JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

How to Improve Docking Accuracy of AutoDock4.2: A Case Study Using Different Electrostatic Potentials

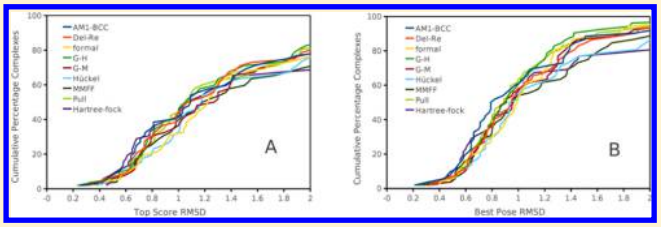
Xuben Hou,[†] Jintong Du,[†] Jian Zhang,[‡] Lupei Du,[†] Hao Fang,*[,†] and Minyong Li*[,†]

[†]Department of Medicinal Chemistry, Key Laboratory of Chemical Biology (MOE), School of Pharmacy, Shandong University, Jinan, Shandong 250012, China

[‡]Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis (MOE), Shanghai Jiao-Tong University School of Medicine (SJTU-SM), Shanghai, 200025, China

Supporting Information

ABSTRACT: Molecular docking, which is the indispensable emphasis in predicting binding conformations and energies of ligands to receptors, constructs the high-throughput virtual screening available. So far, increasingly numerous molecular docking programs have been released, and among them, AutoDock 4.2 is a widely used docking program with exceptional accuracy. It has heretofore been substantiated that the calculation of partial charge is very fundamental for the accurate conformation search and binding energy estimation. However, no systematic comparison of the significances of electrostatic potentials on docking accuracy of AutoDock 4.2 has been determined. In this paper, nine different charge-assigning methods, including AM1-BCC, Del-Re, formal, Gasteiger−Hückel, Gasteiger−Marsili, Hückel, Merck molecular force field (MMFF), and Pullman, as well as the ab initio Hartree−Fock charge, were sufficiently explored for their molecular docking performance by using AutoDock4.2. The results clearly demonstrated that the empirical Gasteiger−Hückel charge is the most applicable in virtual screening for large database; meanwhile, the semiempirical AM1-BCC charge is practicable in lead compound optimization as well as accurate virtual screening for small databases.



1. INTRODUCTION

Compared with time-consuming and high-cost conventional experimental-based techniques, computational chemistry is performing a consequential role currently in drug design and discovery.[1,2] Following this pipeline, there have been tremendous endeavors in finding lead compounds with high speed and low cost, and a case in point is molecular docking that has been well-introduced.[3−7] Molecular docking is a computational approach that attempts to predict preferable noncovalent binding of a small molecule (ligand) to a macromolecule (receptor). It needs to be emphasized that molecular docking is of particular importance since it can be implemented in high-throughput virtual screening to identify brand new chemotype molecules. In recent years, considerable progress has been achieved in developing docking algorithms for predicting ligand−receptor binding modes and virtual screening,[3] such as AutoDock,[8,9] Dock,[10−12] GOLD,[13,14] Glide,[15,16] Surflex-Dock,[17,18] FlexX,[19] Zdock,[20] VoteDock,[21] and so on. Among these, AutoDock has been recognized as the most widely used docking program as per a recent study.[5]

Since its first released in 1990,[8] AutoDock has been proven to be an effective tool for predicting ligand−receptor complex conformations and binding energies with high accuracy. In the past few years, AutoDock has been extensively used for numerous academic, industrial, and governmental proposes.[5] A recent study also indicated that AutoDock showed good success rates when applied to the protein−protein interfaces-(PPIs).[22] Nowadays, most docking programs consider only ligand flexibility with the receptor rigid when fitting small ligands to macromolecules in view of its complexity and high-computational costs. Nevertheless, the recent published version AutoDock4.2 inherited the high accuracy of its former version AutoDock3 and solved this problem by incorporating the explicit conformational modeling of specified receptor side-chains.[23]

In general, molecular docking programs consist of two components: (i) searching for the suitable conformations of ligand−receptor complexes and (ii) scoring of the resulting geometries with respect to binding energies. AutoDock4.2 was associated with an empirical free energy force field with a Lamarckian genetic algorithm (LGA), thus providing a fast prediction of conformation and free energy. In addition, the force field of AutoDock has been calibrated by using a database containing 188 various protein−ligand complexes and has exhibited a standard error of approximately 2−3 kcal/mol in estimating binding free energy.[24]

The score function of AutoDock 4.2 includes five terms:[24] (1) a typical Lennard-Jones 6/12 potential for dispersion/repulsion interactions; (2) a directional 10/12 hydrogen bonding term, where $E(t)$ is dependent on the angle t away from ideal bonding geometry; (3) a Coulombic electrostatic

Received: September 1, 2012
Published: December 17, 2012

potential; (4) a desolvation potential based on the volume (V) of the atoms surrounding a given atom, weighted by a solvation parameter (S) and an exponential term based on the distance; and (5) a term for the loss of torsional entropy upon binding which is directly proportional to the number of rotatable bonds in the molecule (N_{tors}; eq 1)

$$\Delta G_{bind} = \Delta G_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \Delta G_{hbond} \sum_{i,j} E(t) \left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right) + \Delta G_{elec} \sum_{i,j} \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} + \Delta G_{sol} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} + \Delta G_{tor} N_{tor} \quad (1)$$

In the current study, we focus on the effect of different electrostatic potential on ligand. It is clear that ΔG_{elec} is directly based on the atomic charges in ligands. In addition, the AutoDock4.2 force field adopted a charge-based desolvation model and the solvation parameter (S, in eq 1; q_i is the atomic charge) for a given atom is calculated as

$$S_i = (\mathrm{ASP}_i + \mathrm{QASP} \times |q_i|) \quad (2)$$

It is worth mentioning that it is the docking scoring that decides which conformation will be accepted for the next run, when the Lamarckian genetic algorithm is employed in generating new conformations. As a result, the calculation of accurate partial charges on ligands and proteins are expected to have an essential impact on both the docking conformation and the binding energy estimation, possibly leading to the improvement of complex geometry prediction and binding energy calculation. Actually, in our previous research, the influence of various charge-assigning methods on docking results of DOCK5.4 program was well analyzed. Our results disclosed that the quantum mechanics charge yielded reasonable docking results in both binding conformation and free energy.[25]

Electrostatic point charges on the atoms in a molecule can be calculated by multiple semiempirical and empirical methods. It needs to be noted that the ab initio method such as RESP (restrained electro static potential) is pretty accurate when charging a small set of ligands.[26] However, considering the computationally expensive nature of calculating ab initio charges, several fast empirical and semiempirical methods have been well-designed. The very early approaches were the Sanderson equalization of electronegativity[27] and the Del-Re[28,29] empirical models. The Sanderson electronegativity equalization principle gave rise to a multitude of modern charge models including the Gasteiger−Marsili (GM) method.[30,31] The Pullman[32] model was combined by Del-Re with the Hückel[33] method to reproduce dipole moments of aromatic and conjugated compounds. In the same way, the Gasteiger−Hückel (GH) model was introduced by combining the Gasteiger−Marsili σ-charges and the Hückel π-charges. Currently, the AM1-BCC method was released, which combined AM1 atomic charges with bond charge corrections (BCC). AM1-BCC presents a rapid meticulous method to charge organic molecules for general facilitation in molecular simulation with the AMBER force field.[34] Among other charge-assigning models, the formal charge and the Merck molecular force field (MMFF)[35−37] charges should also be mentioned. The charge methods involved in our study were summarized in Supporting Information Table S15. Herein, we describe our studies of several charge methods for their impact on the docking accuracy of AutoDock4.2. We mainly focused on eight diverse semiempirical and empirical charge methods, including AM1 atomic charges with bond charge corrections (AM1-BCC), Del-Re, formal, Gasteiger−Marsili (GM), Gasteiger−Hückel (GH), Hückel, Merck molecular force field (MMFF), and Pullman. Although time-consuming, the ab initio charge method Hartree−Fock (HF) is also involved in our comparisons, intending to explore the impact of accurate charge calculation on docking result in-depth. In total, nine charge-assigning methods were used in this study. To provide a proper benchmark for the comparison of charge performance in our studies, 52 protein−ligand complexes were selected with known crystal structure and available binding energy data. The ligands were extracted from the complexes and then docked back into their host proteins using AutoDock4.2. Then, docking accuracy was evaluated mainly considering the ability to find the correct conformations as well as to accurately estimate binding energy.

2. RESULTS AND DISCUSSION

We would like to present the results on docking accuracy of AutoDock4.2 using nine different charge methods, including AM1-BCC, Del-Re, formal, GM, GH, Hückel, MMFF, Pullman, and Hartree−Fock. Using the default AutoDock4.2 parameters, ten docking experiments were performed on the entire 52 complexes in benchmark database. During each docking run, the receptors were treated as rigid and ligands were considered as flexible.

