PROTEINS: Structure, Function, and Bioinformatics 66:399–421 (2007)

Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking

Sheng-You Huang and Xiaoqin Zou*
Dalton Cardiovascular Research Center and Department of Biochemistry, University of Missouri, Columbia, Missouri 65211

ABSTRACT **One approach to incorporate protein flexibility in molecular docking is the use of an ensemble consisting of multiple protein structures. Sequentially docking each ligand into a large number of protein structures is computationally too expensive to allow large-scale database screening. It is challenging to achieve a good balance between docking accuracy and computational efficiency. In this work, we have developed a fast, novel docking algorithm utilizing multiple protein structures, referred to as ensemble docking, to account for protein structural variations. The algorithm can simultaneously dock a ligand into an ensemble of protein structures and automatically select an optimal protein structure that best fits the ligand by optimizing both ligand coordinates and the conformational variable *m*, where *m* represents the *m*-th structure in the protein ensemble. The docking algorithm was validated on 10 protein ensembles containing 105 crystal structures and 87 ligands in terms of binding mode and energy score predictions. A success rate of 93% was obtained with the criterion of root-mean-square deviation <2.5 Å if the top five orientations for each ligand were considered, comparable to that of sequential docking in which scores for individual docking are merged into one list by re-ranking, and significantly better than that of single rigid-receptor docking (75% on average). Similar trends were also observed in binding score predictions and enrichment tests of virtual database screening. The ensemble docking algorithm is computationally efficient, with a computational time comparable to that for docking a ligand into a single protein structure. In contrast, the computational time for the sequential docking method increases linearly with the number of protein structures in the ensemble. The algorithm was further evaluated using a more realistic ensemble in which the corresponding bound protein structures of inhibitors were excluded. The results show that ensemble docking successfully predicts the binding modes of the inhibitors, and discriminates the inhibitors from a set of noninhibitors with similar chemical properties. Although multiple experimental structures were used in the present work, our algorithm can be easily applied to multiple protein conformations generated by computational methods, and helps improve the efficiency of other existing multiple protein structure(MPS)-based methods to accommodate protein flexibility. Proteins 2007;66:399–421.** © 2006 Wiley-Liss, Inc.

Key words: molecular docking; ligand-protein interactions; protein flexibility; structure-based drug design; protein ensemble

INTRODUCTION

Molecular docking is an important tool for drug discovery. Given a target protein structure molecular docking samples hundreds or thousands of orientations or conformations of a ligand at the putative binding site, evaluates the energy score of each orientation/conformation, and ranks the orientations and conformations according to their scores.[1,2] The ligand pose with the lowest score is predicted as the binding mode. With more and more crystal structures solved for proteins, molecular docking is increasingly used for database screening and combinatorial library design for lead discovery and optimization.

Two major hurdles of molecular docking are scoring function and protein flexibility. The latter is especially challenging because of the large degrees of freedom of a protein. Therefore, although a variety of docking algorithms have been developed to account for ligand flexibility (see Refs. 1 and 2 for review), fewer efforts have been devoted to protein flexibility studies until recently (see Refs. 3–5 for review).

Ligand-induced receptor conformational changes are common in ligand binding, ranging from local rearrangements of side chains to large domain movements. In some cases, even a small change in the receptor conformation would have a remarkable effect on the ligand

The Supplemetary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: NIH; Grant number: DK61529; Grant sponsor: AHA; Grant number: 0265293Z.

*Correspondence to: Xiaoqin Zou, Dalton Cardiovascular Research Centre and Department of Biochemistry, University of Missouri, Columbia, MO 65211. E-mail: zoux@missouri.edu

Received 8 March 2006; Revised 1 August 2006; Accepted 12 August 2006

Published online 9 November 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21214

binding affinity, leading to failure of molecular docking in mode/affinity predictions if this conformational change is not incorporated in docking calculations.[6–8] The ideal method to incorporate protein flexibility would be to search and optimize the full degrees of freedom of the protein simultaneously to those of the ligand, using methods such as Monte Carlo[9] and molecular dynamics (MD)[10] simulations. Unfortunately, such a method is computationally expensive and thus impractical for large-scale docking experiments.[2] Thus, simplified models have been presented to incorporate limited protein motions while keeping computational time practical.

The earliest solution to consider some degree of receptor flexibility is the soft docking method, which implicitly allows for a small degree of overlap between the ligand and the protein by softening interatomic van der Waals repulsions.[11] The first work of explicitly including protein flexibility was achieved by Leach, using a combination of discrete side-chain states.[12] Since then, several improved techniques have been proposed to incorporate side-chain flexibility, including continuous side-chain sampling[13] and rotameric libraries.[14–16] The hinge-bending approach, allowing relative motions of rigid domains, subdomains, or groups of atoms around manually defined hinges in either the ligands or the receptor, mimics the induced-fit process like pliers closing on a screw.[17,18]

Recently, numerous studies have used multiple static receptor structures to represent a flexible receptor in drug design applications (see Refs. 3 and 4 for review). There are two critical challenges in the use of multiple protein structures for docking.[4,19] First, how to obtain reliable predictions of the conformational changes? Second, how to combine/sample these conformations in molecular docking?

One source of multiple protein structures is the experimental structures obtained from X-ray or NMR studies. The pioneer work by Knegtel et al. used a composite grid of interaction energies to incorporate an ensemble of crystal and NMR structures of protein–ligand complexes to account for protein flexibility.[20] The composite grid was constructed by combining the energy grids generated from every protein structure[21] with a weighting scheme, and was then used for standard ligand docking. Osterberg et al. analyzed the composite-grid algorithm in depth with 21 complexes of peptidomimetic inhibitors bound to HIV-1 proteases.[22]

Claussen et al. have developed a new molecular docking tool FlexE to combine multiple protein structures to represent a flexible binding site.[23] In their method, the structures of the ensemble were superimposed. The similar parts were merged, whereas dissimilar parts were treated as separate alternatives. The alternative conformers were combinatorially joined to create new valid protein structures (referred to as "united protein representation") for ligand docking. The test results on 10 ensembles showed that FlexE yields comparable success rates to the merged ranking list of sequential docking with FlexX in terms of binding mode predictions. The average run time for docking a ligand into an ensemble of protein structures with FlexE is a little more than half (0.57-fold) of that for sequentially docking the ligand into all the structures with FlexX.

Shoichet and co-workers have presented a similar but much faster scheme.[19,24] By assuming that the flexible regions of the receptor move independently, only one local protein conformer that was predicted to best fit the initial placement of the ligand was kept for each dissimilar part (i.e., flexible regions). Each selected local conformer was joined with the similar parts (i.e. rigid regions) to form a single, composite representation of the protein for ligand docking. Thus, the algorithm scales linearly rather than exponentially with the degrees of freedom of the receptor. An energy correction term was added to account for receptor conformational changes. Several novel, potent inhibitors were identified for the L99A/M102Q lysozyme mutant and aldose reductase by their algorithm.

An alternative to using multiple protein structures obtained from X-ray or NMR studies is to generate an ensemble of structures through computational methods. Broughton used different conformation snapshots from a short MD simulation of dihydrofolate reductase to construct a weight-averaged composite grid.[25] Test of this method on virtual screening showed improved enrichment in the top 10% of the ranked database. Carlson and coworkers used the structures from MD simulations to develop a receptor-based pharmacophore model for HIV-1 integrase and HIV-1 protease.[26,27] On the basis of the conformational ensemble from MD simulations. Lin et al. presented a relaxed complex method to accommodating receptor flexibility in the search for correct ligand–receptor conformations.[28,29] Kuhn and co-workers also developed a graph-theoretic algorithm to model protein main-chain flexibility.[30] Cavasotto and Abagyan presented the ICM-flexible receptor docking algorithm (IFREDA) to account for protein flexibility in virtual screening.[31] IFREDA was used to generate an ensemble of receptor conformations by docking selected known flexible inhibitors into the flexible receptor through global energy optimization using the biased probability Monte Carlo minimization procedure. Very recently, Cavasotto et al. also introduced a new algorithm based on normal-mode analysis to generate multiple receptor conformations for ligand docking.[32]

Nowadays, the exponentially-expanding protein data bank (PDB)[33] provides a good source of multiple protein conformations. Advances in computational methods and rapid increases in computer power will also allow generation of reliable, well-sampled protein structural ensembles in future. Then, a challenging question of utilizing multiple protein structures is how to efficiently combine the information of the given large amount of protein structures.[3,4] This question is especially important for virtual screening against a large compound database.

In this paper, we developed a fast ensemble docking algorithm to use multiple protein structures to represent a flexible receptor in molecular docking. The algorithm

automatically docks a ligand into the appropriate structure of the protein conformational ensemble, with a comparable computational speed to that for docking a ligand into a single protein structure. The accuracy of our algorithm was tested in terms of binding mode/score predictions and virtual database screening on diverse protein ensembles. The results are comparable to those obtained from sequentially docking the ligand into all the structures of the ensemble and merging the predicted ligand binding modes into one list ligand by re-ranking according to their scores. An application of the algorithm to HIV-1 protease was also presented.

MATERIALS AND METHODS

The Ensemble Docking Algorithm

The key idea of our flexible-receptor docking algorithm is to allow the scoring/optimization procedure of the docking process to automatically select an optimal protein structure from an ensemble to fit the ligand. A cartoon illustration of the ensemble docking algorithm is shown in Figure 1.

Before we explain our ensemble algorithm, we first briefly describe the optimization algorithm for standard rigid-receptor docking. To focus on protein flexibility, we treat the ligand as a rigid body to avoid damping the induced-fit effect, though we use flexible ligands for our virtual screening study and the sequent application. Our studies with flexible ligands described in the Supplementary Material also show that treating the ligands as rigid bodies do not change the results qualitatively.

When a rigid ligand is docked into a single, rigid protein structure, it is placed in the binding site of the protein by a matching/orienting procedure. The tightness of binding is evaluated by a scoring procedure

$$E = E(x, y, z, \theta, \phi, \psi) \quad (1)$$

where x, y, and z represent the coordinates of the center of mass of the ligand, and θ, ϕ, and ψ denote the three Euler angles, respectively. These six parameters, representing six degrees of freedom (three translational and three rotational), define an orientation of the ligand. The optimization procedure of docking is to continuously adjust (x, y, z, θ, ϕ, ψ) until the calculated score converges to a minimum; the corresponding orientation is predicted as the best binding mode. Therefore, if a ligand is docked into each conformation of a protein ensemble consisting of M structures with this standard docking approach, the computational time will be M times the CPU for docking against a single protein structure.

To improve the computational efficiency, our docking algorithm, which uses multiple protein structures to represent a flexible receptor and is referred to as the ensemble docking algorithm, adds the protein conformational state as an additional dimension for parameter optimization. Specifically, our ensemble algorithm reconsiders the scoring procedure as

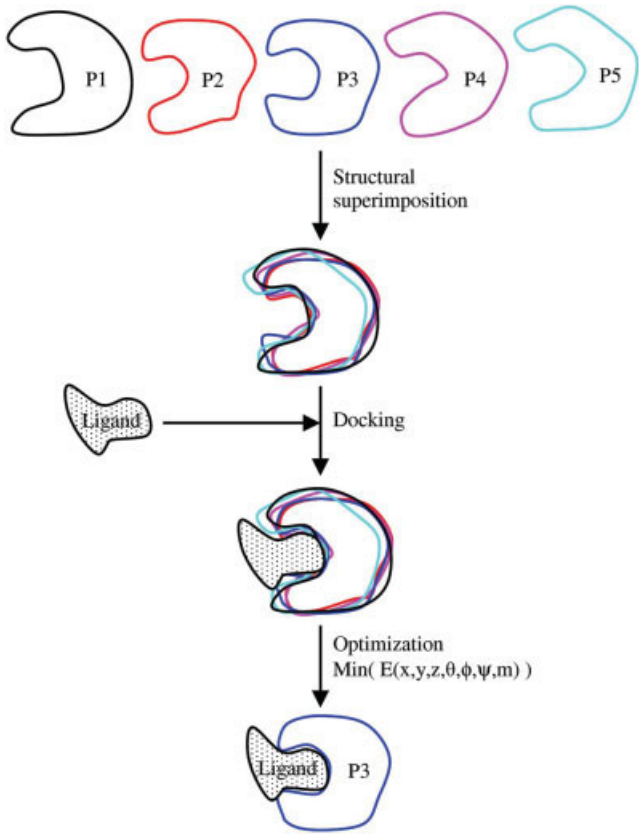


Fig. 1. A cartoon illustration of the ensemble docking algorithm. The ensemble consists of five protein structures denoted by P1, P2, P3, P4, and P5. The optimization procedure is to minimize the binding score $E(x, y, z, \theta, \phi, \psi, m)$ of ligand, where x, y, z, θ, ϕ, and ψ represent six degrees of freedom of the ligand, and m stands for the protein conformational state ranging from P1 to P5. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$$E = E(x, y, z, \theta, \phi, \psi, m) \quad (2)$$

where m stands for the m-th protein structure of the ensemble, an integer ranging from 1 to M. Again, M is the total number of conformations in an ensemble for the target protein.

By adjusting the parameter set (x, y, z, θ, ϕ, ψ, m), our algorithm can simultaneously dock a ligand into a protein conformational ensemble, and allow the optimization program to automatically select an optimal protein conformation guided by calculated scores (Fig. 1).

In other words, our ensemble algorithm performs docking against multiple target structures like docking against a single target structure; the step to select a specific protein conformation is incorporated into the energy optimization of the ligand. The two optimization methods [Eq. (1) vs. Eq. (2)] are expected to take similar run time with the current docking algorithm (described in "Docking algorithm"), as explained below.

