JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

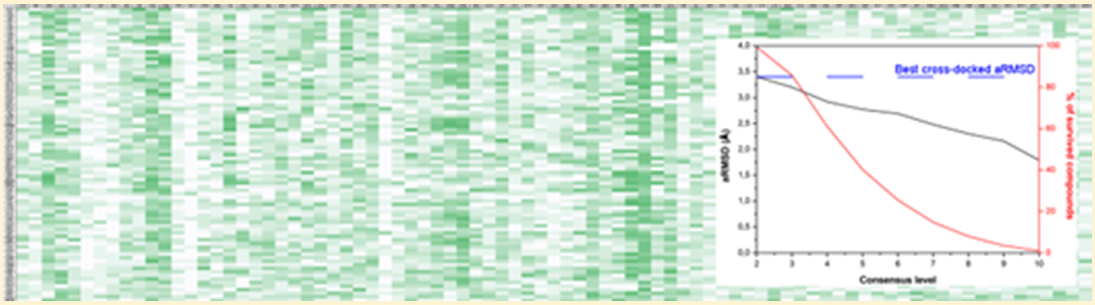
Extensive Consensus Docking Evaluation for Ligand Pose Prediction and Virtual Screening Studies

Tiziano Tuccinardi,*[,†] Giulio Poli,[†] Veronica Romboli,[†] Antonio Giordano,[‡] and Adriano Martinelli[†]

[†]Department of Pharmacy, University of Pisa, 56126 Pisa, Italy

[‡]Sbarro Institute for Cancer Research and Molecular Medicine Center for Biotechnology, Temple University, Philadelphia, Pennsylvania 19122, United States

Supporting Information



ABSTRACT: Molecular docking strategies are one of the most widely used techniques for predicting the binding mode of a ligand and for obtaining new hits in virtual screening studies. In order to improve the accuracy of this approach, we tested the reliability of applying a consensus docking protocol by combining ten different docking procedures. The analysis was carried out in terms of consensus cross-docking and by using an enriched database. The results highlight that from a qualitative point of view consensus docking is able to predict the ligand binding pose better than the single docking programs and is also able to give hints concerning the reliability of the docking pose. With regard to the virtual screening studies, consensus docking was evaluated for three different targets of the Directory of Useful Decoys (DUD), and the obtained results suggest that this approach performs as well as the best available methods found in the literature, therefore supporting the idea that this procedure can be profitably applied for the identification of new hits.

INTRODUCTION

In the structure-based drug design field, protein/ligand docking is one of the most commonly used and profitable techniques. The rapid accumulation of high-resolution 3D structures and homology modeling techniques together with the improvements in docking and scoring functions is making docking-based analysis realistic for wide use, spanning the range from predicting the binding disposition of known ligands to virtual screening (VS) experiments.[1] Usually, molecular docking can be defined as an optimization task to identify the ligand conformation bound to the target with the most favorable binding energy. Unfortunately, this is a challenging task mainly due to the ligand and protein flexibility. One of the key questions regarding the prediction of the geometry of target-ligand complexes is whether it is possible to measure the level of reliability of docking software for predicting the ligand binding pose. In recent years, many studies have been carried out addressing this question, and two different approaches termed *pose identification* and *database enrichment* have been commonly used. To analyze pose identification accuracy, in the self-docking analysis, crystallographic ligand−receptor complexes are used to determine if the docking software can reproduce the native ligand conformation, whereas in the cross-docking analysis the docking of a ligand into the protein structure obtained from a different complex is also evaluated. Compared to self-docking studies, cross-docking analysis better verifies how reliable docking software is because it essentially evaluates how efficiently a docking software package reproduces the experimentally determined binding pose of a ligand by docking it into a protein whose 3D structure was determined in complex with a different ligand. To assess database enrichment, a series of experimentally tested active ligands is included into a large group of decoy molecules that should not be active against the analyzed target to determine if the rank-ordered list of molecules (active and decoys) will contain the known binders among the most favorably scored list elements.

Up to date, many studies have reported the reliability of docking software by using these two different approaches;[2−4] however, only few studies have reported the possibility of using different docking software at the same time (consensus docking). In 2013 Houston and Walkinshaw[5] reported one of the first studies concerning the reliability evaluation of the consensus docking approach. To test this approach, they