First, we discuss the difference of each charge method in predicting the binding conformation of ligands to proteins. Results are listed in the Supporting Information, Table S2 for top score root mean squared distance (RMSD) and Table S3 for best pose RMSD. It is one of the crucial aspects of docking, for the better three-dimensional binding pose proposed, that the more realistic interactions between receptors and ligands will be recreated.[38] On the other hand, we will evaluate the abilities in the in vitro binding affinity calculation using different charge methods. The estimated binding energies data were reported in Supporting Information Table S4 for top score and Table S5 for best pose. In addition, the binding free energies of the ligands to its receptors in testing 52 complexes were calculated using AutoDock4.2 without the Lamarckian genetic algorithm conformation search, in order to identify the effect of nine different atomic charges on the AutoDock4.2 score function.

2.1. Evaluation of Pose Prediction Capabilities.

2.1.1. Top Score RMSD and Best Pose RMSD. The conformation with top score possesses the lowest estimated binding energy; therefore, the "Top Score RMSD" is a significant parameter in assessing the docking accuracy of AutoDock. Figure 1A exhibits the cumulative percentage of complexes as a function of the top score RMSD between the predicted conformation and crystal structure coordinate. It is evident that in terms of top score RMSD, the GH charge carried out best with the highest success rate 82.69%, along with the application of GH charge improved success rate by 15%, compared with the worst Hartree−Fock charges. In case of average RMSD, the results acquired from nine charge methods were all acceptable (below 2 Å), without appreciable distinction between each of them.

dx.doi.org/10.1021/ci300417y | J. Chem. Inf. Model. 2013, 53, 188−200

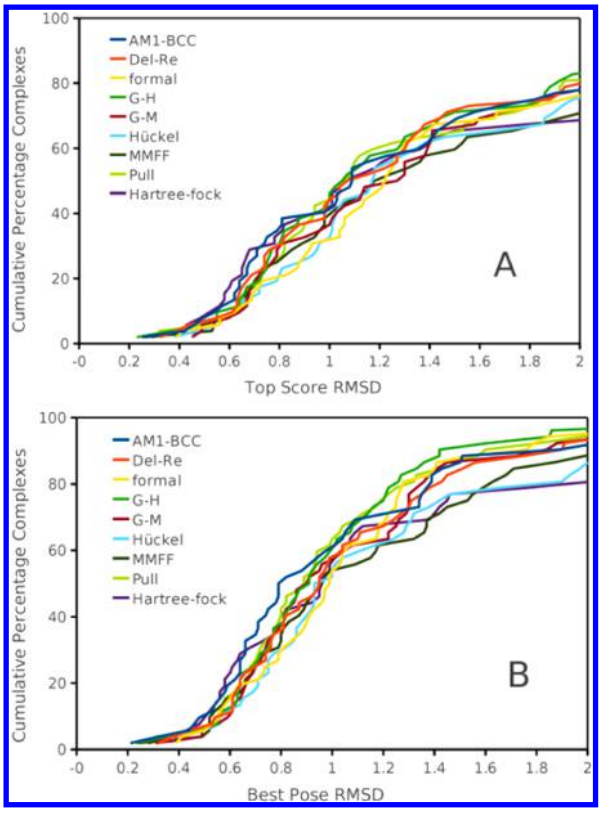


Figure 1. Cumulative percentages of complexes as a function of RMSD in top score conformations (A) and best pose conformations (B).

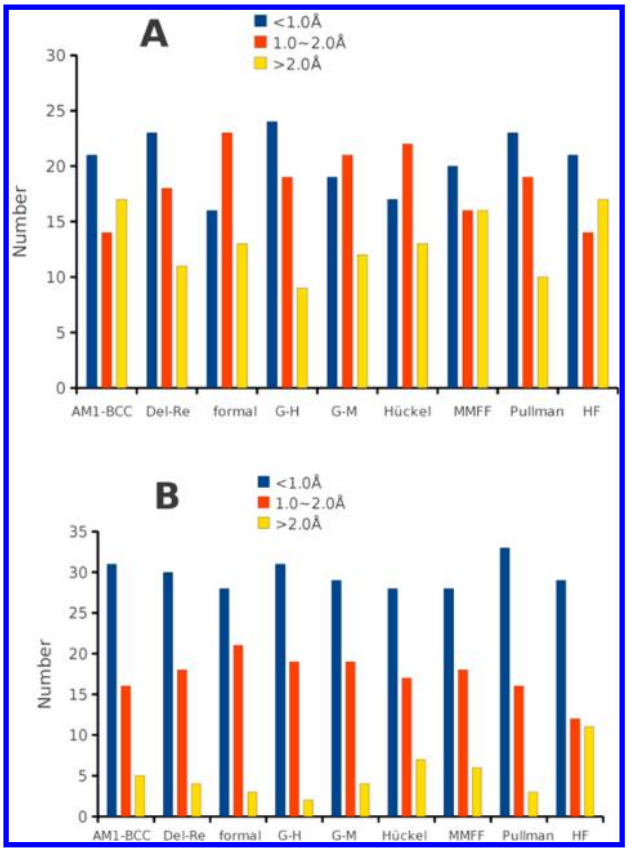


Figure 2. Comparison of different charge methods in terms of docking numbers obtained under 1.0, 1.0−2.0, and below 2.0 Å of RMSD in top score conformations (A) and best pose conformations (B).

As expected, considerable results were observed when it comes to the result of "Best Pose RMSD". A success rate as high as 96.15% in GH charge method was achieved. At the same time, the average RMSD values of all nine charge models were also enhanced by about 0.5 Å as opposed to the top score RMSD. On the other side, Hartree−Fock charge still positioned the worst, with a success rate of 78.85%, and then after that the Hückel charge (86.54%) appears.

For each charge model, their best prediction, moderate prediction, and worst prediction are listed in Figure 2. Manifestly, the GH charge method is among the highest in the number of best predictions and lowest in the number of worst predictions. It should be noted that only success rate can not indicate the ability of each charge method synthetically. As indicated in Figure 2, though the formal charge method possessed a overhead success rate (75% in top score and 94.23% in best pose, it took over the lowest number of best predictions, as well as high success rate from the high numbers of moderate prediction between 1 < RMSD < 2 Å. Supplementally, the Pullman charge had a better appearance in best prediction comparing to the moderate prediction and worst prediction.

Nevertheless, the ability of correct posing by AutoDock programs was measured by the top score results, which seems somewhat inadequate. Moreover, top score conformations are rarely classified simultaneously as to best pose, also their position is rather randomly distributed among all generated poses ordered by the docking score.[38] However, the best pose RMSD value cannot be ignored for its importance in evaluating whether AutoDock could find the most correct pose. Finally, in our study, though top score RMSD was preferred, best pose RMSD was also taken into consideration to evaluate all nine charge methods.

To sum up, all nine charge methods were ordered based on the result of both top score and best pose: GH > Pullman > Del-Re > AM1-BCC > GM > formal > Hückel > MMFF > Hartree−Fock. It is interesting that the GH charge method possesses the highest success rate both in top score RMSD and best pose RMSD while the Hartree−Fock charge method positions are the lowest ones. Generally, GH charges are widely used in structural biology, bioinformatics, and drug design and discovery. Herein, our studies also come to an important conclusion that GH charges are more suitable for accurate docking using AutoDock4.2, for its good accuracy in limited 10 runs, which is critical in high-throughput virtual screening. Moreover, Figure 3 demonstrated that the GH charge was able to find the most correct conformation even when other charge methods all failed. On the other side, the application of the quantum chemical charge-assignment method Hartree−Fock did not obtain better results compared with other semiempirical and empirical charge methods. All results are summarized in Table 1.

2.1.2. Impact of Ligand Flexible on Pose Prediction. In most previously published evaluations, it has been pointed out that the key issue that has to be addressed is to increase the success rate for docking of flexible molecules, because the significant variance between flexible and rigid molecules is well-recognized. Therefore, exploring the ligand flexibility feature may provide a more detailed insight into the docking algorithm, thus pointing out desirable charge method to solve this problem.

dx.doi.org/10.1021/ci300417y | J. Chem. Inf. Model. 2013, 53, 188−200

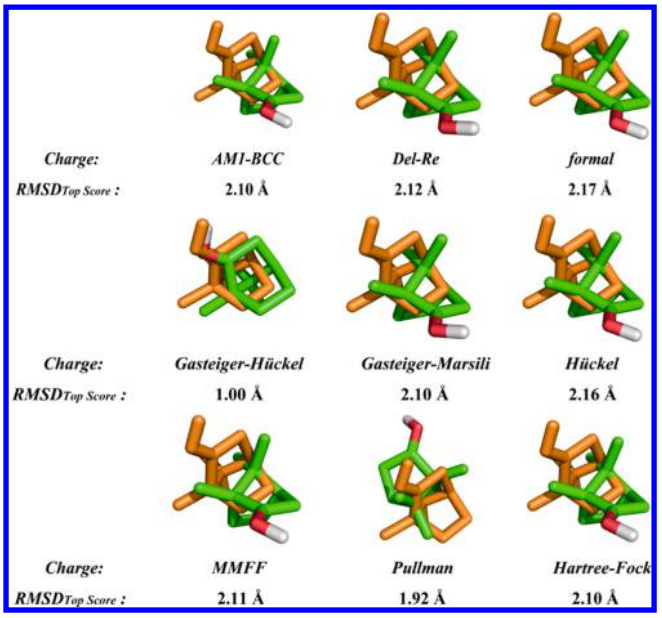


Figure 3. Comparisons between top score conformations and native ligand in different charge methods. PDB entry code: 2CPP. Docked ligands are colored in green, while the native ligand is orange.

