PROTEINS: Structure, Function, and Genetics 47:521–533 (2002)

ConsDock: A New Program for the Consensus Analysis of Protein–Ligand Interactions

Nicodème Paul and Didier Rognan*

Bioinformatic Group, Laboratoire de Pharmacochimie de la Communication Cellulaire, UMR CNRS 7081, Illkirch, France

ABSTRACT **Protein-based virtual screening of chemical libraries is a powerful technique for identifying new molecules that may interact with a macromolecular target of interest. Because of docking and scoring limitations, it is more difficult to apply as a lead optimization method because it requires that the docking/scoring tool is able to propose as few solutions as possible and all of them with a very good accuracy for both the protein-bound orientation and the conformation of the ligand. In the present study, we present a consensus docking approach (ConsDock) that takes advantage of three widely used docking tools (Dock, FlexX, and Gold). The consensus analysis of all possible poses generated by several docking tools is performed sequentially in four steps: (i) hierarchical clustering of all poses generated by a docking tool into families represented by a leader; (ii) definition of all consensus pairs from leaders generated by different docking programs; (iii) clustering of consensus pairs into classes, represented by a mean structure; and (iv) ranking the different means starting from the most populated class of consensus pairs. When applied to a test set of 100 protein–ligand complexes from the Protein Data Bank, ConsDock significantly outperforms single docking with respect to the docking accuracy of the top-ranked pose. In 60% of the cases investigated here, ConsDock was able to rank as top solution a pose within 2 Å RMSD of the X-ray structure. It can be applied as a postprocessing filter to either single- or multiple-docking programs to prioritize three-dimensional guided lead optimization from the most likely docking solution. Proteins 2002;47:521–533.** © 2002 Wiley-Liss, Inc.

Key words: cheminformatics; clustering; docking; drug design; virtual screening

INTRODUCTION

Along with the ever-increasing number of experimentally determined protein three-dimensional (3-D) structures[1] and significant advances in docking/scoring functions,[2,3] virtual screening (VS) of "druglike" libraries[4,5] has become an important computational tool for hit identification.[6–8] However, the current limited accuracy of empirical scoring functions hampers the use of virtual screening in hit optimization procedures. If most docking algorithms[9–21] are able to locate fragments/molecules according to experimental observations (X-ray diffraction and NMR), scoring functions still have difficulties in ranking as top solution the orientation that is the closest to the experimental pose. Two major directions have been followed to optimize scoring (fitness) functions associated with docking tools. The first one consists of the careful calibration of master equations[9,22–27] for predicting intermolecular interactions. Among these equations, potentials of mean force are particularly interesting because they are not based on the direct estimation of interaction energies but on the agreement with statistical rules (distribution of interatomic distances) derived from the analysis of crystal structures.[25,27,28] A QSAR-derived alternative is to rely the ranking of final poses not only on a single function but on a jury decision based in the regression analysis of independently derived interaction energy scores.[29,30] A second possible optimization is the postprocessing of docking results by the use of physicochemical filters (protein-buried ligand volume, size of cavities along the protein–ligand interface, apolar solvent-accessible surface of the ligand, number of close contacts) to eliminate unlikely orientations from the top-ranked solutions.[31] Whatever the strategy used, one should end up with as few docking solutions as possible to start a 3-D-guided hit optimization process that is still manageable by synthetic chemists. In the current study, we propose a novel way of ranking docking poses based on the clustering analysis of solutions given by different docking programs. As recently described for consensus scoring[17,34–36] in the context of database screening, ConsDock is shown to outperform single docking in the ability to rank as top solution a pose closer than 2 Å to the experimentally described orientation.

MATERIALS AND METHODS

Preparation of Protein and Ligand Coordinates

Starting from the experimentally determined 3-D coordinates of 100 protein–ligand complexes (Appendix) deposited in the Protein Data Bank,[1] separate sets of coordinates were generated for the whole receptor, the active site, and its corresponding ligand. Crystallographic water molecules were removed from the protein coordinates. Two

Grant sponsor: Strasbourg-Alsace-Lorraine genopole; Grant sponsor: French Ministry of Research and Technology; Grant sponsor: Fondation pour la Recherche Médicale (FRM, Paris, France).

*Correspondence to: Dr. Didier Rognan, Bioinformatic Group, UMR CNRS 7081, 74, route du Rhin, B.P.24, F-67400 Illkirch, France. E-mail: didier.rognan@pharma.u-strasbg.fr

Received 3 January 2002; Accepted 3 January 2002

copies of the protein coordinates were saved: one without hydrogen atoms (in pdb format) and one with hydrogen atoms (in mol2 format) automatically added by using the BIOPOLYMER module of the SYBYL package.[37] Similarly, two copies (without and with hydrogen atoms, in mol2 format) were saved for each ligand of the test set. Last, active site coordinates (without hydrogen atoms, in pdb format) were generated from the protein–ligand complex by extracting all protein amino acids for which at least one heavy atom was located in a 6.5 Å radius sphere centered on the center of mass of the ligand.[10]

Computation of Descriptors for the Set of 100 Ligands

Six descriptors were calculated by using an in-house script in SYBYL programming language (SPL)[37] that computes (i) the molecular weight, (ii) the number of free rotating bonds (involving heavy atoms only, excluding rings and bonds between sp^2 atoms), (iii) the number of hydrogen bond donors, (iv) the number of hydrogen bond acceptors, (v) the MlogP as calculated by the Moriguchi method,[38] and (vi) the polar surface area, computed by the SAVOL3 program.[39]

Dock4.0 Docking

First, a Connolly surface[40] of each receptor active site was generated by using a 1.4 Å probe radius and further used to generate a set of overlapping spheres that were then clustered according to their spatial distribution. To optimize docking accuracy, spheres located too far away from the known ligand position were eliminated from the finally selected cluster. To compute interaction energies, a 3-D grid of 0.35 Å resolution was centered on the active site. The size of the grid box was chosen to enclose all selected spheres using an extra margin of 6 Å. A typical grid had a size of about 20 × 20 × 20 Å and comprised about 300,000 grid points. Energy scoring grids were obtained by using an all-atom model and a distance-dependent dielectric function ($\epsilon = 4r$) with a 10 Å cutoff. Amber95[41] atomic charges were assigned to all protein atoms. The ligand was then docked into the protein active site by matching sphere centers with ligand atoms. A flexible docking (peripheral search and torsion drive) with subsequent minimization was performed as follows: (i) automatic selection and matching of an anchor fragment within a maximum of 100 orientations, (ii) iterative growing of the ligand using at least 30 conformations (peripheral seeds) for seeding the next growing stage with assignment of energy-favored torsion angles, (iii) simultaneous relaxation of the base fragments as well as of all peripheral segments and final relaxation of the entire molecule. Orientations/conformations were relaxed in 100 cycles of 100 simplex minimization steps to a convergence of 0.1 kcal/mol. The top 30 solutions corresponding to the best Dock energy scores were then stored in a single multi mol2 file.

FlexX1.10 Docking

Standard parameters of the FlexX program[10] as implemented in the 6.72 release of the SYBYL package were used for iterative growing and subsequent scoring of FlexX poses. Active-site atoms were defined as previously described. A receptor description file was automatically defined form the pdb coordinates of the hydrogen-free protein/active site coordinates. Formal charges were assigned to ligand atoms. As previously described for Dock, the top 30 solutions were retained and further stored in a single mol2 file.

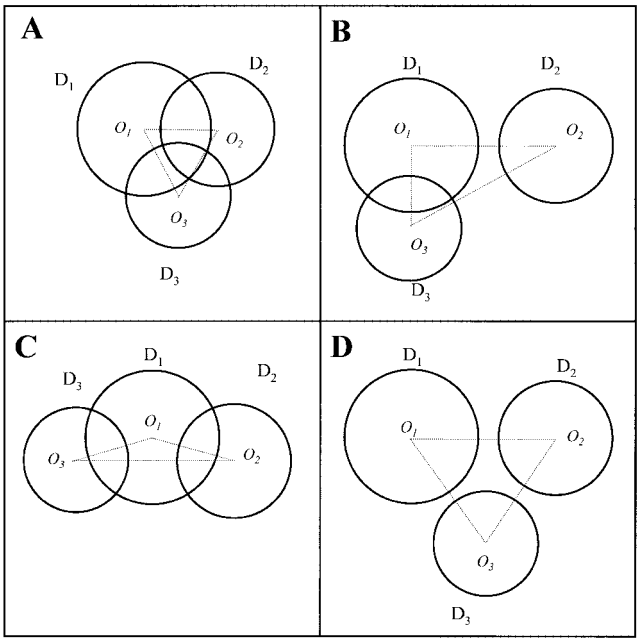

Fig. 1. List of possible cases that may occur during the consensus analysis of clusters originating from three different docking tools. Intersection of three disks D_1, D_2, D_3 of centers O_1, O_2, and O_3, respectively: a consensus solution is found in any case (case A: good triple), if $\max\{\sigma(O_1,O_2),\sigma(O_2,O_3)\} < 2r$ (case B: one pair), if $\sigma(O_2,O_3) < 2r$ (case C: two pairs). In case D (no intersection, at all), no consensus has been found.