Received: July 17, 2014
Published: September 11, 2014

dx.doi.org/10.1021/ci500424n | J. Chem. Inf. Model. 2014, 54, 2980−2986

applied a self-docking evaluation on 228 protein−ligand X-ray complexes by using AUTODOCK 4, VINA, and DOCK 6.5 and verified that the pose prediction success rate (i.e., the number of ligands showing a root-mean-square deviation lower than 2.0 Å between the docked and crystallographic pose) was higher when more than one docking program predicted the same ligand pose.[5] Stimulated by these interesting results, in this paper we report a wide consensus docking reliability analysis that takes into account the consensus of ten different docking procedures, evaluating the cross-docking and the database enrichment analysis for six different targets (three for the cross-docking and three for the database enrichment evaluation). For the cross-docking analysis the cyclin-dependent kinase 2, beta-secretase 1, and estrogen receptor alpha were considered because of the elevated number of reported X-ray ligand-protein complexes and because of their dissimilarity; they do not show any similarity in protein sequence, ligands, and binding sites shape. With regard to the three targets considered for the database enrichment evaluation (neuraminidase, HIV reverse transcriptase, and poly(ADP-ribose) polymerase), they were selected from the Directory of Useful Decoys[6] (DUD) taking into account their reciprocal dissimilarity as well as their diversity with respect to the targets used for the cross-docking analysis. Once we discarded all kinases and steroid receptors among the available DUD packages, we looked for three targets with different roles and functions, showing again different features in terms of protein sequence, ligands, and binding sites shape.

MATERIALS AND METHODS

Protein−Ligand Complex Structures. A total of 83, 25, and 22 human cyclin-dependent kinase 2 (CDK2), beta-secretase 1 (BACE1), and estrogen receptor alpha (ERα) X-ray structures, respectively, were downloaded from the Protein Data Bank[7] and were used in this work. Hydrogen atoms were added by means of the OpenEye Babel software[8] for the proteins and Szybki[9] software for the ligands. The ligands were extracted from the X-ray complexes and then subjected, after a minimization step, to a conformational search (CS) of 1000 steps in a water environment (using the generalized-Born/surface-area model) by means of the Macromodel software.[10] The algorithm used was the Monte Carlo method with the MMFFs force field and a distance-dependent dielectric constant of 1.0. All compounds were then visually checked.

Docking Procedures. For all docking analyses, only the best scored pose was taken into account.

AUTODOCK 4.2.3. AUTODOCK Tools utilities[11] were used in order to identify the torsion angles in the ligands, to add the solvent model and assign the Gasteiger atomic charges to proteins and ligands. The regions of interest used by AUTODOCK[12] were defined by considering the reference ligand as the central group of a grid box of 10 Å in the *x*, *y*, and *z* directions. A grid spacing of 0.375 Å and a distance dependent function of the dielectric constant were used for the energetic map calculations. By using the Lamarckian genetic algorithm, the docked compounds were subjected to 20 runs of the AUTODOCK search using 2 500 000 steps of energy evaluation and the default values of the other parameters.

DOCK 6.5. The molecular surface of the binding site was calculated by means of the MS program,[13] generating the Connolly surface with a probe with a radius of 1.4 Å. The points of the surface and the vectors normal to it were used by the Sphgen program in order to build a set of spheres, with radii varying from 1.4 to 4 Å that describe, from a stereoelectronic point of view, the negative image of the site. Spheres within a radius of 10 Å from the reference ligand were used to represent the site. For each ligand, DOCK 6.5 calculated 500 orientations; of these, the best grid scored was taken into consideration. The grid-based score is based on the nonbonded terms of the molecular mechanic force field. The ligand charge was calculated using the AM1-BCC method, as implemented in the Antechamber suite of Amber 11.[14]

FRED 3.0. FRED[15] requires a set of input conformers for each ligand. The conformers were generated by OMEGA2;[16] in order to avoid bias, the ligands obtained from the CS using Macromodel (see above) were used as starting structures. We applied the following modifications to the default settings of OMEGA2: the energy window was set at 50.0, the maximum number of output conformers was set at 10 000, the time limit was set at 1200, and the root-mean-square deviation (RMSD) value below which two conformations were considered to be similar was set at 0.3 Å. The region of interest for the docking studies was defined in such a manner that it contained all residues which stayed within 10 Å from the ligand in the X-ray structures. The FRED docking method roughly consisted of two steps: shape fitting and optimization. During shape fitting, the ligand was placed into the binding site using a smooth Gaussian potential. A series of three optimization filters was then processed, which consisted in the refinement of the positions of the hydroxyl hydrogen atoms of the ligand, rigid body optimization, and optimization of the ligand pose in the dihedral angle space. In the last optimization step, the Chemgauss3 scoring function was used. After the docking calculation, the poses were scored independently by Chemgauss4.

GLIDE 5.0. The binding site was defined by a rectangular box of 10 Å in the *x*, *y*, and *z* directions centered on the ligand. The possibility of imposing a maximum number of atoms a ligand may have if it was to be docked was deactivated, so that all the ligands were docked independently from the number of their atoms, whereas the GLIDE[17] defaults were used for all other parameters. The GlideScore fitness function is based on Chemscore but includes a steric-clash term, adds buried polar terms to penalize electrostatic mismatches and modifications on other secondary terms. Two docking analyses were carried out using the standard precision (SP) and extra precision (XP) methods. The XP mode is a refinement tool designated to be used only on good ligand poses, the sampling is based on an anchor and refined growth strategy, and the scoring function includes a more complete treatment of some of the SP terms, such as the solvation and hydrophobic terms.

