JOURNAL OF CHEMICAL INFORMATION AND MODELING

Letter

pubs.acs.org/jcim

Reranking Docking Poses Using Molecular Simulations and Approximate Free Energy Methods

G. Lauro,[†] N. Ferruz,[‡] S. Fulle,[§] M. J. Harvey,[∥] P. W. Finn,*[,⊥,#] and G. De Fabritiis*[,‡]

[†]Dipartimento di Farmacia, Università degli Studi di Salerno, Via Giovanni Paolo II 132, 84084 Fisciano (SA), Italy

[‡]Computational Biophysics Laboratory (GRIB-IMIM), Universitat Pompeu Fabra, Barcelona Biomedical Research Park (PRBB), C/Doctor Aiguader 88, 08003 Barcelona, Spain

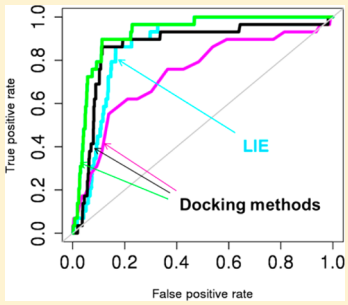
[§]InhibOx Limited, Oxford Centre for Innovation, New Road, Oxford OX1 1BY, United Kingdom

[∥]Acellera, Barcelona Biomedical Research Park (PRBB), C/Doctor Aiguader 88, 08003 Barcelona, Spain

[⊥]Department of Applied Computing, University of Buckingham, Hunter Street, Buckingham MK18 1EG, United Kingdom

[#]InhibOx Limited, Oxford Centre for Innovation, New Road, Oxford OX1 1BY, United Kingdom

Supporting Information



ABSTRACT: Fast and accurate identification of active compounds is essential for effective use of virtual screening workflows. Here, we have compared the ligand-ranking efficiency of the linear interaction energy (LIE) method against standard docking approaches. Using a trypsin set of 1549 compounds, we performed 12,250 molecular dynamics simulations. The LIE method proved effective but did not yield results significantly better than those obtained with docking codes. The entire database of simulations is released.

Computational modeling techniques are becoming well established in the standard toolkit of drug design.[1] In silico virtual screening of large compound libraries is now routine in the early stages of the drug discovery process.[2]

Although the development of universally applicable docking scoring functions still remains a challenge, the family of molecular docking methods[3] has proven very successful, representing an effective compromise between computational cost and quality of results. In most studies, while compounds are usually treated as flexible, the protein counterpart is modeled as a static structure or with very limited flexibility. On the other hand, it is well known that biological macromolecules often show significant movements before and after the binding of small molecules, and accordingly, docking studies using protein conformation ensembles have emerged (ensemble docking).[4] Additionally, the absence of an explicit treatment of the solvating water molecules can strongly influence both sampling and scoring phases. The use of weighted scoring averages from different docking tools has also been shown to be a good technique to improve the ranking.[5−7] These aspects are of great relevance in any virtual screening study, where inaccurate scoring functions could produce a final ranking with a large number of false positives and—more importantly—false negatives, the latter disadvantageously excluded from the subsequent phases of chemical synthesis and biological validation. Thus, high quality prediction of protein−ligand complex poses and binding affinities is critically important in providing data of sufficient accuracy to effectively guide an experimental program, and improvement and development of more accurate computational tools are still needed.[8]

In this context, molecular dynamics (MD) simulation,[9,10] with its innate ability to sample ligand and receptor conformational states, as well as to provide an atomistic treatment of solvent effects, would seem to have much to offer. To date, several different MD approaches for binding energy estimation have been reported.[11,12]

Given the aforementioned theoretical limitations of molecular docking and the recent improvements in the MD field (which have substantially reduced its computational cost, rendering it amenable to use in a high-throughput mode[13]), the main aim of this study was to compare molecular docking against the linear interaction energy[14] (LIE) method for ranking prediction for a large data set of compounds. Although the capabilities of the LIE method have been extensively tested in the past,[15−17] probing its suitability as a high-throughput virtual screening tool is still relatively new. For these reasons, we have considered the binding of 1549 compounds extracted from the trypsin set of the DUD database[18] (see S1 of the Supporting Information PDF file for the library design) and further compared performance against molecular docking results.