Gold1.2 Docking

The active site was defined to encompass any protein atom included in a 10 Å radius sphere centered on the center of mass of the bound-ligand as described in the original PDB entry. Usually, this procedure led to an active site whose dimensions were very similar to that selected by Dock and FlexX. For each of the 10 independent genetic algorithm (GA) runs, a maximum number of 1000 GA operations was performed on a single population of 50 individuals. Operator weights for crossover, mutation, and migration were set to 100, 100, and 0, respectively. To allow poor nonbonded contacts at the start of each GA run, the maximum distance between hydrogen donors and fitting points was set to 5 Å, and nonbonded van der Waals energies were cut off at a value equal to k_{ij} (well depth of the van der Waals energy for the atom pair i,j). The "early-termination" option [applied when the top three solutions are within 1.5 Å root-mean-square deviation (RMSD)] was not selected in the present study to define, as for Dock and FlexX, a set of 30 solutions for each ligand.

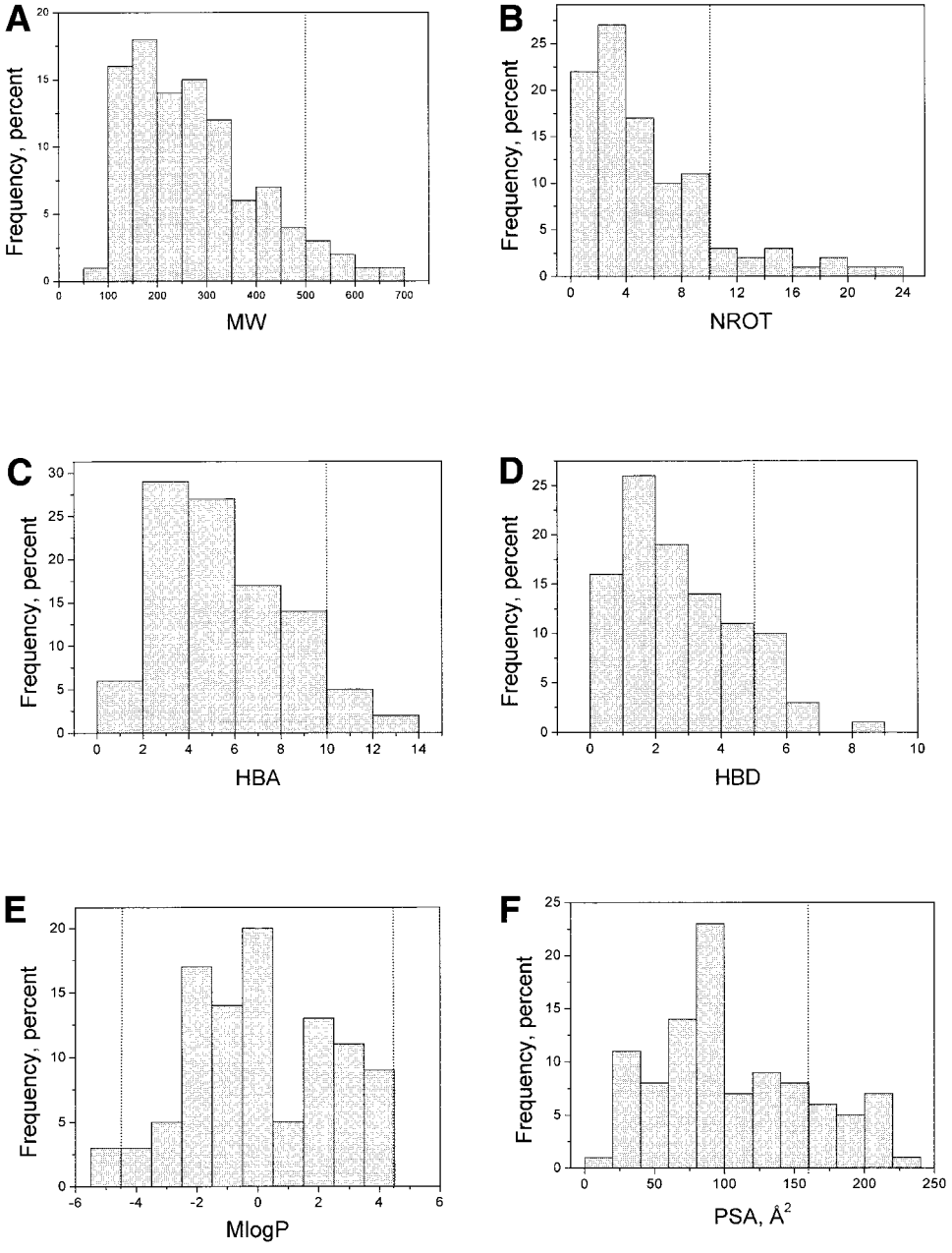

Fig. 2. Physicochemical descriptors of the 100 ligands used in the PDB test set. **A:** Molecular weight. **B:** Number of rotatable bonds. **C:** Number of H-bond acceptors. **D:** Number of H-bond donors. **E:** Calculated MlogP. **F:** Polar surface area (in $Å^2$) of the bioactive conformation. A dotted bar indicate the upper limit, generally accepted to differentiate "druglike"[46,47] molecules from chemicals.

With the above-described sets of parameters, all three docking tools were generally able to dock a ligand within 90–120 s, on a two-processors R12K SGI Origin 200 machine.

ConsDock Analysis

ConsDock runs as a perl postprocessing script that requires as input a multimol2 file of poses selected by each docking tool, a mol2 file of the ligand, and a mol2 or pdb file of the active site. The consensus analysis of all possible poses generated by three docking tools (Dock, FlexX, and Gold) is performed sequentially in four steps:

1. Hierarchical clustering of all 30 poses generated by a docking tool
2. Definition of all consensus pairs originating from different docking programs
3. Clustering of consensus pairs into classes, represented by a mean structure
4. Ranking the different means starting from the most

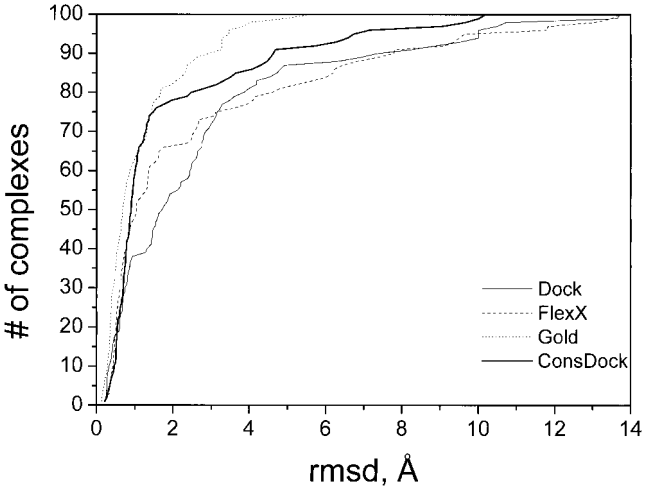

Fig. 3. RMSD (heavy atoms) of the absolute best solution from the experimentally determined pose. Symmetry operators have been taken into account in the RMSD calculation routine to consider heavy atoms belonging to defined chemical group (carboxylate, sulfate, sulfonate, phosphate, phosphonate, guanidine, C-2 symmetrical disubstituted phenyl ring, etc.) as equivalent.

populated class of consensus pairs We will detail each of these four steps in the following section.

Hierarchical clustering

A 3-D representation of a molecule m can be approximated by a point in $\mathbf{R}^{3n+b}$ space, in which n and b represent the number of atomic coordinates x and the number of bonds α, respectively.

$$m = [x_1, x_2, \ldots, x_{3n}, \alpha_1, \alpha_2, \ldots, \alpha_b] \quad (1)$$

Comparing two poses m_1, m_2 of the same molecule can thus be easily achieved by calculating the RMSD σ_{12} between the two sets of coordinates.

$$\sigma(m_1, m_2) = \left[\frac{1}{n}\left(\sum_{i=1}^{N}(x_{1i} - x_{2i})^2\right)\right]^{1/2} \quad (2)$$

with $N = 3n$.

When docking a molecule m to its receptor R, if we denote c_a, c_b two poses of m and c_0 the experimentally determined (from here on called ED) pose of m, we assume that c_a is better than c_b ($c_a \prec c_b$) if $\sigma(c_0, c)_b$. All 30 possible solutions (c_1–c_{30}, c_1 being the top-ranked pose) generated by a docking tool coupled to a scoring function are thus described by a list of poses L.

$$L = \{c_1, c_2, \ldots, c_{30}\} \quad (3)$$

Because the absolute best pose (closest to the ED solution) can be ranked at any position, a hierarchical clustering is performed, starting from the top-ranked pose c_1. For a real positive number r, the first cluster $L_1$is defined as following:

$$L_1 = \{c \in L | \sigma(c, c_1) < r\} \quad (4)$$

For the ensemble of poses L, $L = L_1 \cup L_1^c$, L_1^c being the complementary of L_1 in L. If L_1^c is not empty, by writing $L = L_1^c$, we can get another cluster denoted L_2 from L and continue the hierarchical clustering from the best-ranked pose still available until no pose is left in L. At the end of the clustering, we can define the ensemble of solutions L by the following equation:

$$L = \bigcup_{i=1}^{p} L_i \quad \text{with} \quad p \geq 1 \quad \text{and} \quad L_i \cap L_j = \varnothing \quad \text{for} \quad i \neq j \quad (5)$$

The number of clusters p depends on the choice of r. Thus, we will obtain a list of clusters from a list of poses. For each cluster L_i from the list, we define a radius r_i

$$r_i = \max\{\sigma(c, c_i) \quad \text{with} \quad c_i, c \in L_i \quad \text{and} \quad c_i \prec c \; \forall \, c \in L_i - \{c_i\}\} \quad (6)$$

To speed up calculations, the highest-ranked pose present in a cluster L_i is defined as its leader and will be further used as representative of L_i.