Generally speaking, whether adding m in the energy optimization equation can effectively shorten the compu-

tational time depends on the optimization method used in the docking algorithm. If the docking algorithm uses a complete global optimization approach for energy optimization, the run time for optimizing Eq. (2) is expected to be M times that for optimizing Eq. (1), because the protein conformational variable m is orthogonal to the six degrees of freedom in space. In other words, with a global optimization method that exhaustively searches in the complete parameter space, our ensemble docking algorithm would not increase the computational efficiency significantly. However, because *complete* global optimization methods may not be necessary for molecular docking, docking algorithms usually use local optimization and/or incomplete global optimization (e.g., genetic algorithm) approaches instead.[1] For example, the DOCK 4.0 algorithm,[34] which is used in the present work (see detail in "Docking algorithm"), uses the SIMPLEX local optimization method[35] to optimize ligand orientations/conformations and has been shown to be successful.[34] Our analysis with the ligand–protein complexes in study shows that the number of iteration steps of SIMPLEX is roughly proportional to the number of independent parameters involved in the energy function (data not shown). In other words, the run time of SIMPLEX is roughly proportional to the square of the number of independent parameters.[35] Introducing the variable m to Eq. (1) increases the number of independent (or orthogonal) parameters from 6 to 7. Thus, the fold of increase in computational time on the optimization procedure with SIMPLEX is expected to be roughly $(7/6)^2 = 1.36$, which is insignificant. This point of view is confirmed by our docking calculations (see "computational efficiency of the ensemble docking algorithm" for detail).

Scoring Function

The energy scoring function used in the present study is ITScore, an iterative knowledge-based scoring function recently developed by our group.[36,37] Only intermolecular interactions are included in ITScore. The score is obtained by summing up all protein(P)–ligand(L) atom pair interactions as

$$E = \sum_{\text{P–L atom pair}} u_{ij}(r) \tag{3}$$

where $u_{ij}(r)$ is the pair potential between the protein atom of type i and the ligand atom of type j at the atom pair distance r. A total of 26 atom types were used, based on the definitions provided by the SYBYL software (Tripos).

The pair potentials were derived via an efficient iterative method using 786 protein–ligand complexes extracted from the PDB.[33] Specifically, the initial potential $u_{ij}^{(0)}(r)$ for iteration was taken from a combination of the Lennard-Jones van der Waals potential and a standard potential of mean force. The latter was converted from the atom pair occurrences observed from the known protein–ligand complexes. The main idea of ITScore is to adjust pair potentials by iteration until they correctly discriminate native binding modes from decoy structures.

ITScore was extensively evaluated with about 200 other diverse protein–ligand complexes on native structure identification and binding affinity predictions. It was further assessed with enrichment tests of virtual database screening against four target proteins. The results have shown that ITScore is quite general, accurate, and fast. Details of the theory of ITScore and test studies are given in Refs. 36 and 37.

Structure Superimposition

As previously mentioned, the key idea of our ensemble docking algorithm is to select an optimal conformation out of an ensemble of protein structures through automated optimization via Eq. (2) by searching the parameter space of (x, y, z, θ, ϕ, ψ, m). Accordingly, the space for ligand docking increases from a single protein binding site to the sum of all the binding sites of the protein ensemble, which would significantly decrease the optimization accuracy/efficiency. Considering that the binding sites of different protein structures in an ensemble are often similar, to reduce the ligand sampling space and to make the searching algorithm more efficient, we superimposed the protein structures of an ensemble according to the binding site; the size of the (x, y, z, θ, ϕ, ψ) space for parameter searching for ensemble docking was then reduced to a size similar to the space for single docking. Specifically, we used Ferro and Hermans's algorithm[38] to match two different structures; four backbone carbon atoms in the binding site were selected for matching use. The superimposed protein structures were used for sequent docking evaluations (including scoring and optimization).

Construction of the Reference Protein

Usually, a docking program such as DOCK4.0 requires a protein structure to guide the generation of initial ligand orientations around the binding site, which will be used for later scoring and optimization. However, our ensemble docking algorithm works on multiple receptor structures, and it is neither possible nor reasonable to determine which structure should be used as a representative of the ensemble. Therefore, a reference protein structure should be constructed for generation of initial ligand orientations.

The principle for constructing the reference protein is that the binding pocket should be as large as possible to provide sufficient sampling space for ligands with the least change of the overall shape of the binding site. The procedure is as follows.

First, sphere points were generated for each protein structure of the ensemble by the SPHGEN program in the DOCK software.[39] The sphere points near the binding site of each protein conformation were collected as the reference sphere points. Then, the distances between

the reference sphere points and each residue of every protein structure were calculated, where the distance between a residue and the reference sphere points is defined as the minimum of the distances between every atom in the residue and any of the sphere points. The conformation of the residue with the maximal distance was chosen as a representation of the corresponding residue of the ensemble. The selected conformations of all residues were then combined together as the reference protein. The resulting reference protein has the same sequence as that of the original protein, but each side chain adopts the conformation of the corresponding side chain of one of the protein structures that provides the largest space to the binding site. Lastly, we checked possible clashes between the constructed reference protein and the reference sphere points. If a reference protein atom is within 2.0 Å of any reference sphere point, the atom is removed from the reference protein.

It should be noted that the constructed reference protein is artificial. There may exist some atomic clashes between protein residues. Some of the residues may be incomplete because of atom deletions as mentioned earlier. However, none of these issues matters, because the reference protein is solely used to guide the generation of ligand orientations. The real protein structures rather than the reference protein structure will be used for score calculations and orientation optimization.

Docking Algorithm

We wrote a docking program using the ligand matching method of DOCK4.0,[34] the scoring function ITScore,[36,37] and the ensemble docking algorithm to account for protein structural variations. A clustering filter was added to remove degeneracy of the final orientations. For flexible ligand docking, ligand conformations were generated by the OMEGA software (OpenEye);[40] each ligand conformation was docked into the receptor as a rigid body. The program allows not only single-structure docking but also ensemble docking. The details are described as follows.

First, a set of sphere points representing the negative image of each protein surface were generated for ligand orientation matching. Specifically, for each protein structure (or the reference protein structure), a set of sphere points were generated for the protein surface using the SPHGEN program.[39] The sphere points within 3–5 Å of all the superimposed ligands in the ensemble (depending on ligand sizes) were clustered, and near 50 sphere points were finally selected for each protein structure (or the reference protein structure).

Next, the ligand orientations were generated by using the exhaustive matching algorithm by Ewing and Kuntz[41] where single docking and ensemble docking are based on each single receptor structure and the reference protein structure, respectively. The tolerance distance between ligand atoms and sphere points was set to 0.5 Å during matching. The ligand orientation that matches the most sphere points was first selected for further scoring evaluation. The maximum number of evaluated orientations was set to 500, and the maximum minimization step was set to 200. Equations (1) and (2) were used to calculate ligand energy scores for single docking and ensemble docking, respectively. The grid-based energy calculation algorithm of DOCK[21] was used to improve the computational speed. The grid spacing was set to 0.3 Å, and trilinear interpolation was used for calculating energy scores. The binding energy of each ligand orientation was minimized by the SIMPLEX method.[35] Different ligand orientations were ranked from low to high according to their energy scores. Finally, the orientations were clustered; for two orientations with root-mean-square deviation (rmsd) less than 1.0 Å, only the one with lower score was kept.

Test Data Sets

Ten protein ensembles containing 105 crystal structures from the PDB[33] were used to test our flexible-receptor docking algorithm (Table I). The protein ensembles include four protein kinase subfamilies obtained from the work by Cavasotto and Abagyan[31] (cAPK, CDK2, LCK, and P38), four protein ensembles taken from the work by Claussen et al.[23] (αMMC, DHFR, ricin, and trypsin), and two HIV-1 protease (HIVp) ensembles (HIVpa and HIVpb). The difference between HIVpa and HIVpb is that every conformation in the HIVpb ensemble contains a structural water in the binding site (e.g. HOH301 of 7HVP), whereas the conformations in the HIVpa ensemble do not have any structural water or ion in the corresponding position.

All the structures of a protein ensemble were superimposed by matching four backbone carbon atoms in the binding sites. The bound inhibitors were then separated from the proteins, and a total of 87 ligands were obtained. Water molecules (except the structural water in the HIVpb ensemble) and metal ions were removed from the protein structures. Hydrogen atoms were not considered explicitly, because of the feature of the ITScore scoring function we used in this study.[36,37] Atom types for proteins and ligands were assigned by the SYBYL software (Tripos).

For studies on docking the known ligands into the protein ensembles (See Results section), the ligands were treated as rigid bodies so as to focus on the effect of protein flexibility. Otherwise, the flexibility of the ligands would weaken the induced-fit effect revealed from protein structural variations in the ensemble. We have also shown in the Supplementary Material that treating the ligands as flexible bodies does not change the results qualitatively. For virtual database screening (later section) and the HIV-1 protease application (later section), all ligands (active or inactive) were treated as flexible molecules because the binding modes of the inactive compounds are unknown and also because flexible ligands are commonly used for practical virtual screening and other real applications.

TABLE I. Ten Protein Ensembles Used for Validating Ensemble Docking

Ensemble	PDB code(ligand)	
CAMP-dependent protein kinase (cAPK) (8 pdb, 7 ligands)	1BKX(adn)	1BX6(ba1)
	1FMO(adn)	1STC(sto)
	1YDR(iqp)	1YDS(iqs)
	1YDT(iqb)	1JLU
Cyclin-dependent kinase 2 (CDK2) (13 pdb, 12 ligands)	1AQ1(stu)	1DI8(dtq)
	1DM2(hmd)	1E1X(nw1)
	1E9H(inr)	1FVT(106)
	1FVV(107)	1G5S(i17)
	1H1P(cmg)	1H1Q(2a6)
	1H1S(4sp)	1JSV(u55)
	1HCL	
Lymphocyte-specific kinase (LCK) (5 pdb, 4 ligands)	1QPC(anp)	1QPD(stu)
	1QPE(pp2)	1QPJ(stu)
	3LCK	
Mitogen-activated kinase P38 (7 pdb, 6 ligands)	1A9U(sb2)	1BL6(sb6)
	1BL7(sb4)	1DI9(msq)
	1BMK(sb5)	1M7Q(dqo)
	1P38	
Alpha-momorcharin (αMMC) (7 pdb, 4 ligands)	1AHA(ade)	1AHB(fmp)
	1MRG(adn)	1MRH(fmc)
	1MRI	1MOM
	1AHC	
Dihydrofolate reductase (DHFR) (12 pdb, 12 ligands)	1DHJ(mtx)	1DRA(mtx)
	1DRB(mtx)	1DYH(dzf)
	1DYI(fol)	1DYJ(ddf)
	1JOL(ffo)	1RA2(fol)
	1RA3(mtx)	2DRC(mtx)
	3DRC(mtx)	4DFR(mtx)
Ricin (9 pdb, 5 ligands)	1APG(apg)	1FMP(fmp)
	1IFS(ane)	1IFU(fmc)
	1OBT(amp)	1RTC
	1OBS	2AAI
	1IFT	
Trypsin (16 pdb, 9 ligands)	1PPC(napap)	1PPH(tapap)
	1TNG(amc)	1TNH(fba)
	1TNI(pbn)	1TNJ(pea)
	1TNK(pra)	1TNL(tpa)
	3PTB(ben)	1TPP(apa)[a]
	1MAX	1TLD
	1TPO	1TAW
	2PTC	3PTN
HIV-1 protease without water (HIVpa) (12 pdb, 12 ligands)	1AJV(nmb)	1AJX(ah1)
	1DMP(450)	1G2K(nm1)
	1G35(ahf)	1HVH(q82)
	1HVR(xk2)	1HWR(216)
	1PRO(a88)	1QBS(dmp)
	2UPJ(u02)	7UPJ(inu)
HIV-1 protease with one structural water (HIVpb) (16 pdb, 16 ligands)	1AAQ(psi)	1C70(l75)
	1HBV(gan)	1HIH(c20)
	1HPS(run)	1HPV(478)
	1HPX(kni)	1HTF(g26)
	1HXW(rit)	1ODY(lp1)
	1SBG(im1)	2BPV(1in)
	2BPY(3in)	
	5HVP(acetyl-pepstatin)	
	7HVP(jg365)	8HVP(u85548e)

[a]The ligand of 1TPP (apa) was excluded in this study because of the existence of severe clashes between the protein and the ligand. The protein structure of 1TPP was only used for cross-docking.

Database Preparation

For virtual screening test, a database consisting of 1000 small molecules was constructed by random selection from the compounds in the Available Chemical Directory (ACD, distributed by Molecular Design, San Leandro, CA). The selected compounds contain only 10–50 heavy atoms and consist of no atoms other than C, H, O, N, S, P, F, Cl, Br, and I. Among these molecules, 780 have no formal charge, 160 have a formal charge of 1, and 60 have a formal charge of 2. The database serves as a set of inactive compounds for virtual database screening. The coordinates of the molecules were generated by using the CONCORD program[42] provided in the SYBYL software.

RESULTS

Evaluating the Impact of Protein Flexibility on Single-Receptor Docking Algorithm

Before we presented our flexible-receptor docking results, the influence of ligand-induced conformational changes on ligand binding was evaluated by sequentially docking the ligands into all protein structures with the traditional rigid-receptor docking approach. It also served as a validation of our scoring function and as an assessment of our test sets.

In the rest of the text, we adopted the following commonly used terminologies for rigid-receptor docking: Native docking refers to re-dock a ligand into its co-crystalized protein structure. In contrast, cross docking refers to docking a ligand into each of the superimposed protein structures originally bound with other ligands in the ensemble. Apo-docking refers to docking a ligand into an apo protein structure (namely, the protein structure in free form). Comparisons of the docking results with these three methods allow us to assess how variations in protein structures affect the effectiveness of rigid-receptor docking.

The rmsd criterion

The docking accuracy was first evaluated by calculating the rmsd between the docked ligand and the ligand in the crystal complex structure, after the superposition of the backbone atoms in the binding site. Similar to Ref. 23, the ligand binding mode was considered to be successfully predicted if the rmsd <2.5 Å for any of the top five ranked orientations. The first of the top five orientations with rmsd <2.5 Å was kept for further evaluation. If all the top five orientations have rmsd ≥ 2.5 Å, the best-scored pose was kept for the ligand. This is the default criterion unless otherwise specified.