Given a putative binding process, the LIE approach (as with other approaches such as MM-PBSA[19] and MM-GBSA[20]) considers only the two endpoints of the reversible binding

Published: July 21, 2014

   dx.doi.org/10.1021/ci500309a | J. Chem. Inf. Model. 2014, 54, 2185−2189

cycle: the *free* state (compound solvated in water) and the *bound* state (compound in complex with the protein).

From a theoretical point of view, LIE predicts the binding free energy of compound−protein complexes using the following equation

$$\Delta G_{\text{bind}} = \alpha(\langle U_{\text{c-s}}^{\text{vdW}}\rangle_{\text{b}} - \langle U_{\text{c-s}}^{\text{vdW}}\rangle_{\text{f}}) + \beta(\langle U_{\text{c-s}}^{\text{el}}\rangle_{\text{b}} - \langle U_{\text{c-s}}^{\text{el}}\rangle_{\text{f}}) + \gamma \quad (1)$$

where the brackets indicate the means of the compound-surrounding (c-s) interaction energies (el for electrostatic, vdw for nonpolar, and b and *f* indicate bound and free states of the compound, respectively), which are calculated based on MD trajectories of the two states of the compound (Figure 1). α and

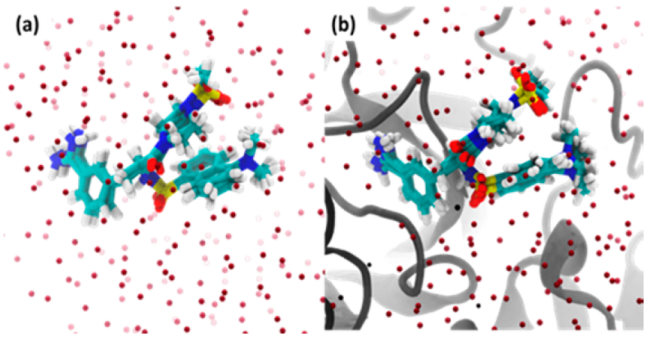


Figure 1. Representation of the two systems considered for a typical LIE calculation. (a) Unbound compound in water. (b) Compound−protein complex. Water molecules are represented in red, compounds in cyan, and proteins in gray.

β are the related scaling factors that could vary depending on the specific system considered, and γ is an additive factor generally weighted to fit the experimental binding affinities.[21]

The initial library contained 1713 compounds from the DUD database[18] comprising 50 ligands and 1663 decoys. From this set, 140 corresponded to duplicates of some compounds in different protonation states. We selected only one representative compound for each of the duplicates giving a set of 1573 molecules. After visual inspection, we discovered some erratic overprotonated amidine compounds (Figure S1, Supporting Information, PDF file). Twelve of the ligands presented this functional group and were also removed from the database.

A total of 1561 compounds were then prepared for MD simulations. The molecules were protonated at physiological pH and parametrized with antechamber 13.[22] The MD simulations were performed using the ACEMD[9] software on the GPUGRID.net[23] using the Amber99SB force field.[24] For each compound, four replicas of *bound* (∼16,000 atoms) and *unbound* (∼1500 atoms) fully solvated systems were run, resulting in 1549 compounds upon retrieval, a total of 123 μs of aggregate simulation time. These same molecules were also considered for docking calculations (see S2 of the Supporting Information PDF file for full details of the methods).

This final library comprised 32 ligands—active compounds (IDs from **1**-**32**)—able to bind trypsin with a known experimental binding affinity and 1517 decoys (IDs **33**-**1549**), namely, putative nonbinding molecules. Molecular docking experiments were performed using AutoDock Vina,[25] Glide,[26] and GOLD[27] (see the Supporting Information txt file for the final set of compounds). The comparison of the coordinates related to the best docking poses obtained from the three docking softwares used showed a modest convergence (see the RMSD plots in S3, S4, and S5 of the Supporting Information PDF file). We used GOLD top-ranked docking poses as starting LIE coordinates of the bound states. It is important to note that these initial structures are crucial and could influence the MD LIE results. In addition, these starting structures were computed following a semi-rigid docking protocol where the receptor is considered as static. Therefore, in order to reduce the influence of this sampling limitation on the MD calculations and let the compounds explore the most stable poses within their conformational space, production runs were done for 10 ns, whereas for each of the four replicas only the last 5 ns were considered for the energy calculations. We refer to this database as ACEMD-DUD-trypsin, which can be obtained upon request.