Definition of consensus solutions

All possible docking solutions S are enclosed in three sets O_d, O_f, and O_g generated by Dock, FlexX, and Gold, respectively.

$$S = |O_d| + |O_f| + |O_g|$$

As previously described, we first defined clusters from O_d, O_f, and O_g for a given real positive number r_1. Let C_d be a cluster of O_d characterized by its leader m_d and a radius r_d, C_f a cluster of O_f with leader m_f and radius r_f and C_g a cluster of O_g with leader m_g and radius r_d. Each cluster can be described by a disk D of center m_0 and radius r. The disk $D(m_0, r_1)$ whose center is m_0 and radius r_1 is defined by:

$$D(m_0, r) = \{m \in C_{m_0} | \sigma(m_0, m) \leq r_1\} \quad (7)$$

If we consider only the intersection between disks defined by clusters, four cases are possible: the intersection between the three disks is not empty [Fig. 1(A)], two disks intersect [Fig. 1(B)], one disk intersects the two others [Fig. 1(C)], the three disks do not intersect each other [Fig. 1(D)].

For a molecule m and any given positive real r, a consensus pose is thus possible if $\sigma(m,m_d) < r$, $\sigma(m,m_f) < r$, $\sigma(m,mg) < r$ For all $i,j \in \{d,f,g\}$ we have $\sigma(m_i,m_j) \leq \sigma(mj,\mathrm{m}) + \sigma(m,mj) < 2r$ So, if we want to find triples $\{m_d, m_f, m_g\}$ whose members have their center in the disk $D(m,r)$, we have to look for leaders distant by $< 2r$. To make sure that we will also select the disks whose intersection is not empty, we will take triples such that:

$$\begin{cases} \sigma(m_d, m_f) < \max\{2r, r_d + r_f\} \\ \sigma(m_d, m_g) < \max\{2r, r_d + r_g\} \\ \sigma(m_f, m_g) < \max\{2r, r_f + r_g\} \end{cases} \quad (8)$$

TABLE I. Comparative Docking Accuracy of Three Docking Tools

	Docking tool			
RMSD	Dock	FlexX	Gold	ConsDock
<0.5 Å	18	18	35	11
<1.0 Å	38	48	62	59
<1.5 Å	45	61	77	75
<2.0 Å	54	66	82	78

Percentage of best solutions (closest to the X-ray pose) within a defined RMSD from the experimentally determined pose.

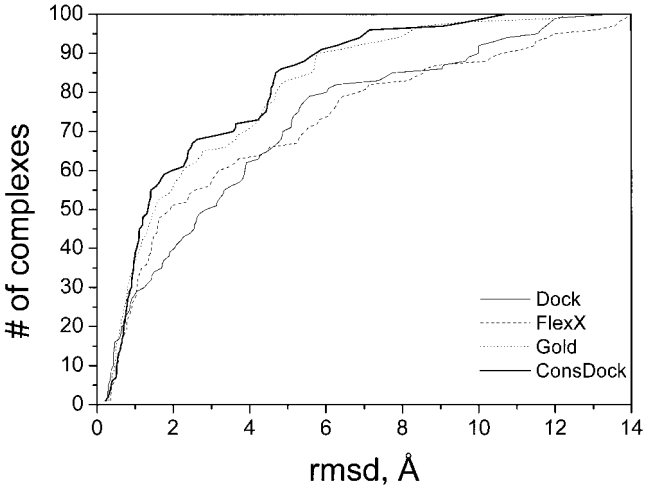

Fig. 4. RMSD (heavy atoms) of the top-ranked solution pose from the experimentally determined pose.

In this case, we will refer to $2r$ as the consensus radius. It measures the accuracy with which the best poses will be chosen from all those available (Fig. 1).

One good triple should be identified each time a leader from each docking tool is closer than r to the ED solution. However, bad triples (far from the ED pose) can nevertheless be found from bad leaders for which the consensus condition is true. To avoid such cases, unrealistic leaders are eliminated if the distance between the center of mass of the ligand (c^L_{mass}) and that of the receptor (c^R_{mass}) is higher than 7.5 Å, a cutoff value that allows a clear discrimination between possible poses (ligand docked in the binding site) and unrealistic ones (ligand docked outside the inner part of the binding site). This simple filter was shown to be very effective in cases where a docking tool was unable to propose any reliable solution within the first 30 poses.

Thus, the success of getting good triples depends on the probability of the respective docking tools to supply good poses. Let p_d be the probability for Dock to supply good poses (RMSD from X-ray lower than 2 Å), p_f for FlexX, and p_g for Gold. If we assume that every docking tool has 60% chances to provide a good pose, $p_d = p_f = p_g = 0.6$. Then, the probability of getting good triples ($p = p_d p_f p_g = 0.216$) is significantly lower than the probability to have bad ones ($q = 1 - p = 1 - p_d p_f p_g = 0.784$). Thus, ConsDock defines consensus poses out of pairs and not of triples. The advantage of this consensus scheme is that good poses are only required for two of the three docking engines. The probability of getting good pairs is then:

$$p = p_d p_f (1 - p_g) + p_d (1 - p_f) p_g + (1 - p_d) p_f p_g + p_d p_f p_g \tag{9}$$

If we take as previously, $p_d = p_f = p_g = 0.6$, then $p = 0.648$ and $q = 0.352$. The probability of getting good pairs is thus far higher than that of getting good triples. If no consensus pairs can be found, all unrealistic poses (center of mass the ligand more than 7.5 Å away from that of the active site) are first eliminated and the largest resulting cluster given by each of the three docking tools is extracted. The one containing the best-ranked pose is further selected for definition of the mean pose. This procedure enables ConsDock to propose at least one solution whatever the consensus situation.

Clustering of consensus pairs: definition of mean poses

Each above-defined consensus pair (c_i, c_j) is represented by its mean:

$$\bar{c}_{ij} = \frac{1}{2}(c_i + c_j) \tag{10}$$

All possible means are then pooled together and clustered as previously described by using a 2 Å consensus radius (r = 1 Å). Because the hierarchical clustering depends on the way means are classified, clustering starts from the means issued from both tools having the most consensus pairs in common. Each cluster of mean (class) is then represented by its mean pose $\bar{\bar{c}}$. When no consensus classes could be defined, the most populated cluster previously defined by an individual docking tool was simply taken as representative of the best possible pose. This allows the definition of a consensus pose for any protein–ligand complex.

Ranking the means: definition of the supermean

All possible classes are ranked according their size (number of enclosed means). The top-ranked class (or supermean $\bar{\bar{c}}_1$) is considered as the most representative of all consensus classes. The supermean is then defined as:

$$\bar{\bar{c}} = \frac{1}{p}\sum_{i=1}^{p} \bar{c}_i \quad \text{with} \quad p = |C_1| \quad \text{and} \quad \bar{c}_i \in C_1 \,\forall\, i = 1 \cdots p \tag{11}$$

To ensure a proper geometry of the supermean, its coordinates are quickly relaxed by 100 steps of steepest-descent energy minimization using the TRIPOS force-field.[42] As possible alternatives, the user can choose (i) the energy-minimized mean that is the nearest to the supermean or (ii) the pose among all 90 available that is the nearest to the supermean.

Statistical analysis of the consensus analysis of our 100 protein–ligand test set indicates that the minimized super-

TABLE II. Quality of the Best-Ranked Pose Predicted by Three Docking Tools

	Docking tool			
RMSD	Dock	FlexX	Gold	ConsDock
<0.5 Å	16	7	15	10
<1.0 Å	28	27	36	39
<1.5 Å	34	43	49	55
<2.0 Å	39	51	56	60

Percentage of top-ranked solutions within a defined RMSD from the experimentally determined pose.

mean generally corresponds to the best choice, selected by default in the current version of ConsDock.

RESULTS AND DISCUSSION

A test set of 100 protein–ligand X-ray structures (Appendix) has been selected according to previously described benchmarks.[10,12,21] Analysis of several descriptors show that the corresponding ligands spread a broad spectrum of physicochemical properties (Fig. 2), for about 90% of them within the frame of what can be considered as "drug likeness".[5]

Two poses were examined for each docking tool. First, we looked at the pose that is the closest to the ED solution. It illustrates the propensity of the docking tool to find a reliable solution, whatever its ranking. Second, the top-ranked pose was examined with respect to the ED solution. It reflects the ability of the related scoring/fitness function to properly rank poses during the incremental flexible docking procedure. From here on, we will consider a tool to be successful if it can propose one solution within a certain RMSD (usually 2 Å) from the ED pose.