GOLD 5.1. The region of interest for the docking studies was defined in such a manner that it contained all residues which stayed within 10 Å from the ligand in the X-ray structures; the "allow early termination" command was deactivated, while the possibility for the ligand to flip ring corners was activated. For all other parameters, GOLD[18] defaults were used and the ligands were subjected to 30 genetic algorithm runs. Four docking analyses were carried out. The four fitness functions implemented in GOLD, i.e., GoldScore (GS), ChemScore (CS), Astex Statistical Potential (ASP), and ChemPLP (PLP), were used.

AUTODOCK VINA 1.1. The input files for the protein and ligands originated from the AUTODOCK Tools utilities for the AUTODOCK calculations were also used for the AUTODOCK VINA[19] calculations, including the grid box dimensions.

The exhaustiveness parameter was set to 10 and the Energy_range to 1, whereas for all other parameters, AUTODOCK VINA defaults were used. The VINA scoring function combines certain advantages of knowledge-based potentials and empirical scoring functions, extracting information from both the conformational preferences of the receptor−ligand complexes and the experimental affinity measurements.

Cross-Docking Analysis. The α carbons of the structures for each target were aligned with each other using a reference structure; then, each ligand was docked into all the available structures of the target from which the ligand was extracted. The docking reliability was evaluated by calculating for each ligand and each target structure the root-mean-square deviation (heavy atoms) between the reference position of the ligand in the experimental target−ligand complex and that predicted by the docking software in the various target structures. The RMSD analysis was carried out using the rms_analysis software of the GOLD suite.

Consensus Docking Evaluation. By applying the ten docking software, ten different binding dispositions (best scored docking pose) resulted from the docking of each ligand into each protein binding site. The RMSD of each of these docking poses against the remaining nine was evaluated by using the rms_analysis software of the GOLD suite. On this basis, for each ligand docked into each protein binding site a 10 × 10 matrix was generated reporting the RMSD results. By using an in-house program these results were clustered, so that among the ten results, all the similar docking poses were clustered together. As clustering algorithm we used the complete-linkage method which is an agglomerative type of hierarchical clustering. This method starts considering each element in a cluster of its own. The clusters are then sequentially combined into larger ones, until all elements are in the same cluster. At each step, the two clusters separated by the shortest distance are combined. We selected an RMSD clustering threshold of 2.0 Å, therefore the so obtained clusters contained the group of poses which are less than 2.0 Å away from all others poses belonging to the same cluster. Finally, the RMSD between the docking poses belonging to the most populated clusters and the experimental disposition of the ligand was evaluated and the average RMSD was reported as the RMSD of the selected cluster.

Virtual Screening Consensus Docking Evaluation. Three targets of DUD[6] were selected for the analysis. On this basis, neuraminidase (NA), HIV reverse transcriptase (HIVRT), and poly(ADP-ribose) polymerase (PARP) were selected with a total of 4484 ligands. The ligands were then docked into their corresponding target structure (1A4G, 1RT1, and 1EFY PDB codes for NA, HIVRT, and PARP, respectively) by using the ten procedures described above (resulting in a total of 44 840 docking evaluations). Results were evaluated by using the enrichment factor (EF) and the area under curve (AUC) of the receiver operator characteristic (ROC) curve.[20] The EF measures the enrichment of the method compared with random selection

$$EF = [t_p/(t_p + f_n)](NC_{tot}/NC)$$

where t_p is the number of high affinity compounds not rejected (true positives); f_n is the number of high affinity compounds rejected during the VS filtering (false negatives); NC_{tot} is the total number of screened molecules of the database; NC is the total number of compounds obtained by the VS protocol. ROC curves are graphic representations of the relation existing between the sensibility and the specificity of a test. It is generated by plotting the fraction of true positives out of the total actual positives versus the fraction of false positives out of the total actual negatives. When the discrimination between actives and decoys is random, the AUC corresponds to 0.5; whereas in the ideal performance where all the active known ligands are ranked before all the decoys, the AUC is very close to 1.0.

Pharmacophore-Based Analysis. The FLAPpharm tool[21] of the FLAP algorithm[22] was used for the pharmacophore model generation of the NA, HIVRT, and PARP targets. For each target, the Glide_XP docking pose of the compounds showing at least a consensus level of 9 was used. These docking results were used as a receptor-based alignment for the pharmacophore generation. The default parameters of FLAPpharm were used and the so obtained three pharmacophoric models were used for screening the enriched data sets constituted by the DUD decoys and the active compounds that were not used for the pharmacophore generation. The GRID probes H, O, N1, and DRY were used to describe shape, hydrogen bond acceptor, hydrogen bond donor, and hydrophobic interactions. The results were evaluated in terms of AUC and of the enrichment factor for 10% of the ranked database ($EF_{10\%}$).