Several previous studies found that α, β, and γ LIE scaling factors are both protein and compound specific[28−31] and that their values can profoundly influence the reliability of the results. For example, the β value (related to the electrostatic contribution, which is particularly important in trypsin recognition by a putative ligand) is generally set to 0.5[14] but can be systematically optimized considering the chemical nature of the compounds investigated. Several efforts were made in this sense in previous studies, but they were mainly oriented to find a precise correlation between calculated and experimental binding affinities on a specific target considered more than to enrich the selection of active compounds in a general virtual screening context. For these reasons, four combinations of LIE scaling factors were chosen based on previous successful studies (S6, Supporting Information PDF file) allowing the analysis of how the modulation of vdW (LIE_1 vs LIE_2) and electrostatic contributions (LIE_1 vs LIE_2 vs LIE_3) influence the final ranking. The best LIE results were obtained using α = 0.18 (Table S2, Supporting Information PDF file), and following a previous study by Almlöf et al.,[31] the starting value of β was set to 0.43 (LIE_3). In addition, different increments or decrements of β were applied depending on the chemical functional groups of the compound under investigation (Table 1). In this way, the systematic over/under estimation of the electrostatic contribu-

Table 1. Summary of Data and Parameters Used for LIE Calculations

summary MD data		LIE parameters[a]		
simulation length	10 ns	α = 0.18	$\Delta\beta_i$ = −0.06	alcohols
			$\Delta\beta_i$ = −0.04	1°, 2° amines
aggregated simulation time	123 μs	$\beta = \beta_0 + ((\sum w_i \Delta\beta_i)/(\sum w_i))$	$\Delta\beta_i$ = −0.02	1° amides
			$\Delta\beta_i$ = −0.03	carboxylic acid
			$\Delta\beta_i$ = +0.02	anions
total simulations	12,250	γ = 0	$\Delta\beta_i$ = +0.09	cations

[a]Obtained from Almlöf et al.

dx.doi.org/10.1021/ci500309a | J. Chem. Inf. Model. 2014, 54, 2185−2189

tions derived from MD simulations is reduced with an internal system-specific mapping of the chemical moieties (Table S1, Supporting Information PDF file). Again, because the main aim of this study was to monitor the relative binding free energy rather than the absolute values, the scaling factor γ was set to 0. The comparisons for the four LIE parameters analyzed are presented in Table S2 of the Supporting Information PDF file.

The LIE ranks were then compared to those obtained from docking calculations (Table 2; see S7 of the Supporting Information PDF file for a complete set of the results). Overall, the LIE and Glide docking approaches reached similar AUC values of 0.87 and 0.86, respectively, whereas GOLD showed a remarkable value of 0.93 and AutoDock Vina exhibited a poorer performance, reaching an AUC value of 0.74 (curves, Figure 2).

Table 2. Comparison of LIE and Docking Methods Investigated in This Study[a]

	LIE	Glide	GOLD	AutoDock Vina
ROC AUC	0.87	0.87	0.93	0.74
EF 2%	3.0	1.5	3.0	6.1
EF 10%	3.7	6.5	7.8	4.3
EF 20%	4.2	4.4	4.5	3.42
% screened (75%)	16.0	11.6	7.0	31.0

[a]The first row compares the area under the ROC curve. The second, third, and fourth rows show enrichment factors at 2%, 10%, and 20% of the ranked database, respectively. The fifth row shows the screening percentages needed in order to recover 75% of the active ligands.