Quality of the Best Possible Pose Proposed by Single and Consensus Docking

First, the present study represents an unprecedented opportunity to critically evaluate the three most widely used docking tools using a similar set of protein–ligand complexes. In our hands, Gold is shown to significantly outperform FlexX and Dock in terms of pure docking accuracy (Fig. 3). At a 1 Å RMSD cutoff, which reflects a very high docking precision, Gold is successful in 62% of the cases, whereas FlexX and Dock have only 48 and 38% success rate, respectively (Table I). The same discrepancy occurs up to a 2 Å cutoff, the lower possible accuracy level for lead optimization purpose. At the latter cutoff, Gold is successful in 82% of all 100 cases, whereas significantly lower performances are obtained by FlexX and Dock (66 and 54%, respectively). All three tools have been parameterized to dock a single ligand at a relatively similar pace (90–120 s/ligand). By using these library screening settings, Gold should be considered as the method of choice if one is interested in getting the best possible solution whatever its rank. It is of interest that our consensus docking analysis by ConsDock is nearly as accurate as the best tool (Gold) it is partly derived from (Fig. 3). At a 2 Å docking accuracy, it is successful in 77% of the cases. Recall that our primary goal was to focus our attention on the top-ranked pose. Thus, consensus docking is not performed at the detriment of the accuracy of the absolute best pose. In terms of pure docking accuracy, ConsDock is much closer to the best tool (Gold) than to the second one (FlexX). Thus, the consensus analysis is biased toward the consideration of the best possible poses. It is surprising that ConsDock is less accurate that any single docking program in finding highly accurate poses (RMSD < 0.5 Å,

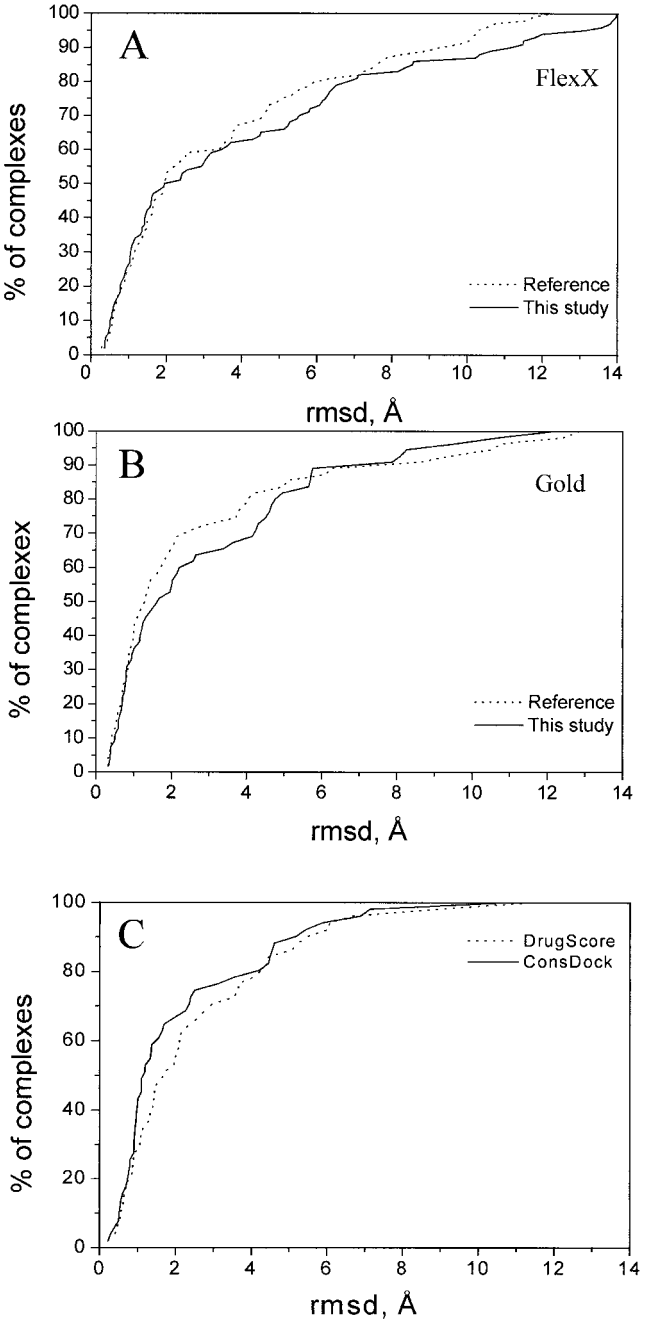

Fig. 5. Comparison of virtual screening parameters (this study) and available benchmarks (reference study) for the docking accuracy (RMSD of the top-ranked solution from the experimentally determined pose) of different docking tools. **A:** FlexX. **B:** Gold. **C:** DrugScore. The comparison with published FlexX, Gold, and DrugScore results was limited to the examination of 100, 55, and 51 common complexes, respectively.

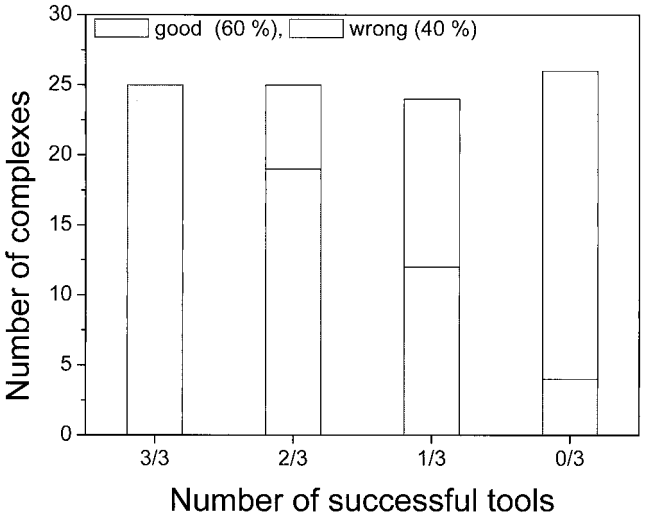

Fig. 6. Statistical analysis of consensus docking for 100 protein–ligand complexes. For each case, the number of successful tools used to derive a consensus (from 0 to 3) is recorded, and the final result of the consensus (successful or not) indicated. Success is considered when the top-ranked pose is predicted within 2 Å RMSD from the X-ray solution.

Table I). We do not really have a clear explanation to this unexpected feature but will look whether this observation is reproducible for another set of PDB entries.

Quality of the Top-Ranked Pose Proposed by Single and Consensus Docking

We next looked at the accuracy with which single and consensus scoring defines the top-ranked solution. Again, Gold was shown to be superior to FlexX and then Dock, when library-screening settings are selected (Fig. 4, Table II). However, the significant difference found between Gold and FlexX when the absolute best pose is considered (Fig. 3) is attenuated when the top-ranked solution is examined. At a 2 Å RMSD precision, Gold is successful in 56% of the cases, whereas FlexX performs well in 51% of the studied complexes. It should also be noticed that selecting a tighter cutoff (1.5 Å RMSD from ED pose) would not have affected that much the performance of both tools (Table II). The advantage of consensus docking over single docking is shown to start for poses that were already well predicted, because ConsDock begins to outperform the three individual docking tools when a 1 Å RMSD from the ED pose is selected as cutoff (Fig. 4). To quantify the advantage of consensus over single docking, we computed the improvement rate IR defined as:

$$\mathrm{IR} = [(\%\mathrm{ConsDock} - \%\ \mathrm{Single})/\%\ \mathrm{Single}] \times 100 \quad (12)$$

where % ConsDock = % of good poses found by ConsDock and % Single = % of good poses found by a single docking tool.

The biggest advantage of consensus over any of the single docking is observed for poses closer than 1.5 Å to the ED solution. The improvement rate of ConsDock over Dock, FlexX, and Gold is then of +61, +28 and +12%, respectively. At a 2 Å RMSD cutoff, the improvement drops slightly (+53, +17 and +7% over Dock, FlexX, and Gold, respectively) to become constant for RMSD values higher than 5 Å.

It could be argued that the small gain obtained with ConsDock with respect to the best individual tool (Gold) does not justify the use of consensus docking. One advantage of ConsDock is that it is successful in 15 of the 44 cases for which Gold could not find any reliable pose within 2 Å RMSD of the ED solution. Furthermore, Gold is limited to active sites in which at least one H-bond donor/acceptor can be detected, a feature that is not necessary in the other two docking tools our consensus docking approach is derived from. Last, it is likely that a consensus docking approach will be less sensitive than single docking to small variations of protein coordinates (induced fit) on ligand binding. To illustrate this phenomenon, we took two PDB entries (1acj, 1ack; see Appendix) corresponding to the same target (acetylcholinesterase) cocrystallized with two different inhibitors (1acj: tacrine, 1ack: edrophonium ion). We next tried to dock edrophonium in the acj active site and tacrine in the ack coordinates. For both ligands, the RMSD of the top-ranked pose from the X-ray solution was significantly lower by using ConsDock (0.89 and 1.13 Å, respectively) than any of the single docking program (from 2.5 to 5.8 Å).