RESULTS AND DISCUSSION

In order to test the consensus docking reliability in terms of qualitative prediction of the ligand binding site disposition, the CDK2, BACE1, and ERα target were considered. As a first step, the reliability of each single docking software was evaluated by means of cross-docking studies. For all complexes of each target, ligands were extracted from their X-ray complex and subjected to conformational search. They were then docked in all the structures of the same target and the obtained docking results were compared with the experimentally determined ligand dispositions (see Materials and Methods for details). This procedure was applied by using ten different docking procedures; as a result, a total of 130 ligand-protein structures were taken into account (see Table 1).

Table 1. Targets Taken into Account for the Cross-Docking Studies

target name	species	no. of structures	docking calculations[a]
cyclin-dependent kinase 2 (CDK2)	human	83	68890
beta-secretase 1 (BACE1)	human	25	6250
estrogen receptor α (ERα)	human	22	4840

[a]Docking calculations: the number of ligand docking studies for each target (considering the ten docking procedures).

Two parameters were taken into account when comparing the reliability of the results obtained from the different docking procedures: (a) the average root-mean-square deviation (aRMSD) of the position of the ligand resulting from the docking with respect to the experimental disposition (for each docking procedure of each target, the aRMSD represented the average of the RMSD values calculated for all the ligands docked into all of the binding sites) and (b) the number of ligands with a reliable docking pose (NLRD i.e., the number of ligands with a RMSD smaller than 2.0 Å). Figure 1 summarizes the main results obtained from the cross-docking studies. Glide

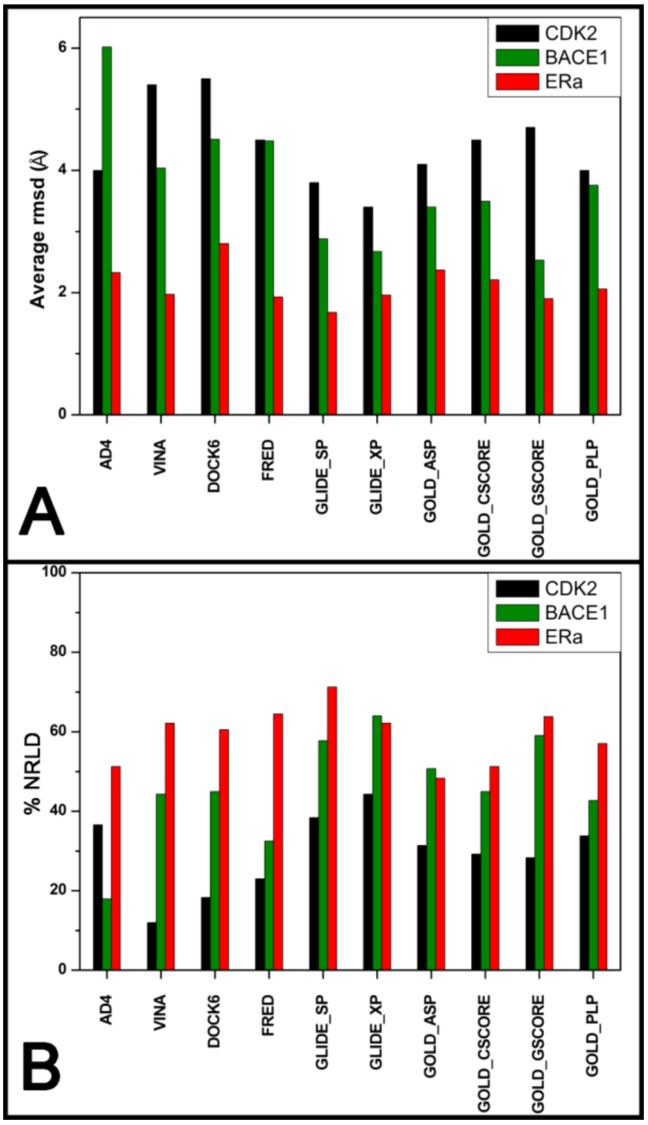


Figure 1. Results of the cross-docking study. For each procedure, the aRMSD (A) and the percentage of poses with a RMSD less than 2 Å (B) are reported.

seemed to be the most performing docking software as it showed the best NLRD value for all three case studies and the best aRMSD for CDK2 and ERα. More in detail, the Glide_XP procedure was the best procedure for CDK2 with an aRMSD of 3.4 Å and an NLRD of 44.3%, Glide_SP was the best one for ERα (aRMSD = 1.7 Å and NLRD = 71.0%), whereas for BACE1 Gold_Gscore showed the best aRMSD (aRMSD = 2.5 Å) and Glide_XP the best NLRD value (NLRD = 64.0%).

