JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

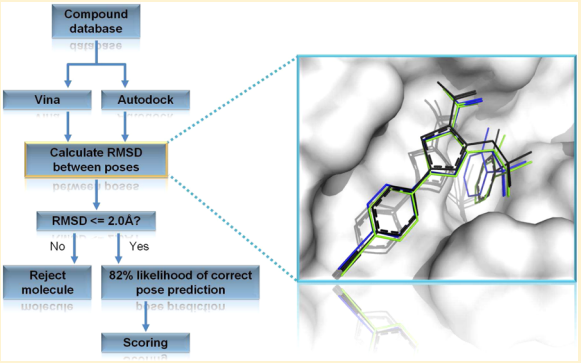
Consensus Docking: Improving the Reliability of Docking in a Virtual Screening Context

Douglas R. Houston* and Malcolm D. Walkinshaw

Institute of Structural and Molecular Biology, University of Edinburgh, Edinburgh, United Kingdom

Supporting Information

ABSTRACT: Structure-based virtual screening relies on scoring the predicted binding modes of compounds docked into the target. Because the accuracy of this scoring relies on the accuracy of the docking, methods that increase docking accuracy are valuable. Here, we present a relatively straightforward method for improving the probability of identifying accurately docked poses. The method is similar in concept to consensus scoring schemes, which have been shown to increase ranking power and thus hit rates, but combines information about predicted binding modes rather than predicted binding affinities. The pose prediction success rate of each docking program alone was found in this trial to be 55% for Autodock, 58% for DOCK, and 64% for Vina. By using more than one docking program to predict the binding pose, correct poses were identified in 82% or more of cases, a significant improvement. In a virtual screen, these more reliably posed compounds can be preferentially advanced to subsequent scoring stages to improve hit rates. Consensus docking can be easily introduced into established structure-based virtual screening methodologies.



INTRODUCTION

Virtual screening is a well-established strategy for discovering protein ligands with numerous successful examples of the technique in the literature.[1−4] The method can be broken down into two steps. First, the compounds to be screened are docked into the active site of a model of the protein structure to give a predicted location, orientation and conformation (hereafter referred to as pose). The second step uses that pose to calculate a predicted affinity of the compound with the protein. The accuracy of this score is dependent on the accuracy of the predicted binding pose.[5] One of the limitations of the virtual screening technique is that significantly different poses can give similar docking scores with no way of distinguishing which one is correct. Incorrect poses, resulting in incorrect affinity predictions, reduce the hit rate of virtual screening. Purchase and testing of such false positives is especially undesirable in a small-scale laboratory setting where resources may be limited.

The problem of predicting the affinity a compound has for a protein is a challenging one. For example, Warren et al. found that none of the standalone scoring functions tested made a useful prediction of ligand binding affinity on a test set of eight proteins.[6] A well-known technique for reducing the error of scoring functions is to combine the results of several different scoring algorithms into a consensus scoring scheme. Consensus scoring is a method whereby the binding affinities of compounds for a particular target are predicted by using more than one scoring algorithm. Several different studies have found that this is superior in accuracy to using a single scoring algorithm alone. For example, Kukol investigated various consensus algorithms and found that, for a given docking pose, a simple combination of AutoDock and Vina scores gave the most consistent performance that showed early enrichment of known ligands for all receptor targets investigated.[7] Chang et al. performed extensive comparisons of Autodock and Vina docking results using the DUD decoy database[8] and, like Kukol, found that taking both of their scores into account improved the overall binding affinity predictions.[9] Cheng et al. compared a number of different standalone scoring algorithms and showed that a combination of them almost always outperformed even the best (although the precise nature of the combinations were found to vary depending on the target).[10]

The structure-based in silico prediction of ligand binding affinity involves two major steps. First, to predict the pose of the ligand (where a correct pose is usually regarded as matching the pose in a cocrystallized protein or enzyme), and second, to predict binding affinities close to experimental observations.[11] Therefore, if scoring algorithms are to correctly predict the affinity of a ligand, they must use the correct binding conformation as input. For example, Peterson et al. found that unless the predicted binding conformations of the nerve agent VX contained certain binding features observed in the cocrystal complexes; no correlation between the affinity predicted by scoring algorithms and experimentally determined affinity was observed.[12]

