand}}^{\text{complex}}|H^g|\Psi_{\text{ligand}}^{\text{complex}}\rangle - \langle \Psi_{\text{ligand}}^{\text{solvent}}|H^g|\Psi_{\text{ligand}}^{\text{solvent}}\rangle + \sigma\Delta A,$$

where $\rho_{\text{ligand}}^{\text{complex}}(\mathbf{r})$ was the quantum mechanical charge density of the ligand calculated in the solvated complex (i.e., obtained by including the direct and solvent-mediated protein fields and the SRF produced by the ligand in the Kohn-Sham Hamiltonian), $\rho_{\text{ligand}}^{\text{solvent}}(\mathbf{r})$ was the quantum mechanical charge density of the ligand in water (obtained by including the reaction field of the ligand in the Kohn-Sham Hamiltonian), $\phi_{\text{protein}}^{\text{complex}}$ was the field generated by the protein in the solvated complex and it included both direct and solvent-mediated contributions, $\phi_{\text{ligand}}^{\text{complex}}$ was the SRF produced by the ligand in the solvated complex, $\phi_{\text{ligand}}^{\text{solvent}}$ was the SRF produced by the ligand in the solvent, and $\langle \Psi_{\text{ligand}}^{\text{complex}}|H^g|\Psi_{\text{ligand}}^{\text{complex}}\rangle - \langle \Psi_{\text{ligand}}^{\text{solvent}}|H^g|\Psi_{\text{ligand}}^{\text{solvent}}\rangle$ compared the energy required to distort the electron cloud of the ligand when it was put in the complex and in water, respectively. ($\Psi_{\text{ligand}}^{\text{complex}}$ and $\Psi_{\text{ligand}}^{\text{solvent}}$ were the wave functions of the ligand calculated in the complex and water, respectively, where H^g was the gas phase Hamiltonian of the ligand.) $\Delta G_{\text{protein}}^{\text{desolv}}$ measured the desolvation penalty of the protein due to ligand binding and we calculated it by taking the difference between the electrostatic energies obtained from two PB calculations: one for the isolated protein in solution, the other for the protein-ligand complex with the charge of the ligand set to zero. The last term described the nonpolar contributions in which the intrinsic surface tension σ was set to 25 kcal/mol Å^2, and ΔA was the change in solvent-accessible surface area associated with protein-ligand binding. (In a recent study, Laio *et al.* introduced an additional function to modify the short-range electrostatic interactions between the quantum and classical atoms to prevent overpolarization.[6] Although we did not do this in our study, such a correction should not alter the qualitative conclusions here because we focused on estimating relative binding affinity involving similar inhibitors rather than doing Car-Parrinello molecular dynamics simulation.)

Our test case was the binding of balanol and its derivatives to protein kinase A (PKA). Figure 1 shows the structure of balanol if X=NH_3^+. We also performed calculations on derivatives in which X was replaced by O, CH_2, S, and SO_2. To improve the odds of cancellation of systematic errors in comparing the binding affinity of similar compounds within the single-conformation approximation, we prepared the ligands in a way that their structures only differed in the geometry of their X groups. For example, when we started with the crystal structure, we first prepared the structure of the PKA-balanol complex using the PDB2PQR program.[7] We then replaced the NH_3^+ group in balanol by another X group and optimized the geometry of the X group with SIESTA using the double-zeta plus polarization basis set (DZP) holding every other atom fixed. To demonstrate that the qualitative conclusions were not sensitive to the choice of conformation, we also prepared two other sets of structures. In one, the structure of balanol was optimized in the MM field of PKA using the ONIOM option in GAUSSIAN03.[8] In the geometry optimization, balanol was treated with the AM1 model, whereas the protein was described by the UFF force field.[9] The derivatives of balanol were constructed in the same way as when the crystal structure was used to generate the balanol reference structure. The other set of structures was generated

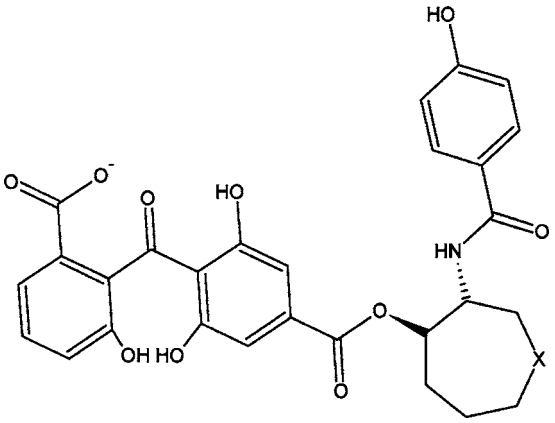

FIG. 1. Structure of balanol if $X=NH_3^+$.

similarly, except that balanol was treated with the Hartree-Fock model using the 3-21G basis set in the ONIOM geometry optimization.

In the binding affinity calculations, we set the dielectric constant of the ligand to 1 because the quantum treatment explicitly accounted for electronic polarization. The dielectric constant of the aqueous solvent was set to 78. Although it is common to choose the dielectric constant of proteins to be 2, 4, or above, we found that these values did not give results consistent with experiment.[10] The top panel of Fig. 2 orders the compounds such that their binding affinity increases from left to right if they follow the experimental trend. Clearly, the curves in the figure do not fit this trend no matter we used the van der Waals or the solvent-accessible surface to define the dielectric boundary, the double-zeta or triple-zeta basis set, and the crystal structure or quantum mechanically optimized ligand structure in the protein field (AM1/UFF ONIOM model). On the other hand, we could follow the experimental trend much better if we used a dielectric constant of 1.5 no matter which one of the three sets of structures was used (bottom panel of Fig. 2). Furthermore, using a dielectric constant of 1 worsens the results somewhat (results not shown). We also calculated the binding affinity using the "traditional" classical electrostatics model in which the dielectric constants of both the ligand and the protein were set to be the same. Here, because the ligand was treated classically, one had to guess a value for its dielectric constant. We found that as long as the dielectric constants of the protein and the ligand were chosen to be the same, setting them to a value between 1 and 2 all gave similar trend as the QM/MM/PBSA model (results not shown). Together, the results from all these calculations suggest that the dielectric constants of the protein and the ligand are probably about the same and they range between 1 and 2.

