WILEY InterScience® DISCOVER SOMETHING GREAT

PROTEINS STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

FLIPDock: Docking flexible ligands into flexible receptors

Yong Zhao and Michel F. Sanner*

Department of Molecular Biology, TPC26, The Scripps Research Institute, La Jolla, CA 92037-1000

ABSTRACT

Conformational changes of biological macromolecules when binding with ligands have long been observed and remain a challenge for automated docking methods. Here we present a novel protein–ligand docking software called FLIPDock (Flexible LIgand–Protein Docking) allowing the automated docking of flexible ligand molecules into active sites of flexible receptor molecules. In FLIPDock, conformational spaces of molecules are encoded using a data structure that we have developed recently called the* Flexibility Tree *(FT). While the FT can represent fully flexible ligands, it was initially designed as a hierarchical and multiresolution data structure for the selective encoding of conformational subspaces of large biological macromolecules. These conformational subspaces can be built to span a range of conformations important for the biological activity of a protein. A variety of motions can be combined, ranging from domains moving as rigid bodies or backbone atoms undergoing normal mode-based deformations, to side chains assuming rotameric conformations. In addition, these conformational subspaces are parameterized by a small number of variables which can be searched during the docking process, thus effectively modeling the conformational changes in a flexible receptor. FLIPDock searches the variables using genetic algorithm-based search techniques and evaluates putative docking complexes with a scoring function based on the AutoDock3.05 force-field.

In this paper, we describe the concepts behind FLIPDock and the overall architecture of the program. We demonstrate FLIPDock's ability to solve docking problems in which the assumption of a rigid receptor previously prevented the successful docking of known ligands. In particular, we repeat an earlier cross docking experiment and demonstrate an increased success rate of 93.5%, compared to original 72% success rate achieved by AutoDock over the 400 cross-docking calculations. We also demonstrate FLIPDock's ability to handle conformational changes involving backbone motion by docking balanol to an adenosine-binding pocket of protein kinase A.

Proteins 2007; 68:726–737.
© 2007 Wiley-Liss, Inc.

Key words: docking; induced fit; protein flexibility; rotamer library; HIV-1 protease; protein kinase A.

INTRODUCTION

Large biological macromolecules are known to undergo conformational changes when binding ligand molecules. While most modern automated docking procedures treat the ligand molecules as flexible, the receptor is often held rigid because the representation of its flexibility in a brute force method (i.e., allowing each atom to move) is computationally expensive. The incorporation of the receptor flexibility in the automated docking process still remains one of the main challenges for the computational prediction of ligand–receptor interactions and protein–protein interactions.[1,2]

One technique for addressing this problem is to dock a ligand against an ensemble of receptor conformations,[3] obtained from either experimental sources or by computational means.[4–8] Although some encouraging results have been reported, these approaches leave the choice of the receptor conformations, as well as their number, to the user. Molecular dynamics (MD) trajectories have been used to generate receptor conformations for docking.[9,10] While the receptor is fully flexible during such MD simulations, substantial backbone rearrangement such as domain motions usually occur at time scales that are beyond today's typical MD simulations capabilities. Affinity grid maps, providing a discrete representation of the active site, have been calculated from various receptor conformations and combined into a hybrid map to represent a flexible receptor.[11] This approach also requires picking certain conformations of the receptor. Claudio and Abagyan[12] used Monte Carlo (MC)-based optimization to generate a discrete set of receptor conformations and performed flexible ligand–rigid receptor docking. They sampled the receptor conformational spaces and captured several loop motions. Normal mode analysis has also been used to generate discrete receptor conformations for flexible ligand, rigid receptor docking.[13]

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

Grant sponsor: NIH; Grant numbers: BISTI, GM65609, 2003-2007, NIH RR08605.
*Correspondence to: Michel F. Sanner, Department of Molecular Biology, TPC26, The Scripps Research Institute, 10550 North Torrey Pine Rd., La Jolla, CA 92037-1000.
E-mail: sanner@scripps.edu
Received 29 November 2006; Revised 3 January 2007; Accepted 17 January 2007
Published online 23 May 2007 in Wiley InterScience (www.interscience.wiley.com).
DOI: 10.1002/prot.21423

© 2007 WILEY-LISS, INC.

Various ways of softening the van der Waals (vdw) potentials have also been explored. The repulsive terms created by close contacts can be made less penalizing, the vdw radii can be manipulated over the course of the simulation,[14,15] or entire side chains can be removed.[16] These approaches essentially enlarge the docking pocket, and therefore tend to increase the number of false positive predictions. Usually these unphysical modifications are corrected in a post-docking step which is performed on a relative small number of promising poses.

Several groups have combined soft potentials with structural refinement in order to predict key interactions between the ligand and the receptor. Sherman et al.[14] first docked the ligands into rigid receptor using a softened energy function and then minimized the protein–ligand complex. The refined protein structure was then docked using a hard potential function. This two-step approach captured all the key interactions between 21 ligand–receptor complexes. A similar approach is reported by Mizutani et al.[15] In this work the protein cavity is first enlarged by offsetting vdw radii and the structure is optimized in the subsequent step. Approaches that combine docking and molecular dynamics were summarized in a recent review.[17]

In the approaches described above, the conformational space of the receptor is not explored during the docking procedure, but rather during a pre- or post-docking stage. This is largely due to the fact that using a brute-force representation of a flexible receptor involves too many degrees of freedoms. We have recently developed the *Flexibility Tree* (FT)[18] data structure which encodes conformational subspace of biological molecules using a small number of variables, thus dramatically reducing the computational costs of modeling flexible macromolecules. The FT data structure allows focusing on motions that modify the active site, which are more relevant to the binding of a given ligand. The hierarchical and multiresolution nature of the FT enables the conformational change of a receptor to be parameterized using a small number of variables which can be searched concurrently with the conformation, position, and orientation of a ligand molecule.

In this paper, we present the docking program FLIPDock (Flexible LIgand–Protein Docking). We briefly describe the FT data structure and illustrate how the problem of docking flexible molecules can be represented using a pair of FTs. We demonstrate FLIPDock's ability to predict the correct binding modes for molecular complexes in which rigid–receptor docking methods are known to fail.

METHODS

The architecture of FLIPDock

The three main ingredients of a docking program are: the representation of the ligand–receptor complex, the search method, and the method for ranking putative solutions.[19] In FLIPDock we use FT data structures[18] to represent the conformational spaces of both the ligand and the receptor. The solution space (conformations of the molecules, relative position and orientation of the ligand relative to the receptor) is searched using a genetic algorithm (GA) combined with a divide and conquer approach. The putative dockings are evaluated using an atomic pairwise scoring function based on the AutoDock 3.05 force-field.[20]

The FT was designed to encode relatively small, yet complex subspaces of a protein's conformational space by combining and nesting a variety of motion descriptors.[18] A FT is built by recursively partitioning a molecule or molecular system into molecular fragments moving relative to each other. Interdomain motion descriptors (hinge, shear, twist, screw, etc.) and intradomain motion descriptors (rotameric side chains, normal modes, essential dynamics, etc.) can be assigned to any molecular fragment. For example, Figure 1 shows a FT for a generic protein. Suppose this protein is composed of two domains that move relative to each other. The initial domain subdivision allows us to assign a "local perturbation" motion descriptor (e.g. the combination of a small translation and a constrained rotation)[18] to domain B. This motion descriptor jitters domain B relative to domain A. Both domains are further subdivided into *rigid* and *moving* regions, allowing the assignment of a hinge motion to the *moving* regions and normal modes perturbations to the *rigid* regions. Finally, critical side chains in the active site are split out of the *rigid* regions to allow them to adopt rotameric conformations.