In order to verify the effects of the consensus approach on the docking evaluations, the ten docking results (deriving from the docking calculations by using the ten different docking procedures) obtained for each ligand into each protein binding site were clustered together for searching common binding modes (see Materials and Methods for further details). Supporting Information Figure S1 shows how the different docking poses clustered well for each ligand−protein combination; on average, for CDK2 four different docking poses clustered together for each ligand−protein combination, five for BACE1 and seven for ERα. Starting from a consensus level of two (i.e., taking into account all the ligand−protein combinations that showed at least two out of ten docking poses clustered together) we then evaluated the effects of the consensus on the aRMSD of the docking results. For all three targets we observed an increase in the docking quality prediction by taking into account an increasing consensus level value. For CDK2 a consensus level of two resulted in an aRMSD of 3.4 Å with a population of 99.4% of all the ligand−protein combinations. Increasing the consensus level value, the number of ligand−protein combination decreased; however, the aRMSD progressively increased. At consensus level 10, only 1% of the starting ligand−protein combinations was taken into account (this means that only 1% of the ligand-protein combinations showed a complete agreement among all the ten docking procedures), but the aRMSD value was 1.8 Å (see Figure 2 and Table S1), which was about 2-fold better than that obtained by using the best docking procedure in the cross-docking analysis. Very similar results were also obtained for BACE1 and ERα; in all cases, by increasing the consensus level, a lower number of ligand−protein combinations were considered, but better aRMSD values were obtained. Furthermore, the aRMSD obtained with the consensus docking approach was in all three cases a better value than that obtained with the best docking procedure in the cross-docking analysis.

In order to verify if the consensus level shown by different ligands was influenced by the main ligand properties, the average consensus level shown by each ligand was correlated with the number of its heavy atoms and torsion angles. Interestingly, as shown in Figure S2, this analysis strongly supported the hypothesis that there was no correlation between the properties of the ligands and their average consensus level, thus suggesting that a high consensus level was not dependent on a low number of heavy atoms or a low number of free torsion angles.

From this analysis it is possible to suggest that the consensus docking analysis can be profitably used for evaluating the

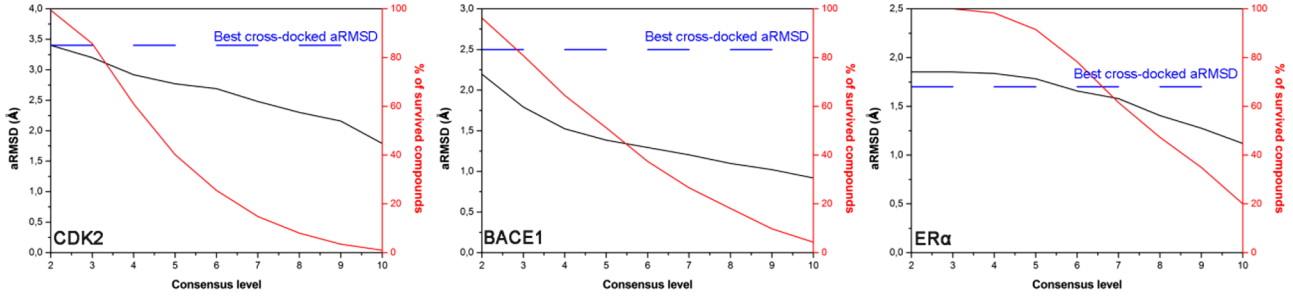


Figure 2. Statistical results of the consensus docking. The blue line indicates the aRMSD value obtained by the best cross-docking results with a single docking procedure, the black line indicates the aRMSD of the consensus docking, and the red line is the percentage of survived compounds.

dx.doi.org/10.1021/ci500424n | J. Chem. Inf. Model. 2014, 54, 2980−2986

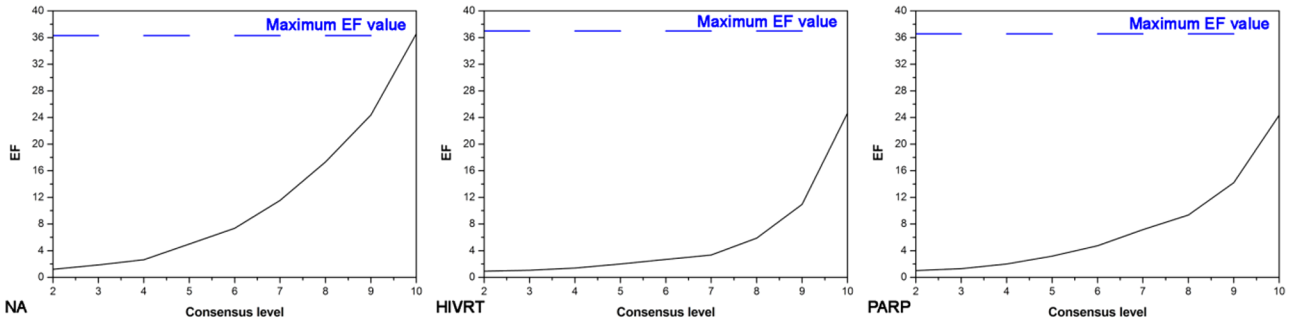


Figure 3. EF plotted versus consensus level for the three data sets. The blue line indicates the maximum EF that can be reached.