Table 1. Average RMSD and Success Rate for Different Charge Methods

	best pose		top score	
charge	average RMSD (Å)	success rate (%)	average RMSD (Å)	success rate (%)
AM1-BCC	1.014	90.38	1.672	76.92
Del-Re	1.082	92.31	1.759	78.85
formal	1.118	94.23	1.770	75.00
GH	1.039	96.15	1.746	82.69
GM	1.126	92.31	1.677	76.92
Hückel	1.251	86.54	1.748	75.00
MMFF	1.203	88.46	1.837	69.23
Pullman	1.057	94.23	1.762	80.77
Hartree–Fock	1.323	78.85	1.795	67.31

The total test set was split into three subsets, depending on the rotatable bond number of the ligand: (small set) rotatable bond number < 5; (middle set) rotatable bond number ranging from 6–10; (large set) rotatable bond number > 10. As disclosed in Figure 4, it is apparent that some charge methods were unsuitable in special subsets. In terms of top score RMSD (Figure 4A), MMFF charge method performed the worst in "small set" and "middle set", with success rates of 68.18% and 64.71%, while others were higher than 70%. But in "large groups", it is the Hückel and Hartree–Fock charge method that performed the worst with success rates of 61.54% and 53.85%.

There comes a similar conclusion when it comes to best pose RMSD (Figure 4B). The MMFF charge method performed the worst in the small set, while Hückel and Hartree–Fock charge methods perfromed poorly in the large set. Although we often got a higher success rate when calculating with best pose RMSD instead of top score RMSD, it is surprising that all charge methods performed a success rate as high as 100% in the middle set, which indicated that the AutoDock4.2 program is quite suitable for middle size molecules with rotatable bond number ranging from 6–10. In line with the conclusion in the

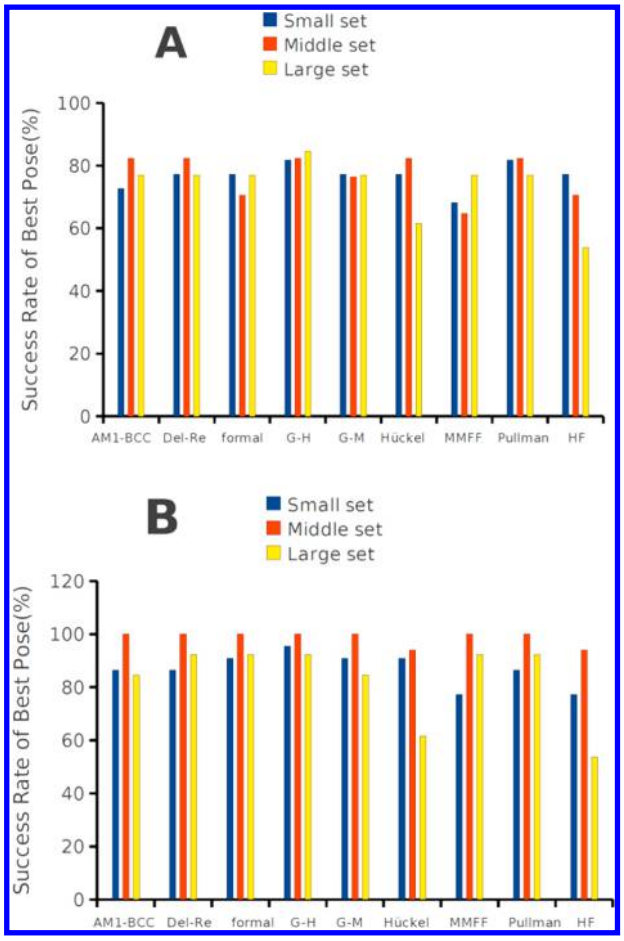


Figure 4. Success rate of different charge methods in three subsets: (A) top score conformations; (B) best pose conformations.

whole test set, statistical data here also exhibited that the GH charge method has the best accuracy when docking molecules of all size in AutoDock4.2. Moreover, MMFF charge possesses the lowest success rate both in the small and middle sets, and this can also explain its bad performance in whole test set.

2.2. Evaluation of Scoring Functions. Another important ability of molecular docking programs is to calculate the strength of binding the ligand into the target. Currently, many types of score functions were proposed, and a case in point is the empirical AutoDock scoring function, which is among the most successful ones by using relatively simple equations to calculate a docking score. The main advantage of this approach is its speed, which makes it possible to virtually screen millions of compounds in a short time. Our goal was to evaluate the influence of different charge methods on binding energy prediction, for the fact that correct prediction of binding energy is crucial in the LGA conformation search.

2.2.1. Correlation with the Experimental Binding Affinities. Table 2 and Figure 5 reveal the correlation between experimental binding energies and calculated energies by using AutoDock4.2. As described in Table 2, the correlation is rather limited in any of the nine charge models, which indicated the fact that it is quite difficult to predict binding free energy with high accuracy. It has been proven that when comparing the experimentally derived binding affinities with the calculated docking score, unreasonable correlations were usually obtained.[39] Moreover, this drawback is the important

dx.doi.org/10.1021/ci300417y | J. Chem. Inf. Model. 2013, 53, 188–200

Table 2. Correlation between Docking Energy Scores and Binding Energies

charge	correlation coefficient value	
	best pose	top score
AM1-BCC	0.331	0.294
Del-Re	0.107	0.172
formal	0.147	0.128
GH	0.123	0.142
GM	0.023	0.124
Hückel	0.043	0.077
MMFF	0.162	0.185
Pullman	0.069	0.148
Hartree−Fock	0.124	0.173

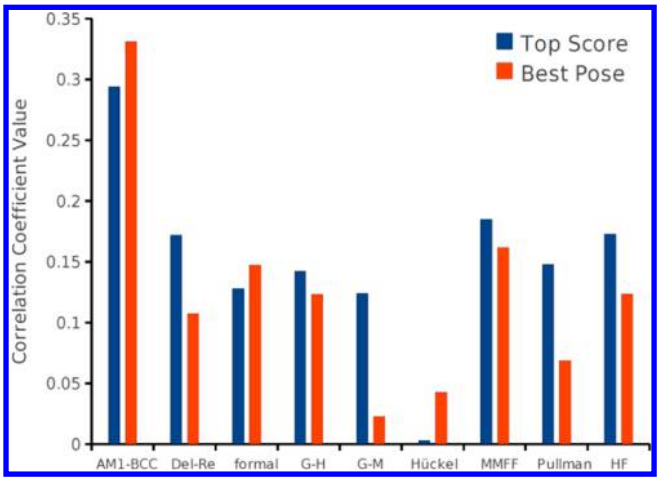


Figure 5. Correlation between Docking Energy Scores and Binding Energies.

reason that restricts the further application of docking software in the drug design process.

Although weak correlations were observed in this part, the correlation coefficient of AM1-BCC charge model has indicated significant superiority among all nine charge models, while Hückel charge model performed the worst. Generally speaking, using empirical methods to calculate electrostatic potentials has the advantage of high speed; however, these methods also possess the drawback of low accuracy. For example, in the GM charge method, as opposed to the semiempirical AM1-BCC, inorganic compounds such as metal ions, which frequently present in diverse functional proteins, are not taken into consideration. On the other side, the AM1-BCC method is an accurate semiempirical charge method based on a large validation set and is able to predict experimental homo- and heterodimer hydrogen-bond energies. Moreover, several published papers have reported that docking accuracy was increased when using the semiempirical method to calculate partial charges of the ligand atoms.[25,40] Further discussion of how different charge methods were treated in the scoring function of AutoDock4.2 is shown in section 2.2.3. In conclusion, the AM1-BCC charge method was able to dock large ligands more successfully than GH charge, while Hartree−Fock and other empirical charges failed (Figure 6).

When it turns to the result of best pose, the accuracy of binding energy estimation increased in Del-Re, GH, GM, MMFF, Pullman, and Hartree−Fock charges, while is decreased in AM1-BCC, formal, and Hückel charges. In general, the enhancement is more significant as to the reduction, especially when it comes to the GM charge. Results herein were not surprising because Bikadi et al. have already proved that there is rarely weak relationships betweent good pose prediction and accurate binding energy estimation.[41] Furthermore, the conformation of Best Pose is more similar with native ligand as to the conformation of top score; thereafter, the best pose conformations were able to evaluate the AutoDock score function simply, with little influence of deviation in geometry. Thus, the estimated binding energies of best pose conformations can also indicate the accuracy of the AutoDock scoring function.