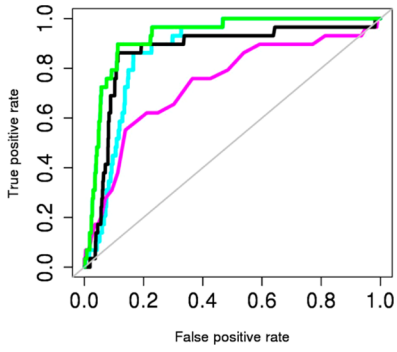


Figure 2. Receiver operating characteristic (ROC) curves for the four methods. Areas under the ROC curves are expressed in Table 2. Glide is shown in black, LIE in cyan, GOLD in green, and AutoDock Vina in magenta.

ROC curves describe the trade-off between selectivity and sensitivity throughout the whole database. However, the common practice is to advance only a small number of the top-ranked compounds for the following steps of the pipeline. Therefore, it is useful to consider performance measures that focus on the first portion of the rank-ordered scoring lists. The enrichment factor (EF) and true positive rate (TPR) evaluate the quality of the ranking methods at different percentages of the rank-ordered database list that are selected for screening.[32] EF is defined as

$$\mathrm{EF} = \frac{\mathrm{Hits}_S/N_S}{\mathrm{Hits}_T/N_T} \quad (2)$$

where Hits_S is the number of ligands (active compounds) in the sampled subset, Hits_T is the total number of ligands in the database, and N is the number of compounds.

TPR simply measures the proportion of active ligands in the chosen subset. Figure 3 shows a comparison of these statistics up to 20% of the database. It can be appreciated how for the low percentages of screening (at 5%), docking (GOLD) outperforms LIE when considering both EF (Figure 3a) and TPR (Figure 3b). At higher percentages of screening, the best docking methods and the LIE method perform similar, while differences with AutoDock Vina are more evident, especially when comparing TPR at 15−20% of the ranked database (Figure 3b). Quantitatively, Table 2 shows a comparison of EFs at the 2%, 10%, and 20% levels, whose theoretical maximum values are 48.4, 10, and 5, respectively. As expected from Figure 3, at the 2% level, Autodock Vina outperforms the other methods (EF 6.1), whereas LIE, GOLD, and Glide show lower values (3.0 for the first two and 1.5 for the latter). However, as shown in Figure 3, at higher percentage screening levels, the performance of the methods begin to converge. At 10%, Glide shows enrichment closer to GOLD, with an EF of 6.5 compared to GOLD's 7.8, while the other methods still lag behind (3.7 for LIE, 4.3 for AutoDock Vina). At the highest percentage considered of 20%, LIE, GOLD, and Glide are comparable (4.2, 4.5, and 4.4, respectively), and only AutoDock Vina shows a lower performance (3.4).

As noted above, identifying actives near the top of the rank-order list ("early enrichment") is important because in most real-world virtual screening scenarios only a relatively small number of compounds are selected for experimental testing because of practical and budgetary constraints. In this regard, it is noteworthy to mention that although showing a moderate performance at high percentages of screening, Autodock Vina

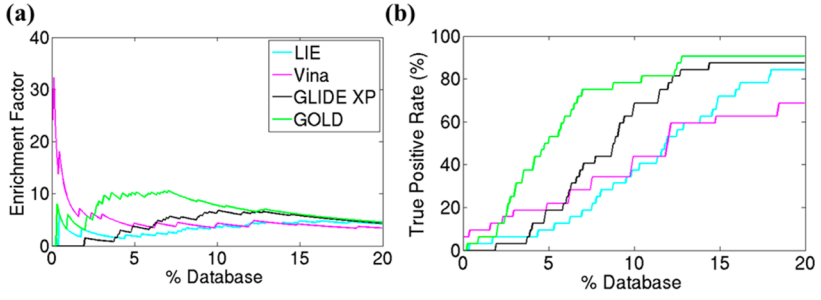


Figure 3. Comparison of enrichment factors (a) and true positive rates (b) for the four methods at different percentages of the screened database. The results show in general terms a better performance of GOLD at low percentages of screening but similar enrichment performance for LIE, Glide, and GOLD at percentages of 15% or above. Interestingly, Autodock Vina was the only method that detected true ligands in the first two positions of the ranking, leading to a sharp peak of 32.1% in panel a.

dx.doi.org/10.1021/ci500309a | J. Chem. Inf. Model. 2014, 54, 2185−2189

was remarkably the only method recovering true ligands at the first two positions of the ranking and identifying a total of three ligands in the top 10 positions. On the other hand, LIE and Glide sit toward the bottom of the performance range as shown with the other docking methods (Figure 4). If the first set of 50 compounds (~3% database) progressed to the following biophysical assays, GOLD would have recovered 10 ligands, whereas LIE and Glide would have recovered two and one compounds, respectively.