Received: August 22, 2012
Published: January 27, 2013

dx.doi.org/10.1021/ci300399w | J. Chem. Inf. Model. 2013, 53, 384−390

The known benefits of combining scoring algorithms prompted our investigation into whether improvements in pose prediction, as distinct from affinity prediction, could be seen when the results of docking algorithms, rather than scoring algorithms, are combined. As there are no references to this idea in the literature, the precise methodology and selection of programs to study was problematic. Why consensus scoring consistently performs better than the best standalone scoring algorithm has been the subject of discussion. It has been reported that combining dissimilar types of scoring functions results in a greater increase in accuracy than combining similar types.[10] For example, a knowledge-based scoring method should in theory be combined with an empirical method so that each scoring approach compensates for the other's weaknesses. In addition to the complementarity of dissimilar scoring classes, it is also thought that programs calibrated using different training data sets are complementary as essentially they are able to draw on a larger body of statistical information.

Currently, two of the best of the freely available docking programs are Autodock[13] and Vina.[14] Utilizing a Lamarckian Genetic Algorithm, Autodock is a semi-empirical docking method that has been thoroughly reviewed in the literature over the years and has produced some notable successes.[15,16] Although the first version was released over two decades ago, continual updates have kept the package competitive.[17] Vina is a more recent release and as such has not been as extensively tested in the literature as Autodock. Nevertheless, its combination of empirical and knowledge-based scoring functions with an Iterated Local Search global optimizer has been shown to perform at least as well as Autodock and with reduced runtime.[14] To test the consensus docking idea, we selected Autodock and Vina as they employ significantly different docking methodologies (i.e., different atom typing schemes, different global optimizers, different local optimizers, and different scoring functions[13,14,18]) and have employed different collections of crystal complexes and binding data to weight their optimization algorithms. The intention was to investigate whether using these to apply a consensus strategy to the docking problem would result in improvements in posing accuracy and prediction of binding modes. A third docking program, DOCK,[19] was employed to test whether any improvements seen were more generally applicable.

MATERIALS AND METHODS

Selection of Test Set of Crystal Complexes. The PDBbind-CN database is a large collection of protein−ligand crystal complexes with associated experimentally determined binding data.[20] The version at time of writing (v2010) contained 5075 protein−ligand entries of which 2061 possess good quality structural and binding data. A subset of 231 of these have been selected by the database caretakers as the "core set", which aims to provide a collection of high-quality complexes that represents a broad cross-section of the database as a whole and which is of a size that is appropriate for docking and scoring studies. This preprepared core set of complexes was selected for study (see Supporting Information for ligand set properties).

Preparation of Structures for Docking. Water molecules and other heteroatoms were removed, and the program PDB2PQR 1.6[21] was used to assign position-optimized hydrogen atoms, utilizing the additional PropKa[22] algorithm with a pH of 7.4 to predict protonation states. The removal of heteroatoms was required due to the automated nature of the preparation, which in turn was necessitated by the large number of complexes involved. It is expected that some of the incorrect pose predictions made by the docking programs may be attributable to this, but we found that it did not have a significant impact on the overall docking success rates. The Autodock Tools 1.5.4 utility prepare_receptor4.py was used to assign Gasteiger charges to atoms. Hydrogen atoms were assigned to ligand structures using OpenBabel 2.3.0,[23] utilizing the -p option to predict the protonation states of functional groups at pH 7.4. The Autodock Tools utility prepare_ligand4.py was used to assign Gasteiger charges and rotatable bonds. As Autodock and Vina both use the same pdbqt format for their input, the same prepared files could be used for each.

Docking Trials. Each ligand was docked into its corresponding receptor using Autodock 4.2.3 and separately using Vina 1.1.2. The size of the box that defines the search space for both programs was set at 4 Å around the ligand (i.e., 8 Å was added to each of the maximum x, y, and z dimensions of the ligand), with the center of the ligand defining the center of the search space. The Autogrid grid point spacing was set at 0.2 Å, which is higher resolution than the default of 0.375 Å, and in some cases, this has been found to improve docking accuracy. The Autodock parameter file specified 10 Lamarckian genetic algorithm runs for each docking. This is significantly below the number recommended in the documentation of 50−100; however, experience has shown that the increase in speed more than compensates for the very small reduction in accuracy. The number of energy evaluations was set dynamically, in that ligands with fewer rotatable bonds were allowed fewer minimizations, with a minimum of 125,000 for completely rigid molecules and a maximum of 10,000,000 for molecules with 10 or more rotatable bonds. A small number of ligands were found to have more than the 32 rotatable bond limit hardcoded into Autodock, and these complexes were removed from the test set, leaving a total of 228. Random initial placement of the ligand (translation, rotation, and torsion angles) was specified in the Autodock docking parameter file.