Comparison of the Current Study With Existing Benchmarks

We have compared the way we used individual tools with respect to available benchmarks for two main reasons: (i) the performance of ConsDock is directly related to the accuracy of the individual tools it is derived from; thus, it is important to optimally use each of the docking tool, and (ii) consensus docking is here envisaged by using virtual library-screening settings to reduce the time necessary for docking in triplicate. We should then ascertain that the docking speed (90–120 s/ligand/tool) used in the current study is not gained at the detriment of docking accuracy. This analysis has been limited to FlexX and Gold because benchmarks are only available for the latter docking programs.[43–45] To avoid biases from the examination of different training sets, we only compared the accuracy of each tool (current study vs. reference study) for protein–ligand complexes that are common to both datasets. When looking at the performance of each docking tool to predict a top-ranked pose within 2 Å RMSD from the ED solution (Fig. 5), we see that our implementation of library-screening parameters is not detrimental to the accuracy of FlexX [Fig. 5(A)]. However, a loss of performance (−18% for poses where RMSD < 2 Å) is noticed for Gold [Fig. 5(B)]. This observation is expected because optimal usage of Gold implies very different genetic algorithm settings (population size, number of operations, etc.), resulting in higher accuracy (66% of good poses within 2 Å instead of 56% in the current study) but at the price of a much reduced speed (ca. 30 min/ligand). We believe that this slight drop in accuracy is still acceptable with respect to the considerable CPU time that has been spared. In any case, the automatic removal of water

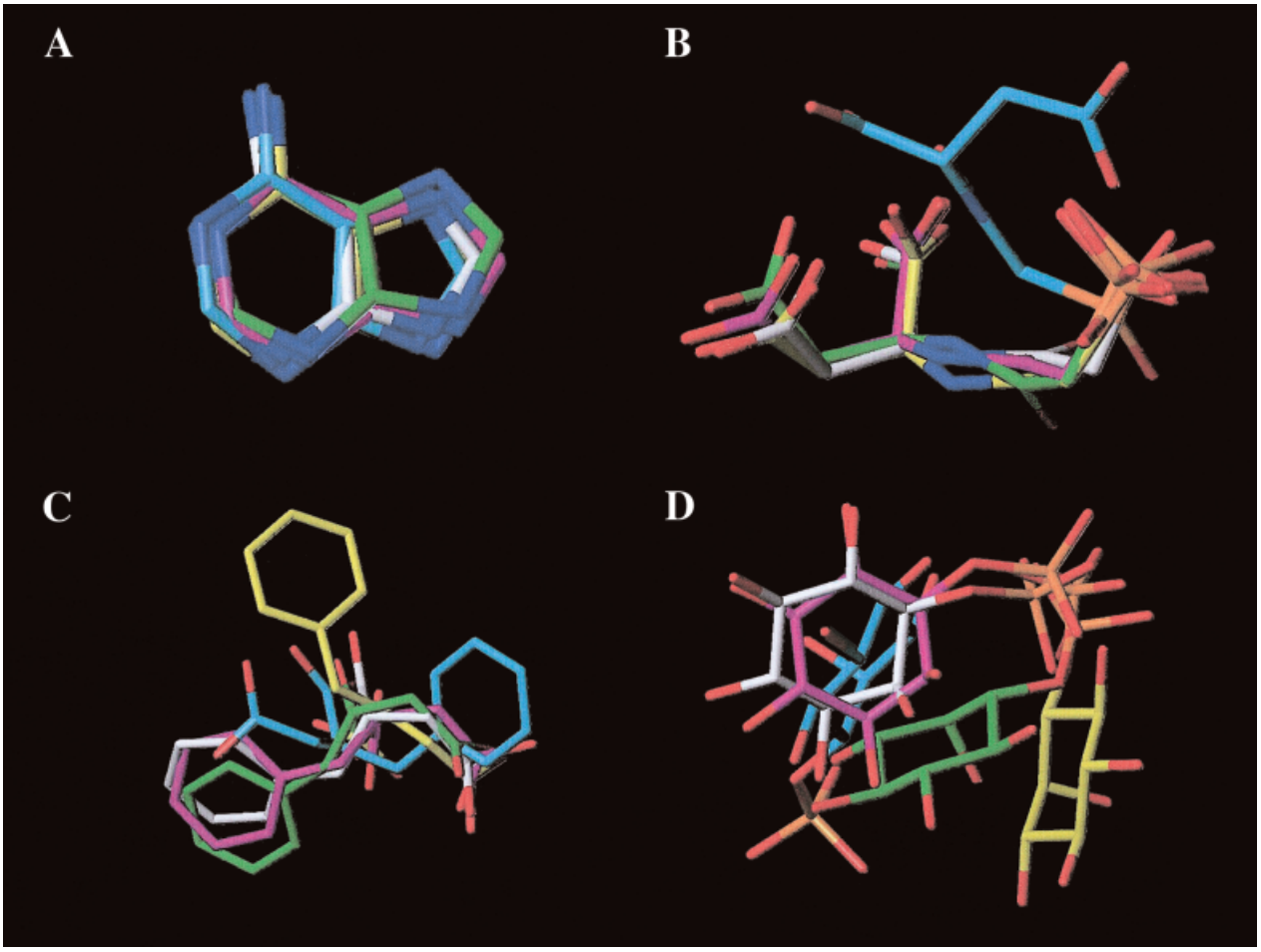

Fig. 7. Overview of four possible test cases occurring in successful consensus docking (supermean within 2 Å of the X-ray solution). **A:** All three docking tools are successful (1aha). **B:** Two docking tools are successful (1acm). **C:** One docking tool is successful (1hyt). **D:** None of the three docking tools is successful (1imb). Top-ranked poses predicted by Dock, FlexX, Gold, and ConsDock are overlaid to the X-ray solution and displayed by using the following color coding: oxygen, red; nitrogen, blue; phosphorus, orange. Carbon atoms of the Dock, FlexX, Gold, ConsDock, and X-ray pose are colored in cyan, yellow, green, magenta, and white, respectively. RMSD values are given in the Appendix.

molecules as well as the absence of ligand minimization before docking is rather well tolerated for both docking tools. It is possible that the absence of ligand minimization could favor poses sharing some dihedral angles with the X-ray conformation although incremental construction of ligands during the flexible docking should not favor predefined starting conformations. Furthermore, energy minimization of ligands contained in a large database screened for molecular docking is a very intensive task that is rarely performed. Because we wanted to investigate VS settings for the three docking programs, no ligand was minimized before docking.

We next compared our consensus scoring procedure to a recently described postprocessing filter (DrugScore) that uses as scoring function a potential of mean force[27] instead of an empirical binding free energy equation. DrugScore was stated to be significantly superior to either FlexX or Dock docking/scoring in identifying the top-ranked pose within 2 Å RMSD from the X-ray solution. When applied to a set of protein–ligand complexes present in both the DrugScore and our data set (51 complexes in total), ConsDock is shown to outperform DrugScore for the accuracy of the top-ranked pose [Fig. 5(C)]. In 67% of the common cases investigated by DrugScore and ConsDock, our consensus docking routine is able to predict the top-ranked pose with an RMSD from the X-ray structure lower than 2 Å. ConsDock shows, for this peculiar set of complexes, an improvement of 13.5% over DrugScore.

Statistical Analysis of Consensus Clustering

Of the 100 complexes investigated in the current study, we made a statistical analysis of consensus docking for the number and identity of individual tools used to generate consensus and whether the consensus pose was considered as successful. Eight situations can theoretically occur (Fig. 6), depending on the number of successful tools (from 0 to 3) and the quality of the resulting consensus (successful or not). As expected, a consensus is always successful (RMSD of the top-ranked pose lower than 2 Å) when it is derived out of three tools for which a reliable solution could also be found (Fig. 6). This happened for 25 protein–ligand complexes and is examplified by 1aha [Fig. 7(A)]. When only two tools worked properly (total of 25 complexes), a