The GH charge, which possesses the highest accuracy in predicting the correct pose, only received middle accuracy in binding energy predication among all nine charge methods. This result is not confusing for the fact that many empirical charge methods devote much more efforts to the improvement of speed, which no wonder lead to the sacrifice of accuracy. Accordingly, it seems that there is no direct transition from good pose prediction to good energy estimation when using different charge methods. This conclusion is also comparable with that of Spyrakis et al.[42] and Plewczynski et al.[38]

We found that the application of formal charge, which is the simplest charge method, resulted in a similar accuracy value compared with other empirical charge methods including Del-Re, GH, and MMFF, and even better than GM and Hückel charge methods. It seems that though optimized by large test set, many empirical charge methods did not reveal superiority to the simplest formal charge when applied in predicating binding energies of ligand to macromolecules.

In addition, we summarized the error between the estimated binding energy and experimental ones, intending to analyze each charge method in-depth. Figure 7 reflects that overestimation was usually obtained in Hartree−Fock, AM1-BCC, and MMFF charge methods (all represent good correlation in former section), while underestimation was obtained in formal and Hückel charges. The underestimated in formal and Hückel charges may also result from their lack in finding detailed information of each charged atom. GH, GM, and other other empirical charge methods performed better with lower error in estimating binding energies. The better performances of GH and GM here may have some relationship with the fact that force field of AutoDock4.2 was calibrated and validated by using a large collection of protein complexes with ligands adding Gasteiger charges. The different performance of nine charge methods due to the diverse atomic charge assigned on the same ligand, which finally lead the difference in binding energy calculation. Further discussion could be seen in section 2.3.

2.2.3. Impact of Ligand Flexible on Binding Energy Estimation. Once again, the ligand flexibility was taken into consideration, intending to get more details for each charge method. We can smoothly come upon an equivalent result that AM1-BCC charge performs with significant superiority, compared to other charge methods, and then the MMFF and Hartree−Fock charges follow. It is appealing that the AM1-BCC and Hartree−Fock charges perform much better in middle size groups with rotatable bond number from 6−10 and improved about 0.2−0.3 as to the result in the whole test set. The MMFF charge presented the best result in small size and was the second highest in the middle size group, but it failed to dock large ligands. In Figure 8A, the Pullman and Hückel charges obtained the best correlation coefficient in the large size

dx.doi.org/10.1021/ci300417y | J. Chem. Inf. Model. 2013, 53, 188−200

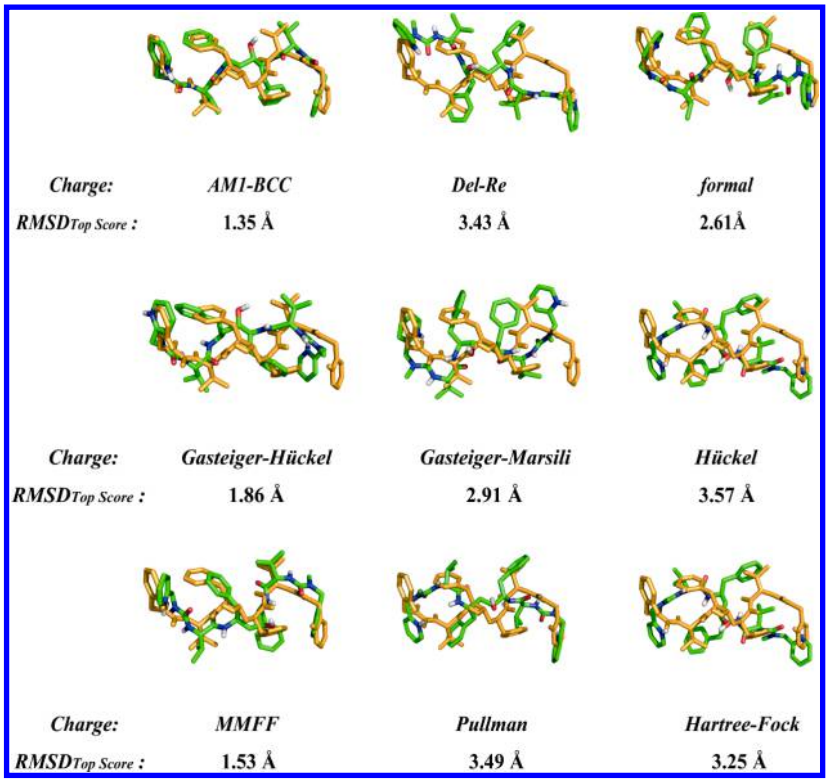


Figure 6. Comparisons between Top Score Conformations and Native Ligands in Different Charge Method. PDB code: 1HJV. Docked ligands are colored in green, while the native ligand is orange.

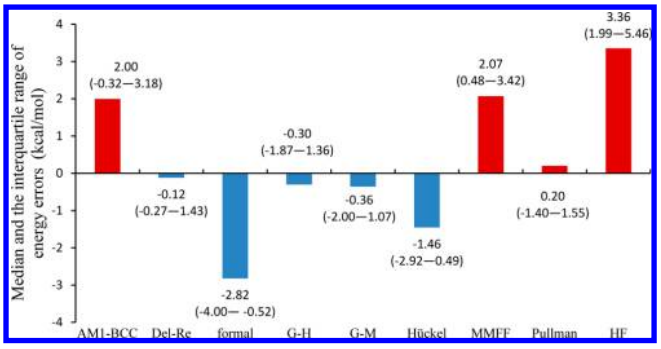


Figure 7. Median and the interquartile range of binding energy prediction errors using different charge methods (top score result).

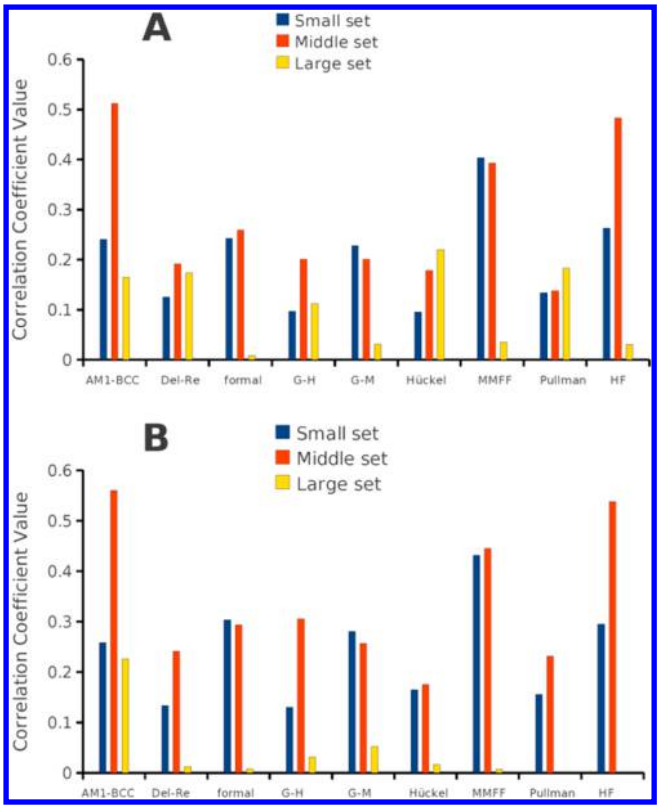


Figure 8. Correlation coefficient of different charge methods in three subsets: (A) top score conformations; (B) best pose conformations.

group, whereas they were much worse in both the small and middle groups. Similar results can be found in both Del-Re and GH charges. Other charge methods including formal and GM were mediocre, with little difference from results in the whole test set. In this case, results were similar for top score and best pose, except in the large group, which is better in top score than best pose.

It should be mentioned that no charge method, even the ab initio Hartree−Fock, could handle the large group and the correlation coefficient values were lower than 0.2. In fact, considering the restriction in computational resources, the score functions in current docking programs are relatively simple, intending to get better performance within a short time. Consequently, for large ligands that could create more contacts with a protein target, the accurate calculation is much obstinate. Furthermore, AutoDock4.2 scoring function does not include the ADME properties of ligands de facto, which is the crucial aspect of in vivo ligand activity. It is much unpliable for large molecules to penetrate a cell membrane then arrive at their protein targets, which can considerably decrease pre se activity and lead to the overestimation of large molecules in the

dx.doi.org/10.1021/ci300417y | J. Chem. Inf. Model. 2013, 53, 188−200

AutoDock4.2 scoring function. This may explain the fact that correlations of large groups are much weaker in best pose (Figure 8B) than that in top score (Figure 8A), for the conformations of best pose are more similar with native ligand structures.

2.3. Estimating Binding Affinities in Crystal Structure. As shown in the results above, docking results were quite different when using different charge assigning methods. So, it is important to detect how atomic charge works in the score function. In this section, we focused on how different atomic charges influence the docking scores of AutoDock4.2. The binding free energies were estimated using AutoDock4.2 score function with the structure of ligand remaining the same as that in crystal structure. Results of 52 complexes were collected in Supporting Information Tables S6−S14.