It should be noted that in the literature, the success of a docking program on binding mode prediction is usually measured by the rmsd between the experimentally-observed ligand orientation and the top-ranked ligand orientation with a threshold of 2.0 Å.[43] The increased threshold of 2.5 Å accounts for the slight influence of structural superimpositions.[23,31] That is, the cross-docking binding mode cannot be observed experimentally, and thus the predicted ligand orientation can only be compared with the ligand orientation extracted from its native complex structure that is superimposed with the corresponding non-native protein structure. This reference ligand orientation depends on the way of superimposition, and therefore would slightly affect the rmsd values. Second, the rmsd of the top-ranked orientation is often used to assess the performance of a docking tool because in practical virtual screening, it is time-consuming to inspect several poses of a large database of ligands. However, the correct, predicted pose of a ligand depends not only on the docking algorithm but also highly on the scoring function.[23] No current scoring functions can always rank the experimentally observed orientations to the first place.[44] Because the scoring function issue is beyond the scope of the present work, we considered the top five orientations instead of the top one for docking accuracy evaluations.[23] As shown in the Supplementary Material, a more strict criterion that defines the top orientation with an rmsd <2.0 Å as a success does not change the relative performances of different docking experiments presented as follows.

Table II gives a summary of the success rates of single rigid-receptor docking in terms of binding mode predictions against 10 protein ensembles. The fact that 100% of the ligands were correctly docked into their corresponding native protein structures (namely, native-docking) using the modified DOCK4.0 program[34] and the ITScore scoring function[36,37] (see Materials and Methods for detail) validates the docking algorithm and scoring function we used. When the ligands were docked into the protein structures complexed with other ligands in the same ensemble (cross docking), most of the success rates decrease, reflecting the impact of the ligand-induced conformational changes on ligand binding modes. The impact is especially prominent for P38, of which the cross-docking success rate is as low as 37%. The effect of induced-fit on docking is not obvious for DHFR and trypsin, of which the cross-docking success rates still reach 100 and 95%, respectively. Compared with cross docking, the binding mode predictions are generally much worse when the ligands were docked into ligand-free structures (apo-docking). The best apo-docking predictions are for LCK and αMMC, yielding a success rate of 75% each. The worst cases are P38 and ricin, where apo-docking totally misses the true binding modes, demonstrating the necessity of incorporating protein flexibility in ligand docking. DHFR, HIVpa, and HIVpb were excluded from the apo-docking tests because no free-form structure is available for DHFR with the same sequence and neither for dimeric HIV-1 proteases. The overall poorer performance of apo-docking than cross docking are in agreement with the findings in Ref. 45. Apo-crystal structures lack ligand-induced conformational changes, whereas different ligand-bound (holo) crystal structures may share some common induced structural variations, leading to the difference in docking performance. When considering all possible

TABLE II. Comparison of the Success Rates in Terms of Binding Mode Predictions Between Single Rigid-Receptor Docking (Including Native-Docking, Cross-Docking, Apo-Docking, or the Average of the Three), Sequential Docking (Merged Ranking List or SEQ.), and Ensemble Docking (ENS.) on 10 Protein Ensembles[a]

Ensemble	No. of ligands	Success rate in terms of binding mode prediction (%)					
		Native-docking	Cross-docking	Apo-docking	Average	SEQ.	ENS.
cAPK	7	100	76	71	79	100	100
CDK2	12	100	76	42	75	83	92
LCK	4	100	83	75	85	100	100
P38	6	100	37	0	40	33	67
αMMC	4	100	83	75	82	100	100
DHFR[b]	12	100	100	n/a	100	100	100
Ricin	5	100	55	0	36	80	80
Trypsin	9	100	95	65	84	89	89
HIVpa[b]	12	100	82	n/a	83	100	100
HIVpb[b]	16	100	80	n/a	81	100	100
Average	*8.7*	*100*	*77*	*47*	*75*	*89*	*93*

[a]The criterion of successful binding mode prediction is that any of the top five orientations for the ligand has an rmsd <2.5 Å.
[b]No apo-docking results for DHFR, HIVpa, and HIVpb because of the absence of the corresponding ligand-free conformations.

TABLE III. Comparison of the Success Rates in Terms of Both Binding Mode and Energy Score Predictions Between Single Rigid-Receptor Docking, Sequential Docking (SEQ.), and Ensemble Docking (ENS.) on 10 Protein Ensembles

Ensemble	No. of ligands	Success rate in terms of both pose and score predictions (%)					
		Native-docking	Cross-docking	Apo-docking	Average	SEQ.	ENS.
cAPK	7	100	43	57	52	100	100
CDK2	12	100	51	8	51	83	92
LCK	4	100	83	75	85	100	100
P38	6	100	23	0	31	33	67
αMMC	4	100	83	58	75	100	100
DHFR	12	100	87	n/a	88	100	100
Ricin	5	100	50	0	33	80	80
Trypsin	9	100	83	52	72	89	89
HIVpa	12	100	28	n/a	34	100	100
HIVpb	16	100	24	n/a	29	100	94
Average	*8.7*	*100*	*56*	*36*	*55*	*89*	*92*

single docking results, namely, combining native docking, cross docking, and apo-docking, the success rates of mode prediction are worse than that of native docking except for DHFR (see the "Average" column in Table II), ranging from 36 to 85% for different protein ensembles.

The scoring criterion

It has been shown by Cavasotto and Abagyan that scoring is more sensitive to the induced-fit effect than mode prediction.[31] Therefore, we further evaluated the performance of scoring predictions with different docking methods, which is especially important for virtual database screening. Ideally, one should use the experimental binding free energies as a reference.[22] Yet, not all experimental affinity data are available for the protein–ligand complexes that we used. Plus, the energies predicted by the ITScore scoring function need to be scaled (by a factor of ~12) for direct comparison with measured affinities, even though ITScore yields very good correlations in predicting binding energies.[36,37] Therefore, for each ligand, the energy score from native-docking was used as a reference to evaluate the energy scores obtained from cross docking. Analogous to the rmsd criterion (i.e., failure if rmsd >2.5 Å), a cross-docking score was defined to be "bad" if it was worse by a set threshold value than the corresponding native-docking score, representing a failure of cross docking from the scoring point of view. Otherwise, the cross-docking score was considered "good", and cross docking was successful. The energy threshold was arbitrarily set to 10 (roughly 1 kal/mol) according to our previous experience with the enrichment tests using ITScore.

Table III shows that the success rates significantly decrease with cross docking and apo-docking if both rmsd and scores rather than rmsd alone are considered (cf. Table II). For example, the cross-docking success rate decreases from 100 to 87% for DHFR, and from 82 to 28% for HIVpa. Another example is CDK2, for which the apo-docking success rate decreases from 42 to 8%. Thus, binding mode prediction alone seems to be insufficient to assess docking accuracy.

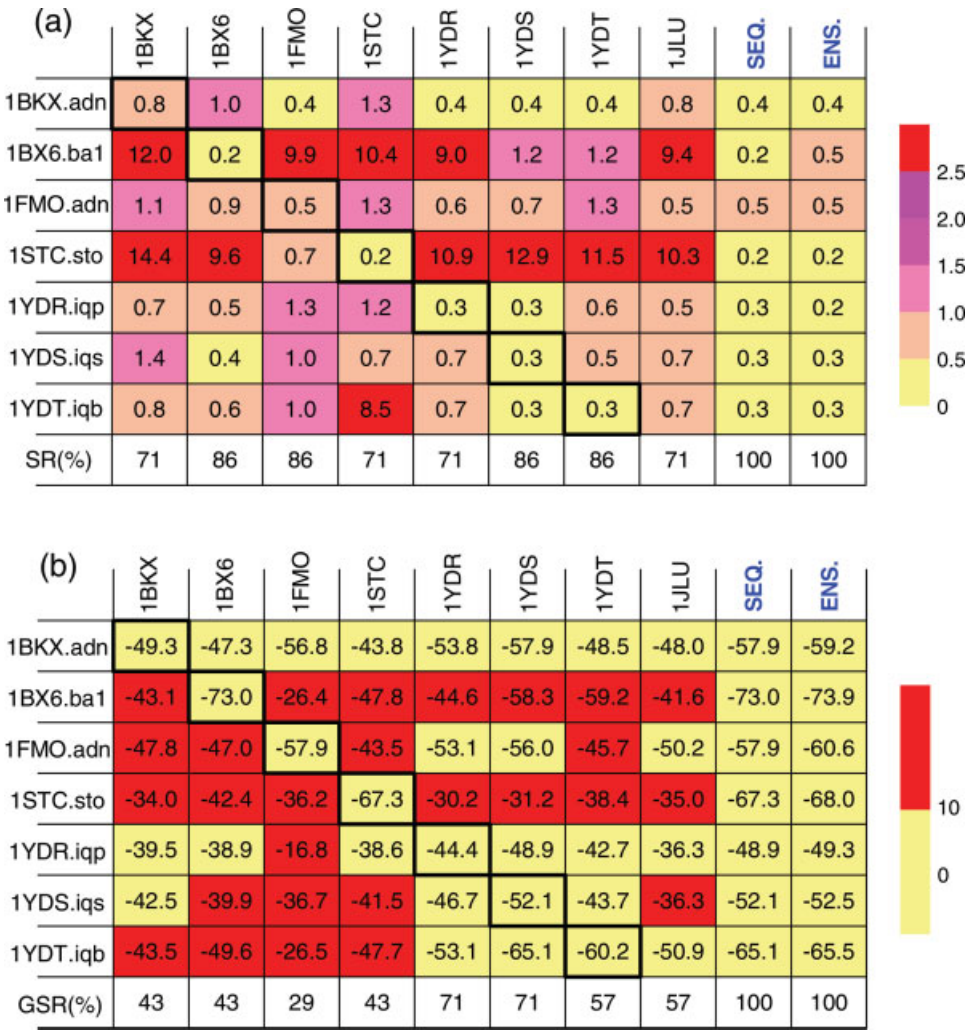


Fig. 2. Cross-docking on cAPK. The rows refer to different ligands, and the columns to different protein structures, sequential docking (merged ranking list or SEQ.) and ensemble docking (ENS.). The cells with darker border represent native-docking. (**a**) The color-scaled table of the rmsd values (Å). SR stands for the success rate of binding mode predictions. (**b**) The color-scaled table of the calculated binding scores. The color of the cell (*i*,*j*) is scaled by the energy difference $E(i, j)–E(i, i)$. Here, $E(i, j)$ represents the calculated energy score for the *i*-th ligand docked into the *j*-th protein structure, and $E(i, i)$ is the score for the *i*-th native complex. Cells with good scores are colored in yellow, and cells with bad scores in red. See the text for more detail. GSR, the rate of good scores. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Detailed analysis of the binding modes and scores were done for several typical protein ensembles, each containing 7–12 ligands. The results are shown in Figures 2–5. Taking Figure 2 as an example, the cells with darker border represent native-docking, the cells in the column of 1JLU (an unliganded crystal structure of cAPK) represent apo-docking, and the other cells (except the cells in the last two columns, which will be explained in the next section) represent cross docking.

The upper panels of Figures 2–5 summarize the rmsd results, and lower panels present the energy scores. The red cells stand for failure of docking in terms of the rmsd criterion or energy criterion. Notably, some ligands can be reasonably docked into all or most protein structures in the corresponding ensemble. For example, 1BKX.adn (adenosine), complexed with 1BKX, was docked into any structure of cAPK with a correct binding mode and good energy score. Further examinations show that in these cases most of the protein residues interacting with the ligand share similar conformations with the native protein structure bound by the same ligand. It is also seen from Figures 2–5 that there are more red cells in panel B than in panel A, again indicating that energy scoring is much more sensitive to protein structural variations than binding mode prediction, as explained later.

The phenomenon that some ligands can be docked into non-native structures with a good pose but a bad score, for example, docking the ligand 1YDS.iqs to 1FMO in Figure 2, usually occurs to those proteins with a binding site of a well-defined shape. The well-shaped

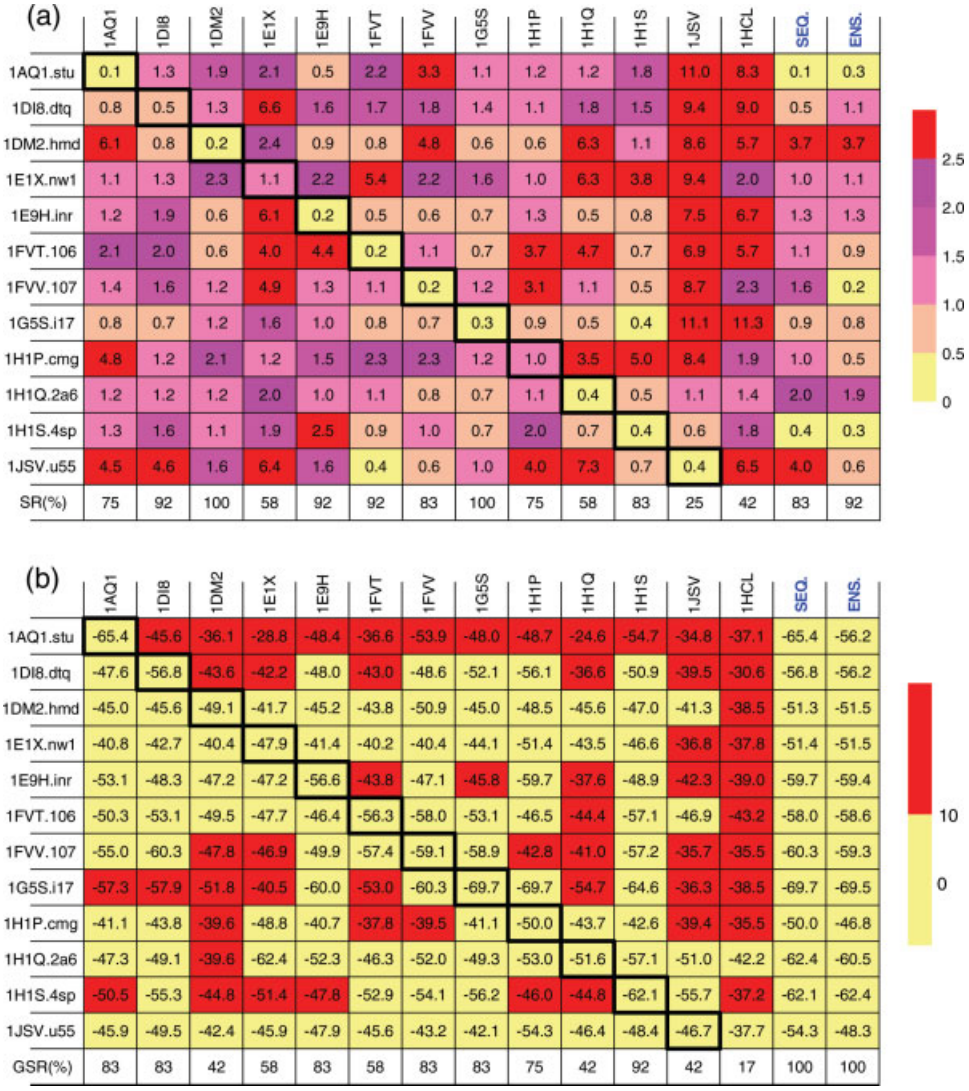


Fig. 3. Cross-docking on CDK2. The legend is the same as that of Figure 2. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