One potential application of this model is to perform lead optimization in a drug design process in which one chemically modifies a drug lead in many ways to see which derivatives might have better binding affinity. As long as the modifications are not too large, the approximations made in this model should be reasonable. Moreover, one does not need to be right all the time for such an application. A computational approach that can reduce the number of compounds that need to be synthesized and screened is already very useful. For systems of the size studied here, each binding affinity calculation only took ~35 min with a single Xeon 3.6 GHz processor. Thus, many computational chemical modifications can be easily done with modern multiprocessor computer clusters.

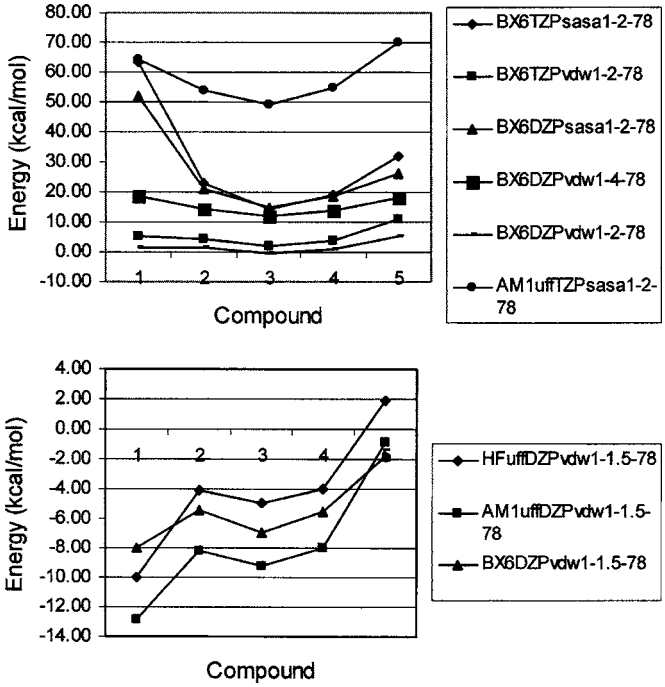

FIG. 2. Top panel: Binding affinity of balanol and its derivatives to protein kinase A when the dielectric constant of the protein was set to 2 or 4. To decode the curve, BX6 stands for crystal structure (Ref. 11), DZP and TZP stand for double-zeta and triple-zeta polarization basis sets, respectively, vdw and sasa denote van der Waals and solvent-accessible surfaces, respectively, which were used to define the dielectric boundary in PB calculations, AM1uff means that the PKA-balanol structure was obtained by optimizing the geometry of balanol using the AM1 model in the UFF force field of PKA, and 1-2-78, etc., indicates that the dielectric constant of the ligand, protein, and solvent were 1, 2, and 78, respectively. Bottom panel: Same as the top panel except that the dielectric constant of the protein was set to 1.5. In addition, HFuff means that the PKA-balanol structure was obtained by optimizing the geometry of balanol using the HF/3-21G model in the UFF force field of PKA.

This research was supported in part by a Research Award from the University of Missouri-Saint Louis, a Research Board Award from the University of Missouri System, and a NIH grant to one of the authors (C.F.W.).

[a]Also at the College of Chemistry and Chemical Engineering, Shenzhen University, Shenzhen 518060, China.
[b]Author to whom correspondence should be addressed.

[1] J. M. Soler, E. Artacho, J. D. Gale, A. Garcia, J. Junquera, P. Ordejon, and D. Sanchez-Portal, J. Phys.: Condens. Matter **14**, 2745 (2002).
[2] M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon, Comput. Phys. Commun. **62**, 187 (1991); J. D. Madura, J. M. Briggs, R. C. Wade *et al.*, *ibid.* **91**, 57 (1995).
[3] M. L. Wang and C. F. Wong, J. Phys. Chem. A **110**, 4873 (2006).
[4] S. Baroni, A. Dal Corso, S. de Gironcol *et al.*, http://www.pwscf.org
[5] A. D. MacKerell, D. Bashford, M. Bellott *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[6] A. Laio, J. VandeVondele, and U. Rothlisberger, J. Chem. Phys. **116**, 6941 (2002).
[7] T. Dolinsky, J. Nielsen, J. McCammon, and N. Baker, Nucleic Acids Res. **32**, W665 (2004).
[8] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 03,

Revision C.02, Gaussian, Inc., Wallingford, CT, 2004.
[9] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, J. Am. Chem. Soc. **114**, 10024 (1992).
[10] Y. S. Lai, J. S. Mendoza, G. E. Jagdmann *et al.*, J. Med. Chem. **40**, 226 (1997).
[11] N. Narayana, T. C. Diller, K. Koide, M. E. Bunnage, K. C. Nicolaou, L. L. Brunton, N. H. Xuong, L. F. Ten Eyck, and S. S. Taylor, Biochemistry **38**, 2367 (1999).