The program DOCK 6.5[23] was also used for comparison. Ligand and protein files were prepared in the same way as for Autodock and Vina. The DOCK utilities Sphgen, Sphere_selector, and Showsphere were used to define the protein binding site, and Grid was used to calculate energy grids. For individual parameter settings please consult the example parameter files in the Supporting Information.

RMSD Comparison of Docking Results. For each complex, the first Vina result (out of the nine, it outputs by default) and the lowest energy member of the largest Autodock cluster were selected as the docking result for each program. Ligand docking poses were compared to the crystallographic poses using the root mean square deviation calculation method described previously.[14] This method is able to take symmetry into account, for example, by correctly assigning a low RMSD to a molecule with 2-fold symmetry that has been rotated 180° during docking. The same method was used to compare the Autodock and Vina docking poses.

RESULTS AND DISCUSSION

A total of 228 ligands were docked into their respective protein structures using Autodock and Vina (see Supporting Information for ligand set properties). Of those 228 trials, 122 of the Autodock runs and 141 of the Vina runs were found to be correctly posed (RMSD $\leq$ 2.0 Å between the docked and the crystallographic pose). A total of 118 of the ligands were

dx.doi.org/10.1021/ci300399w | J. Chem. Inf. Model. 2013, 53, 384−390

found to be docked similarly by Autodock and Vina (RMSD ≤ 2.0 Å), and 97 of these binding mode predictions were correct (i.e., the RMSD between docked pose and crystallographic pose was less than 2.0 Å, see Figure 1 for an example). Out of the

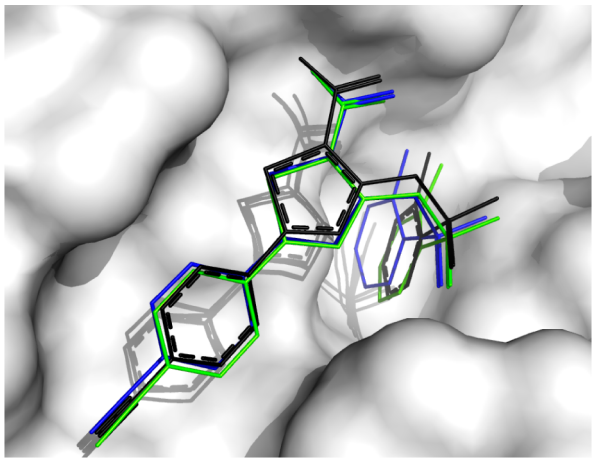


Figure 1. Comparison of docking and crystallographic binding poses. The structure of RNA-dependent RNA polymerase in complex with a thiophene-based non-nucleoside inhibitor (PDB ID: 2D3U) is shown. The atomic surface of the protein is colored white. The ligand pose as found in the crystal structure is shown in black. The ligand position after docking by Autodock is shown in blue, and the Vina docking in green. The RMSD between the coordinates for the Autodock pose and the crystal structure is 0.75 Å, and for the Vina pose, it is 0.40 Å. The RMSD between the Autodock and Vina pose is 0.70 Å. Thus, in a virtual screening scenario, this molecule would progress from the docking stage to the scoring stage.

118 cases where Autodock and Vina both agreed on a particular binding mode, 21 of these were found to be docked incorrectly (i.e., the RMSD between docked pose and crystallographic pose was greater than 2.0 Å, see Figure 2 for an example). By using both programs together, a correct pose is predicted in 82% of cases. The success rate of each program alone was found in this trial to be 55% for Autodock and 64% for Vina. Thus, combining the results from the two programs provides a way of improving confidence in the docking results at the cost of rejecting some "true positive" predictions from the individual programs. However, this cost is less important in a virtual screening context; what is vital in such a situation is the quality of the results that progress to the scoring and acquisition phases. The decrease in false positives (and concomitant decrease in resource wastage) that would be expected from focusing only on the most accurately docked compounds (i.e., those compounds that in testing showed an 82% success rate) would more than compensate for any potential increase in false negatives.