2.3.1. General Performance of Nine Charge Methods. The estimated binding free energies were calculated following eq 3. In the current study, the unbound structure of ligand is the same as the bound state (crystal structure), so the E_{Internal} is equal to E_{Unbound}, and they do not contribute to the total energy. On the other hand, the $E_{\text{Torsional}}$ is calculated only based on the number of torsinal bond in ligand, so this term remains the same in each complex. It is clear that there are significant differences between nine charge methods in the estimated binding free energies, so the difference should come from the $E_{\text{Intermolecular}}$, including energies of dispersion/repulsion (vdW), hydrogen bonding (hbond), desolvation potential (desolv), and electrostatic interactions (elec). It should be mentioned that the output file of AutoDock4.2 only provides the total value of vdW, hbond, desolv. But as we have discussed before, only energies of desolvation and electrostatic potential were related to atomic charges, so the differences in $E_{\text{vdW+hbond+desolv}}$ of nine charge assigning methods could represent the differences in E_{desolv}.

$$E_{\text{binding}} = E_{\text{Intermolecular}} + E_{\text{Internal}} + E_{\text{torsional}} - E_{\text{Unbound}} \quad (3)$$

$$E_{\text{Intermolecular}} = E_{\text{vdW+hbond+desolv}} + E_{\text{elec}} \quad (4)$$

Results have shown that there were indeed differences in calculating $E_{\text{vdW+hbond+desolv}}$ and E_{elec}, and the former one is more sensitive to different charge methods (see Surporting Information Table S6−S14). AM1-BCC charge methods also performed best in binding energy calculation of crystal structure, while other charge methods shown no significant difference. It is interesting that when we try to remove the E_{elec} form the total energy, all charge methods provided more accurate results, especially GH, GM, and Pullman. It seems that the method used in calculating electrostatic interactions in AutoDock4.2 score function still need to be improved for accurate predicting binding energies.

Then we evaluated the energy errors in estimating binding energies of crystal structures and found that AM1-BCC, MMFF, and Hartree−Fock charge methods always underestimate the binding energy by about 3−6 kcal/mol; however, some energies were overestimated when using GH, GM, Del-Re, and Pullman charges. The underestimation and overestimation were equal in formal and Hückel (see Supporting Information Figure S1). Similar results in docking studies have already been shown in previous sections (Figure 7). It is not surprising that some charge methods behaved similar to each other when we consider the basis of each charge method. The AM1-BCC have been parametrized against the HF/6-31G* electrostatic potential while the parameters used the in MMFF are also based on HF/6-31G* calculation dates. GH combines the σ charge of GM and the π charge of Hückel, and Pullman also combined Hückel π charge with Del-Re σ charge. The little difference of predicting binding energies in GM, compared with GH, and Pullman, compared with Del-Re, indicated that the introduction of the Hückel π charge slightly effects the accuracy of binding energy estimation. The similarity in the Hückel and formal charge methods may come from the fact that both of them only assign fixed charges on special atoms, and in many cases, they compute the same atomic charge.

Generally, the more overestimations in GH charge, compared with the AM1-BCC method, could make GH obtain more accurate results, because both of them often underestimate the E_{binding}. But it is of particular relevance here that, though better predication could be found in 42 complexes using GH charge method as to AM1-BCC, GH failed in the remaining 10 complexes with E_{binding} overestimated. The energy data were more dispersed in GH and this indeed resulted in bad performance of GH when using the correlation coefficient value (R^2). Similar failures could be found in other empirical charge methods. We further checked the ligands in the ten failed complexes and found that most of them contained some special chemical groups (usually phosphate). So there comes another significant conclusion that the empirical charge method is more sensitive to the ligand structure, whereas the AM1-BCC is more stable. Then, we deleted eight complexes in which the ligand possesses phosphate groups, and the correlation coefficient value (R^2) of all empirical charge method increased, while AM1-BCC and MMFF remained more or less unchanged. Only the Hartree−Fock (6-31G*) charge method decreased. This result further strengthened our former conclusion that empirical charge methods are more sensitive to differernt ligand structures compared with AM1-BCC and MMFF, and often lead to a bad result when used in the docking of special ligand, like ATP (contain phosphate groups). Moreover, this kind of unsuitable may limit the use of AutoDock4.2 in biological fields. Fortunately, the AM1-BCC and MMFF have been proved to get better results in ligand with phosphate groups, and the AM1-BCC charge method has also been identified as the most accurate in formal docking studies in section 2.1.

2.3.2. How Atomic Charge Affect the Binding Energy Calculation. It has already been shown that when comparing the estimated binding energy of each charge method, empirical charge method (for example GH) often leads to a higher estimated binding energies compared to these quantum chemical based method AM1-BCC, MMFF, and HF (6-31G*). In order to explore this feather, we evaluated the relationship of atomic charges with calculated energies on every atom in ligand and found out that it is the different charge assigned to atoms in function groups (O, N, P atom and neighbor C atoms) that lead to the differences in binding energy calculation. For example, in 2QWD different atomic charges were assigned on O, N, and C atoms using AM1-BCC and GH charge method, and finally, the GH computed a more accurate binding energy (Table 3; Figure 9). In general, the more positive charge on C1 in carboxyl group, more negative charge on N1 in amino group and more negative charges on O atoms in carbonyl and hydroxyl group using the AM1-BCC charge method contribute most to the errors in binding energy estimation.

dx.doi.org/10.1021/ci300417y | J. Chem. Inf. Model. 2013, 53, 188−200

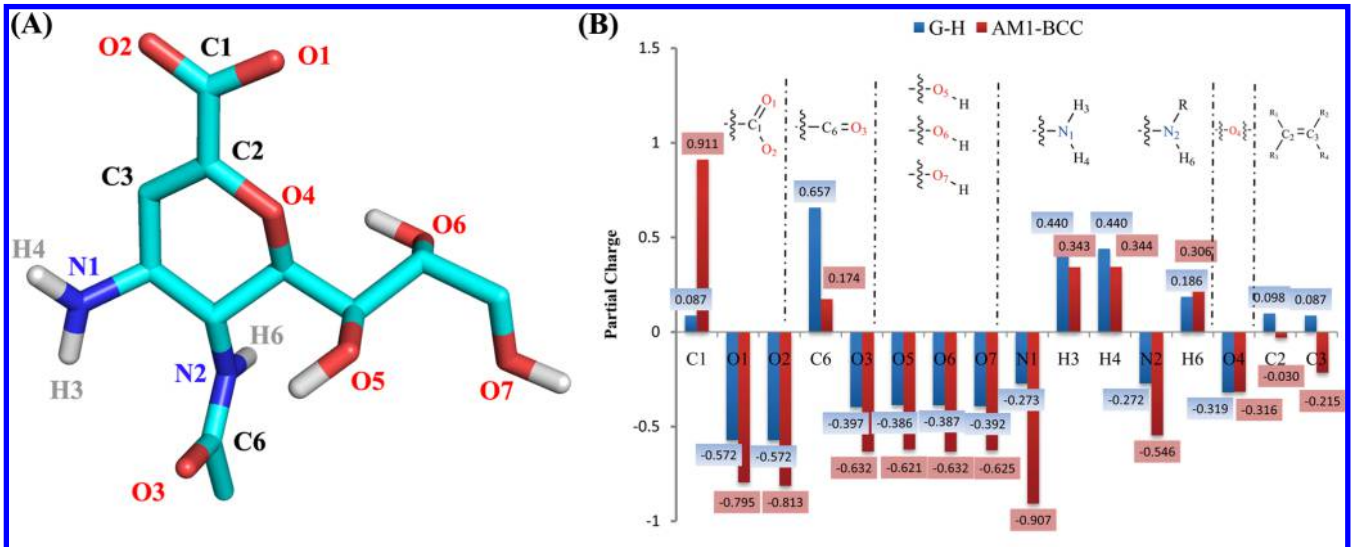
Table 3. Correlation between Binding Energy Calculation of Crystal Structure and Experimental Data

	correlation coefficient value		
charge	$E_{binding}$ of crystal structure	$E_{binding} - E_{elec}$	$E_{binding}$ (omit ligands that possess phosphates)
AM1-BCC	0.351	0.421	0.361
Del-Re	0.248	0.383	0.306
formal	0.292	0.365	0.326
GH	0.267	0.415	0.341
GM	0.289	0.422	0.346
Hückel	0.253	0.358	0.360
MMFF	0.250	0.359	0.261
Pullman	0.264	0.406	0.339
Hartree–Fock	0.216	0.262	0.169

However, when it comes to ligands containing special function groups like phosphate, the AM1-BCC perform better than GH. For example, in 1M0Q both AM1-BCC and GH charge overestimated the binding energy, but the result of AM1-BCC is more close to experimental data (Table 4). As shown in Figure 10, the different atomic charge of P1 and P2 contribute the most in the overestimation of GH. In addition, the polar hydrogen atoms H2 in amino group and H11 in hydroxyl group also pushed total energy away from experimental data in GH. It is not surprising because AM1-BCC has already been successfully used in the nucleic acids

Table 4. Estimated Energy using AM1-BCC and GH Charge Method in 2QWD and 1M0Q

	estimated energy (kcal/mol)		
PDB code	AM1BCC	GH	experimental energy (kcal/mol)
2QWD	−2.06	−6.67	−6.62
1M0Q	−5.02	−8.88	−3.03