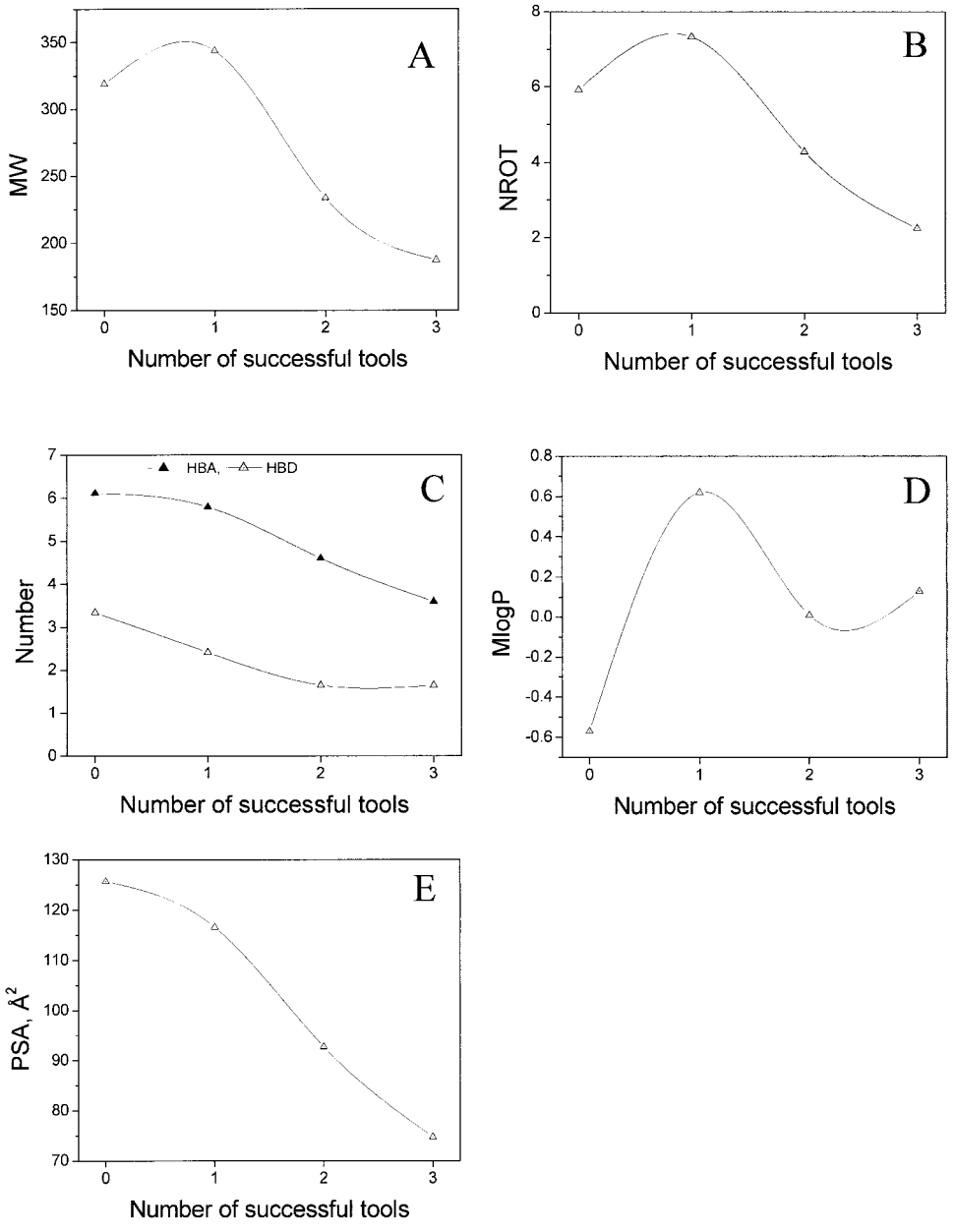

Fig. 8. Averaged properties of ligands successfully docked by 0, 1, 2, or 3 docking tools. **A:** Molecular weight. **B:** Number of free rotating bonds. **C:** Number of H-bond donors and acceptors. **D:** MlogP. **E:** Polar surface area.

successful consensus could be obtained in 76% of these cases (Fig. 6). This situation is examplified by 1acm [Fig. 7(B)]. It is of interest that lower performance of individual tools does not preclude for the prediction of a good consensus. In 24 complexes, only one of the three tools was able to place the top-ranked solution within 2 Å of the X-ray solution. Nevertheless, a good consensus pose was found in 12 (50%) of these 24 complexes [e.g., 1eed, Fig. 7(C)]. Last, 26 ligands could not be successfully docked by any of the three tools. It is of interest that a good consensus could be obtained for 4 of these 26 hard cases [e.g., 1imb, Fig. 7(D)]. The good performance of ConsDock in difficult cases (less than two successful docking tools) can be easily explained by (i) the tendency of our approach to favor good poses by filtering out unrealistic solutions for which the center of mass of the ligand is located far away from the inner part of the active site and (ii) the special treatment of cases for which no consensus pairs could be defined. Hence, the failure of any single docking tool to rank at the top of the list a pose within 2 Å RMSD from the ED solution does not mean that a solution ranked at the bottom of this list cannot be a good one.

Looking at simple physicochemical properties (molecular weight, number of rotatable bonds, number of H-bond

TABLE III. Relationships Between the ConsDock Score and the Docking Accuracy of the Predicted Top-Ranked Pose

ConsDock score	RMSD, Å	Number of ligands
0	5.44	1
1	3.88	10
2	3.77	12
3	3.57	15
4	2.03	11
5	1.85	17
6	1.45	34

Docking accuracy is expressed as the average RMSD from the X-ray structure of all ligands sharing the same ConsDock score.

donors and acceptors, MlogP, and polar surface area) of ligands whose top-ranked pose that has been docked with a 2 Å RMSD accuracy, with either none, one, two, or all three docking tools helps to delineate some trends required for successful docking (Fig. 8). It is of interest that there is for most of the above-described properties, a clear discrimination between the set of physicochemical parameters favoring docking (at least two accurate docking tools) and those unsuitable for our consensus analysis (Fig. 8). In the set of 100 complexes investigated here, 88% of the ligands successfully docked by at least two docking tools are properly docked by ConsDock. By analogy to the well-known "rule of 5," which helps discriminating between potential druglike molecules and chemicals,[46,47] we propose a set of rules favoring successful docking (ConsDock score) depending on physicochemical properties of the ligands to dock, encoded in the following function:

$$\text{ConsDockscore} = S_{MW} + S_{NRT} + S_{HBD} + S_{HBA} + S_{M\,\log P} + S_{PSA} \quad (13)$$

where $S_{MW} = 0$ if MW > 300 or 1 if MW ≤ 300
where $S_{NRT} = 0$ if NRT > 6 or 1 if NRT ≤ 6
where $S_{HBD} = 0$ if HBD > 3 or 1 if HB ≤ 3
where $S_{HBA} = 0$ if HBA > 5 or 1 if HBA ≤ 5
where $S_{M\log P} = 0$ if $|\text{MlogP}| > 3$ or 1 if $|\text{MlogP}| \leq 3$
where $S_{PSA} = 0$ if PSA > 110 Å^2 or 1 if PSA ≤ 110 Å^2

S_n is here the score associated with the parameter n (MW: molecular weight; NRT: number of free rotating bonds; HBD: number of H-bond donors; HBA: number of H-bond acceptors; MlogP: logP calculated by the Moriguchi method[38]; PSA: polar surface area). Clustering the 100 ligands according to the computed ConsDock score shows a clear relationship between the latter score and the docking accuracy with which this ligand could be docked (Table III). Well-docked ligands generally have a ConsDock score between 3 and 6; 77% of the ligands whose top-ranked pose has been predicted within 2 Å RMSD of the X-ray structure have a ConsDock score of at least 3. The ConsDock score could be used in VS either as a preprocessing tool for eliminating molecules that are difficult to dock or as a postprocessing filter for favoring molecules predicted to be more easily docked.

CONCLUSIONS

The present study proposes a new approach for analyzing docking results based on the consensus analysis of the output of several docking programs. The idea to develop a consensus docking analysis tool derives from the recent discovery that consensus scoring significantly outperforms any single scoring function in the identification of potential hits from the protein-based virtual screening of chemical databases. As recently demonstrated,[34] the main effect of consensus scoring is that the value predicted from repeated sampling tends to be closer to the true value. When applied to a set of 100 protein–ligand complexes, consensus docking is shown to outperform any single docking tool in the quality (RMSD from the true solution) of the top-ranked pose. ConsDock can be applied as a postprocessing filter to single or multiple docking programs to significantly reduce the number of possible solutions, and thus prioritize lead optimization from the top-ranked solution. It has been here applied to the three most widely used flexible docking programs (Dock, FlexX, and Gold) but can be easily applied to the consensus analysis of other docking tools[13–21] in either single mode (docking of a single ligand) or database mode (docking of a 3-D database). Our implementation of ConsDock is perfectly compatible with the consensus analysis of docking results for large libraries (ca. 100,000 molecules). Even if ConsDock is basically more suited to hit optimization than hit identification, it could be used as a virtual screening post-processing tool. Using 32 processors of a R14K SGI Origin3800, a 100,000 ligands-containing database can be docked in triplicate (Dock, FlexX, and Gold) within a week. ConsDock postprocessing can thus be distributed in parallel over several processors (e.g., 32) to achieve the analysis of 9 million poses ($30 \times 100{,}000 \times 3$) generated by the three docking tools within 3 h. Because ConsDock does not score the selected poses, the most likely solution for each ligand of the database has to be rescored independently with single or multiple scoring functions to enable the selection of top scorers for experimental testing.

ACKNOWLEDGMENTS

D.R. thanks the Centre Informatique National de l'Enseignement Supérieur (CINES, Montpellier, France) for allocation of computing time. We also acknowledge Dr. Holger Gohlke (The Scripps Research Institute, La Jolla, CA) for supplying the DrugScore benchmarks.