As shown in Figure 3, varying the RMSD cutoff used to select the matching Vina and Autodock poses results in varying false negative and false positive rates. A balance must be found between these two undesirable outcomes; too lenient a cutoff results in large numbers of molecules being classed as matching, which would increase the number of incorrectly docked molecules, while too strict a cutoff results in larger numbers of correctly docked molecules being rejected. It appears that the optimal choice is a cutoff of approximately 2.0 Å. Interestingly, this is the same value often presented in the literature to

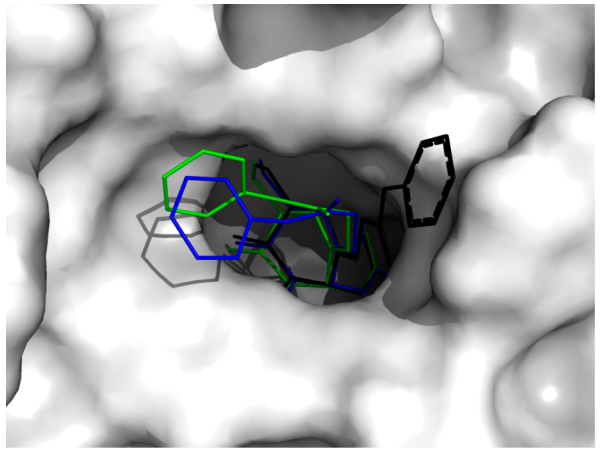


Figure 2. Comparison of docking and crystallographic binding poses in the case of a false positive. The structure of O-GlcNAcase in complex with the inhibitor NButGT (PDB ID: 2VVS) is shown. The atomic surface of the protein is colored white. The ligand pose as found in the crystal structure is shown in black. The ligand position after docking by Autodock is shown in blue, and the Vina docking in green. The RMSD between the coordinates for the Autodock pose and the crystal structure is 2.93 Å, and for the Vina pose, it is 2.95 Å. Both results would be classified as docking failure with an RMSD cutoff of 2.0 Å. However, the RMSD between the Autodock and Vina pose is only 0.62 Å, meaning that this molecule, along with its incorrect docking poses, would progress to the scoring stage. Thus this is an example of a "false positive" result of the technique.

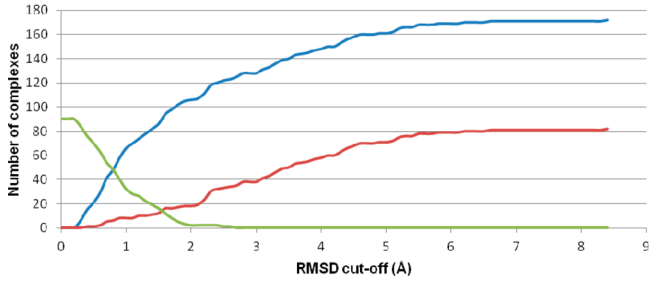


Figure 3. Distribution of consensus docking statistics according to RMSD cutoff. The blue line represents those cases where the RMSD between the Autodock and Vina docking pose is less than the cutoff. The red line represents those cases where the RMSD between the Autodock and Vina docking pose is less than the cutoff, and the RMSD between the Autodock or Vina and the crystallographic pose is greater than 2.0 Å (incorrect pose prediction). The green line represents those cases where the RMSD between the Autodock and Vina docking pose is greater than the RMSD cutoff, and the RMSD between the Autodock or Vina docking poses and the crystallographic pose is less than 2.0 Å (correct pose prediction).

identify correctly docked ligands or to cluster different docking poses of the same molecule.[24−28]

It is reported in the literature that the success of docking software in correctly predicting binding poses is dependent on the flexibility of the molecule in question. The conformational space that must be sampled for molecules with more rotatable bonds is greater than that for more rigid molecules, reducing the chances of successfully predicting the correct pose. Because of this effect, it is generally recommended that any library used for docking-based virtual screening should be prefiltered to remove very flexible molecules. It was expected that limiting the flexibility of docked ligands by filtering would increase the instances of Autodock and Vina producing similar poses, and

dx.doi.org/10.1021/ci300399w | J. Chem. Inf. Model. 2013, 53, 384−390

that this might in turn increase those instances where the predicted poses both significantly differ from the crystallographic pose (i.e., false positives). Therefore, the effect of varying the maximum number of rotatable torsions allowed in the ligand set was explored.