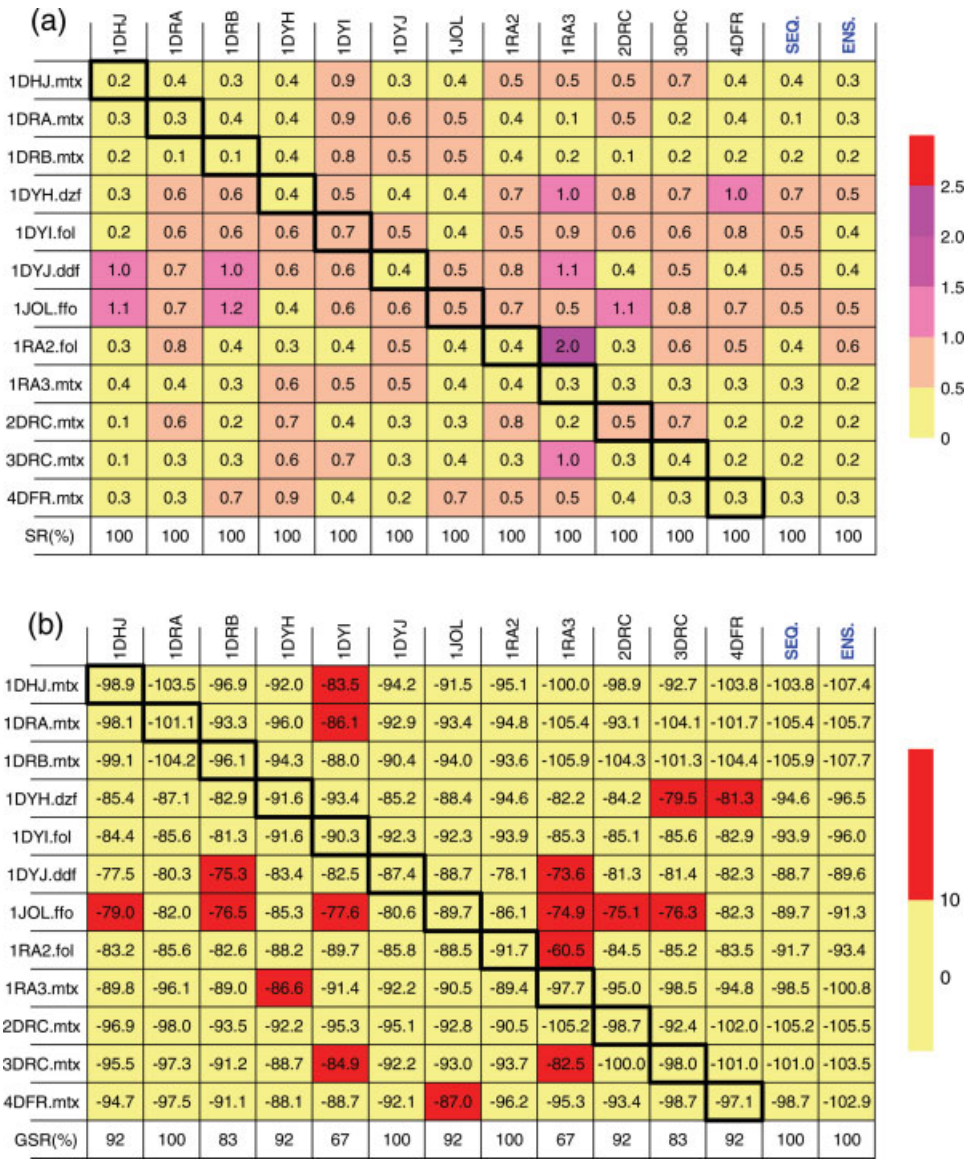
binding site can hold the ligand well even with a poor energy score, because docking the ligand elsewhere would be scored even worse. Therefore, both rmsd and scores need to be considered when cross docking proteins such as DHFR and HIV proteases (Figs. 4 and 5).

Ensemble Docking Algorithm to Consider Protein Structural Variations

The last section has demonstrated the necessity of incorporating protein flexibility in ligand docking. Here, we present a new docking algorithm, referred as ensemble docking, which simultaneously docks a ligand into an ensemble of multiple protein structures. The simultaneous, automated optimization procedure incorporates protein movements reflected in the ensemble structures into ligand docking at no significant increase in computational time.

We docked all the ligands into the corresponding protein ensembles with our ensemble docking algorithm. The results are summarized in the last columns of Tables II and III. The detailed rmsd values and corresponding ranks are listed in Table IV.

To evaluate the performance of ensemble docking, we also performed "sequential docking" for all the ligands of all the ensembles.[23] Specifically, each ligand was sequentially docked into all the structures of the corresponding protein ensemble, and all the predicted binding modes were merged into one list and re-ranked according to their scores (referred to as "merged ranking list"). The clustering analysis that was used in the docking algorithm (see Materials and Method) was also applied to remove similar orientations in the merged ranking list. Only the top five orientations were kept for each ligand (similar to the strategies described in "The rmsd criterion" section). Theoretically speaking, sequential docking considers all conformational



(a)

	1DHJ	1DRA	1DRB	1DYH	1DYI	1DYJ	1JOL	1RA2	1RA3	2DRC	3DRC	4DFR	SEQ.	ENS.
1DHJ.mtx	0.2	0.4	0.3	0.4	0.9	0.3	0.4	0.5	0.5	0.5	0.7	0.4	0.4	0.3
1DRA.mtx	0.3	0.3	0.4	0.4	0.9	0.6	0.5	0.4	0.1	0.5	0.2	0.4	0.1	0.3
1DRB.mtx	0.2	0.1	0.1	0.4	0.8	0.5	0.5	0.4	0.2	0.1	0.2	0.2	0.2	0.2
1DYH.dzf	0.3	0.6	0.6	0.4	0.5	0.4	0.4	0.7	1.0	0.8	0.7	1.0	0.7	0.5
1DYI.fol	0.2	0.6	0.6	0.6	0.7	0.5	0.4	0.5	0.9	0.6	0.6	0.8	0.5	0.4
1DYJ.ddf	1.0	0.7	1.0	0.6	0.6	0.4	0.5	0.8	1.1	0.4	0.5	0.4	0.5	0.4
1JOL.ffo	1.1	0.7	1.2	0.4	0.6	0.6	0.5	0.7	0.5	1.1	0.8	0.7	0.5	0.5
1RA2.fol	0.3	0.8	0.4	0.3	0.4	0.5	0.4	0.4	2.0	0.3	0.6	0.5	0.4	0.6
1RA3.mtx	0.4	0.4	0.3	0.6	0.5	0.5	0.4	0.4	0.3	0.3	0.3	0.3	0.3	0.2
2DRC.mtx	0.1	0.6	0.2	0.7	0.4	0.3	0.3	0.8	0.2	0.5	0.7	0.2	0.2	0.2
3DRC.mtx	0.1	0.3	0.3	0.6	0.7	0.3	0.4	0.3	1.0	0.3	0.4	0.2	0.2	0.2
4DFR.mtx	0.3	0.3	0.7	0.9	0.4	0.2	0.7	0.5	0.5	0.4	0.3	0.3	0.3	0.3
SR(%)	100	100	100	100	100	100	100	100	100	100	100	100	100	100

(b)

	1DHJ	1DRA	1DRB	1DYH	1DYI	1DYJ	1JOL	1RA2	1RA3	2DRC	3DRC	4DFR	SEQ.	ENS.
1DHJ.mtx	-98.9	-103.5	-96.9	-92.0	-83.5	-94.2	-91.5	-95.1	-100.0	-98.9	-92.7	-103.8	-103.8	-107.4
1DRA.mtx	-98.1	-101.1	-93.3	-96.0	-86.1	-92.9	-93.4	-94.8	-105.4	-93.1	-104.1	-101.7	-105.4	-105.7
1DRB.mtx	-99.1	-104.2	-96.1	-94.3	-88.0	-90.4	-94.0	-93.6	-105.9	-104.3	-101.3	-104.4	-105.9	-107.7
1DYH.dzf	-85.4	-87.1	-82.9	-91.6	-93.4	-85.2	-88.4	-94.6	-82.2	-84.2	-79.5	-81.3	-94.6	-96.5
1DYI.fol	-84.4	-85.6	-81.3	-91.6	-90.3	-92.3	-92.3	-93.9	-85.3	-85.1	-85.6	-82.9	-93.9	-96.0
1DYJ.ddf	-77.5	-80.3	-75.3	-83.4	-82.5	-87.4	-88.7	-78.1	-73.6	-81.3	-81.4	-82.3	-88.7	-89.6
1JOL.ffo	-79.0	-82.0	-76.5	-85.3	-77.6	-80.6	-89.7	-86.1	-74.9	-75.1	-76.3	-82.3	-89.7	-91.3
1RA2.fol	-83.2	-85.6	-82.6	-88.2	-89.7	-85.8	-88.5	-91.7	-60.5	-84.5	-85.2	-83.5	-91.7	-93.4
1RA3.mtx	-89.8	-96.1	-89.0	-86.6	-91.4	-92.2	-90.5	-89.4	-97.7	-95.0	-98.5	-94.8	-98.5	-100.8
2DRC.mtx	-96.9	-98.0	-93.5	-92.2	-95.3	-95.1	-92.8	-90.5	-105.2	-98.7	-92.4	-102.0	-105.2	-105.5
3DRC.mtx	-95.5	-97.3	-91.2	-88.7	-84.9	-92.2	-93.0	-93.7	-82.5	-100.0	-98.0	-101.0	-101.0	-103.5
4DFR.mtx	-94.7	-97.5	-91.1	-88.1	-88.7	-92.1	-87.0	-96.2	-95.3	-93.4	-98.7	-97.1	-98.7	-102.9
GSR(%)	92	100	83	92	67	100	92	100	67	92	83	92	100	100

Fig. 4. Cross-docking on DHFR. The legend is the same as that of Figure 2. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

variations in the given protein structures and thus can serve as a control to evaluate flexible-receptor docking algorithms using the same scoring function.[23] The results of sequential docking are presented in the second last columns of Tables II and III.

It is seen from Tables II and III that ensemble docking matches sequential docking in success rates for every protein ensemble except HIVpb, no matter which criterion is used. For HIVpb, the success rates are the same for the two docking methods in terms of mode predictions (100%), but different in terms of mode plus score predictions (94% for ensemble docking and 100% for sequential docking). For the other protein ensembles, both ensemble docking and sequential docking match native-docking, yielding successful predictions in binding modes and scores for all the ligands. On average, compared with single, rigid-receptor docking (the column of "Average"), the ensemble docking algorithm increases the success rate by 18% (from 75 to 93%) with the rmsd criterion, and by 37% (from 55 to 92%) with the rmsd + score criterion, respectively (Tables II and III).

It is notable that for CDK2 and P38, ensemble docking yields even higher success rates, which can be understood as follows. When a ligand is docked into a protein structure, there probably exist some wrong posed but well-scored orientations ("false positives") because of the influence of the scoring function. These false positives lower the rank of the correct pose because of their better scores. In the present work, the top five orientations were considered to calculate the success rate. In the case of ensemble docking, which samples 500 orientations, the number of false positives might happen to be limited and therefore the correct pose might still be

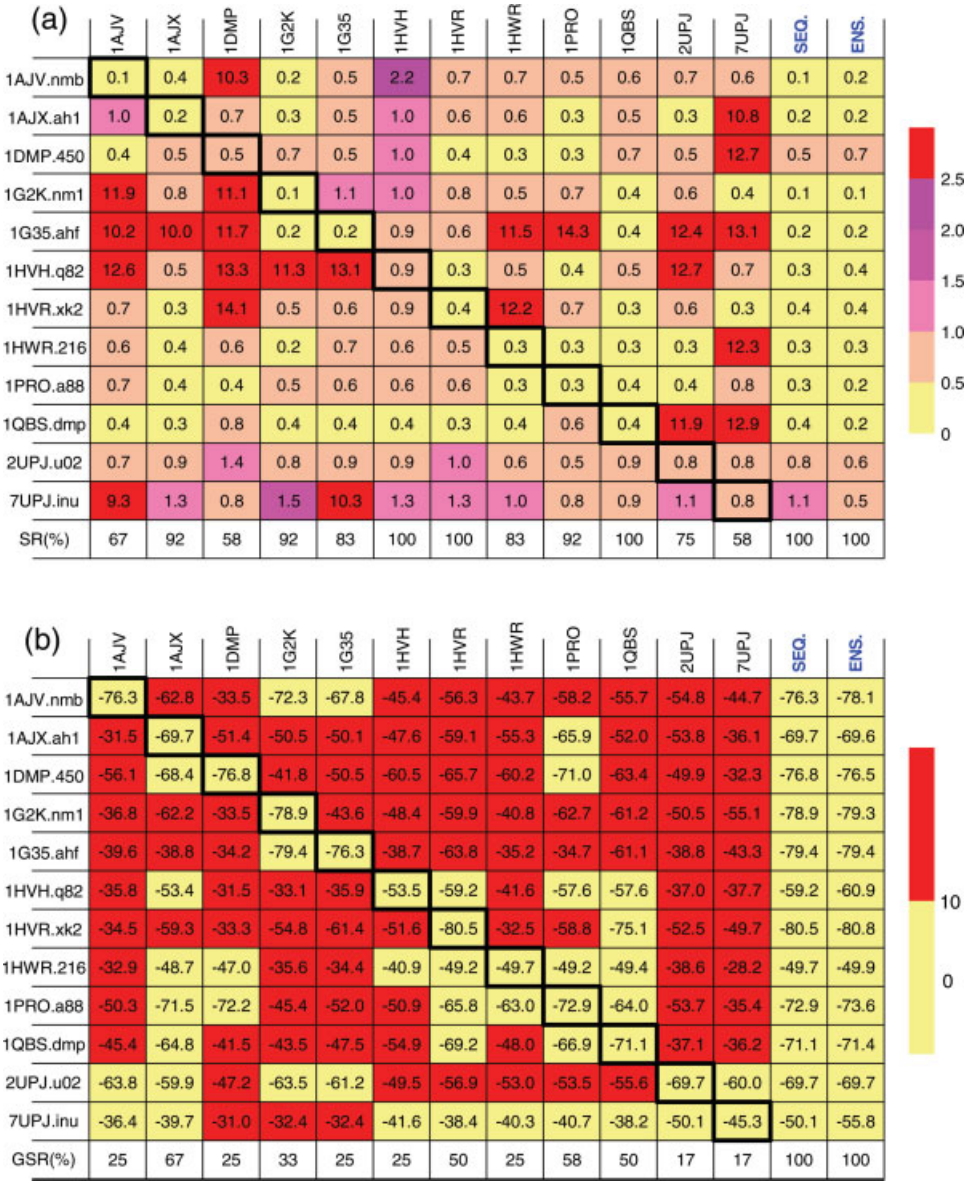


Fig. 5. Cross-docking on HIVpa. The legend is the same as that of Figure 2. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ranked within the top five. However, in the case of sequential docking, which samples $M \times 500$ orientations, the chance of generating false positives is increased and the rank of the correct pose may thus be lowered. Use the CDK2 ensemble as an example. Sequential docking ranked the correct pose of 1JSV.U55, one of the 12 ligands in the ensemble of CDK2, as No. 7, and thus was considered to fail to predict the binding mode of 1JSV.U55 according to our definition. However, ensemble docking ranked the correct pose of 1JSV.U55 as No. 2 and was described as a success. As a result, ensemble docking has a higher success rate than sequential docking for CDK2. Similar findings were observed for P38. It should be emphasized that the better performance of ensemble docking here does not essentially demonstrate the advantage of ensemble docking over sequential docking, but rather reflects the difference of ligand orientation sampling and the influence of the scoring function.