Table 2. Consensus Docking Results for the Enriched Database Analysis

	NA			HIVRT			PARP		
consensus level	actives	decoys	EF	actives	decoys	EF	actives	decoys	EF
1	49	1745	1.0	40	1439	1.0	33	1178	1.0
2	49	1440	1.2	35	1366	0.9	33	1161	1.0
3	47	874	1.9	34	1146	1.1	33	902	1.3
4	44	565	2.6	29	742	1.4	32	554	2.0
5	41	259	5.0	25	438	2.0	29	304	3.2
6	30	119	7.4	19	242	2.7	23	154	4.7
7	24	52	11.5	11	110	3.4	18	74	7.2
8	18	20	17.3	10	53	5.9	11	32	9.3
9	10	5	24.4	8	19	11.0	7	11	14.2
10	2	0	36.6	6	3	24.7	4	2	24.4
AUC	0.90			0.67			0.89		

reliability of a ligand docking pose. In particular, a docking pose that shows a higher consensus level will be more reliable than other poses. Furthermore, if for a ligand it is not possible to identify a good consensus level, it is not possible to consider the docking results as reliable and additional studies are necessary, such as molecular dynamic simulations for better assessing the ligand disposition.

As a second step of the consensus docking evaluation, we tested this approach by using enriched databases. To this aim, three different targets of DUD[6] were randomly selected and subjected to the consensus docking analysis. The ligands belonging to the three data sets were subjected to docking calculations by using the ten procedures, and the results were clustered as reported above. The performance evaluation was carried out using well-established accepted metrics, such as enrichment factor (EF) and the AUC of the ROC curve. The DUD data sets were built for reaching a maximum EF value of 35.0−37.0; as shown in Figure 3, the increase of the consensus level determined a general increase of the EF to values that corresponded to more than 65% of the maximum EF that can be reached.

With regard to the AUC, in all three cases it showed good results with values ranging from 0.67 to 0.90 (see Table 2).

Very recently, Arciniega and Lange reported the evaluation of the docking data feature analysis (DDFA) as a new approach for evaluating virtual screening experiments.[23] This approach was tested by using DUD and the obtained results highlighted that it performed with similar good results as the best available methods reported in literature. Since the authors reported the EF_{max} and the AUC values, it was possible to carry out a direct comparison between the two methods. As shown in Figure 4, the two approaches showed similar results for both EF and AUC parameters, therefore confirming the reliability of the virtual screening consensus docking approach.

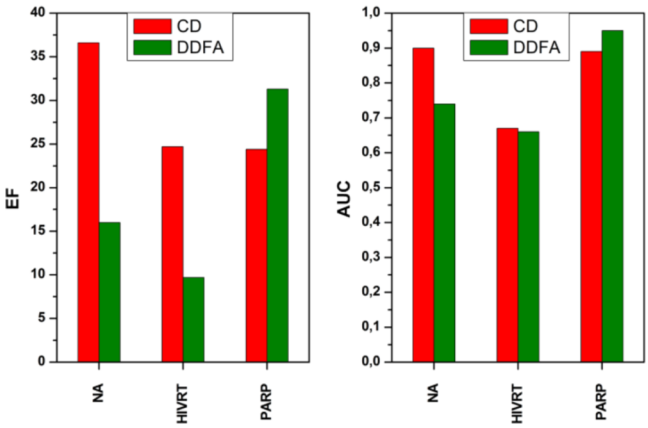


Figure 4. Direct comparison between the virtual screening consensus docking (CD, red) and the docking data feature analysis (DDFA, green) approach. EF (left) and AUC (right) were analyzed.

Finally, the improved reliability of the docking poses obtained with the consensus docking analysis was also tested for the generation of pharmacophoric hypotheses. Even if crystallographic structures of the target are available, it is common practice to use pharmacophoric methods for reinforcing the available information and supporting binding interaction hypotheses. The docked compounds that showed at least a consensus level of nine in the previous enriched databases analysis were used for the generation of three pharmacophoric models (one for each target) and the pharmacophore models were then tested for their ability of