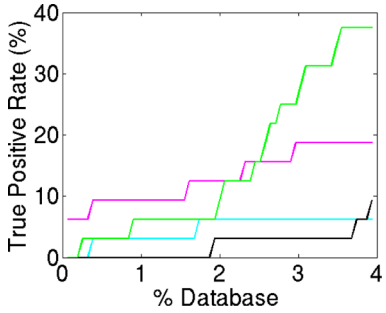


Figure 4. Comparison of true positive rates for the four methods at the early stages of screening, up to 4%.

Finally, in order to obtain an idea of the performance in retrieving a high percentage of active compounds, we have compared the percentages of the database required to be screened in order to recover 75% of active ligands. As indicated by the last row of Table 2, LIE shows a moderate value of 16.0% ,while the best docking code needs only 7.0% of the database to be screened.

In conclusion, we have reported a virtual screening application using the linear interaction energy (LIE) method on a large data set of compounds. Specifically, LIE demonstrated moderate predictive capabilities for true positives on the top-ranked database when compared with the best docking methods, and although showing satisfactory performance when screening more than 15% of the library, docking scoring functions show equal or better statistics, especially if we compare the number of compounds needed to recover 75% of active ligands. Considering the cost of setting up and performing the simulations, the data here seem to indicate minimal usefulness of the LIE method in identifying active ligands among a large set of related decoys at least for a protein−ligand with low flexibility, as is the case in this study. To give a quantitative indication of the differences in computational costs, we estimated 32 h of calculation time for docking experiments (Glide, XP mode) on an Intel Xeon E3-1245v2 @ 3.40 GHz vs 8 days for MD calculations on a 100 GPU cluster.

It is possible that the performance of LIE could be improved by further optimization of the set of LIE scaling factors (α, β, and γ). However, our aim was to investigate the potential performance of LIE as applied in a typical blind screening protocol. Specifically, in order to make the LIE docking comparison unbiased, we reproduced a screening procedure assuming no information a priori on the investigated systems. Interestingly, our results indicate that best LIE performance was obtained when the β scaling factors were dynamically modulated on the chemical functional groups of the compounds. This suggests that new, accurate, ligand-based computations of LIE scaling factors could be validated in a robust multi-target scenario.

Nevertheless, it is also possible that there are better ways to use the MD trajectory data of this study than the method used by us. In order to facilitate development of these methods, we release all the ACEMD-DUD-trypsin simulation data for further studies, in the hope that further work may improve on these results.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Details of library design, simulation and docking protocols, erratic compounds, comparison with other LIE parameters, and further analysis of the simulations. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: paul.finn@buckingham.ac.uk (P.F.).
*E-mail: gianni.defabritiis@upf.edu (G.D.F.).

Present Address

S.F.: BioMed X Innovation Center, Im Neuenheimer Feld 583, 69120 Heidelberg, Germany.

Author Contributions

G.L. and N.F. contributed equally

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

G.L. acknowledges support from the University of Salerno and Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), PRIN 2009 'Design, conformational and configurational analysis of novel molecular platforms'. N.F. acknowledges support from Generalitat de Catalunya (FI-Agaur). G.D.F. acknowledges support by the Spanish Ministry of Science and Innovation (ref. BIO2011-27450). S.F. acknowledges support from the EC Marie Curie Initial Training Network "STARS" (grant agreement 238490). We are also grateful to all of the volunteers of GPUGRID who donated GPU computing time to the project. We finally thank H. Gutierrez de Teran for useful discussions on the method.

■ REFERENCES

(1) Hasselgren, C.; Muthas, D.; Ahlberg, E.; Andersson, S.; Carlsson, L.; Noeske, T.; Stålring, J.; Boyer, S. Chemoinformatics and Beyond. In *Chemoinformatics for Drug Discovery*; Bajorath, J., Eds.; John Wiley & Sons, Inc: Hoboken, NJ, 2013; pp 267−290.