Each motion descriptor has a set of variables (e.g., the angle of a hinge, the index of a rotameric side chain, the amplitude of a normal mode, etc.). The range of these variables and the hierarchical combination of the corresponding motion descriptors define the conformational subspace of the molecule encoded in the FT. These variables can be searched efficiently using standard global search techniques. Every motion descriptor in a FT provides a mechanism for randomizing its motion variables within their allowed range. The FT provides a hierarchical and multiresolution representation of molecular flexibility suitable for investigating, evaluating and analyzing motion information obtained from sources as diverse as: experiments, simulations, computations, or the biologist's intuition.

In Figure 2 we illustrate how a FT can be used to represent a flexible ligand molecule. The tree is defined by identifying rigid moving fragments starting from a fixed core fragment. These rigid fragments are connected to each other by single bonds; hence their relative motion can be represented by an "*axis rotation*" motion descriptor in which the axis is set to be the rotatable bond. This approach is equivalent to an internal-coordinates parameterization of the ligand conformation. It is noteworthy however, that the motion objects in the FT are generic enough to enable biasing the rotation to mimic a torsional potential for instance. This could be achieved by configur-

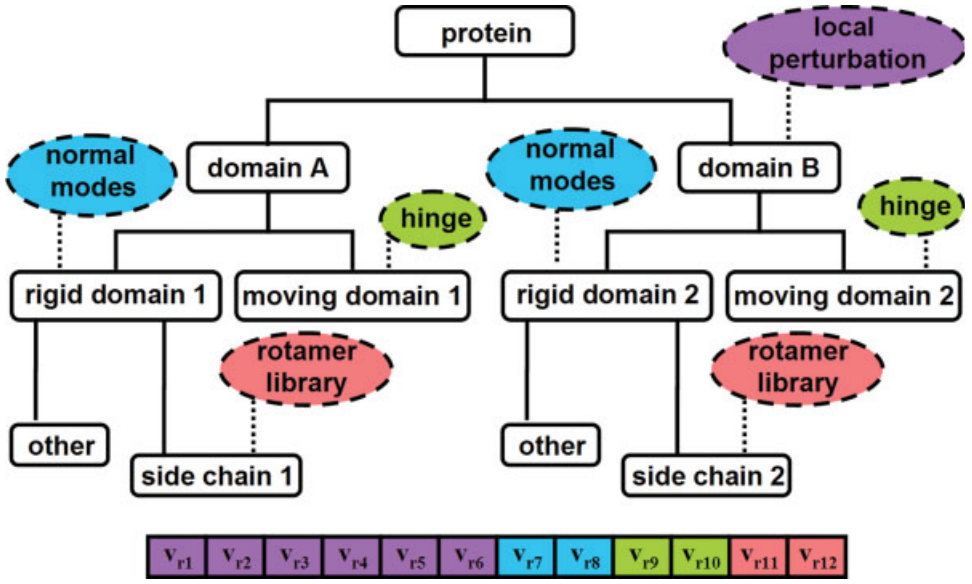

Figure 1

A FT data structure encoding some flexibility in a protein. FT nodes are depicted as boxes and motion descriptors are depicted in dashed ovals. Motion variables are listed at the bottom: six variables (v_{r1} v_{r6}) for the local perturbation, one variable for each: normal mode (v_{r7}, v_{r8}), hinge (v_{r9}, v_{r10}) and rotamer (v_{r11}, v_{r12}) motion objects. The main purpose here is to illustrate the variety of motion descriptors, their hierarchical nesting and their combinations. In docking studies the FT is constructed in a way to emphasize receptor flexibility that is known or assumed to be important for biological activity. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

ing the motion descriptor's randomization mechanism with a multimodal distribution rather than a uniform one.

The two FTs shown in Figures 1 and 2 encode the conformational spaces of a receptor and a ligand molecule, respectively. The conformation of the receptor can be described with 12 variables. The conformation of the ligand is described by four rotation–angle variables. To parameterize the problem of docking this flexible ligand

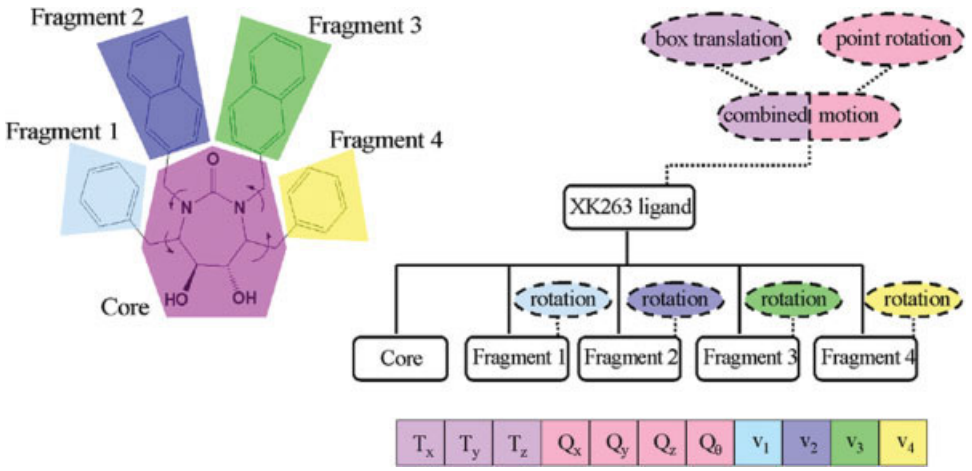

Figure 2

Representing a flexible ligand using a FT structure. If we allow four single bonds to be rotatable, the XK263 ligand is partitioned into five fragments. The four fragments move relative to the fixed core. The position of these four fragments relative to the core is described by four variables: v_1–v_4. By assigning the combined motion to the FT root node, the ligand can undergo free rotations encoded by the quaternion Q_x, Q_y, Q_z, Q_θ, and a constrained translation (T_x, T_y, T_z). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

DOI 10.1002/prot

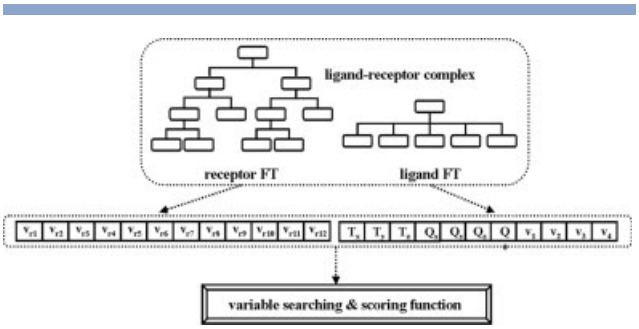

Figure 3

In FLIPDock the problem of docking a flexible ligand into a flexible receptor is encoded using two FT data structures. Randomizing the motions variables in the FTs provides random ligand–receptor conformations and relative positions and orientations. These putative dockings are evaluated using a scoring function. FLIPDock optimizes the motions variables to minimize the given scoring function.

into the flexible receptor, we need to encode the position and orientation of the ligand relative to the receptor. This is achieved by assigning a combination of a "*box translation*" and a "*point rotation*" to the root node of the ligand's FT (Fig. 2). The "*box translation*" will be configured to cover the active site of the receptor and will constrain the center of the ligand molecule. This combined rotation and translation of the ligand adds another seven variables (3 for the translation and 4 for the quaternion representing the rotation) for a total of 23 variables which define the space to be searched. Randomizing this set of variables yields a putative docking, i.e. a random conformation of both molecules (within the conformational spaces encoded by the FTs) and a random position and orientation of the ligand molecule relative to the receptor (Fig. 3).