dx.doi.org/10.1021/ci500424n | J. Chem. Inf. Model. 2014, 54, 2980−2986

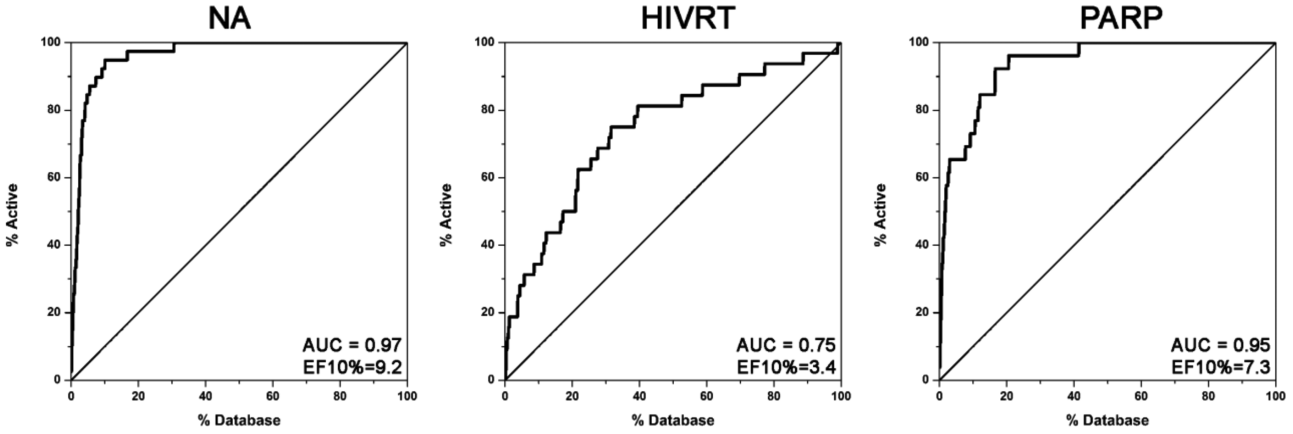


Figure 5. Filtering results for the NA, HIVRT, and PARP enriched database by using the pharmacophoric models.

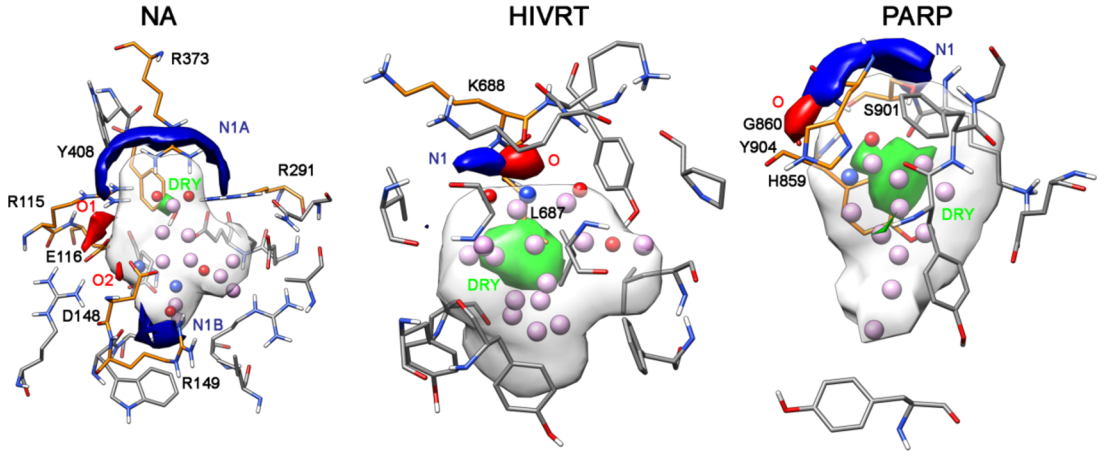


Figure 6. NA, HIVRT, and PARP pharmacophore models represented with the common donor/acceptor/hydrophobic atom locations and pharmacophoric interaction fields (PIFs). The N1, O, and DRY regions are colored blue, red, and green, respectively. The models are overlapped to the corresponding crystallographic receptor structures, and the residues that match with the PIFs are colored orange.

screening the enriched data sets used above. The construction of the models was developed by using the FLAPpharm program.[21] This software, starting from an already set of aligned molecules, is able to generate a pharmacophoric hypothesis by scoring the atom disposition according to the GRID molecular interaction fields (MIFs).[24] A FLAP pharmacophoric model can be defined as a pseudomolecule identified by a pattern of atom-based pharmacophoric points, derived from the common atomic locations of the aligned compounds as well as of field-based pharmacophoric points representing the hotspots obtained by the condensation of the pharmacophore interaction fields corresponding to the common shared portions of the MIFs of the aligned compounds. Interestingly, all the three pharmacophore models showed a good reliability. In particular, for the NA and PARP targets the enriched VS results showed an AUC value greater than 0.90 and $EF_{10\%}$ values of 9.2 and 7.3 for NA and PARP, respectively (see Figure 5). The pharmacophore model obtained for the HIVRT target showed decent results, even if worse than that obtained for the other two targets, with an AUC value of 0.75 and an $EF_{10\%}$ of 3.4.

As the alignment of the ligands for obtaining the pharmacophore models was performed using the structures docked into the corresponding receptor crystallographic structures, it would be useful to check for matching between the residues of the target proteins and the pharmacophoric interaction fields. In Figure 6 the binding sites of the NA, HIVRT, and PARP targets overlap with their corresponding pharmacophore model. As regards NA, the large N1A region corresponded to the interaction with R115, R291, and R373 and the N1B region corresponded to R149; the O1 and O2 regions corresponded to E116 and D148, respectively, whereas the small hydrophobic region DRY corresponded to Y408. For HIVRT, the large DRY region corresponded to the hydrophobic interaction with L687 whereas the N1 and O regions corresponded to the nitrogen and oxygen backbone of K688, respectively. Finally, the large DRY region of the PARP pharmacophore model corresponded to the hydrophobic interaction with Y904, the O region corresponded to the interaction of the oxgen backbone atom of G860 and the large N1 region corresponded to the interaction with the nitrogen backbone of H859 and the hydroxy group of S901.