Figure 4 shows the results of varying the average flexibility of the ligand set. As expected, the instances of docking failure are

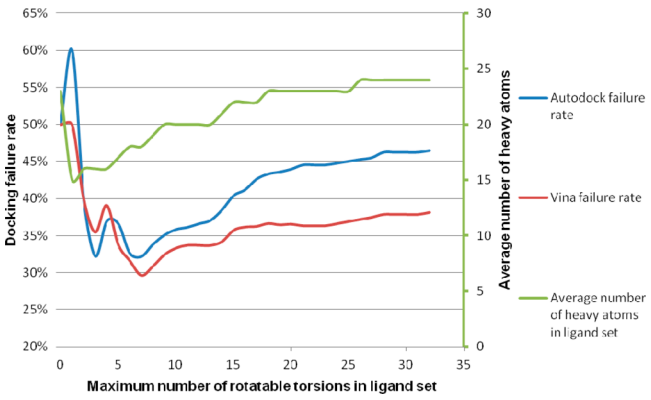


Figure 4. Effect of ligand flexibility on docking success rate. This graph shows that for both Autodock and Vina there is an optimal cutoff for ligand flexibility, which itself tends to vary with molecule size. Both programs had difficulty correctly predicting the binding poses of molecules that were too rigid (probably due to the small fragment-like nature of these molecules) or too flexible.

high for very flexible molecules. However, the docking failure rates for both Vina and Autodock are also high for very rigid molecules. This was surprising, as in general, the smaller conformational search space required for such molecules means docking programs should have an increased likelihood of finding the correct pose. It was hypothesized that this might be due to the generally smaller nature of such compounds, as both Vina and Autodock tend to be less reliable when docking small fragment-like molecules. Figure 4 indeed shows that as the number of rotatable torsions in the ligands decreases, so does molecule size. There appears to be a certain size for molecules that maximizes docking reliability; as molecule size increases beyond a certain point, docking reliability falls again. There is an optimum where flexibility is low enough to prevent an explosion of conformational search space, but size (and complexity) is high enough to allow a distinct binding conformation. However, comparing Figure 4 with Figure 5 shows that when the consensus docking technique is used to identify incorrectly docked ligands, the deleterious effect of high molecule flexibility on docking accuracy disappears.

What are the potential drawbacks to the consensus docking approach? Apart from a minor increase in computational cost (docking with Vina in addition to Autodock only takes approximately 10% longer than Autodock alone), the main drawback is that of false negatives. If the Autodock and Vina docked poses differ, it is impossible to identify which is the correct prediction; the two different docked poses would appear almost equally likely. Only in those cases where Autodock and Vina agree does accuracy increase. However, false positives are particularly undesirable when resources for compound acquisition are limited; use of this technique during a virtual screening project would reduce wastage on acquisition and testing of inactive compounds.