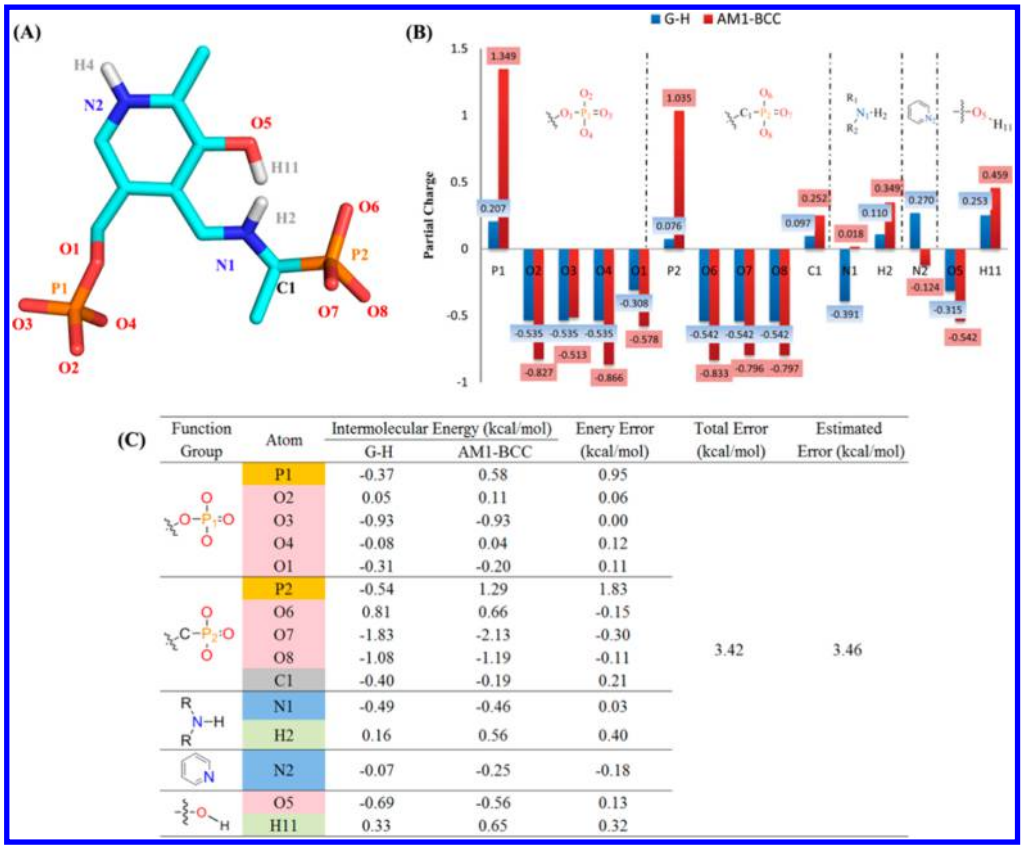


(C)

Function Group	Atom	Intermolecular Energy (kcal/mol) G-H	Intermolecular Energy (kcal/mol) AM1-BCC	Enery Error (kcal/mol)	Total Error (kcal/mol)	Estimated Error (kcal/mol)
	C1	-0.41	0.72	1.13		
	O1	-1.73	-1.88	-0.15		
	O2	-1.18	-1.26	-0.08		
	C6	-0.22	-0.45	-0.23		
	O3	-0.65	-0.51	0.14		
	O5	0.10	0.25	0.15		
	O6	0.28	0.55	0.27		
	O7	0.08	0.38	0.30		
	N1	0.38	2.04	1.66	4.19	4.61
	H3	-0.76	-0.71	0.05		
	H4	-1.29	-1.10	0.19		
	N2	0.05	0.44	0.39		
	H6	-0.56	-0.59	-0.03		
	O4	0.07	0.07	0.00		
	C2	-0.18	-0.18	0.00		
	C3	-0.41	-0.01	0.40		

Figure 9. Different Atomic Charge and Intermolecular Energy Using AM1-BCC and GH in 2QWD.

dx.doi.org/10.1021/ci300417y | J. Chem. Inf. Model. 2013, 53, 188−200



Function Group	Atom	Intermolecular Energy (kcal/mol) G-H	Intermolecular Energy (kcal/mol) AM1-BCC	Enery Error (kcal/mol)	Total Error (kcal/mol)	Estimated Error (kcal/mol)
	P1	-0.37	0.58	0.95		
	O2	0.05	0.11	0.06		
	O3	-0.93	-0.93	0.00		
	O4	-0.08	0.04	0.12		
	O1	-0.31	-0.20	0.11		
	P2	-0.54	1.29	1.83		
	O6	0.81	0.66	-0.15		
	O7	-1.83	-2.13	-0.30		
	O8	-1.08	-1.19	-0.11	3.42	3.46
	C1	-0.40	-0.19	0.21		
	N1	-0.49	-0.46	0.03		
	H2	0.16	0.56	0.40		
	N2	-0.07	-0.25	-0.18		
	O5	-0.69	-0.56	0.13		
	H11	0.33	0.65	0.32		

Figure 10. Different atomic charge and intermolecular energy using AM1-BCC and GH in 1M0Q.

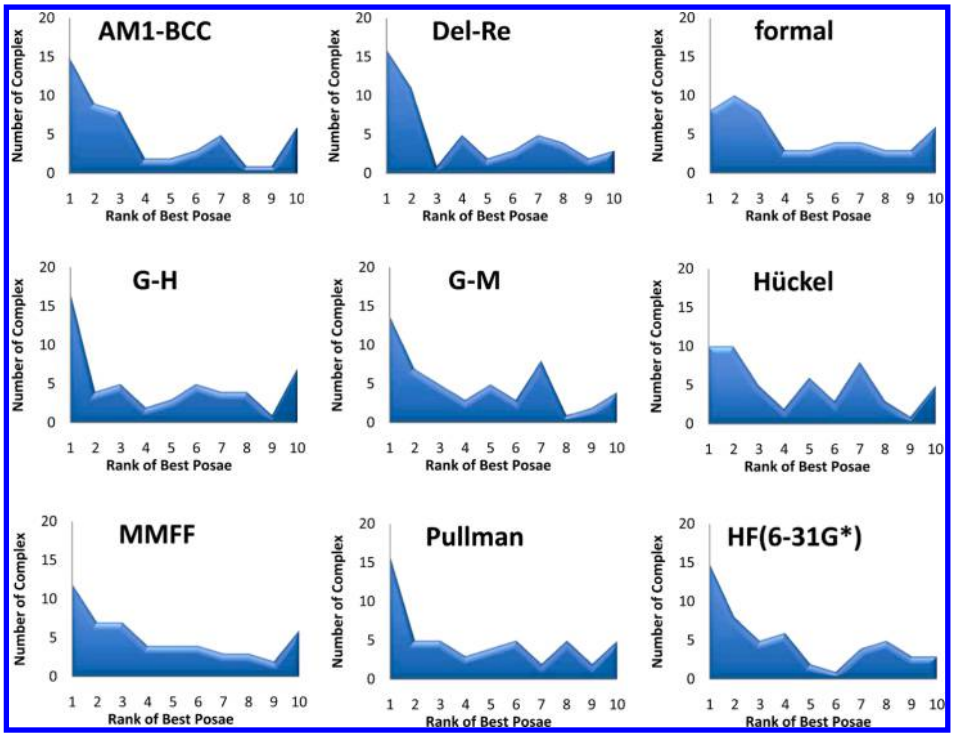


Figure 11. Numbers of complexes in each rank of best pose conformation using different charge methods.

docking,[43] and another study has shown that the positive charge for the phosphorus atom is more suitable in simulation.[44] In addition, Illingworth and co-workers also found that Gasteiger charge was not suitable in docking $NADP^+$ (nicotinamide adenosine diphosphate), when compared with a QM/MM based induced charge methods.[45]

Similar results could be found in other charge models. The MMFF and Hartree−Fock charge methods calculated lower binding energies compared to empirical method (Del-Re, etc.),

dx.doi.org/10.1021/ci300417y | J. Chem. Inf. Model. 2013, 53, 188−200

due to the different atomic charge assigned to O, N, C, and P atoms.

2.4. Evaluation of Pose Ranking Functions. Finally, we focused on the question that whether AutoDock4.2 scoring function could order conformations correctly based on their RMSD value; in other words, whether the Best Pose could obtain the highest score. It is relatively influential because scoring functions of docking program were not only used to calculate binding energies but also to guide algorithms during docking process. As mentioned in former sections, though genetic algorithms are applied to generate new conformations, it is the scoring function that decides which conformations will be accepted for the next run, whereas the weak scored ones are discarded.