(2) McGregor, J. M.; Luo, Z.; Jiang, X. Virtual Screening in Drug Discovery. In *New Frontiers in the Post-Genomic Era. Drug Discovery Research*; Huang, Z., Eds.; John Wiley & Sons, Inc: Hoboken, NJ, 2007; pp 63−88.

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

(4) Huang, S.-Y.; Zou, X. Ensemble docking of multiple protein structures: Considering protein structural variations in molecular docking. *Proteins Struct. Funct. Bioinf.* **2007**, *66*, 399−421.

(5) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100−5109.

(6) Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of consensus scoring strategies for evaluating computational models of protein−ligand complexes. *J. Chem. Inf. Model.* **2009**, *46*, 380−391.

(7) Betzi, S.; Suhre, K.; Chétrit, B.; Guerlesquin, F.; Morelli, X. GFscore: A general nonlinear consensus scoring function for high-throughput docking. *J. Chem. Inf. Model.* **2006**, *46*, 1704−1712.
(8) Ou-Yang, S. S.; Lu, J. Y.; Kong, X. Q.; Liang, Z. J.; Luo, C.; Jiang, H. Computational drug discovery. *Acta Pharmacol. Sin.* **2012**, *33*, 1131−1140.
(9) Harvey, M. J.; Giupponi, G.; De Fabritiis, G. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632−1639.
(10) Phillips, J. C. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(11) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184−10189.
(12) Shan, Y. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.
(13) Giorgino, T.; Harvey, M. J.; de Fabritiis, G. Distributed computing as a virtual supercomputer: Tools to run and manage large-scale BOINC simulations. *Comput. Phys. Commun.* **2010**, *181*, 1402−1409.
(14) Aqvist, J.; Samuelsson, J. E.; Medina, C. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.* **1994**, *7*, 385−91.
(15) Wang, J.; Dixon, R.; Kollman, P. A. Ranking ligand binding affinities with avidin: a molecular dynamics-based interaction energy study. *Proteins* **1999**, *34*, 69−81.
(16) Aqvist, J.; Marelius, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 613−626.
(17) Almlöf, M.; Brandsdal, B. O.; Åqvist, J. Binding affinity prediction with different force fields: Examination of the linear interaction energy method. *J. Comput. Chem.* **2004**, *25*, 1242−1254.
(18) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(19) Miller, B. R. MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **2012**, *8*, 3314−3321.
(20) Greenidge, P. A.; Kramer, C.; Mozziconacci, J.-C.; Wolf, R. M. MM/GBSA binding energy prediction on the PDBbind data set: Successes, failures, and directions for further improvement. *J. Chem. Inf. Model.* **2013**, *53*, 201−209.
(21) Su, Y.; Gallicchio, E.; Das, K.; Arnold, E.; Levy, R. M. Linear interaction energy (LIE)models for ligand binding in implicit solvent: Theory and application to the binding of NNRTIs to HIV-1 reverse transcriptase. *J. Chem. Theory Comput.* **2007**, *3*, 256−277.
(22) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic atom type and bond type perception in molecular mechanical calculations. *J. Mol. Graph. Model.* **2006**, *25*, 247−260.
(23) Fabritiis, G. D. GPUGRID.net. http://www.gpugrid.org/.
(24) Case, D. A. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(25) Trott, O.; Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.
(26) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein−ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177−6196.
(27) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.
(28) Gutiérrez-de-Terán, H.; Åqvist, J. Linear Interaction Energy: Method and Applications in Drug Design. In *Computational Drug Discovery and Design*; Baron, R., Ed.; Springer: New York, 2012; pp 305−323.
(29) Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient β in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins Struct. Funct. Bioinf.* **1999**, *34*, 395−402.
(30) Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 27−35.
(31) Almlöf, M.; Carlsson, J.; Åqvist, J. Improving the accuracy of the linear interaction energy method for solvation free energies. *J. Chem. Theory Comput.* **2007**, *3*, 2162−2175.
(32) Li, H. An effective docking strategy for virtual screening based on multi-objective optimization algorithm. *BMC Bioinf.* **2009**, *10*, 58.