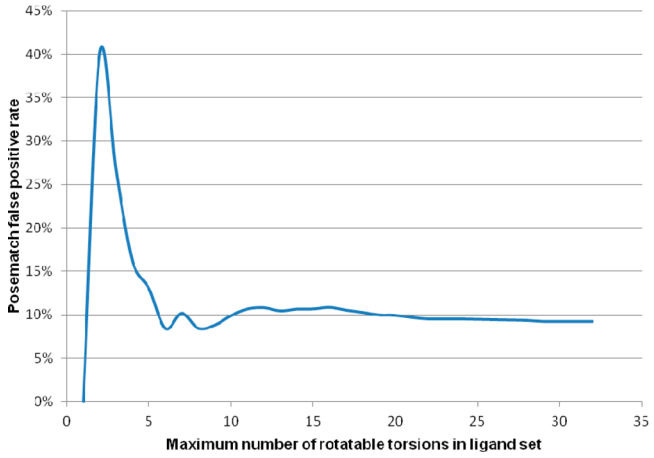


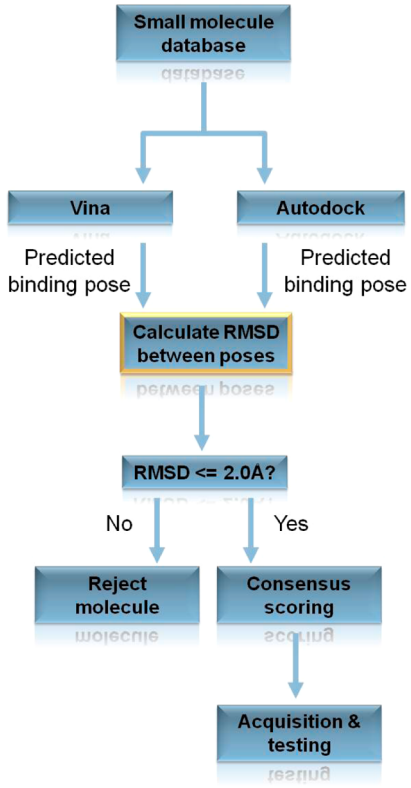
Figure 5. Effect of ligand flexibility on consensus docking false positive rate. This graph shows that as predicted, ligands with very low flexibility were more likely to be docked in the same pose by both Autodock and Vina, and that this would increase the false positive rate, i.e., the number of cases where Autodock and Vina produced similar docking poses that were not similar to the crystallographic pose. However, this effect seems to decline very rapidly with increasing ligand flexibility, such that by the time ligands with up to five rotatable torsions are included in the set, Autodock and Vina pose agreement is correctly implying increased likelihood of docking success and is effectively filtering out incorrectly docked ligands.

Placing the consensus docking method after the docking but before the scoring stages of a virtual screening workflow (Scheme 1) would result in a reduction in the proportion of molecules acquired for assay on the basis of scores that were calculated from an erroneous docking pose. Ultimately, this should result in a decrease in the numbers of compounds that show no activity in the primary assay. For example, in a hypothetical virtual screen resulting in the acquisition of 100 compounds, our results show that the use of Vina alone would result in 36 of those compounds being selected using incorrect pose predictions. If only compounds that were docked similarly by both Vina and Autodock were allowed to progress to the scoring stage, this figure would drop to 18.

One concern is that the consensus docking method may be biased with regards to the types of molecules it rejects. We have already seen that more flexible molecules are more likely to be rejected due to the docking programs' reduced ability to correctly predict the binding poses of these. This is a desirable outcome, but it was thought necessary to verify that the method was not also rejecting molecules that would score highly in scoring algorithms (and therefore be more likely to exhibit high affinity for the target). A comparison of the scores of the PDBbind ligands that Autodock and Vina dock similarly was made with those ligands rejected by the consensus docking. The top 10% best-scored ligands of the consensus posed ligands had an average Autodock score of −13.0 kcal/mol and an average Vina score of −11.7 kcal/mol. The top 10% best-scored ligands of the non-consensus posed ligands had an Autodock score of −10.9 kcal/mol and a Vina score of −10.0 kcal/mol. Therefore, we see that in fact the consensus docking method is slightly biased toward molecules that are scored more highly. If we consider this effect in a virtual screening context, if only those ligands that Autodock scores highly (predicted K_i < 1 μM) are examined (as might be the case in a virtual screen), we see that 29% are incorrectly docked. After

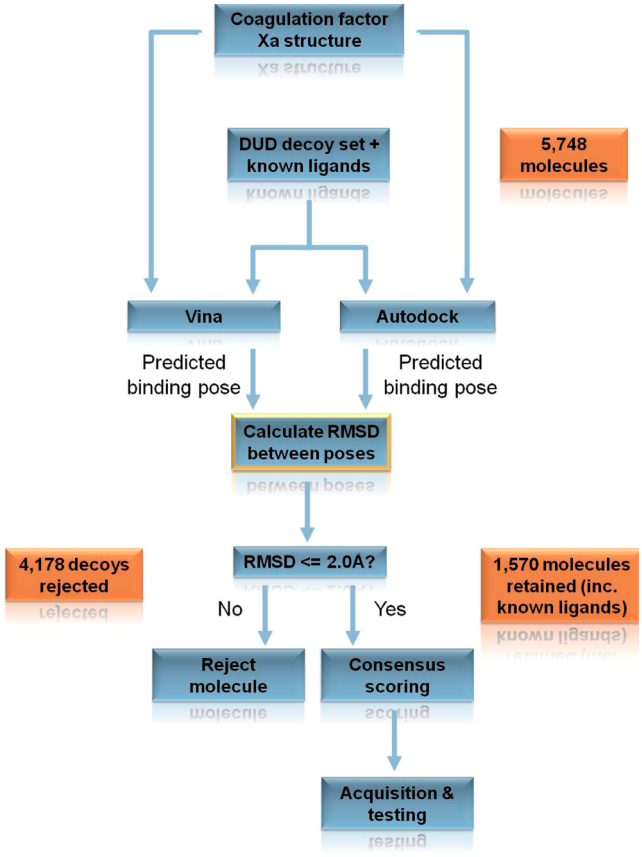
dx.doi.org/10.1021/ci300399w | J. Chem. Inf. Model. 2013, 53, 384−390