CONCLUSIONS

The consensus docking approach herein evaluated was able to give important improvements from a qualitative and a quantitative viewpoint. The consensus cross-docking analysis clearly highlighted that a high level of pose consensus statistically improved the reliability of the docking results. A docking pose that shows a higher consensus level will be more likely than the other poses. On the contrary, the lack of consensus for a ligand should be considered as an alert and the

dx.doi.org/10.1021/ci500424n | J. Chem. Inf. Model. 2014, 54, 2980−2986

binding disposition of that ligand should be carefully analyzed in order to identify the most reliable docking pose. From a quantitative viewpoint our results strongly support the hypothesis that virtual screening consensus docking can be profitably used for identifying new hit compounds. The main lack of this approach is the required computing time, because by using this method a whole data set of molecules should be subjected to all the docking procedures. However, in order to simplify this approach it is possible to apply a hierarchical strategy in which the whole data set is subjected to docking calculations by using the two faster docking methods, then only the compounds that show consensus among the two procedures are subjected to the third docking procedure and so on. By using this strategy, with the current CPU available it is possible to apply a VS consensus docking approach with an acceptable computing time. Overall, consensus docking is a simple approach that can support the user in both defining the correct pose of a ligand and identifying new hits in a virtual screening campaign.

■ ASSOCIATED CONTENT

Supporting Information

Clusterization level for the different ligand−protein combination, statistical results of the consensus docking, and correlation between the number of heavy atoms and torsionals against the average consensus level of each ligand. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: +39 0502219595. E-mail: tiziano.tuccinardi@farm.unipi.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Many thanks are due to Prof. Maurizio Botta for the use of the GLIDE program in his computational laboratory (University of Siena, Italy).

■ REFERENCES

(1) Tuccinardi, T. Docking-based virtual screening: recent developments. *Comb. Chem. High Throughput Screen* **2009**, *12*, 303−314.
(2) Tuccinardi, T.; Botta, M.; Giordano, A.; Martinelli, A. Protein kinases: docking and homology modeling reliability. *J. Chem. Inf. Model.* **2010**, *50*, 1432−1441.
(3) Liebeschuetz, J. W.; Cole, J. C.; Korb, O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J. Comput. Aided Mol. Des.* **2012**, *26*, 737−748.
(4) Mukherjee, S.; Balius, T. E.; Rizzo, R. C. Docking validation resources: protein family and ligand flexibility experiments. *J. Chem. Inf. Model.* **2010**, *50*, 1986−2000.
(5) Houston, D. R.; Walkinshaw, M. D. Consensus docking: improving the reliability of docking in a virtual screening context. *J. Chem. Inf. Model.* **2013**, *53*, 384−390.
(6) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.
(7) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(8) *Babel*, version 2.2; OpenEye Scientific Software: Santa Fe, NM, USA, 2010.
(9) *Szybki*, version 1.3; OpenEye Scientific Software: Santa Fe, NM, USA, 2010.
(10) *Macromodel*, version 9.7; Schrödinger Inc: Portland, OR, 2009.
(11) Sanner, M. F. Python: a programming language for software integration and development. *J. Mol. Graph. Model.* **1999**, *17*, 57−61.
(12) 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*, 2785−2791.
(13) *DOCK*, version 6.5; Molecular Design Institute: University of California: San Francisco, CA, 1998.
(14) Case, D. A.; Darden, T. A.; Cheatam, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER*, version 11; University of California: San Francisco, CA, 2010.
(15) *FRED*, version 3.0.0; OpenEye Scientific Software: Santa Fe, NM, USA, 2013.
(16) *OMEGA*, version 2.4.6; OpenEye Scientific Software: Santa Fe, NM, USA, 2013.
(17) *Glide*, version 5.0; Schrödinger Inc: Portland, OR, 2009.
(18) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609−623.
(19) 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.
(20) Poli, G.; Tuccinardi, T.; Rizzolio, F.; Caligiuri, I.; Botta, L.; Granchi, C.; Ortore, G.; Minutolo, F.; Schenone, S.; Martinelli, A. Identification of New Fyn Kinase Inhibitors Using a FLAP-Based Approach. *J. Chem. Inf. Model.* **2013**, *53*, 2538−2547.
(21) Cross, S.; Baroni, M.; Goracci, L.; Cruciani, G. GRID-based three-dimensional pharmacophores I: FLAPpharm, a novel approach for pharmacophore elucidation. *J. Chem. Inf. Model.* **2012**, *52*, 2587−2598.
(22) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.* **2007**, *47*, 279−294.
(23) Arciniega, M.; Lange, O. F. Improvement of virtual screening results by docking data feature analysis. *J. Chem. Inf. Model.* **2014**, *54*, 1401−1411.
(24) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.