Detailed results for cAPK, CDK2, DHFR, and HIVpa are given in Figures 2–5. It is notable that unlike single-structure docking, which often identifies the correct modes but gives poor scores (as discussed in "The scoring criterion"), ensemble docking and sequential docking always yield good scores for the right poses. This phenomenon is especially remarkable for DHFR and HIVpa, as seen from Figures 4 and 5 where there are many more red cells in Panel B than in Panel A except the last two columns (representing sequential and ensemble docking, respectively). For example, in Figure 5, re-docking 1HVR.xk2 into 1HVR (native-docking) yields an

TABLE IV. Results of Ensemble Docking

		First solution		Rank 1			First solution		Rank 1
Ensemble	Ligand	rmsd (Å)	Rank	rmsd (Å)	Ensemble	Ligand	rmsd (Å)	Rank	rmsd (Å)
cAPK	1BKX.adn	0.35	1	0.35		3DRC.mtx	0.16	1	0.16
	1BX6.ba1	0.49	1	0.49		4DFR.mtx	0.29	1	0.29
	1FMO.adn	0.46	1	0.46	Ricin	1APG.apg	0.17	1	0.17
	1STC.sto	0.17	1	0.17		1FMP.fmp	1.84	1	1.84
	1YDR.1qp	0.22	1	0.22		1IFS.ane	0.39	1	0.39
	1YDS.1qs	0.33	1	0.33		1IFU.fmc	0.57	1	0.57
	1YDT.1qb	0.30	1	0.30		1OBT.amp	2.46	342	3.71
CDK2	1AQ1.stu	0.34	2	2.78	Trypsin	1PPC.napap	0.53	1	0.53
	1DI8.dtq	1.06	1	1.06		1PPH.tapap	0.29	1	0.29
	1DM2.hmd	0.74	21	3.71		1TNG.amc	0.55	1	0.55
	1E1X.nw1	1.06	1	1.06		1TNH.fba	0.55	2	10.13
	1E9H.inr	1.30	1	1.30		1TNI.pbn	2.02	7	9.76
	1FVT.106	0.94	1	0.94		1TNJ.pea	2.12	5	10.30
	1FVV.107	0.24	1	0.24		1TNK.pra	1.54	2	9.59
	1G5S.i17	0.80	1	0.80		1TNL.tpa	2.23	4	10.00
	1H1P.cmg	0.54	1	0.54		3PTB.ben	1.96	1	1.96
	1H1Q.2a6	1.92	1	1.92	HIVpa	1AJV.nmb	0.19	1	0.19
	1H1S.4sp	0.30	1	0.30		1AJX.ah1	0.15	1	0.15
	1JSV.u55	0.62	2	4.50		1DMP.450	0.68	1	0.68
LCK	1QPC.anp	0.61	2	5.59		1G2K.nm1	0.08	1	0.08
	1QPD.stu	0.66	1	0.66		1G35.ahf	0.16	1	0.16
	1QPE.pp2	0.30	1	0.30		1HVH.q82	0.36	1	0.36
	1QPJ.stu	0.46	1	0.46		1HVR.xk2	0.37	1	0.37
P38	1A9U.sb2	1.08	1	1.08		1HWR.216	0.33	1	0.33
	1BL6.sb6	0.63	3	6.94		1PRO.a88	0.24	1	0.24
	1BL7.sb4	0.34	1	0.34		1QBS.dmp	0.22	1	0.22
	1BMK.sb5	0.59	8	5.64		2UPJ.u02	0.59	1	0.59
	1DI9.msq	0.44	28	9.54		7UPJ.inu	0.49	1	0.49
	1M7Q.dqo	0.43	5	9.00	HIVpb	1AAQ.psi	0.46	1	0.46
αMMC	1AHA.ade	0.48	1	0.48		1C70.175	0.28	1	0.28
	1AHB.fmp	0.89	1	0.89		1HBV.gan	0.32	1	0.32
	1MRG.adn	2.44	1	2.44		1HIH.c20	0.09	1	0.09
	1MRH.fmc	0.67	1	0.67		1HPS.run	0.69	1	0.69
DHFR	1DHJ.mtx	0.29	1	0.29		1HPV.478	0.56	1	0.56
	1DRA.mtx	0.28	1	0.28		1HPX.kni	0.23	1	0.23
	1DRB.mtx	0.19	1	0.19		1HTF.g26	0.42	1	0.42
	1DYH.dzf	0.54	1	0.54		1HXW.rit	0.19	1	0.19
	1DYI.fol	0.36	1	0.36		1ODY.lp1	0.41	1	0.41
	1DYJ.ddf	0.44	1	0.44		1SBG.im1	0.08	1	0.08
	1JOL.ffo	0.50	1	0.50		2BPV.1in	0.19	1	0.19
	1RA2.fol	0.56	1	0.56		2BPY.3in	0.26	1	0.26
	1RA3.mtx	0.24	1	0.24		5HVP.acetyl*	0.47	1	0.47
	2DRC.mtx	0.20	1	0.20		7HVP.jg365	0.32	1	0.32
						8HVP.u85548e	0.50	1	0.50

"Rank 1" refers to the top-ranked orientation of each ligand. "First solution" refers to the first orientation among the ranked orientations that satisfies the rmsd criterion (< 2.5 Å).

rmsd of 0.4 Å and a score of −80.5. Cross-docking the same ligand into 7UPJ identifies the correct binding mode (rmsd = 0.3 Å), but gives a poor score (−49.7). In contrast, both sequential docking (0.4 Å, –80.5) and ensemble docking (0.4 Å, –80.8) make good binding mode as well as energy score predictions. The difference results from the fact that ensemble docking and sequential docking have incorporated protein conformational changes in the ensemble, which is absent in cross docking of a single protein structure.

As seen from Figures 2–5, ensemble docking only fails to predict the binding mode of 1DM2.hmd for CDK2, and sequential docking fails only in the cases of 1DM2.hmd and 1JSV.u55 (Panel A of Fig. 3). The three failures result from the fact that several wrong but well-scored poses lower the rank of the corresponding near-native binding mode below the top five. For example, in the merged ranking list of sequential docking for 1DM2.hmd, a near-native binding mode (rmsd of 0.2 Å) was ranked at the 8-th place. Similarly, a pose with

TABLE V. Comparison of the Average Run Time for Docking a Ligand With Different Algorithms[a]

Ensemble	No. of structures	Docking one ligand (seconds)		
		Single	SEQ.	ENS.
cAPK	8	1.40	11.21	0.87
CDK2	13	1.15	14.96	0.90
LCK	5	1.21	6.04	1.12
P38	7	1.03	7.18	0.87
α MMC	7	0.58	4.04	0.41
DHFR	12	2.37	28.48	1.80
Ricin	9	0.98	8.85	0.94
Trypsin	16	0.59	9.38	0.53
HIVpa	12	2.82	33.80	2.65
HIVpb	16	2.94	47.02	3.25
Average	*10.5*	*1.51*	*17.10*	*1.33*

[a]The run time was for generation and optimization of ligand orientations. All calculations were done on a Personal Computer with 3.2 GHz Pentium IV CPU and 3.0 GB RAM.

rmsd of 0.7 Å was ranked No. 7 for 1JSV.u55. This phenomenon occurs more often to sequential docking than to ensemble docking because the former involves more orientational sampling and thus is likely to include more improperly-scored wrong poses.

Computational Efficiency of the Ensemble Docking Algorithm

The above cross docking experiments have demonstrated that the performance of ensemble docking matches that of sequential docking, an important criterion to evaluate flexible-receptor docking algorithms. Another crucial, and maybe more challenging criterion is the computational efficiency.

Table V lists the average run time for docking a ligand into multiple protein structures with the ensemble docking algorithm. For comparison, Table V also lists the average run time for docking a ligand into a single, rigid protein structure and the average accumulated run time for sequentially docking a ligand into all protein structures of the ensemble. The run time is spent on generation and optimization of ligand orientations, which excludes the preparation time for the one-time, precalculated energy grid. All calculations were done on a personal computer with 3.2 GHz Pentium IV CPU and 3.0 GB RAM.

Table V shows that ensemble docking takes much less time than sequential docking. The table also shows that the accumulated run time for sequential docking increases linearly with the number of protein structures in the ensemble. With an average of 10.5 structures in an ensemble, our ensemble docking algorithm (1.33 s) is about 12.9-fold faster than sequential docking (17.10 s), and comparable to that of single-receptor docking (1.51 s).

Note in Table V that the average run time of ensemble docking is even less than that of single docking. The result may seem a little surprising because the optimization procedure of ensemble docking is expected to take longer time because of the addition of one more parameter. The time difference can be understood as follows. The run time in the present work is spent on two processes, matching/orienting and scoring/optimization. Indeed, ensemble docking is slightly more computationally expensive than single docking on the second process—scoring/optimization of ligand orientations, about 1.5-fold slower. However, ensemble docking spends less time on the first process—matching/orienting for ligand orientations—than single docking, resulting from the nature of the present matching algorithm, which can be explained as follows. As pointed out in Ref. 41, the distance matching algorithm that DOCK4.0 uses will spend more time on molecules that are complementary in shape to the receptor because they have more internal distances in common and thus increase the number of matches between the hot spots in the binding site ("spheres") and these molecules. Here, the docked molecules are usually more similar in shape to the binding site of the single protein structures than that of the reference protein used in ensemble docking because of the artificially enlarged binding pocket of the reference protein. Consequently, single docking usually has more matches than ensemble docking during the generation of ligand orientations and thus consumes more cpu time on this process, although ultimately only top 500 matched orientations are used for later scoring/optimization in both docking algorithms. In summary, to understand the computational efficiency of ensemble docking, one needs to consider that the total run time is determined by two processes in the docking algorithm. If the early process—matching/orienting—were to consume a dominant portion of the total cpu time, then ensemble docking would be up to five times faster than single docking. On the other hand, if the later process—scoring/optimization—were to consume a dominant portion of the cpu time, then single docking would be about 1.5 times faster than ensemble docking. In the present work, the two processes consume comparable cpu time, so on average the resulting run time of ensemble docking is less than that of single docking. It should be emphasized that the comparison on run time between ensemble docking and single docking is not an absolute measure, because the run time of the process for matching/orienting depends on individual protein ensemble as well as the docking algorithm. Therefore, when the run time of the early process is close for two docking methods, single docking will be expected to be faster than ensemble docking (e.g. ensemble HIVpb) (Table V).

Overall, our ensemble docking algorithm yields a run time comparable to that of single rigid-receptor docking, independent of the number of structures in the ensemble. As shown in the Supplementary Material, flexible ligand docking does not change the conclusion.

Virtual Database Screening With the Ensemble Docking Algorithm

The third important criterion to evaluate a flexible-receptor docking algorithm is virtual database screening. Virtual database screening is a commonly used approach in computational structure-based drug design. After

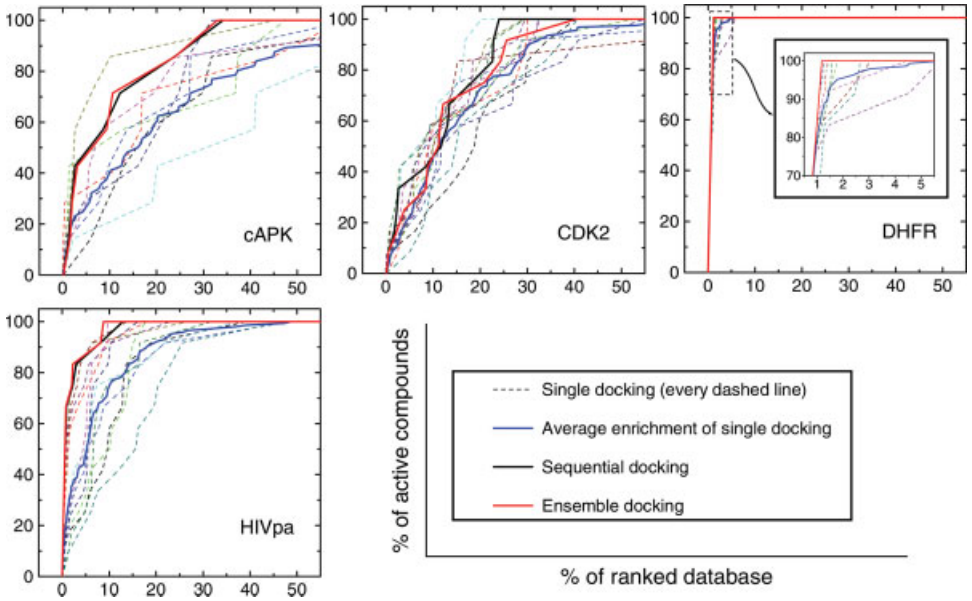


Fig. 6. Comparison of different docking methods on enrichment tests against four typical proteins. The dashed lines stand for single rigid-receptor docking with every structure in a protein ensemble. The blue solid line indicates the average enrichment of all single docking results. The black solid line represents sequential docking, whereas the red solid line represents ensemble docking. The inset in Panel DHFR is an enlargement of the region enclosed by the dashed box.

screening, potentially potent compounds are selected out of a large database for further experimental tests, which hopefully would dramatically reduce the time and cost of drug discovery. Using only one rigid protein structure for screening would have missed potential inhibitors.