Search engines and scoring functions

Automated docking is an optimization process in which we search the space of putative dockings and evaluate the solutions according to a scoring function. FLIPDock is designed as a component-based framework. New search techniques and scoring functions can be added to the framework, thus facilitating the interchange and mixing of various search techniques and scoring functions. Currently, FLIPDock provides a simple GA search engine as well as a two step "divide and conquer" genetic algorithm (DAC-GA). During a simple GA search, the center of a ligand is constrained to a 3-dimensional subspace called the *docking box*. In a DAC-GA search the docking box is partitioned into smaller boxes, typically 3 Å cubes and a short GA optimization is carried out in each one of these small boxes. In a second step, a GA optimization is carried out over the original docking box using a population containing the best individuals (motion variable sets) from the short optimizations carried out on the small boxes. A comparison between the simple GA and DAC-GA is given in the "Results and Discussions" section and demonstrates the advantage of DAC-GA over simple GA.

Scoring functions are also components in the FLIPDock framework. The simplest scoring function currently available in FLIPDock is the RMSD score, which measures how close the putative ligand–receptor complex is from a given reference complex conformation. While this scoring function is not useful for real docking problems (since it requires knowledge of the answer and the energy landscape is too simple) it is helpful for tasks such as debugging. More importantly, the existence of two interchangeable scoring functions allows us to demonstrate and verify the ability to easily interchange scoring functions in the FLIPDock framework.

The scoring function used for the docking experiments reported here is based on the AutoDock v3.05 force-field.[20] This empirical force-field takes into consideration four nonbonded interactions: dispersion/repulsion, hydrogen bonding, electrostatics and desolvation upon binding. In AutoDock, grid maps are used to speed up the numerical evaluation of the interaction energy between the ligand and the receptor. Such grid maps cannot be precomputed in FLIPDock since the receptor changes conformation during the docking. Instead, the atomic pairwise interactions are evaluated based on the AutoDock force-field. The FLIPDock score is the sum of the internal energy (IE) of the ligand, the internal energy of the receptor and the interaction energy (E) between ligand and receptor [Eq. (1)]. If only part of the receptor is flexible, the internal energy of the receptor can be obtained by adding the internal energy of the rigid part (a constant), the internal energy of the flexible part, and the interaction energies between the rigid and the moving part of the receptor. Since the conformational changes in ligand are limited to modifying torsion angles, the bond length and bond angles remain unchanged. Hence the 1–2 (covalent bond) and 1–3 pairwise interactions are not computed during the docking calculation. The inclusion of 1–4 interactions in the score is optional. Nonbonded interactions between pairs of atoms that are further than a user-specified distance cutoff (default value: 8.0 Å) are not evaluated. Note that a FLIPDock score is a sum of several empirical energetic terms used to guide the variable optimization and does not correlate with binding energy or other experimental data (Fig. 4).

$$\text{Score} = E_{\text{receptor}-\text{ligand}} + \text{Internal Energy}_{\text{ligand}} + \text{Internal Energy}_{\text{receptor}} \quad (1)$$

$$\text{Score} = \text{IE}_{\text{ligand}} + \text{IE}_{\text{flexible receptor}} + \text{IE}_{\text{rigid receptor core}} + E_{\text{ligand}-\text{core}} + E_{\text{ligand}-\text{flexible}} + E_{\text{core}-\text{flexible}} \quad (2)$$

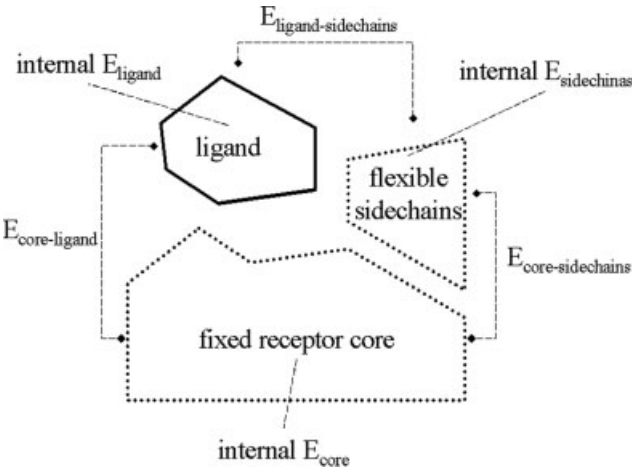

Figure 4

The FLIPDock score is composed of empirical interaction energy (E) and internal energy (IE) terms. A receptor and a ligand are depicted as dashed and solid blocks respectively. When a part of the receptor is rigid the internal energy of the rigid part of the receptor is constant and omitted from the score.

Implementation

The FLIPDock program is implemented using Python programming language[21] (http://www.python.org). The scoring function is the most computation intensive module. It is written in C++ and wrapped to be callable from Python. We have integrated FLIPDock with our molecular viewer PMV[22] and the visual programming environment Vision[23] to provide a graphical interface for constructing FTs, setting up docking experiments and graphically analyzing the results.

RESULTS

In this section, we demonstrate using FLIPDock for docking problems in which conformational changes in the receptor are known to lead automated docking procedures using a rigid representation of the receptor to fail. First we revisit a cross docking experiment for HIV protease inhibitors in which side chains in the active site need to change conformation to accommodate various ligands. Next, we demonstrate FLIPDock's ability to handle backbone motion in the receptor by docking balanol to an adenosine bound protein kinase A (PKA) conformation. Finally, we compare the performance of the two search methods currently available in FLIPDock and show that the DAC-GA outperforms the simple GA.

Cross-docking 20 HIV protease inhibitors

Here we revisit a cross docking study carried out earlier by Österberg et al.[11] in which they used 21 crystallographic structures of HIV protease (HIV-PR) with different ligands bound. The cross docking experiment consists in docking each ligand against each of the 21 conformations of the HIV-PR receptor. The 1HVR complex was excluded from our cross docking test as the inhibitor displaces a water molecule present in 20 other complexes. Cross docking with AutoDock, without using averaged grid map representations, was shown to fail (i.e. ligand RMSD >2.0 Å) in 112 (28%) of the 400 dockings[11,24] because of structural variance in the receptor induced by the different inhibitors, especially in the conformations of the two arginine 8 (ARG8).