Scheme 1. Suggested Placement of the Consensus Docking Technique in a Virtual Screening Workflow[a]



[a]Rejecting molecules on the basis of pose similarity would maximize the number of correctly docked compounds moving forward to the scoring and acquisition phase.

Scheme 2. Results of Example Virtual Screen Using Coagulation Factor Xa and Its DUD Decoy Set[a]



[a]In addition to the increase in pose prediction accuracy, the consensus docking method enriches the results even before the scoring stages.

the consensus docking filter has been applied, only 8% of these incorrectly docked molecules are left.

An example virtual screen that utilizes the consensus docking method demonstrates this beneficial side effect. The PDBbind 2010 core set contains three crystal structures of coagulation factor Xa in complex with inhibitors of different affinities. Both Vina and Autodock are able to correctly predict the binding pose of these molecules; therefore, in a virtual screening context, these molecules would pass through to the scoring stages. However, we also used Autodock and Vina to dock the 5745 DUD decoy compounds available for this target (which are selected for the similarity of their physical properties to known ligands but dissimilar topology[8]). Out of these, only 1567 were predicted by Vina and Autodock to bind in a similar conformation. Therefore, as part of a virtual screen, 4178 decoy compounds (73%) would be rejected, significantly reducing the list of molecules that would need to be graded by the subsequent consensus scoring method (Scheme 2). This enrichment effect is in addition to the primary purpose of using consensus docking, namely its ability to assign greater confidence to correctly posed molecules.

It would appear then that the use of Autodock and Vina in conjunction can significantly increase the accuracy of pose prediction, when those programs are in agreement. It was expected that this general principle ought to be applicable to all docking programs. We used the program DOCK to test this. Table 1 summarizes the results of these combinations. DOCK was able to correctly predict (RMSD $\leq$ 2.0 Å) the binding pose of 58% of the PDBbind ligands. When only those ligands that agreed (RMSD $\leq$ 2.0 Å) with Autodock's predicted pose are considered, 88% of the ligands are found to be docked correctly, and when only those ligands that agreed (RMSD $\leq$ 2.0 Å) with Vina's predicted pose are considered, 86% are found to be docked correctly. Utilizing all three programs together results in 92% of the predicted poses being correct in the retained set of ligands, although this is at the expense of a significantly higher proportion of rejected molecules. These results suggest that the consensus docking method could be useful for virtual screening approaches that utilize docking programs other than Vina and Autodock.

CONCLUSION

Comparing Autodock and Vina docking poses and rejecting those that were not sufficiently similar increased the proportion of correctly docked ligands in the non-rejected set to 82% compared to 64% for the best docking program alone. This is a very significant improvement in docking performance. It would appear that using two different docking programs together for any docking studies should improve the accuracy of the predicted ligand binding mode. In addition, this dual approach could be trivially applied to virtual screening studies, perhaps by docking the compound library first with Vina (as it is around an order of magnitude faster than Autodock) and then docking the top ranked compounds using Autodock and comparing the predicted binding poses. Those that agree should have a higher

dx.doi.org/10.1021/ci300399w | J. Chem. Inf. Model. 2013, 53, 384−390

Table 1. Proportion of Correctly Docked Ligands That Pass through to the Scoring Stages of a Putative Virtual Screen with and without the Use of the Consensus Docking Technique[a]