To assess the performance of our ensemble docking algorithm, small-scale virtual database screening tests have been carried out against four typical protein ensembles (cAPK, CDK2, DHFR, and HIVpa), which contain sufficient ligand-bound structures. For each ensemble, the database is composed of active compounds, consisting of all the known ligands in the ensemble, and 1000 inactive compounds, extracted randomly from the Available Chemical Directory (see Materials and Methods).

Every ligand (active or inactive) in the database was allowed to be flexible, with multiple conformations generated via the OMEGA software, Version 1.0b6 (OpenEye).[40] Specifically, the default parameters of OMEGA v1.0b6 were used except the following two parameters: MAXCONFS, the maximum number of output conformers per molecule, was set to 20; RMS, the rmsd threshold used for duplicate removal, was set to 1.0. The corresponding default values of MAXCONFS and RMS in OMEGA v1.0b6 are 50 and 0.8 Å, respectively. Here, we restricted the maximum number of output conformers by increasing the value of RMS in order to save computational time for every database screening. A diverse set of conformers of up to 8 Å of rmsd were obtained for each ligand. Certainly, MAXCONFS and EWINDOW, the energy cutoff from the global minima for kept conformers, can be set to higher values in order to sample more conformational space of the ligands. Our control studies described in the Supplementary Material and in later section have shown that the use of larger values of MAXCONFS and EWINDOW does not change the results significantly.

Next, each generated conformer was treated as a rigid body and docked into a protein ensemble. Only intermolecular interactions were considered. In the previous sections, the top five ranked orientations were used for a given ligand conformation for rmsd evaluations in order to minimize possible influence of the scoring function. Here, only the best scored orientation was kept for each ligand conformation because the binding modes of the inactive compounds are unknown. Then, only the best scored conformation was kept for each ligand to compare with other ligands in the database for ranking.

A widely used index of success for virtual screening is the enrichment factor, defined as the percentage of found known inhibitors in a certain top percentage of the ranked database. It should be noted that the computed enrichments here are expected to be high because we actually performed cross docking experiments in which the bound protein structures of the inhibitors are included in the database, even though high enrichments are true for both ensemble docking and sequential docking and thus do not affect our comparison between these two algorithms. Figure 6 shows the enrichment results of the ensemble docking algorithm for four typical ensembles, each of which contains at least seven known active ligands. Detailed data are presented in Table VI, showing the enrichment factors at several typical database percentages (1, 2, 5, and 10%). For comparison, the results of single rigid-receptor docking (including their average enrichments) and sequential docking are also given in Figure 6 and Table VI. Figure 6 shows

TABLE VI. Comparison of the Enrichments at Four Top Percentages (1, 2, 5, and 10%) of the Ranked Database Between Single Rigid-Receptor Docking, Average of Single Docking (AVE.), Sequential Docking (SEQ.), and Ensemble Docking (ENS.) on Four Typical Ensembles

	Enrichment at the top *X*% of the ranked database (%)															
Ensemble	Single rigid-receptor docking													AVE.	SEQ.	ENS.
cAPK (%)	1BKX	1BX6	1FMO	1STC	1YDR	1YDS	1YDT	1JLU								
1	0.0	28.6	14.3	0.0	0.0	14.3	14.3	0.0						8.9	0.0	0.0
2	0.0	28.6	42.9	0.0	14.3	28.6	14.3	14.3						17.9	28.6	28.6
5	0.0	28.6	42.9	14.3	28.6	57.1	28.6	14.3						26.8	42.9	42.9
10	28.6	28.6	42.9	14.3	57.1	71.4	28.6	42.9						39.3	57.1	57.1
CDK2 (%)	1AQ1	1DI8	1DM2	1E1X	1E9H	1FVT	1FVV	1G5S	1H1P	1H1Q	1H1S	1JSV	1HCL			
1	8.3	0.0	0.0	8.3	8.3	0.0	8.3	8.3	8.3	16.7	16.7	8.3	0.0	7.1	8.3	8.3
2	8.3	0.0	8.3	8.3	16.7	8.3	8.3	8.3	8.3	16.7	25.0	8.3	0.0	9.6	16.7	8.3
5	8.3	25.0	16.7	16.7	41.7	33.3	16.7	25.0	16.7	16.7	41.7	16.7	8.3	21.8	33.3	25.0
10	25.0	50.0	50.0	33.3	50.0	50.0	58.3	50.0	58.3	50.0	58.3	33.3	33.3	46.2	41.7	41.7
DHFR (%)	1DHJ	1DRA	1DRB	1DYH	1DYI	1DYJ	1JOL	1RA2	1RA3	2DRC	3DRC	4DFR				
1	83.3	83.3	83.3	83.3	83.3	83.3	83.3	83.3	75.0	75.0	75.0	58.3		79.2	83.3	83.3
2	100.0	100.0	100.0	100.0	100.0	91.7	100.0	100.0	83.3	83.3	100.0	83.3		95.1	100.0	100.0
5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	91.7	100.0	100.0	100.0		99.3	100.0	100.0
10	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0		100.0	100.0	100.0
HIVpa (%)	1AJV	1AJX	1DMP	1G2K	1G35	1HVH	1HVR	1HWR	1PRO	1QBS	2UPJ	7UPJ				
1	16.7	41.7	16.7	8.3	16.7	16.7	50.0	16.7	50.0	50.0	0.0	8.3		24.3	66.7	66.7
2	16.7	58.3	16.7	25.0	33.3	25.0	75.0	25.0	58.3	66.7	16.7	8.3		35.4	66.7	66.7
5	25.0	66.7	25.0	41.7	41.7	50.0	83.3	33.3	75.0	83.3	25.0	16.7		47.2	83.3	83.3
10	50.0	91.7	41.7	66.7	75.0	100.0	91.7	83.3	83.3	91.7	75.0	33.3		73.6	91.7	100.0

that the ensemble docking algorithm always gives comparable results to sequential docking, and performs better than most of single, rigid-receptor docking runs.

For cAPK, among the top 10% of the ranked compounds, ensemble docking identifies 57.1% of the known inhibitors, matches sequential docking, and performs better than single, rigid-receptor docking except for 1YDS. The worst rigid-receptor docking occurs to 1STC, in which case only 14.3% of the known inhibitors are found (Table VI and Fig. 6), indicating the strong impact of the induced-fit on virtual database screening.

For CDK2 (Table VI and Fig. 6), ensemble docking is close to sequential docking in identifying the known ligands. Figure 6 also shows that the overall performance of ensemble docking (red line) is slightly better than the average enrichment of single rigid-receptor docking (blue line). Docking with the apo structure (1HCL) performs the worst and cannot identify any known inhibitor until at the top 5% of the ranked database.

For DHFR, all the docking algorithms yield good enrichments and find all the known inhibitors within the top 6% of the ranked database. Ensemble docking and sequential docking still perform better than single docking (the inset of Panel DHFR in Fig. 6). The high enrichments may be attributed to the following two reasons. First, most of the listed active ligands are strong inhibitors of DHFR and are therefore readily identified from the database. Second, seven of the given active ligands of DHFR are methotrexate (mtx), the other five ligands share very similar structures to mtx, leading to similar ligand-induced conformational changes among the protein structures of the ensemble and thus similar docking results. Our results with DHFR on binding mode and energy score predictions and enrichment tests suggest that this ensemble of 12 ligand-bound protein structures may not be a sufficiently challenging test set for the evaluation of a flexible-receptor docking algorithm. Moreover, as pointed out by McGovern and Shoichet,[45] a holo conformation might bias the docking screen, tending to bind the compounds similar to the ligand in the holo complex. Therefore, despite the high enrichments, this DHFR ensemble would allow only ligands with a similar scaffold to mtx to be docked and may be inappropriate for new drug discovery.

For HIVpa, ensemble docking and sequential docking perform the best and identify 100 and 91.7% of the known inhibitors at the top 10% of the ranked database, respectively (Table VI and Fig. 6).

In summary, ensemble docking matches sequential docking. Despite a few exceptions, overall, ensemble docking performs better than single rigid-receptor docking. Single docking using apo structures (e.g. 1HCL of CDK2) performs less satisfactorily, which is in agreement with the findings in Ref. 45.

An Application to HIV-1 Protease

In the previous sections, we have validated our ensemble docking algorithm by cross docking experiments on 10 diverse protein ensembles in terms of docking accuracy and computational efficiency. In this section, we will evaluate our algorithm on a more realistic protein ensemble in which the corresponding bound protein structures of the known ligands are excluded. The HIV-1 protease was selected for this test because many HIVp complexes are available in the PDB. Here, the protein ensemble consists of 10 structures extracted from 10 HIVp complexes (PDB codes: 1AAQ, 1C70, 1HIH, 1HPS, 1HPV, 1HTF, 2BPV, 5HVP, 7HVP, 8HVP). The corresponding bound ligands were excluded in this test. The structural water was kept as part of the protein as aforementioned. Ten inhibitors extracted from 10 other HIVp complexes were selected as the active ligands [Fig. 7(a)]. To evaluate the ability of our ensemble docking algorithm in distinguishing known inhibitors from decoy compounds, we also constructed a test set consisting of 12 random, diverse molecules from the ACD, which have similar chemical properties to the known inhibitors and serve as the inactive compounds [Fig. 7(b)].

All the active and inactive ligands were docked into the single-rigid structures with the single docking approach and into the HIVp structural ensemble with the ensemble docking method. The ligands were treated as flexible molecules, with their multiple conformations generated with the OMEGA software, as described in earlier Section.

Figure 8 shows the results of predicting binding modes. Here, we adopted a more commonly- used criterion for calculating docking accuracy. Namely, a prediction is defined as a success if the top-ranked orientation has an rmsd below 2.0 Å.[43] It can be seen from Figure 8 that the ensemble docking algorithm correctly predicts the binding modes of all the inhibitors, yielding a success rate of 100%. In contrast, the success rates of single-structure docking are lower (ranging from 10 to 90%, Fig. 8) because of the limitation of the rigid-receptor treatment. The higher success rate of ensemble docking results from the consideration of protein structural variations.

We also evaluated the performance of ensemble docking on energy score prediction. The results are shown in Table VII. The columns represent different docking runs (single-structure docking or ensemble docking). The rows list the active and inactive ligands according to their binding scores, from low to high.

To further quantitatively evaluate the efficiency of different docking methods in discriminating inhibitors from noninhibitors, the following measure is defined, similar to Ref. 46:

$$EF = 100\% \times \frac{\langle \text{Rank(inactive)} \rangle - \langle \text{Rank(active)} \rangle + N_{\text{cmpd}}/2}{N_{\text{cmpd}}} \quad (4)$$

where N_{cmpd} is the number of the tested compounds, $\langle \text{Rank(inactive)} \rangle$ and $\langle \text{Rank(active)} \rangle$ are the average ranks of the inactive and active compounds, respectively. The value of the efficiency factor EF can vary from 0 to

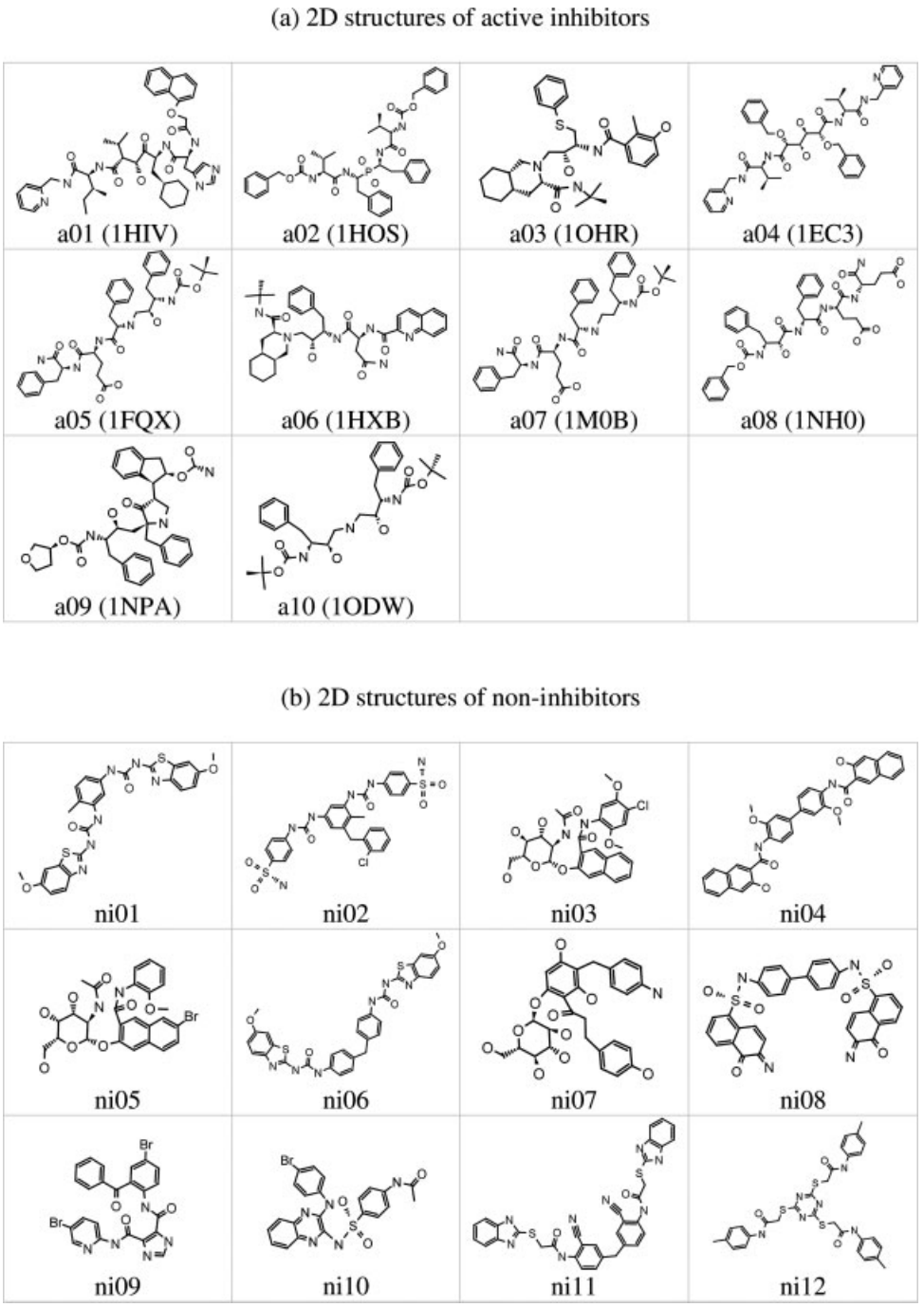


Fig. 7. The 2D structures of the compounds used in an application to HIV-1 protease. (**a**) The active inhibitors; the corresponding PDB codes are shown in parentheses. (**b**) The selected compounds from the ACD with similar chemical properties to the active inhibitors and serve as inactive compounds.