We first repeated Österberg's cross docking experiment and using a HIV-PR model in which the side chains of ARG8 were made flexible. In Figure 5, we show an example of FTs for one of these 400 dockings. These FTs were built for docking the ligand from the 1HBV complex into the receptor conformation taken from the 1HPS complex. The rotatable bonds in the ligand dictate the partitioning of the ligand into fragments used to build the FT. Note that, while fragments 1,3 and 4 move relative to the core, fragment 2 moves relative to fragment 1. This is easily specified in the FT by making fragment 1 the reference node for fragment 2 in the FT.

The FT for the HIV-PR receptor was constructed to allow the side chains of residues ARG8 from both chains to choose alternative rotameric conformations obtained from Dunbrack's backbone independent rotamer library.[25] Rotamers that overlap with the rigid receptor

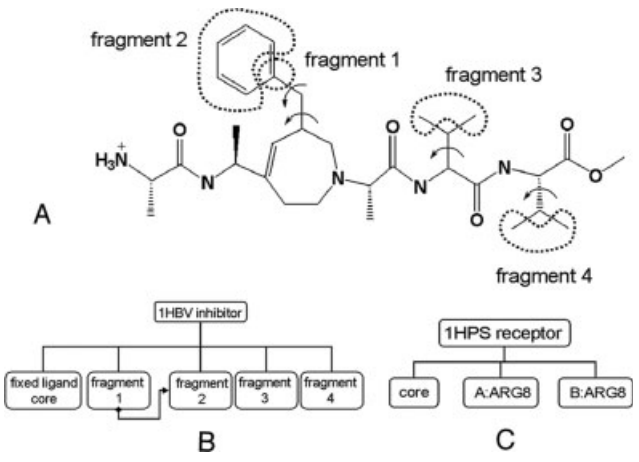

Figure 5

*FTs used to represent a cross-docking problem. Motion descriptors are not shown. (**A**) We allow four single bonds in the SB203238 ligand (inhibitor in 1HBV) to be freely rotatable. The molecular fragments moved by the rotations are marked in dashed regions. (**B**) The FT data structure for 1HBV inhibitor. Note that the fragment 2 moves relative to fragment 1 while other fragments move relative to the fixed ligand core. We set fragment 1 as the reference of fragment 2 to encode this torsion nesting. All the fragments undergo "axis rotation" motion. (**C**) The FT data structure for 1HPS receptor.*

DOI 10.1002/prot

Table I

Only the Two ARG8 Sidechains are Allowed to Adopt Alternative Conformations. The Ligand Positions are Reproduced in 96.25% of 400 Cross Docking Cases. The Only Case with RMSD > 4.0 Å (receptor: 1HVL, ligand 1HTG) Occurred because Motion Variables are Not Well Optimized by the Simple GA

Proteases \ Ligands	1HBV	1HEF	1HEG	1HIH	1HIV	1HPS	1HTE	1HTF	1HTG	1HVI	1HVJ	1HVK	1HVL	1HVS	1SBG	4HVP	4PHV	5HVP	8HVP	9HVP
1HBV	0.90	0.61	1.05	1.21	1.63	1.12	0.40	1.24	2.34	1.21	1.13	1.23	1.17	0.53	0.71	1.54	1.04	0.94	1.80	0.57
1HEF	1.60	0.87	1.68	0.62	0.55	0.96	0.50	0.93	1.21	0.86	1.49	0.92	0.79	0.42	0.67	1.39	0.64	1.02	1.45	0.39
1HEG	1.51	0.98	1.36	0.68	0.86	1.13	0.26	1.29	0.68	0.93	1.14	1.06	0.93	0.71	0.72	3.34	0.89	0.71	1.43	0.80
1HIH	1.08	1.32	1.43	1.07	1.32	0.88	0.81	0.58	0.85	0.89	1.12	0.92	0.70	0.68	0.88	1.40	1.11	1.09	1.89	0.92
1HIV	1.44	1.11	1.04	0.83	0.61	0.72	0.71	0.78	1.09	1.32	0.70	1.03	1.13	0.59	1.21	1.48	0.63	0.91	1.70	0.56
1HPS	1.60	1.22	1.44	0.97	0.68	0.89	0.70	0.81	0.93	1.08	1.12	1.33	1.15	0.34	1.12	3.14	0.46	0.79	1.48	1.13
1HTE	1.87	1.72	1.41	0.81	0.96	1.11	0.50	1.32	0.73	0.95	0.95	1.24	1.23	0.76	0.86	3.19	1.06	1.14	1.85	0.55
1HTF	1.64	1.01	1.38	1.27	1.49	0.55	0.13	0.70	2.62	1.63	1.84	0.94	1.35	0.92	0.51	1.22	0.55	0.99	1.27	0.61
1HTG	1.07	1.03	1.46	0.96	0.55	0.44	0.52	0.84	0.52	0.83	1.46	0.75	0.75	0.55	0.33	1.62	0.51	0.94	1.43	0.64
1HVI	1.61	1.28	0.56	1.23	1.63	1.36	0.59	1.18	0.82	0.75	1.14	1.01	0.69	0.99	0.75	2.67	0.74	0.97	1.78	1.13
1HVJ	1.68	1.21	1.39	1.13	1.49	1.44	1.42	0.56	1.30	0.78	0.96	0.47	1.09	0.79	1.13	1.45	1.04	1.07	1.61	0.71
1HVK	1.52	1.57	1.33	0.90	1.52	1.24	0.61	0.59	0.84	1.13	0.30	0.73	0.65	0.64	0.80	1.50	0.91	1.31	1.86	1.26
1HVL	1.88	1.18	1.46	1.29	1.54	1.59	0.48	0.45	6.33	0.78	0.57	0.77	1.08	0.27	1.08	1.55	0.59	1.18	2.12	1.22
1HVS	1.72	1.20	1.56	1.07	1.28	1.16	0.43	1.22	2.71	0.90	0.82	1.10	0.70	0.38	1.27	1.48	0.84	1.27	1.63	0.78
1SBG	0.91	0.83	1.33	0.53	1.52	0.38	0.66	0.82	2.60	0.79	1.81	0.87	1.31	0.82	0.38	1.46	0.95	1.29	1.90	0.42
4HVP	1.93	1.40	1.23	1.19	0.72	1.55	0.69	0.91	2.63	0.31	1.18	1.08	0.40	0.75	1.22	3.39	0.78	0.75	1.54	0.76
4PHV	1.01	1.34	1.62	0.98	1.13	0.76	0.36	0.75	1.05	1.23	1.21	1.11	1.16	0.57	0.58	1.50	0.78	0.85	1.57	0.51
5HVP	0.57	1.29	1.70	1.00	1.03	0.61	0.49	0.52	0.61	1.16	1.08	1.04	0.75	1.02	0.78	3.26	0.81	0.95	1.89	0.65
8HVP	2.16	1.27	0.81	0.81	0.77	0.98	0.77	0.97	2.72	1.02	0.94	0.87	0.64	0.83	0.75	1.40	0.56	0.98	1.74	0.75
9HVP	1.12	1.78	1.07	0.62	1.42	0.99	0.48	1.08	0.76	0.95	1.43	1.51	1.11	0.77	0.86	1.22	1.03	0.87	1.56	0.50

< 2 Å	2Å ~ 4 Å	> 4 Å

core were removed from the searching space in a preprocessing step.