REFERENCES

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
2. Tame JRH. Scoring functions: a view from the bench. J Comput Aided Mol Des 1999;13:99–108.
3. Gohlke H, Klebe G. Statistical potentials and scoring functions applied to protein-ligand binding. Curr Opin Struct Biol 2001;11: 231–235.
4. Oprea TI. Property distribution of drug-related chemical databases. J Comput Aided Mol Des 2000;14:251–264.
5. Pickett SD, McLay IM, Clark DE. Enhancing the hit-to-lead properties of lead optimization libraries. J Chem Inf Comput Sci 2000;2:263–272.
6. Hopkins SC, Vale RD, Kuntz ID. Inhibitors of kinesin activity

from structure-based computer screening. Biochemistry 2000;39: 2805–2814.
7. Böhm HJ, Boehringer M, Bur D, Gmuender H, Huber W, Klaus W, Kostrewa D, Kuehne H, Luebbers T, Meunier-Keller N, Mueller F. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. J Med Chem 2000;13:2664–2674.
8. Grüneberg S, Wendt B, Klebe G. Subnanomolar inhibitors from computer screening: a model study using human carbonic anhydrase II. Angew Chem Int Ed 2001;40:389–393.
9. Gehlhaar DK, Verkhivker GM, Rejto PA, Sherman CJ, Fogel DB, Fogel LJ, Freer ST. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. Chem Biol 1995;2:317–324.
10. Rarey M, Kramer B, Lengauer T, Klebe G. A fast flexible docking method using an incremental construction algorithm. J Mol Biol 1996;261:470–489.
11. Ewing TJA, Kuntz ID Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem 1997;18:1175–1189.
12. Jones G, Wilett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J Mol Biol 1997;267:727–748.
13. McMartin C, Bohacek R. QXP: powerful, rapid computer algorithms for structure-based drug design. J Comput Aided Mol Des 1997;11:333–344.
14. Totrov M, Abagyan R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins 1997;1:215–220.
15. Baxter CA, Murray CW, Clark DE, Westhead DR, Eldridge MD. Flexible docking using Tabu search and an empirical estimate of binding affinity. Proteins 1998;33:367–382.
16. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ. Automated docking using a Lamarckian genetic algorithm and empirical binding free energy function. J Comp Chem 1998;19:1639–1662.
17. Charifson PS, Corkery JJ, Murcko MA, Walters WP. 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.
18. Liu M, Wang S. MCDOCK: a Monte Carlo simulation approach to the molecular docking problem. J Comput Aided Mol Des 1999;13: 435–451.
19. Hou TJW, Chen L, Xu X. Automated docking of peptides and proteins by using a genetic algorithm combined with a tabu search. Protein Eng 1999;12:639–647.
20. Perola E, Xu K, Kollmeyer TM, Kaufmann SH, Prendergast FG, Pang YP. Successful virtual screening of a chemical database for farnesyltransferase inhibitor leads. J Med Chem 2000;43:401–408.
21. Diller DJ, Merz KM Jr. High throughput docking for library design and library prioritization. Proteins 2001;43:113–124.
22. Böhm HJ. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. J Comput Aided Mol Des 1994;8:243–256.
23. Eldridge M, Murray CW, Auton TA, Paolini GV, Lee RP. Empirical scoring functions. I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. J Comput Aided Mol Des 1997;11:425–445.
24. Wang R, Liu L, Lai L, Tang Y. SCORE: a new empirical method for estimating the binding affinity of a protein-ligand complex. J Mol Model 1998;4:379–384.
25. Muegge I, Martin YC. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. J Med Chem 1999;42:791–804.
26. Rognan D, Laumoeller SL, Holm A, Buus S, Tschinke V. Predicting binding affinities of protein ligands from three-dimensional coordinates: application to peptide binding to class I major histocompatibility proteins. J Med Chem 1999;42:4650–4658.
27. Gohlke H, Hendlich M, Klebe G. Knowledge-based scoring function to predict protein-ligand interactions. J Mol Biol 2000;295:337–356.
28. Mitchell JBO, Laskowski RA, Alex A, Thornton JM. BLEEP: potential of mean force describing protein-ligand interactions. I. Generating potential. J Comput Chem 1999;11:1165–1176.
29. Head RD, Smyte ML, Oprea TI, Waller CL, Green SM, Marshall GR. VALIDATE: a new method for the receptor-based prediction of binding affinities of novel ligands. J Am Chem Soc 1996;118: 3959–3969.
30. Terp GE, Johansen BN, Christensen IT, Jorgensen FS. A new concept for multidimensional selection of ligand conformations (MultiSelect) and multidimensional scoring (MultiScore) of protein-ligand binding affinities. J Med Chem 2001;14:2333–2343.
31. Stahl M, Bohm HJ. Development of filter functions for protein-ligand docking. J Mol Graph Model 1998;16:121–132.
32. Kearsley SK, Sallamack S, Fluder EM, Andose JD, Mosley RT, Sheridan RP. Chemical similarity using physiochemical property descriptors. J Chem Inf Comput Sci 1996;36:118–127.
33. Ginn CMR, Turner DB, Willett P, Ferguson AM, Heritage TW. Similarity searching in files of three-dimensional chemical structures: evaluation of the EVA descriptor and combination of rankings using data fusion. J Chem Inf Comput Sci 1997;37:23–37.
34. Bissantz C, Folkers G, Rognan D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. J Med Chem 2000;43:4759–4767.
35. Stahl M, Rarey M. Detailed analysis of scoring functions for virtual screening. J Med Chem 2001;44:1035–1042.
36. Wang R, Wang S. How does consensus scoring work for virtual library screening? An idealized computer experiment. J Chem Inf Comput Sci 2001;41:1422–1426.
37. SYBYL 6.72 release, Tripos Inc., St. Louis, MO 63144, USA.
38. Moriguchi L, Hirono S, Liu Q, Nakagome K, Matstushita Y. Simple method of calculating octanol/water partition coefficient. Chem Pharm Bull 1992;40:127–130.
39. Pearlman RS, Skell JM, Deanda F. Laboratory for Molecular Graphics and Theoretical Modelling, College of Pharmacy, Univerity of Texas, Austin, TX 78712.
40. Connolly ML. Solvent accessible surface of proteins and nucleic acids. Science 1983;221:709–713.
41. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. J Am Chem Soc 1995;117:5179–519742.
42. Clark M, Cramer RD III, van Opdenbosch N. Validation of the general purpose TRIPOS 5.2 force-field. J Comput Chem 1989;10: 982–1012.
43. Performance of Flexx on a test set of 200 protein-ligand complexes can be found at http://cartan.gme.de/FlexX/html/flexx-eval.html
44. Performance of Gold on a test set of 99 protein-ligand complexes can be found at http://www.cdc.cam.ac.uk/prods/gold/rms_tab.html
45. Kramer B, Rarey M, Lengauer T. Evaluation of the FlexX incremental construction algorithm for protein-ligand docking. Proteins 1999;37:228–241.
46. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Delivery Rev 1997;23:3–25.
47. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 2001;46:3–26.