100%. If all the inhibitors are ranked favorable over the noninhibitors, the efficiency factor EF is 100%, indicating the complete agreement between the predictions and the experimental findings. In contrast, if all the noninhibitors are ranked better than the inhibitors, the efficiency factor EF is 0%. An equal distribution of the inhibitors in the ranked list of compounds gives an efficiency factor of 50%, representing random selection. The higher the EF value, the better the efficiency of identifying known inhibitors. Table VII gives the efficiency factors for each docking run. It can be seen from this table that ensemble docking discriminates all the inhibitors from the rest compounds by ranking the inhibitors in the top of the list, giving an efficiency factor of 100%. However, single-structure docking gives much lower efficiency factors, ranging from 25.8 to 84.2%. For some

	1AAQ	1C70	1HIH	1HPS	1HPV	1HTF	2BPV	5HVP	7HVP	8HVP	ENS.
a01	13.9	13.5	14.8	8.6	15.2	13.8	13.3	16.9	0.5	15.7	0.5
a02	0.9	0.8	0.6	0.6	0.8	13.0	0.6	0.5	9.9	15.4	0.5
a03	0.8	0.5	0.5	0.3	11.0	11.9	0.6	11.7	0.3	12.5	0.4
a04	0.8	0.3	0.8	14.1	15.4	0.6	0.8	0.4	0.2	0.3	0.3
a05	0.2	0.8	0.8	0.6	0.7	14.9	0.4	0.7	0.7	13.6	0.5
a06	6.9	0.5	1.2	0.3	0.5	12.3	0.3	0.6	10.9	0.8	0.2
a07	11.7	0.7	1.0	0.6	0.6	14.0	10.2	0.5	13.0	14.0	0.5
a08	3.7	3.9	0.7	4.0	0.5	3.9	3.9	4.1	0.3	0.4	0.4
a09	9.9	0.6	0.6	0.6	11.7	11.6	0.7	0.7	13.4	15.1	0.4
a10	0.4	0.8	0.5	0.2	11.3	13.2	15.2	12.2	12.8	13.6	0.7
SR(%)	50	80	90	70	50	10	60	60	50	30	100

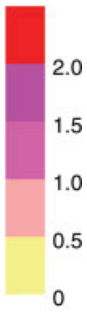


Fig. 8. The binding modes predicted by single-structure docking and ensemble docking on a HIV-1 protease ensemble. The rows refer to different ligands, and the columns represent different protein structures. Note that the corresponding bound ligands in the protein ensemble were excluded in this test. SR stands for the success rate of binding mode predictions (see the text for definition).

structures such as 1AAQ, 1HPV, 1HTF, 7HVP, and 8HVP, the efficiency factors are below 50% and worse than that of random selection (Table VII). These results again demonstrate the limitations of single-structure docking and the efficiency of the ensemble docking algorithm.

Effect of Full Ligand Flexibility on Ensemble Docking

Lastly, we will briefly discuss about the effect of using the full set of ligand conformers on the docking results. Specifically, we used one of the latest versions of OMEGA 1.8.1 (OpenEye) to generate a very wide range of ligand conformers, with the energy cutoff parameter, EWINDOW, set to 14.0 kcal/mol and the maximum number of output conformers, MAXCONFS, set to 200. In the earlier section, EWINDOW was set to 3.0 kcal/mol and MAXCONFS was set to 20, though a diverse set of ligand conformers (rmsd of up to 8 Å) were still obtained with this set of narrower OMEGA parameters. Following the procedures described in the earlier section, each ligand conformer was docked as a rigid body, and the scores of different conformers were merged and ranked for each ligand.

Three typical protein ensembles (cAPK, DHFR, and HIVpa), each consisting of 8–12 protein structures, were used for docking tests of known ligands. Docking accuracies were evaluated in terms of the rmsd and scoring criteria described in the earlier section. Table VIII lists the success rates of different docking methods for each protein ensemble. For comparison, the success rates for rigid-ligand docking given in Tables II and III are also listed in Table VIII. Table VIII shows that the success rates of native docking are still high, serving as a validation of the ITScore scoring function used in the present work. The table also shows that when considering a wide range of ligand conformations, the relative performances of different docking methods are consistent with those for treating the ligands as rigid bodies. Namely, the performance of ensemble docking matches that of sequential docking, and considerably better than that of cross docking.

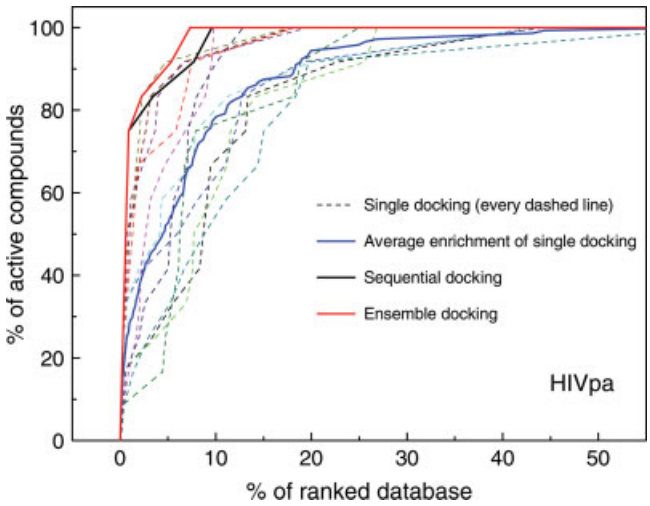


Fig. 9. Comparison of different docking methods on the enrichment test against the HIVpa ensemble using a wider range of ligand conformers (EWINDOW = 14.0 kcal/mol and MAXCONFS = 200). The figure legends are identical to those for Figure 6.

TABLE VII. A Ranked List of the Docked Ligands (Including Active Inhibitors and Inactive Noninhibitors) According to the Computed Binding Energies From 10 Single-Structure Docking Runs and the Corresponding Ensemble Docking Experiment

1AAQ		1C70		1HIH		1HPS		1HPV		1HTF		2BPV		5HVP		7HVP		8HVP		ENS.	
Lig.	Score	Lig.	Score	Lig.	Score	Lig.	Score	Lig.	Score	Lig.	Score	Lig.	Score	Lig.	Score	Lig.	Score	Lig.	Score	Lig.	Score
a04	−54.7	a04	−79.0	a07	−69.1	a02	−73.9	a02	−56.3	a04	−51.6	a05	−68.4	a04	−62.8	a01	−78.5	a04	−64.8	a04	−75.4
a05	−53.9	a05	−64.9	a02	−68.8	a06	−65.6	ni08	−51.9	ni08	−49.7	a02	−65.8	a05	−61.4	a08	−58.9	a08	−56.2	a01	−73.5
ni11	−50.6	a10	−64.2	a05	−66.1	a05	−63.7	a08	−50.0	ni09	−46.0	a06	−62.3	a02	−60.9	a04	−58.6	ni08	−55.3	a02	−69.8
ni06	−49.6	a02	−62.5	a06	−56.2	a07	−62.3	a01	−49.9	ni07	−45.8	a03	−58.5	a08	−53.0	a05	−55.0	ni09	−49.5	a05	−65.4
a02	−49.4	a09	−59.5	a03	−55.1	a09	−60.3	ni09	−48.9	ni04	−45.7	ni08	−54.1	ni10	−52.9	ni08	−54.0	ni01	−48.0	a06	−63.2
ni01	−49.1	a08	−57.6	a04	−54.7	a03	−58.1	ni06	−47.4	ni10	−45.0	ni10	−52.2	ni08	−52.7	ni10	−52.7	ni10	−47.1	a07	−61.6
ni08	−48.6	a03	−56.8	a09	−53.8	ni04	−56.0	ni04	−46.2	a07	−44.7	ni06	−52.0	ni06	−52.4	ni11	−50.3	ni04	−46.5	a03	−60.0
ni10	−48.2	ni08	−55.1	ni08	−53.1	ni05	−55.2	ni12	−46.1	ni06	−44.3	a07	−50.4	ni04	−52.2	ni05	−50.3	a07	−46.4	a09	−59.8
ni09	−47.8	ni11	−53.6	ni03	−50.6	ni08	−54.2	ni05	−45.9	ni05	−44.2	ni11	−48.8	ni09	−51.1	ni04	−50.1	a05	−46.2	a10	−59.6
ni03	−47.5	ni07	−52.1	ni10	−50.0	ni11	−53.8	ni10	−45.7	a02	−43.9	ni09	−48.5	ni11	−49.5	ni09	−49.5	ni06	−45.5	a08	−57.7
ni07	−46.4	a07	−51.6	a10	−49.7	a10	−52.7	a05	−45.4	ni11	−42.9	ni03	−48.4	a09	−49.4	ni03	−49.4	ni05	−45.2	ni04	−54.5
ni04	−45.9	ni10	−51.1	ni05	−49.0	ni03	−51.8	ni03	−44.6	ni02	−42.0	ni07	−47.4	a06	−48.8	ni06	−48.6	ni11	−44.6	ni06	−54.4
ni05	−45.8	ni09	−48.3	a08	−48.8	ni09	−49.8	ni07	−44.5	ni12	−41.9	ni01	−47.2	ni07	−48.5	ni02	−47.8	a06	−44.4	ni08	−54.3
a03	−45.3	ni06	−48.3	ni11	−48.7	ni10	−49.5	ni11	−44.3	a08	−41.8	a04	−46.8	ni01	−48.2	ni07	−46.2	ni03	−43.6	ni09	−53.5
a10	−45.1	ni05	−47.3	ni07	−48.4	ni06	−48.8	ni02	−44.2	a01	−41.1	ni05	−46.3	ni02	−48.1	a03	−46.0	ni02	−42.8	ni10	−51.9
ni02	−44.8	ni03	−46.3	ni06	−47.9	ni12	−48.7	a04	−43.1	a10	−40.2	a09	−46.0	ni05	−47.2	a07	−44.1	ni07	−42.8	ni07	−51.8
a08	−44.0	ni12	−46.1	ni09	−47.7	a08	−47.7	a10	−42.2	ni03	−39.0	ni04	−45.9	ni03	−46.8	ni01	−43.9	a01	−42.5	ni03	−51.8
ni12	−44.0	a06	−46.0	ni01	−47.1	a01	−47.7	a07	−41.6	a05	−38.7	a08	−45.4	ni12	−46.6	a10	−42.9	a02	−41.9	ni11	−50.5
a07	−41.7	ni04	−45.4	ni02	−44.6	ni02	−47.7	ni01	−41.3	ni01	−38.0	a01	−44.4	a01	−45.5	ni12	−41.8	ni12	−40.2	ni05	−50.4
a01	−39.5	ni02	−44.8	ni12	−44.5	ni07	−47.1	a06	−39.8	a06	−37.5	ni02	−43.7	a07	−45.2	a02	−41.4	a10	−40.1	ni12	−49.6
a09	−38.6	ni01	−44.5	ni04	−41.3	ni01	−45.2	a03	−39.6	a03	−36.7	ni12	−42.0	a10	−43.3	a06	−40.6	a09	−38.6	ni01	−48.0
a06	−36.9	a01	−42.9	a01	−37.0	a04	−43.9	a09	−34.2	a09	−34.4	a10	−41.9	a03	−40.4	a09	−40.1	a03	−38.5	ni02	−45.8
EF(%)	32.5	EF(%)	80.0	EF(%)	84.2	EF(%)	71.7	EF(%)	35.0	EF(%)	25.8	EF(%)	56.7	EF(%)	50.0	EF(%)	44.2	EF(%)	36.7	EF(%)	100.0

The first 10 columns refer to single-structure docking, and the last column to ensemble docking. The active inhibitors are colored in gray. The last row lists the efficiency factors of discriminating inhibitors from noninhibitors for each docking run.

TABLE VIII. Comparison of the Success Rates in Terms of rmsd and rmsd + score Criteria[a] Between Single Docking, Sequential Docking (SEQ.), and Ensemble Docking (ENS.) on Three Typical Protein Ensembles, With Each Ligand Treated as a Flexible Molecule and Represented by up to 200 Conformers Generated by the OMEGA Software

		Success rates (%)[b]											
		Native-docking		Cross-docking		Apo-docking		Average		SEQ.		ENS.	
Ensemble	Criteria												
cAPK	rmsd	**86**	100	**57**	76	**57**	71	**61**	79	**86**	100	**86**	100
	rmsd + score	**86**	100	**43**	43	**43**	57	**48**	52	**86**	100	**86**	100
DHFR[c]	rmsd	**100**	100	**98**	100	**n/a**	n/a	**98**	100	**100**	100	**100**	100
	rmsd + score	**100**	100	**90**	87	**n/a**	n/a	**91**	88	**100**	100	**100**	100
HIVpa[c]	rmsd	**100**	100	**76**	82	**n/a**	n/a	**78**	83	**100**	100	**100**	100
	rmsd + score	**100**	100	**27**	28	**n/a**	n/a	**33**	34	**100**	100	**100**	100

[a]The rmsd criterion for the success rate is that any of the top five conformations of the ligand has an rmsd <2.5 Å.
[b]The success rates are listed in the left column of each method category and highlighted in bold font. For comparison, the success rates of docking the rigid ligands are listed in the right column of each method category.
[c]No apo-docking results for DHFR and HIVpa because of the absence of the corresponding ligand-free conformations.