We used the same energy potential (i.e., the AutoDock v3.05 force field) and the same set of torsions in the ligands as in Österberg et al.[11] We also limited the number of energy evaluations to be less than 10^6. The Larmarckian GA search engine used in the reference study is more sophisticated than the DAC-GA we used here, as DAC-GA does not perform local searches. Since the goal was to test whether two flexible ARG8 side chains are sufficient for enabling cross dockings, we used relative small cubic docking boxes (6 Å on a side and centered at the known docking site) in order to reduce the computational cost. The docking boxes were further partitioned into eight cubes (3 Å on a side) for the DAC-GA. A simple GA searched each small box using the following parameters: 50 generations, population size = 10 times number of motion variables, replacement rate = 80%, mutation rate = 30%, crossover probability = 50%. In the second step of DAC-GA, the initial populations contained the best individuals from each small box and randomized individuals. The GA search was carried out with the following parameters: population size = 10 times number of motion variables, replacement rate = 10%, mutation rate = 3%, crossover probability = 50% and up to 1000 generations. The GA was allowed to terminated early if the search converged (i.e., the deviation of the scores in the last generation is less than 10^{-4}).

We carried out five DAC-GA runs for each cross docking and use the top scored pose as the prediction. FLIP-Dock successfully predicted (i.e. ligand RMSD < 2.0 Å) 385 (96.25%) of the 400 dockings (Table I).

Cross docking the ligand from 1HTG into the receptor taken from 1HVL failed with an RMSD of 6.33 Å. After 10 additional independent docking calculations for this pair, the best RMSD achieved was 1.14 Å. This clearly indicates that the failure is due to the search engine which was not able to optimize the variables in the first 5 runs.

The 5 additional cases in which the RMSD is larger than 3.0 Å all involved the 4HVP ligand, the most flexible one in this cross docking experiment (18 torsions). One of these cases (receptor: 4HVP, ligand: 4HVP) is shown in Figure 6. The ligand position and orientation are well predicted; however, the ARG6 side chain in the ligand

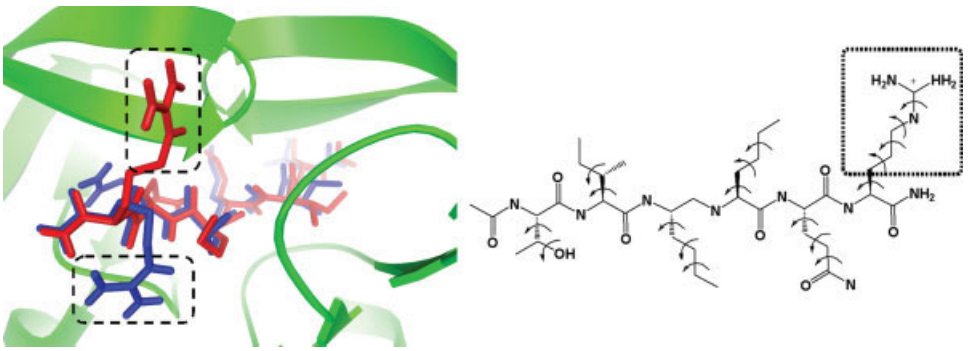

Figure 6

Redocking the 4HVP inhibitor. The crystal structure of the 4HVP inhibitor is shown in blue while the FLIPDock prediction is depicted in red. The ARG6 side-chain is modeled with five nested rotatable bonds (in dashed square). FLIPDock placed the ARG6 side chain close to the "flap" of the receptor, while experimental data shows an alternative position. The overall RMSD of ligand atoms is 3.65 Å.

assumes an alternative position, which offsets the RMSD to higher values (between 3 and 4 Å). These scenarios were found in all the five failed cases with RMSD $>$ 3.0 Å. This problem was also observed by Österberg et al.[11] These alternate conformations in which the ligand's ARG6 side chain moves toward the flap of HIV-PR are favored by the AutoDock force field (i.e., ligand–receptor interaction energy is −16.46 kcal/mol while the same term is −15.15 kcal/mol using the crystal structure).

This experiment demonstrated that modeling the flexibility of the ARG8 in the two chains using rotamers enables the successful crossdocking of most of the 20 HIV-PR inhibitors. However, for this particular problem, the number and the identity of side chains that need to be flexible in the cross docking experiments (ARG8 in both chains) were known before hand. In most docking problems this information is not known *a priori*. The prediction of flexible side-chains using a single conformation is likely to lead to a larger number of side-chains that ought to be made flexible in the receptor's active site. To emulate such a scenario and to demonstrate FLIPDock's ability to cope with numerous flexible side-chains, we applied the procedure outlined below to blindly identify flexible side chains from a single structure of HIV-PR and repeated the cross-docking experiment with a model of HIV-PR in which all side chains identified by the procedure were made flexible.

To identify flexible side chains in HIV-PR's active site, we used a single conformation of HIV-PR (1HVR) in which the ligand has C2 symmetry. This choice was made with the expectation that the amino acids identified to be the most flexible would be the same in both chains. First we removed the ligand and minimized the structure using AMBER.[26] The "dist" program in the CONCOORD suite[27] was then used to identify all interatomic interactions in the structure and classify them according to the strength of the interaction. Covalent bonds form the tightest interactions and long-range nonbonded interactions are among the weakest. Based on the interaction strength, a specific geometric freedom is assigned by "dist" to each pair, i.e. covalent bonds have less freedom (e.g. 0.03 Å) while the nonbonded atoms pairs have larger freedom (e.g. 8.0 Å). We then defined the geometric constraint of an atom as the sum of the inverse of the geometric freedom from all pairs involving this atom. The geometric constraint of a residue was calculated as the average of the geometric constraints of the atoms in the residue. The residues within 6 Å of the XK263 ligand are shown in Figure 7, colored by geometric constraints of residues. Residues with the lowest geometric constraints are predicted to be the most flexible. Given the C2 symmetry of the HIV-PR

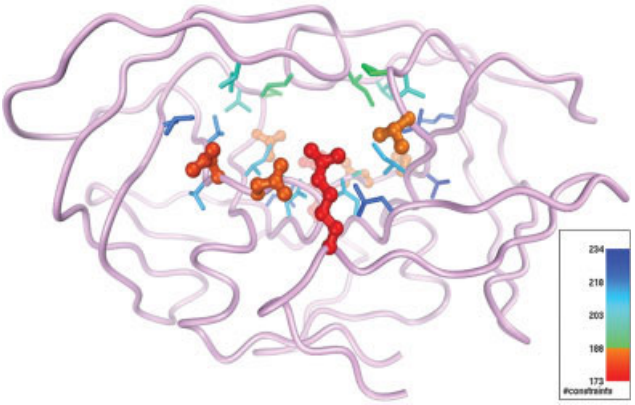

Figure 7

Side chains near the HIV-PR active site are colored by the geometric constraints.