Combination	No. of ligands where RMSD between pose prediction and crystallographic pose ≤2.0 Å (correctly docked)	No. of ligands where RMSD between program pose predictions ≤2.0 Å (and pass to scoring stage)	No. of ligands where RMSD between pose predictions ≤2.0 Å and RMSD between predictions and crystallographic pose ≤2.0 Å (correctly docked and programs agree on pose)	No. of ligands where RMSD between pose predictions ≤2.0 Å and RMSD between predictions and crystallographic pose ≥2.0 Å (incorrectly docked but programs agree on pose)	% of ligands passing to scoring stage that are correctly docked	No. of ligands rejected prior to scoring stage
Autodock	122	n/a	n/a	n/a	55	0
DOCK	132	n/a	n/a	n/a	58	0
Vina	141	n/a	n/a	n/a	64	0
Autodock + Vina	n/a	118	100	18	82	110
Vina + DOCK	n/a	125	107	18	86	103
Autodock + DOCK	n/a	109	96	13	88	119
Autodock + Vina + DOCK	n/a	83	76	7	92	145

[a]Total number of ligands (and crystal complexes) was 228.

chance of being correct, and so would then be passed through to the scoring stages.

■ ASSOCIATED CONTENT

Supporting Information

Details of the properties of the ligands in the PDBbind 2010 core set of crystal structures complexes and example DOCK parameter files. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +44 131 650 7358. E-mail: dhouston@staffmail.ed.ac.uk.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Dr. Paul Taylor for his valuable assistance.

■ ABBREVIATIONS

RMSD, root-mean-square deviation; PDB, protein data bank

■ REFERENCES

(1) Badrinarayan, P.; Sastry, G. N. Virtual high throughput screening in new lead identification. *Comb. Chem. High Throughput Screen* **2011**, *14*, 840−860.

(2) Kitchen, D. B.; Decornez, H.; Furr, F. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−949.

(3) McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494−502.

(4) Seifert, M. H.; Kraus, J.; Kramer, B. Virtual high-throughput screening of molecular databases. *Curr. Opin. Drug Discovery Dev.* **2007**, *10*, 298−307.

(5) Kolb, P.; Irwin, J. J. Docking screens: Right for the right reasons? *Curr. Top. Med. Chem.* **2009**, *9*, 755−70.

(6) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−31.

(7) Kukol, A. Consensus virtual screening approaches to predict protein ligands. *Eur. J. Med. Chem.* **2011**, *46*, 4661−4664.

(8) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−801.

(9) Chang, M. W.; Ayeni, C.; Breuer, S.; Torbett, B. E. Virtual screening for HIV protease inhibitors: A comparison of AutoDock 4 and Vina. *PLoS One* **2010**, *5*, e11955.

(10) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **2009**, *49*, 1079−1093.

(11) Udatha, D. B.; Sugaya, N.; Olsson, L.; Panagiotou, G. How well do the substrates KISS the enzyme? Molecular docking program selection for feruloyl esterases. *Sci Rep.* **2012**, *2*, 323.

(12) Peterson, M. W.; Fairchild, S. Z.; Otto, T. C.; Mohtashemi, M.; Cerasoli, D. M.; Chang, W. E. VX hydrolysis by human serum paraoxonase 1: A comparison of experimental and computational results. *PLoS One* **2011**, *6*, e20335.

(13) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1−5.
(14) 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.
(15) Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **2004**, *47*, 1879−1881.
(16) Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin. Drug. Discovery* **2010**, *5*, 597−607.
(17) 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**, *6*, 1145−1152.
(18) 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*, 1639−1662.
(19) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: Combining techniques to model RNA-small molecule complexes. *RNA* **2009**, *15*, 1219−30.
(20) Wang, R.; Fang, X.; Lu, Y.; Yang, C. Y.; Wang, S. The PDBbind database: Methodologies and updates. *J. Med. Chem.* **2005**, *48*, 4111−4119.
(21) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, 522−525.
(22) Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins* **2005**, *61*, 704−721.
(23) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33.
(24) Guilbert, C.; James, T. L. Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J. Chem. Inf. Model.* **2008**, *48*, 1257−1268.
(25) Baxter, C. A.; Murray, C. W.; Waszkowycz, B.; Li, J.; Sykes, R. A.; Bone, R. G.; Perkins, T. D.; Wylie, W. New approach to molecular docking and its application to virtual screening of chemical databases. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 254−62.
(26) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287−303.
(27) Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model.* **2002**, *20*, 281−95.
(28) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337−56.