We identified the "number of first hits" to reflect the frequency of best pose conformation rank first in energy estimation. As presented in Figure 11, we can come across distinctive charge methods achieved similarly in this evaluation, with only moderate results. In our study, five charge methods, including GH, Del-Re, Pullman, AM1-BCC, and Hartree−Fock, have indicated preponderance among all nine charge methods. However, the performance here was not very reasonable, considering the fact that the GH charge, which performed best, barely got 17 first hits among all 52 complexes. The bad result in this section may have some relationship with limited docking runs in current studies. In our study, default parameters of AutoDock4.2 were used with only 10 docking runs, considering the suitability when applied in high-throughput virtual screening approaches. Therefore, the advantages of LGA were not tapped fully. However, increasing the number of docking runs would be a good idea in future studies, along with the improvement of computational sources. Generally, the application of GH and AM1-BCC charge methods leads to plausible results as to other charge methods, which is in line with former evaluations.

3. CONCLUSIONS

In summary, our goal was to explore the effect of different charge-assignment methods on docking accuracy of AutoDock4.2 by using a benchmark database containing 52 complexes. Nine charge methods, including the semiempirical charge AM1-BCC, empirical charge Del-Re, formal, Gasteiger−Marsili, Gasteiger−Hückel, Hückel, Merck molecular force field, Pullman, and the ab initio Hartree−Fock, were selected for their popularity in current drug design. In our study, both pose prediction and scoring capabilities were well measured.

Generally, in terms of pose prediction, all nine charge methods represent applicable accuracy. The GH charge method stood by the highest success rate in pose prediction (82.69% in top score; 96.15% in best pose) and also grasped the highest number in best prediction (<1 Å). Moreover, GH charge behaved the best in all sizes of ligands, even large ligands with high flexibility. We should underline that the GM charge, which is often used to calculate ligand charge in AutoDock studies, was mediocre among all nine charge methods. The order of all nine charges could be described as GH > Pullman > Del-Re > AM1-BCC > GM > formal > Hückel > MMFF > Hartree−Fock.

In the evaluation of binding energy estimation, a relatively weak correlation was obtained between the docking score and in vitro measured activity of ligands when using semiempirical, empirical and even ab initio charge methods. To some extent, all nine charge methods involved in our studies were not able to calculate binding energies accurately. It should be mentioned that partial charges are hard to calculate for the fact that while partial atomic charges are very intuitive, they do not really exist. Partial charges are only simplified models of the true electrostatic potential energy in molecules. In general, the semiempirical charge method AM1-BCC, ab initio charge method Hartree−Fock and empirical charge method MMFF performed better correlations, with significant superiorities in comparison with other charge methods. We should also highlight that, although Hartree−Fock is commonly considered to have high accuracy, good docking results as expected were not received. For that reason, in terms of the time consumption and low accuracy, the ab initio Hartree−Fock method is not suitable in docking studies using AutoDock4.2. The performance of each charge method in different size of ligand was considered afterward. The AM1-BCC performed better in both the middle and large groups, while Hartree−Fock performed well in the middle group, and MMFF in the small group. Furthermore, studies in this section manifested that current charge methods failed to predicate the binding strength of large size ligands with high flexibility.

The results in estimating binding energy using crystal structure has shown that all nine charge method intended to underestimate the binding energy, especially the AM1-BCC, MMFF, and Hartree−Fock. Differnet charge methods lead to different calculated binding energies because, in the AutoDock 4.2 score function, the atomic charges play an important role in the calculation of desolvation potential (E_{desolv}) and electrostatic interactions (E_{elec}). It is interesting that when we try to remove E_{elec} from total scores, there were slight improvments in predicted binding energies of all nine charge methods. It should be mention that the authors of AutoDock 4.2 have already declare that the result was not significantly affected if the electrostatic term was removed during the optimization of the force field.[24] So the method to calculate electrostatic interactions still need to be improved. In addition, we also found that AM1-BCC is very stable in different complexes (always underestimated the binding energies by about 3−6 kcal/mol), while GH is more sensitive to different ligand strucutres. These could partly explain the better performance of AM1-BCC in predicted binding energy than that of GH.

A more detailed study on relationships of atomic charges and score calculations was performed. We further come to a conclusion that the different atomic charge assigned on atoms, in special function groups, does contribute to the errors in total energy calculation. Of particular relevance here that when it comes to ligands containing phosphate groups, it seems that emprical charge methods (GH charge or others) could not handle them well, while AM1-BCC and MMFF achieved. Here we come to another conclusion that the AM1-BCC and MMFF charge method are more suitable in docking molecules containing phosphate, which is very important in the further applications of AutoDock4.2 in biological fields. Previous study has also shown that AutoDock was successfully used in nucleic acids docking, when using AM1-BCC to calculate the atomic charge of ligand.[43]

In this way, how to correctly choose charge method is very important for those studies using the AutoDock4.2 program. We have already proven that there are significant differences in docking accuracy of AutoDock4.2 when using different charges. Because of the outstanding performance in pose prediction with limited ten docking runs, the Gasteiger−Hückel charge method is recommended when AutoDock4.2 is implemented to

virtually screen millions of compounds (e.g., The ZINC database[46]). Meanwhile, the GH charge method also has the advantage of being fast and high popularity in current drug discovery. On the other hand, when we intend to carry out lead compound optimization approaches or accurate virtual screening a small amount compounds, more docking runs are preferred in order to obtain higher accuracy. At this point, accurate binding energy calculations become more important for the fact that in conformation generations using LGA, it is the calculated binding energies that decide which pose will be sent to the next LGA run. So the semiempirical AM1-BCC charge method is recommended for its high accuracy in not only pose prediction but also binding energy estimation. Moreover, when the AutoDock4.2 is employed in the biological field, which contains many phosphate groups, the AM1-BCC charge is recommended again. In addition, despite the fact that the MMFF charge method obtained promising results in scoring function evaluation, especially in small size ligands, the lowest success rate in pose prediction (69.23%) restricts its further usage in virtual screening.

4. MATERIALS AND METHODS

4.1. Benchmark Database. The benchmark database used in our study was selected from a test set that have been used in a recent papers as well as the training set of the AutoDock scoring function.[41,47] For most cases, the protein structure is extracted from X-ray crystal of a protein−ligand complex, because it allows people to omit protein structural changes during the process of binding the same ligand. Additionally, the position of the active site is easy to identify in a protein−ligand complex. Crystal structures of all 52 protein−ligand complexes involved in this study were obtained from the Brookhaven Protein Data Bank with experimental binding affinities taken from the PDBBind Database[48](see Supporting Information, Table S1). To perform an effective redocking procedure, several criteria were described as follows:

(i) Structural diversity in both ligands and proteins for ensuring that the benchmark is typical and possesses universal applicability.
(ii) No covalent binding between protein and ligand.
(iii) Crystallographic resolution must be lower than 3.2 Å for previous studies have shown that using poor resolution structures may result in incorrect conformations of ligands.[49]

Furthermore, the binding site was checked carefully when the ligand was found to interact with more than one asymmetric unit, for ensuring the biological unit in our study.

4.2. Preparation of Ligands and Proteins. In the present work, the receptors were treated rigid, while the ligands were flexible. The active rotatable bond ranges from 0 to 32. All ligands were extracted from crystal structures of complexes with bond types and atom types assigned manually, and hydrogens were added if missing. Empirical charge assignments were performed using the method of Del-Re, formal, GM, GH, Hückel, MMFF, and Pullman in the SYBYL 7.3 package.[50] Semiempirical charge assignments were calculated using the AM1-BCC method in the Chimera program.[51] The Gaussian 09 program was used to perform standard Hartree−Fock (HF/6-31G*) charges.[44] All charge-assignment methods were carried out with no geometry optimization, in order to keep the initial crystal coordinates of the ligand.

For proteins, extra structures such as sulfate, covalently linked sugars, and halogens were deleted. It is important to mention that keeping the water molecules in or around the active binding site is essential in improving docking accuracy.[52] Therefore, in this case, water molecules in or around the active binding site were conserved while others were removed for minimizing computational time. Furthermore, some cofactors, including ATP, HEME, and NADPH, were kept as well as metal ions, and their atom types and bond types were assigned manually. Gasteiger partial charges were computed, and hydrogen atoms were added in protein residues in AutoDockTools. No extra optimization of the protein structures was performed in this case.

4.3. Evaluation Methods. Commonly, evaluating the performance of docking programs needs to consider two characteristics: (i) the ability to determine the most similar conformation compared with the ligand in holo form crystal structure and (ii) the capacity to accurately estimate the binding energy between ligands and proteins.

At first, several researchers tried visual inspection of the obtained ligand conformations to evaluate the docking result. However, the empirical approach makes it difficult to compare their results with others. That is one of the important reasons why more and more researchers choose RMSD (root mean square distance) as the main benchmark in describing docking accuracy. Although it may not be ideal, it is the only widely acceptable and reliable value with high comparability. Furthermore, docking programs propose many solutions with single ligand imported, and thereafter, we checked not only the RMSD of conformation with highest docking score (top score), but also RMSD of conformation closest to native structure (best pose).[38] To fully evaluate the AutoDock4.2 program, we also calculated the number of successfully docked pairs (success rate), which is defined as the ratio of complexes for which top score or best pose conformations are below the given threshold in comparison with all evaluated pairs. In this case, the RMSD must be below 2 Å. In general, an RMSD value less than 2 Å is widely accepted as accurate in molecular docking,[53] probably because the resolution in X-ray crystal structure analysis is often around 2 Å, and higher precision than it is not meaningful. In addition, the best (RMSD < 1 Å), moderate (1 Å < RMSD < 2 Å), and the worst predictions (RMSD > 2 Å) were defined.