DOI 10.1002/prot

Table II

RMSD of Ligand Atoms. FLIPDock Docks 20 HIV_PR Inhibitors into 20 Receptor Conformations, Allowing 8 Sidechains (ARG8, ASP29, ASP30 and VAL82 from Both Chains) to be Flexible During Docking Process. The Cross Docking Success Rate is 93.5% (RMSD < 2.0 Å)

Proteases \ Ligands	1HBV	1HEF	1HEG	1HIH	1HIV	1HPS	1HTE	1HTF	1HTG	1HVI	1HVJ	1HVK	1HVL	1HVS	1SBG	4HVP	4PHV	5HVP	8HVP	9HVP
1HBV	1.08	0.98	0.63	0.81	1.42	0.64	0.62	6.83	0.83	0.92	1.29	0.89	0.90	0.79	0.81	1.87	0.96	1.08	1.47	0.55
1HEF	1.68	0.89	1.55	0.90	0.61	0.93	0.78	0.67	0.50	0.77	0.89	1.14	0.81	0.49	0.59	1.53	1.20	1.04	1.61	0.83
1HEG	1.18	0.80	0.82	1.12	0.54	1.02	0.64	0.52	0.87	1.28	1.21	0.89	1.21	0.82	0.81	3.00	0.59	0.98	1.57	0.42
1HIH	0.90	0.70	1.90	0.99	1.31	0.96	0.58	0.58	0.51	0.93	0.94	0.94	1.00	0.52	0.79	3.82	0.81	1.08	2.10	1.03
1HIV	1.52	2.01	1.22	1.20	0.64	0.89	1.24	0.87	0.68	1.26	0.96	0.76	1.31	0.78	1.23	3.35	0.53	1.10	1.97	1.30
1HPS	1.63	1.30	2.16	1.13	0.67	0.43	0.51	0.57	0.41	1.06	1.26	1.17	1.09	0.79	0.68	1.28	0.86	0.82	1.50	0.76
1HTE	2.54	1.85	1.10	1.21	1.42	0.67	0.49	1.15	0.41	0.85	1.25	0.88	0.94	0.87	1.17	3.25	1.20	1.11	2.00	0.90
1HTF	1.38	1.28	0.57	1.11	1.43	0.98	0.17	0.98	0.71	1.18	1.33	1.10	0.91	0.74	0.40	1.56	1.02	1.21	1.47	0.58
1HTG	1.44	0.78	0.86	0.49	0.73	0.55	0.57	1.12	0.52	0.76	1.14	0.86	1.13	0.76	0.80	1.60	0.36	0.99	1.30	0.60
1HVI	1.66	1.08	0.97	1.24	1.39	0.36	0.60	0.54	0.96	0.83	0.80	1.11	0.96	0.67	0.97	2.35	1.12	1.12	1.70	1.24
1HVJ	2.09	1.18	0.69	1.18	1.32	1.19	0.77	2.08	0.97	1.13	1.13	1.07	0.86	0.38	1.31	2.62	1.13	1.24	1.74	1.06
1HVK	2.49	1.31	1.55	0.88	0.28	0.43	0.64	0.63	0.49	0.70	0.97	0.94	0.66	0.64	1.10	3.21	0.60	1.24	2.05	1.31
1HVL	2.10	1.58	1.36	1.11	1.34	1.32	0.63	2.98	0.62	1.35	0.95	0.49	0.91	0.77	1.19	1.68	0.96	1.23	2.36	1.37
1HVS	1.55	1.14	1.58	1.17	1.27	1.29	0.52	1.01	0.67	0.76	1.20	0.73	0.84	0.82	1.19	1.73	0.98	1.26	1.80	0.84
1SBG	1.15	1.05	0.92	0.70	0.92	0.38	0.50	1.08	0.41	0.92	0.78	1.04	1.10	1.00	0.64	1.60	0.85	1.19	1.50	0.58
4HVP	1.55	1.23	1.43	0.98	0.75	0.74	0.82	1.82	2.49	0.82	1.23	0.72	1.06	0.77	1.30	3.65	0.75	0.76	1.60	0.93
4PHV	0.91	0.96	1.42	0.87	0.53	0.81	0.46	0.37	0.44	1.31	1.00	1.00	1.14	1.15	0.26	2.79	1.07	0.68	1.42	0.79
5HVP	2.22	0.91	1.53	1.18	0.87	0.80	0.71	6.05	0.27	1.16	1.38	0.78	0.87	0.90	0.75	1.61	0.83	0.83	1.93	0.83
8HVP	2.06	1.23	0.82	0.66	0.51	0.89	0.62	4.91	0.94	1.11	0.91	1.23	1.05	1.08	1.34	1.92	1.13	1.16	1.72	1.27
9HVP	0.90	1.67	0.96	1.21	1.39	0.83	0.39	0.87	0.60	1.16	1.13	1.09	1.04	0.79	0.74	1.28	0.78	0.75	1.49	0.79

< 2 Å	2Å ~ 4 Å	> 4 Å

homodimer, we expected symmetrical residues in the two chains would have similar geometric constraints. It turned out to be true for the top 8 residues: ARG8, ASP29, ASP30, and VAL82 from both chains. ARG8 were predicted to be the most flexible residues. The results of this method are consistent with previous experimental studies: Ala et al.[28] suggested that ASP29, ASP30, ARG8, among others, are key residues for molecular recognition of cyclic urea HIV-PR inhibitors; Tong and coworkers[29] also observed conformational changes of ASP29 and ASP30; Melnick et al.[30] identified a series of nonpeptide HIV-PR inhibitors. They observed the two VAL82 residues, together with THR80 and PRO81, are flexible upon ligand binding.

We repeated the cross-docking study making the eight side chains with the lowest geometric constraints flexible. We used a *docking box* of size 9 Å × 9 Å × 18 Å centered on the same point as Österberg's grid map (2.4, 7.1, −8.5). Each cross-docking test was repeated 10 times using the DAC-GA search engine. In order to make a fair comparison with the reference study, we used the same set of torsions as Österberg in the ligands and we use the same energy potential. However, our search engine (DAC-GA) is simpler than the one used in AutoDock and we did not perform clustering after the 10 runs but simply used the best scoring result as our prediction. The predicted ligand poses (the best scored result in 10 runs) are compared to experimental data, as shown in Table II.

FLIPDock was successful in 374 cases (93.5%) demonstrating FLIPDock's ability to handle multiple flexible side chains in the receptor. In the reference study, where rigid receptor representations were used, the success rate after 10 runs for each pair was 72%. Since we included more energetic terms in FLIPDock, as shown in equation (2), the final FLIPDock scores and the scores reported in the reference study are not directly comparable.

Note that when we allow more flexibility in the receptor, the success rate drops slightly (from 96.25% with only ARG8 flexible to 93.5% with four side chains flexible in each chain). This might be because the more pliable active sites allow the ligand molecules to bind in alternative modes. The 1HTF inhibitor failed to reproduce the poses found in the crystal structure when it was docked into three alternative HIV-PR conformations (RMSD > 4.0 Å). It is the smallest among all the 20

ligands. The scoring function used in this experiment cannot distinguish alternative binding modes from the one observed experimentally (the best score has an RMSD of 6.83 Å, the second best score is only 0.04 worse and has an RMSD of 1.31 Å).