Then, we repeated the virtual screening study on the HIVpa ensemble, using the wider set of OMEGA parameters for the generation of ligand conformers. The results are shown in Figure 9. Comparing to the panel for HIVpa in Figure 6, the enrichment curves in the two figures are similar.

In summary, even including wildly different ligand conformations still demonstrates the effectiveness of the ensemble docking algorithm.

DISCUSSION

One useful approach to account for receptor flexibility in molecular docking is to use multiple protein conformations obtained from X-ray or NMR studies, or molecular simulations.[3,4] One critical question for the use of multiple protein conformations is how to efficiently combine these structures in docking calculations.[3,4] A straightforward option is to dock a ligand sequentially into all the protein conformations. However, this kind of sequential docking approach is computationally expensive and thus impractical for large database screening. The present study attempts to address this question.

Direct comparison between our ensemble docking algorithm with other flexible-receptor docking algorithms using multiple experimental protein conformations is difficult because of the use of different sampling and optimization procedures as well as scoring functions.[43] Therefore, we discuss only the differences from other related work as follows.

Unlike the composite-grid methods, which combine different conformations into a single representation of an average energy grid and thus may lead to loss of geometric accuracy,[20,22,25] our ensemble docking algorithm treats each conformation of the ensemble separately during the energy calculations and pose optimization. In other words, our algorithm keeps all the structural information of different conformations, and thus has an accuracy comparable to the sequential docking approach. Second, without using the united protein representation[23] or the composite protein representation,[19] our approach, in principle, can cope with larger movements than side-chain and loop movements, as long as these movements are reflected in the given structural ensemble. Third, a valuable feature of FlexE[23] is its ability to generate new validate protein structures that do not present in the initial ensemble, though this feature also renders its computational speed. The work by Wei et al.[19] significantly improves the computational speed by assuming that the mobile regions of the receptor move independently and thus keeping only the best-fit local conformation for each flexible region. As a comparison, by simultaneously docking a ligand into multiple protein conformations, our ensemble docking algorithm is the fastest among the three algorithms, with a speed comparable to that of docking into a single structure. Thus, the speedup of the algorithm increases linearly with the number of structures. The speedup is especially prominent with NMR structures or computer-generated structures, which involve a large amount of structures. Finally, we have found that rmsd alone seems to be insufficient to evaluate the success of predictions on binding. Energy score analysis and enrichment tests are also important assessment criteria.

It is noticeable that there exists a small overlap between the test sets in the present study and the training set for the development of the ITScore scoring function.[36] The test sets (except the two HIVp ensembles) were directly taken from the literature on protein flexibility studies,[23,31] and consist of 105 protein structures, including 87 protein–ligand complexes, as shown in Table I. Among them, 12 complexes were found in the training set of ITScore: 6 for the CDK2 ensemble (PDB codes: 1DI8, 1DM2, 1E1X, 1E9H, 1FVT, and 1JSV), 2 for LCK (1QPC and 1QPE), and 4 for P38 (1A9U, 1BL6, 1BMK, and 1M7Q). The training set of ITScore was used to derive the pair potentials for the scoring function by reproducing the ligand binding conformations in the training set.[36] The slight overlap of the ligand–protein complexes (~13.8%) could partly contribute to the

high success rates (100%) of native docking given in Tables II and III, though it should not be statistically important enough to alter any of our conclusions. More importantly, native docking with 7 out of 10 protein ensembles listed in Table I that do not contain any complexes in the training set of ITScore still yielded success rates of 100%, further suggesting that the overlap should have little effect on the present study.

There are several limitations in our method. Compared with FlexE and Wei et al.'s algorithm, the ensemble docking method cannot generate conformations beyond those observed in the given ensemble. Yet, the diversity of protein structures is important for the discovery of new drugs.[45] One solution is to improve the conformational sampling quality of the ensemble, either by increasing the number of experimental structures, or by combining with some conformational generation algorithms (e.g., FlexE, Wei et al.'s algorithm, IFREDA, etc.). Second, an empirical energy correction term needs to be added in the future to account for receptor conformational changes.[19,47] A third limitation is the optimization method, which is also an important issue for all docking programs. Although the issue is out of the scope of this work, we will still discuss how it may affect the performance of our algorithm, as follows. Same as DOCK4.0, the present study uses SIMPLEX as the optimization tool. Thus, the ligand sometimes may be trapped in a local minimum because SIMPLEX is essentially a local minimization method. In fact, different local minima (and hence different energy scores) were seen in the sequential and ensemble docking experiments in the present work. As a result, occasional wrong predictions often originate from the optimization method rather than from the ensemble algorithm itself. There are two possible solutions to this problem. One solution is to use a relatively global minimization method such as genetic algorithm, which would not be affected by local minima. However, global minimization methods are computationally expensive. Because the DOCK4.0 software has been shown to perform quite well on identifying the correct ligand binding mode out of hundreds of ligand orientations with SIMPLEX, we therefore adopted SIMPLEX as the energy optimization tool in this study. The second solution is to choose a set of appropriate docking parameters to enhance the performance of the docking algorithm, for example, the number of sphere points, the maximum number of orientations, the maximum optimization steps, and so forth.

CONCLUSION

We have presented a fast ensemble docking algorithm to consider protein structural variations in ligand docking. In this algorithm, each protein conformation is treated separately and an integer conformational variable is introduced. By adjusting the protein conformational variable and the ligand coordinates via an energy optimization procedure, the ensemble docking algorithm can dock a ligand simultaneously into multiple protein structures and automatically select an optimal protein conformation. Our calculations have shown that the accuracy of the ensemble docking algorithm is comparable to that of the merged ranking list of sequential docking against all the structures. Our algorithm can incorporate any conformational change observed in the experimental/theoretical structures, ranging from local side-chain motions to large backbone and domain movements.

Specifically, combined with ITScore, an iterative knowledge-based scoring function recently developed by our group, the ensemble docking algorithm has been extensively evaluated on 10 protein ensembles of 105 crystal structures and 87 ligands. In the case of binding mode predictions, considering the top five orientations, the flexible-receptor ensemble algorithm identified 93% of the correct poses with a criterion of rmsd <2.5 Å, which performs much better than single apo structure docking (47%), considerably better than single rigid-structure docking (75%), and comparably to sequential docking (89%). Similar trends were observed in energy score predictions and the enrichment tests of virtual database screening. We further applied the ensemble docking algorithm to a more realistic ensemble of HIV-1 protease in which the corresponding bound protein structures of the inhibitors are excluded in the ensemble. The results show that the algorithm not only predicts the correct binding modes of the known inhibitors, but also discriminates the inhibitors from other compounds with similar chemical properties. In addition to docking accuracy, the ensemble algorithm also showed to be computationally efficient. The run time is comparable to that of ligand docking into a single structure, and does not depend on the number of conformations of an ensemble.

Though we have used multiple experimental structures in the present study, our algorithm can also be used for docking against a conformational ensemble generated via computational methods such as MD/MC simulations and above-mentioned united protein representation approaches. The idea can also be applied to improve the efficiency of a recently proposed new protocol for modeling ligand/receptor induced fit effect.[8] In this protocol, the ligand conformations and receptor conformations are sampled separately. Flexible ligands are docked into multiple induced-fit protein conformations, and the conformations are improved iteratively during docking calculations.

ACKNOWLEDGMENTS

We thank Dr. Jaya Pamidimukkala for valuable comments on the manuscript. Support to XZ from OpenEye Scientific Software Inc. (Santa Fe, NM), Tripos, Inc. (St. Louis, MO), and MDL Information Systems, Inc. (San Leandro, CA) is gratefully acknowledged.

REFERENCES

1. Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 2003;32:335–373.

2. Halperin I, Ma B, Wolfson H, Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 2002;47:409–443.
3. Carlson HA. Protein flexibility is an important component of structure-based drug discovery. Curr Pharm Des 2002;8:1571–1578.
4. Teodoro ML, Kavraki LE. Conformational flexibility models for the receptor in structure based drug design. Curr Pharm Des 2003;9:1635–1648.
5. Teague SJ. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2003;2:527–541.
6. Murray CW, Baxter CA, Frenkel AD. The sensitivity of the results of molecular docking to induced fit effects: application to thrombin, thermolysin and neuraminidase. J Comput Aided Mol Des 1999;13:547–562.
7. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. J Med Chem 2004;47:45–55.
8. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem 2006;49:534–553.
9. Apostolakis J, Pluckthun A, Caflisch A. Docking small ligands in flexible binding sites. J Comput Chem 1998;19:21–37.
10. Mangoni M, Roccatano D, Di Nola A. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. Proteins 1999;35:153–162.
11. Jiang F, Kim SH. Soft docking: matching of molecular surface cubes. J Mol Biol 1991;219:79–102.
12. Leach AR. Ligand docking to proteins with discrete side-chain flexibility. J Mol Biol 1994;235:345–356.
13. Abagyan R, Totrov M, Kuznetsov D. ICM—a new method for protein modeling and design–applications to docking and structure prediction from the distorted native conformation. J Comput Chem 1994;15:488–506.
14. Desmet J, Wilson IA, Joniau M, De Maeyer M, Lasters I. Computation of the binding of fully flexible peptides to proteins with flexible side-chains. FASEB J 1997;11:164–172.
15. Schaffer L, Verkhivker GM. Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side-chain optimization. Proteins 1998;33:295–310.
16. Frimurer TM, Peters GH, Iversen LF, Andersen HS, Moller NP, Olsen OH. Ligand-induced conformational changes: improved predictions of ligand binding conformations and affinities. Biophys J 2003;84:2273–2281.
17. Sandak B, Nussinov R, Wolfson HJ. A method for biomolecular structural recognition and docking allowing conformational flexibility. J Comput Biol 1998;5:631–654.
18. Sandak B, Wolfson HJ, Nussinov R. Flexible docking allowing induced fit in proteins: insights from an open to closed conformational isomers. Proteins 1998;32:159–174.
19. Wei BQ, Weaver LH, Ferrari AM, Matthews BW, Shoichet BK. Testing a flexible-receptor docking algorithm in a model binding site. J Mol Biol 2004;337:1161–1182.
20. Knegtel RM, Kuntz ID, Oshiro CM. Molecular docking to ensembles of protein structures. J Mol Biol 1997;266:424–440.
21. Meng EC, Shoichet BK, Kuntz ID. Automated docking with grid-based energy approach to macromolecule-ligand interactions. J Comput Chem 1992;13:505–524.
22. Osterberg F, Morris GM, Sanner MF, Olson AJ, Goodsell DS. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins 2002;46:34–40.
23. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395.
24. Ferrari AM, Wei BQ, Costantino L, Shoichet BK. Soft docking and multiple receptor conformations in virtual screening. J Med Chem 2004;47:5076–5084.
25. Broughton HB. A method for including protein flexibility in protein-ligand docking: improving tools for database mining and virtual screening. J Mol Graph Model 2000;18:247–257.
26. Carlson HA, Masukawa KM, Rubins K, Bushman FD, Jorgensen WL, Lins RD, Briggs JM, McCammon JA. Developing a dynamic pharmacophore model for HIV-1 integrase. J Med Chem 2000;43:2100–2114.
27. Meagher KL, Carlson HA. Incorporating protein flexibility in structure-based drug discovery: using HIV-1 protease as test case. J Am Chem Soc 2004;126:13276–13281.
28. Lin JH, Perryman AL, Schames JR, McCammon JA. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. J Am Chem Soc 2002;124:5632–5633.
29. Lin JH, Perryman AL, Schames JR, McCammon JA. The relaxed complex method: accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers 2003;68:47–62.
30. Zavodszky MI, Lei M, Thorpe MF, Day AR, Kuhn LA. Modeling correlated main-chain motions in proteins for flexible molecular recognition. Proteins 2004;57:243–261.
31. Cavasotto CN, Abagyan RA. Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol 2004;337:209–225.
32. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc 2005;127:9632–9640.
33. 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.
34. Ewing TJA, Makino S, Skillman AG, Kuntz ID. DOCK 4.0: search strategies for automated molecular docking of flexible molecule database. J Comput Aided Mol Des 2001;15:411–428.
35. Nelder JA, Mead R. A simplex method for function minimization. Comput J 1965;7:308–313.
36. Huang S-Y, Zou X. An iterative knowledge-based scoring function to predict protein-ligand interactions: I. Derivation of interaction potentials. J Comput Chem 2006;27:1866–1875.
37. Huang S-Y, Zou X. An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. J Comput Chem 2006;27:1876–1882.
38. Ferro DR, Hermans J. A different best rigid-body molecular fit routine. Acta Cryst A 1977;33:345–347.
39. Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE. A geometric approach to macromolecule-ligand interactions. J Mol Biol 1982;161:269–288.
40. OMEGA, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, 2001. Available at http://www.eyesopen.com/.
41. Ewing TJA, Kuntz ID. Critical evaluation of search algorithms for automated molecular docking and database screening. J Comput Chem 1997;18:1175–1189.
42. Rusinko A, III, Sheridan RP, Nilakantan R, Nilakantan KS, Bauman N, Venkataraghavan R. Using CONCORD to construct a large database of three-dimensional coordinates from connection tables. J Chem Inf Comput Sci 1989;29:251–255.
43. Cole JC, Murray CW, Nissink JWM, Taylor RD, Taylor R. Comparing protein-ligand docking programs is difficult. Proteins 2005;60:325–332.
44. Charifson P, Corkery J, Murcko M, 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.
45. McGovern SL, Shoichet BK. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. J Med Chem 2003;46:2895–2907.
46. Muegge I, Martin YC, Hajduk PJ, Fesik SW. Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. J Med Chem 1999;42:2498–2503.
47. Barril X, Morley SD. Unveiling the full potential of flexibile receptor docking using multiple crystallographic structures. J Med Chem 2005;48:4432–4443.