Another purpose of our study was to assess the accuracy of binding energy estimation. In AutoDock4.2, the binding energy of each output confirmation was determined by its scoring function. The most commonly used method was the comparison of the experimental values with the estimating energies by calculating correlations. In our study, the quality of AutoDock score function was checked by calculating the correlation between the estimated binding energy and experimental values of top score conformations. Nevertheless, sometimes the Top Score conformation does not present the best geometry, so we also calculate the correlation using the best pose conformations. In addition, when Lamarckian genetic algorithm (LGA) is used in generating new conformations, it is the scoring function of AutoDock4.2 that determins which conformation will be send to the next LGA run. So, correct scoring of the best pose conformation will lead to a more accurate docking result. Thus, we also defined the "number of first rate", which presents the percentage of best pose conformation ranking first in all resulting conformations.

To further evaluate how different atomic charges influence the total energy in the AutoDock4.2 scoring function, the

binding energy of ligand in crystal structure were calculated in order to avoid the influence of LGA conformation search.

4.4. Docking Setting. In view that we are comparing the difference between charge methods, it is substantial to guarantee each charge method was treated equally. For each complex, 10 docking experiments were performed by using the Lamarckian genetic algorithm in AutoDock4.2, with other default parameters set. The initial position, torsion, and orientation of candidate ligands were set randomly, and all rotatable torsions are released during docking runs. The AutoGrid program was employed in the construction of grid parameter files and atom-specific affinity maps. To make sure that the entire binding site was detected, atom-specific affinity map files were created using grid boxes depending on the shape of each ligand in crystal structure. Meanwhile, the grid boxes were centered on the root of ligand with a spacing of 0.375 Å. The RMSD between docked ligand structures, and the crystal ligand structures were calculated after each docking run.

■ ASSOCIATED CONTENT

Supporting Information

Descriptions of the 52 ligand−protein complexes selected in our benchmark database, and the complete statistical results produced by all docking tests. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Tel.: +86-531-8838-2731 (H.F.); +86-531-8838-2076 (M.L.). E-mail: haofangcn@sdu.edu.cn (H.F); mli@sdu.edu.cn (M.L.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from the National Program on Key Basic Research Project of China (No. 2013CB734002), the National Natural Science Foundation of China (No. 30901836, 81001362, and 21172133), the Program for New Century Excellent Talents in University (No. NCET-11-0306), the Shandong Natural Science Foundation (No. JQ201019), and the Independent Innovation Foundation of Shandong University, IIFSDU (No. 2010JQ005 and 2012JC002).

■ REFERENCES

(1) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* **2002**, *47* (4), 409−443.

(2) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43* (25), 4759−4767.

(3) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335−373.

(4) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004**, *56* (2), 235−249.

(5) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **2006**, *65* (1), 15−26.

(6) Hermann, J. C.; Marti-Arbona, R.; Fedorov, A. A.; Fedorov, E.; Almo, S. C.; Shoichet, B. K.; Raushel, F. M. Structure-based activity prediction for an enzyme of unknown function. *Nature* **2007**, *448* (7155), 775−779.

(7) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004**, *57* (2), 225−242.

(8) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8* (3), 195−202.

(9) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **1996**, *9* (1), 1−5.

(10) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161* (2), 269−288.

(11) Morris, G. M.; Goodsell, D. S.; Huey, R.; Olson, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des.* **1996**, *10* (4), 293−304.

(12) Moustakas, D. T.; Lang, P. T.; Pegg, S.; Pettersen, E.; Kuntz, I. D.; Brooijmans, N.; Rizzo, R. C. Development and validation of a modular, extensible docking program: DOCK 5. *J. Comput.-Aided Mol. Des.* **2006**, *20* (10−11), 601−619.

(13) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48* (20), 6504−6515.

(14) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52* (4), 609−623.

(15) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739−1749.

(16) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47* (7), 1750−1759.

(17) Jain, A. N. Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21* (5), 281−306.

(18) Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.* **2003**, *46* (4), 499−511.

(19) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, *37* (2), 228−241.

(20) Chen, R.; Li, L.; Weng, Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins* **2003**, *52* (1), 80−87.

(21) Plewczynski, D.; Lazniewski, M.; von Grotthuss, M.; Rychlewski, L.; Ginalski, K. VoteDock: consensus docking method for prediction of protein-ligand interactions. *J. Comput. Chem.* **2011**, *32* (4), 568−581.

(22) Krüger, D. M.; Jessen, G.; Gohlke, H. How Good Are State-of-the-Art Docking Tools in Predicting Ligand Binding Modes in Protein−Protein Interfaces? *J. Chem. Inf. Model* **2012**, *52*, 2807−2811.

(23) 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* (16), 2785−2791.

(24) 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* (6), 1145−1152.

(25) Tsai, K. C.; Wang, S. H.; Hsiao, N. W.; Li, M.; Wang, B. The effect of different electrostatic potentials on docking accuracy: a case study using DOCK5.4. *Bioorg. Med. Chem. Lett.* **2008**, *18* (12), 3509−3512.

(26) Fox, T.; Kollman, P. A. Application of the RESP methodology in the parametrization of organic solvents. *J. Phys. Chem. B.* **1998**, *102* (41), 8070−8079.

(27) Sanderson, R. T. Electronegativity and bond energy. *J. Am. Chem. Soc.* **1983**, *105*, 2259−2261.
(28) Del Re, G. A simple MO-LCAO method for the calculation of charge distributions in saturated organic molecules. *J. Chem. Soc.* **1958**, 4031.
(29) Del Re, G.; Pullman, B.; Yonezawa, T. Electronic structure of the α-amino acids of proteins: I. Charge distributions and proton chemical shifts. *Biochim. Biophys. Acta* **1963**, *75*, 153.
(30) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity-a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(31) Gasteiger, J.; Marsili, M. A new wodel for calculating atomic charges in molecules. *Tetrahedron Lett.* **1978**, *19*, 3181−3184.
(32) Berthod, H.; Giessner-Prettre, C.; Pullman, A. Sur les rôles respectifs des électrons σ et π dans les propriétés des dérivés halogénés des molécules conjuguées. Application à l'étude de l'uracile et du fluorouracile. *Theor. Chem. Acta.* **1967**, *8*, 212−222.
(33) Purcell, W. P.; Singer, J. A. A brief review and table of semiempirical parameters used in the Hueckel molecular orbital method. *J. Chem. Eng. Data* **1967**, *12*, 235.
(34) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23* (16), 1623−41.
(35) Halgren, T. A. Merck molecular force field. 2. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.
(36) Halgren, T. A. Merck molecular force field. 1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(37) Halgren, T. A. Merck molecular force field. 3. Molecular geometries and vibrational frequencies for MMFF94. *J. Comput. Chem.* **1996**, *17*, 553−586.
(38) Plewczynski, D.; Lazniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **2011**, *32* (4), 742−755.
(39) Wang, R.; Lu, Y.; Fang, X.; Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein−ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (6), 2114−2125.
(40) Wang, J.; Kollman, P. A.; Kuntz, I. D. Flexible ligand docking: A multistep strategy approach. *Protiens* **1999**, *36* (1), 1−19.
(41) Bikadi, Z.; Hazai, E. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminform.* **2009**, *1*, 15.
(42) Spyrakis, F.; Amadasi, A.; Fornabaio, M.; Abraham, D. J.; Mozzarelli, A.; Kellogg, G. E.; Cozzini, P. The consequences of scoring docked ligand conformations using free energy correlations. *Eur. J. Med. Chem.* **2007**, *42* (7), 921−933.
(43) Holt, P. A.; Chaires, J. B.; Trent, J. O. Molecular docking of intercalators and groove-binders to nucleic acids using Autodock and Surflex. *J. Chem. Inf. Model.* **2008**, *48* (8), 1602−1615.
(44) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(45) Illingworth, C. J.; Morris, G. M.; Parkes, K. E.; Snell, C. R.; Reynolds, C. A. Assessing the role of polarization in docking. *J. Phys. Chem. A.* **2008**, *112* (47), 12157−12163.
(46) Irwin, J. J.; Shoichet, B. K. ZINC-A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177−182.
(47) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19* (14), 1639−1662.
(48) Wang, R.; Fang, X.; Lu, Y.; Yang, C.; Wang, S. The PDBbind Database: Methodologies and updates. *J. Med. Chem.* **2005**, *48* (12), 4111−4119.
(49) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727−48.
(50) *SYBYL 7.3*; Tripos International: St. Louis, Missouri; 2006.
(51) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. Chimera-A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605−1612.
(52) Thilagavathi, R.; Mancera, R. L. Ligand-protein cross-docking with water molecules. *J. Chem. Inf. Model.* **2010**, *50* (3), 415−21.
(53) Onodera, K.; Satou, K.; Hirota, H. Evaluations of molecular docking programs for virtual screening. *J. Chem. Inf. Model.* **2007**, *47* (4), 1609−18.