Most experimentally determined ligand poses could be found among the top three predictions of the 10 runs (98% of the 400 cross-dockings). The only three cases for which all 10 FLIPDock runs failed (the ligand RMSD >2.0 Å) involved the 4HVP inhibitor. The ARG6 side chain of the ligand was found to be flipping up in these cases, similarly to what we had observed in the previous experiment with only ARG8 side chain flexible (Fig. 6).

Ideally, the RMSD of the eight moving side chains in the receptor should also be below 2 Å. However, the RMSD values obtained for the moving part of the receptor are currently not as good as the one for the ligand for a couple of reasons. First, the rotameric side chain closest to the crystal structure may have steric clashes with the ligand, forcing the side chain into the next best rotameric position. This leads to a rather large RMSD deviation for such a side chain. Second, the force field for the receptor's internal energy is different from the one used to solve the crystal structure and one should first minimize the crystal structure using the same force field before evaluating the RMSD values in the receptor.

Docking balanol into an adenosine bound pocket

Crystal structures of protein kinase A (PKA) binding balanol (PDB:1BX6) and adenosine (PDB: 1FMO, 1BKX) are available. Previous studies[10,12] have shown that balanol cannot be docked successfully into the PKA conformation taken from the adenosine-binding complexes due to steric clashes with the GLY-rich flap (GLY50-VAL57). Here we use FLIPDock to dock balanol to the adenosine-binding PKA conformation (1FMO). We first superimposed the backbone atoms of the two adenosine bound structures (1FMO and 1BKX), which revealed a hinge-like displacement of the GLY-rich flap [Fig. 8(A)]. The two nitrogen atoms from GLY50 and VAL57 were chosen to define the hinge axis. We then built a FT for the adenosine-binding PKA allowing the hinge motion of the GLY-rich flap [Fig. 8(B)]. In order to eliminate clashes between the moving flap and the rest of the kinase, we allowed the following three residues to choose alternative rotameric positions: ARG56 and PHE54 on the flap and the LYS78 near the flap. Note that the flexibility information was derived from the two adenosine-binding conformations. No a priori knowledge of balanol was assumed. This FT encodes several combined motions: the flap moves relative to the fixed region, while the two side chains move relative to the GLY-rich flap and the LYS78 side chain moves relative to the fixed core. The balanol structure is shown in Figure 8(C) with the rotatable bonds indicated by arrows.

We docked balanol 10 times against the flexible PKA model. The RMSD between best scoring ligand and balanol's position in the native structure (1BX6) was 1.40 Å [Fig. 8(D)]. This demonstrated that combining flexible side chains with a hinge motion is sufficient to describe the PKA flexibility needed for accommodating balanol. The LYS78, ARG56, and PHE54 side chains have taken alternative positions to avoid steric clashes. The predicted binding pose reproduces the hydrogen bonds to LYS72 and ASP184, as found in the native balanol bound complex (1BX6). It forms hydrogen bond with VAL123, instead of GLN121. Additional hydrogen bonds to PHE54, GLU91, and GLU170 are also observed in the FLIPDock prediction.

GA versus DAC-GA

FLIPDock currently provides two search engines: simple GA and DAC-GA. In order to compare these two search methods we docked the 20 ligands from the data set used by Öesterberg[11] to 9HVP. This particular receptor conformation was chosen because it is challenging: i.e., several inhibitors failed to dock against this conformation of HIV-PR in the reference study.[11] We used a FT for the receptor in which eight side chains were allowed to assume rotameric conformations and carried out 10 independent docking trials for each of the 20 ligands. The simple GA searches the 9 Å × 9 Å × 18 Å docking box. In the DAC-GA, the same box is divided into 54 smaller boxes (cubes of 3 Å on a side). In every small box, we allow the simple GA to use high mutation rate and high replacement rate (30% and 80% respectively) to explore the solution space more aggressively. Each of these simple GA calculations terminates after 50 generations. In a second step, the DAC-GA searches the original docking box (i.e. 9 Å × 9 Å × 18 Å), using an initial population containing the best prediction found in each of the smaller boxes. In this step we reduce the mutation rate (3%) and replacement rate (10%) and we allow the GA to evolve more generations. In order to make a fair comparison between the two search techniques, the same number of docking evaluations (10^6) were carried out for both GA schemes. Both GAs stop when searching converged (standard deviation of scores in the population is less than 10^{-4}), or maximum number of putative dockings (10^6) is reached.

In Figure 9 we compare the results produced by simple GA and DAC-GA. Figure 9(A) shows the RMSD of the ligand with the best score out of 10 separate runs. If we define success as a docking in which the RMSD of the ligand is below 2.0 Å, the simple GA search fails in four cases (1HEF, 1HTG, 1HVI, and 4PHV) while the DAC-