APPENDIX

	Dock			Flex X			Gold			ConsDock		
PDB code	RMSD1[a]	rank[b]	RMSD2[c]	RMSD1	rank	RMSD2	RMSD1	rank	RMSD2	RMSD1	rank	RMSD2
1aaq	13.65	20	11.76	1.51	20	1.56	2.21	1	2.21	1.76	1	1.76
1abe	0.31	8	0.35	0.41	7	3.04	0.17	5	0.34	0.29	1	0.29
1acj	0.33	1	0.33	2.61	6	3.19	0.36	19	2.58	0.28	2	2.63
1ack	0.83	8	3.87	0.51	8	1.04	0.64	26	1.96	0.52	2	1.41
1acm	2.99	16	4.85	0.48	11	0.58	0.31	2	0.71	0.78	1	0.78
1aha	0.26	7	0.36	0.51	1	0.51	0.26	14	0.39	0.35	1	0.35
1apt	8.86	29	9.03	6.01	6	6.23	2.36	13	8.26	6.57	1	6.57
1atl	2.79	20	3.33	0.99	7	1.48	1.4	2	6.4	1.09	1	1.09
1azm	0.81	17	2.62	1.62	30	2.37	1.32	25	1.68	0.97	4	1.2
1baf	0.7	11	0.9	6.81	27	7.08	1.44	10	4.76	1.01	3	4.66
1bbp	ns[d]	—	ns	13.04	6	13.08	0.67	5	2.25	0.99	1	0.99
1cbs	ns	—	ns	1.32	2	2.39	1.52	3	5.32	1.48	4	2.31
1cbx	0.61	7	0.98	1.04	6	1.43	0.67	4	0.76	0.94	1	0.94
1cil	0.27	1	0.27	2.35	2	2.99	1.95	21	5.73	1.28	7	2.39
1com	2.63	19	5.15	0.72	1	0.72	0.69	25	4.35	0.83	9	4.45
1coy	0.28	4	0.65	1.04	1	1.04	0.21	28	0.7	0.66	1	0.66
1cps	0.61	8	0.64	1.32	24	1.65	0.63	1	0.63	0.73	12	0.8
1dbb	0.86	1	0.86	0.7	16	0.78	0.51	7	0.59	0.53	1	0.53
1dbj	1.54	19	3.22	0.78	6	1.17	0.38	3	0.78	0.43	6	5.57
1did	0.81	14	4.27	1.05	5	4.51	2.08	29	4.96	0.71	3	4.49
1die	3.48	16	5.47	2.67	30	4.48	0.29	19	3.38	2.8	6	4.61
1drl	2.74	12	6.27	0.78	30	5.78	0.37	4	1.58	0.45	3	1.35
1dwd	4.81	14	4.82	0.89	6	6.13	3.46	8	4.25	4.55	1	4.55
1eap	4.72	1	4.72	4.15	17	5.33	3.85	2	8.1	4.46	1	4.46
1eed	8.24	20	11.89	1.36	30	1.42	4.95	21	12.13	1.31	1	1.31
1ebp	1.62	2	1.7	11.79	13	11.8	1.21	25	2.15	1.55	1	1.55
1etr	4.92	2	5.11	2.66	21	2.93	2.31	29	5.72	4.29	1	4.29
1fkg	1.6	24	1.91	4.79	11	6.07	3.27	7	7.88	4.68	1	4.68
1fki	0.28	1	0.28	0.39	1	0.39	0.34	1	0.59	0.37	1	0.37
1frp	0.7	3	0.84	0.64	8	8.58	0.4	2	0.63	0.78	2	5.88
1ghb	2.19	9	3.89	11.82	30	11.14	0.72	16	5.76	1.96	1	1.96
1glp	3.85	1	3.85	6.43	9	7.11	1	5	5.69	3.64	1	3.64
1glq	6.29	11	10.4	6.22	19	6.29	3.28	18	4.66	6.28	1	6.28
1hdc	2.84	2	2.85	9.49	19	13.58	3.26	24	9.59	9.66	1	9.66
1hfc	10.72	12	10.73	1.38	13	2.42	1.66	6	2.75	1.23	3	2.5
1hri	9.52	16	9.65	9.35	29	10.22	1.3	5	1.44	10.18	1	10.18
1hsl	1.55	9	2.09	0.51	1	0.51	0.58	21	1.53	0.76	2	1.69
1hyt	3.15	12	5.08	0.77	11	3.62	0.64	2	0.94	0.91	1	0.91
1icn	3.07	10	9.73	18.57	26	19.79	1.65	2	2.65	3.12	1	3.12
1igj	1.91	10	2.45	4.09	27	8.15	1.44	5	4.15	2.47	1	2.47
1imb	2.44	23	5.56	0.73	10	5.23	0.83	7	3.65	0.92	1	0.92
1ive	2.59	11	2.69	3.08	9	5.54	1.04	27	1.22	2.36	1	2.36
1lah	0.46	1	0.46	0.34	3	0.37	0.29	2	0.38	0.51	1	0.51
1lcp	2.19	19	2.54	1.35	1	1.35	0.39	9	1.01	1.1	1	1.1
1ldm	2.43	6	2.55	0.52	20	0.56	0.76	16	0.8	0.7	1	0.7
1lic	2.38	16	3.31	5.21	3	5.28	2.5	2	4.64	3.32	3	5.44
1lmo	ns	—	ns	6.19	22	6.44	5.59	24	6.81	6.73	2	7.01
1lna	4.53	20	9.08	0.85	14	1.42	1.47	15	1.54	1.02	6	1.33
1lst	0.42	3	0.46	0.68	4	0.89	0.21	6	0.82	0.69	1	0.69
1mcr	2.64	27	4.47	7.78	13	11.99	2.93	15	5.69	4.62	1	4.62
1mdr	3.12	6	3.91	0.63	14	0.91	0.35	14	0.37	0.51	1	0.51
1mmq	6.89	15	7.73	0.65	2	0.93	0.5	1	0.5	0.62	1	0.62
1mrg	0.74	2	5.09	0.63	1	0.63	0.16	19	0.55	0.54	2	4.82
1mrk	0.94	7	1.24	2.56	1	2.56	0.53	7	0.96	0.56	1	0.56
1mup	1.46	2	1.71	3.11	7	3.47	1.31	28	4.51	1.3	11	4.54
1nco	0.87	3	0.88	9.63	11	11.48	4.02	1	4.02	0.85	1	0.85
1pbd	0.44	1	0.44	0.44	14	1.61	0.34	5	1.33	0.73	1	0.73
1poc	7.33	5	7.37	3.58	28	16.06	1.68	2	2	7.14	1	7.14
1rne	13.71	27	15.48	18.72	7	18.86	3.27	10	5.62	5.73	1	5.73
1rob	0.67	22	1.04	7.33	15	8.55	0.71	21	1.17	0.91	1	0.91

APPENDIX (Continued)

	Dock			Flex X			Gold			ConsDock		
PDB code	RMSD1[a]	rank[b]	RMSD2[c]	RMSD1	rank	RMSD2	RMSD1	rank	RMSD2	RMSD1	rank	RMSD2
1snc	1.28	23	2.6	4.59	20	8.92	0.87	6	2.35	1.37	1	1.37
1srj	1.32	1	1.32	2.48	8	6.82	0..64	4	1.02	1.05	6	6.88
1stp	0.62	1	0.62	0.52	12	0.98	0.48	15	0.51	0.51	1	0.51
1tdb	1.88	2	1.89	1.6	16	10.33	0.95	6	8.5	1.19	1	1.19
1tka	2.94	24	6.01	1.34	3	1.59	1.34	3	2.03	1.06	12	1.4
1tng	0.29	1	0.29	0.5	10	1.95	0.34	25	1.87	0.46	2	1.64
1tnl	1.42	1	1.42	0.55	2	1.07	0.57	22	0.92	0.7	1	0.7
1tph	0.61	12	1.5	0.43	16	0.67	0.37	15	0.62	0.87	3	1.11
1tpp	3.28	14	3.13	1.14	25	1.45	1.33	2	1.45	0.94	6	0.97
1ukz	2.49	20	12.01	0.38	4	0.65	1.1	20	2.78	0.77	2	1.19
1ulb	0.36	7	0.43	0.47	10	1.05	0.23	15	0.68	0.32	1	0.32
1wap	0.25	1	0.25	0.37	1	0.37	0.31	22	0.45	0.5	1	0.5
1xid	2.11	17	2.27	3.94	13	4.31	0.77	9	4.31	4.08	3	4.22
1xie	2.48	12	3.82	0.93	17	5.46	0.44	2	0.5	0.53	4	5.18
2ada	0.42	1	0.42	0.45	6	0.77	0.53	16	0.8	0.39	1	0.39
2ak3	1.71	24	3.15	0.83	3	1.92	0.46	2	3.64	0.52	4	1
2cgr	1.47	9	2.2	1.03	11	1.13	0.39	1	0.39	0.79	1	0.79
2cht	0.9	5	2.03	0.53	1	0.53	0.35	14	0.75	0.56	3	1.1
2cmd	1.76	1	1.76	0.57	28	3.71	0.48	8	1.39	0.97	13	3.57
2ctc	0.49	6	0.77	0.63	26	1.8	0.53	4	1.25	0.85	3	1.4
2dbl	3.23	3	4.6	1.31	1	1.31	0.9	3	1.16	0.95	1	0.95
2gbp	0.68	9	0.74	0.35	27	0.85	0.36	8	0.47	0.54	1	0.54
2lgs	4.18	6	5.28	1.21	26	1.35	0.77	18	1.36	1.06	1	1.06
2phh	0.68	8	3.5	0.47	1	0.47	2.61	11	4.32	0.71	2	2.38
2plv	10.49	26	11.55	7.81	30	8.33	2.35	21	7.38	9.07	1	9.07
2r07	1.42	9	11.63	1.63	19	11.49	1.26	6	3.87	1.24	19	3.63
2sim	3.98	9	4.23	0.87	3	1.6	0.68	9	1.3	0.61	2	0.96
4aah	0.28	2	0.29	0.45	2	0.55	0.27	16	0.46	0.61	1	0.61
3cpa	2.78	20	7.56	1.75	25	3.11	0.85	1	0.85	1.37	11	2.26
3hvt	0.59	1	0.59	9.11	30	10.61	0.36	19	1.09	0.76	1	0.76
3ptb	0.22	6	0.37	0.55	3	0.8	0.61	17	0.87	0.84	1	0.84
3tpi	0.74	1	0.74	0.67	19	1.09	0.41	1	0.41	0.69	1	0.69
4cts	1.46	10	3.73	0.35	27	0.5	0.56	15	3.48	0.89	3	1.04
4dfr	4.18	9	5.37	0.47	4	1.03	1.13	1	1.13	0.65	1	0.65
4fab	0.79	9	1.43	2.48	16	6.34	0.74	2	4.8	0.87	11	4.42
4phv	0.39	9	0.42	0.24	3	0.36	0.14	7	0.31	10.01	3	10.7
6abp	0.39	3	0.42	0.24	3	0.36	0.14	7	0.31	0.22	1	0.22
7tim	1.84	17	6.06	0.6	1	0.6	0.48	4	0.48	0.9	2	0.9
8atc	3.6	20	5.31	0.6	16	0.78	0.39	2	0.5	0.75	1	0.75
8gch	2.74	18	4.86	5.65	19	5.82	3.46	15	4.68	3.5	2	4.55

[a]RMSD in Å of the best pose (closest to the X-ray structure).
[b]Rank of the best pose.
[c]RMSD in Å of the top-ranked pose from the X-ray structure.
[d]No solutions.