DOI 10.1002/prot

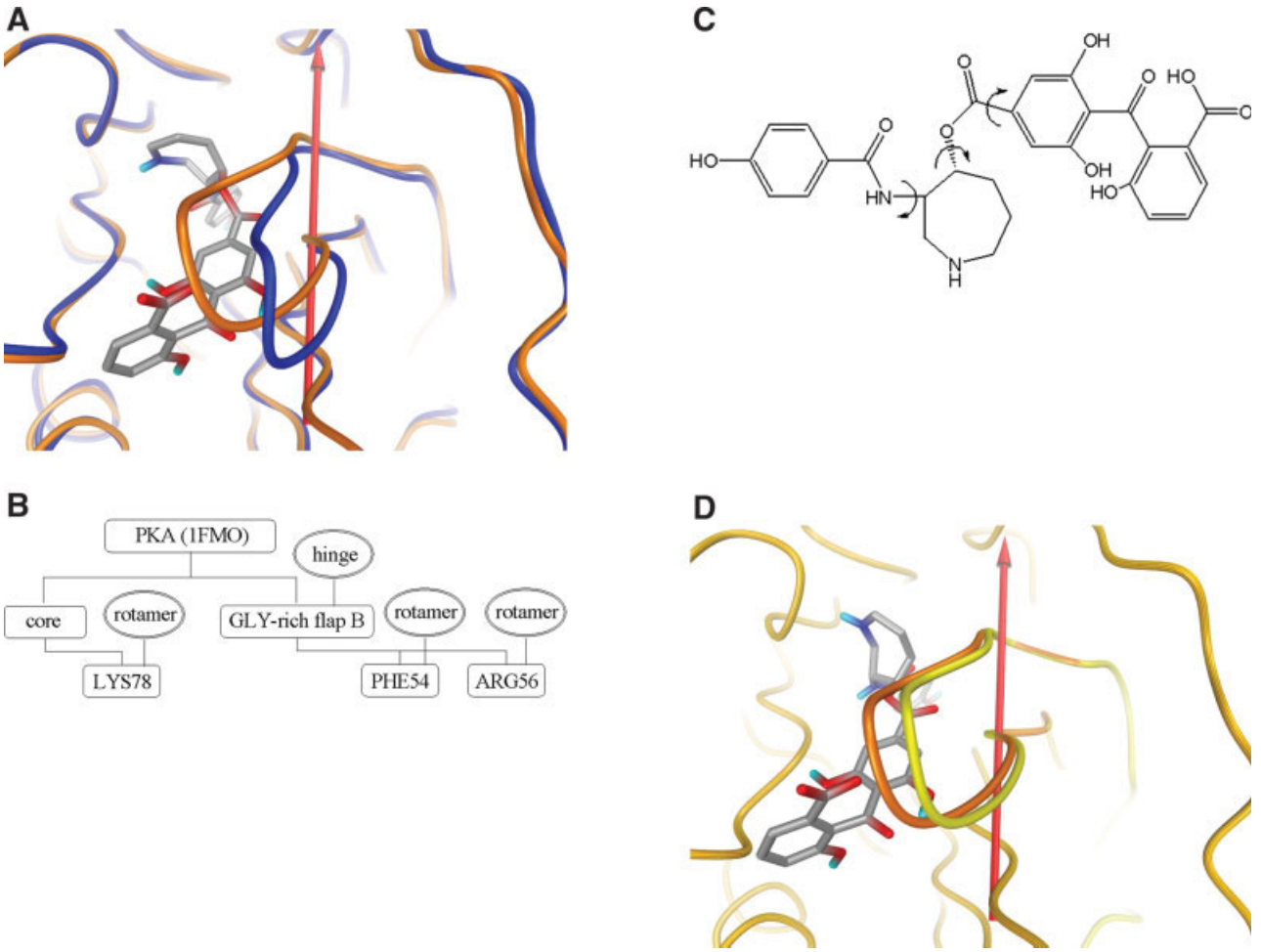

Figure 8

*(**A**) Two adenosine-bound PKA conformations (1FMO: orange and 1BKX: blue) are superimposed. A hinge-like displacement of GLY-rich loop is observed. After superimposing the balanol-bound PKA conformation (1BX6, structure not shown) to 1FMO, the ligand, depicted as tubes, has steric clash with the 1FMO conformation. (**B**) FT built for PKA. The GLY-rich flap in PKA is allowed to undergo a hinge motion. LYS78 and ARG56 are made flexible to eliminate potential clashes. PHE54 on the flap are also modeled by rotamer library since it could make contact with the ligand. (**C**) The balanol structure and the three rotatable bonds. (**D**) Combining a hinge motion (axis in red) with three flexible side chains enables docking balanol to the nonnative receptor conformation. The FLIPDock predicted PKA conformation is shown in yellow with the GLY-rich loop opening up to accommodate the binding ligand. The RMSD between the prediction and the native ligand position (as in 1BX6) is 1.4 Å.*

GA always succeeds. In Figure 9(B) we show the average RMSD over the 10 solutions obtained from independent docking calculations. The low average values for DAC-GA indicate that this method finds the right solution more consistently than the simple GA. A second comparison of the two GAs carried out for another receptor [i.e., 20 ligands against the 1HVI receptor (see data in supplemental materials)], showed similar results.

DISCUSSION AND CONCLUSION

FLIPDock predicts interactions between a flexible ligand and a flexible receptor. We have demonstrated FLIPDock's ability to dock flexible ligands to receptors requiring conformational changes, including backbone motion. Two FT data structures are used to describe a protein–ligand complex. The FT is able to describe fully flexible molecules such as ligands. It can also be used to selectively encode some flexibility of large biological molecule by nesting high-level motion descriptors such as hinges, normal modes, rotamers, etc. When relevant conformational changes in a macromolecule can be described by such high level motion descriptors, the small number of variables used to parameterize the receptor's flexibility makes it possible to explore the receptor's conformational space during the docking process, as opposed to methods depending on pre- or post-docking procedures.

A major advantage of FLIPDock is its versatility both in the description and encoding of docking problems using FTs and in the combination of search methods as well as

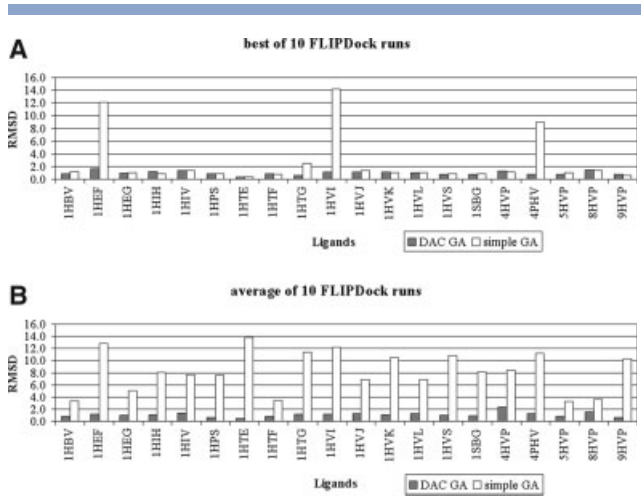

Figure 9

*(**A**) The RMSD of the docked ligand with the best score out of the 10 runs are shown for both simple GA and DAC GA. (**B**). The average RMSD over the 10 runs using the simple GAs and DAC-GA are shown.*

scoring functions. Representing the docking problem using FT data structures provides users with an unprecedented level of control in the description of the receptor's flexibility. For instance, a domain can undergo a simple hinge described by a single variable, or the user may chose to add a local perturbation to explore the neighborhood of the hinge at the expense of six additional variables. A side chain can be modeled using rotameric conformations or by treating each bond as explicitly rotatable. A flexible ligand prepared in AutoDock3.0 format (PDBQ) can be automatically converted into FT format.

The FLIPDock approach also allows the user to control the allocation of computational resources to the representation of particular motions. A FLIPDock calculation is an optimization problem in which the complexity depends on the number of variables to be optimized. Current GA-based searching methods can handle limited number of variables. It was shown[31] that AutoDock can dock the ligands with up to 12 torsion angles using Larmarckian GA (19 variables, including ligand rotation and translation). In this work, the number of variables in the HIV-PR cross-docking experiment ranges from 19 to 33. FLIPDock allows focusing on conformational changes that are directly relevant to the interaction with a ligand. More importantly, it enables the user to allocate degrees of freedom to be explored within the budget of the number of variables that can be realistically searched using the available resources. For instance, fixing the translation of a ligand in FLIPDock is as simple as replacing the combined translation/rotation motion in the ligand's FT by a rotation motion object. This can be done interactively using the mouse in a Vision-based graphical editor we have developed. The FT data structure was developed to allow the addition of new motion descriptors. For instance, one could design a local perturbation for rotamers, where all side chain torsion angles could be randomized within some bounds. The combination and nesting of motion descriptors provides a powerful way to encode very complex conformational spaces. Such FTs can be used to encode hypothesis about motions of parts of the receptor as well as information obtained form a variety of sources. A successful FLIPDock setup should balance the flexibility needed and the computational expense involved.

While the FT and therefore FLIDPock is not designed to determine which part of a receptor should be made flexible or which motion descriptor best describes specific conformational changes, it provides a general platform for the representation and combination of motion information that is available.[18] A user can define a FT that combines domain motion, backbone motions as well as side chain flexibility as demonstrated in the balanol docking example. When internode motions (e.g. hinges) are used, the molecular geometry at the boundary of the molecular fragments may be distorted. If the ligand is close enough to interact with this boundary region (e.g. the atoms near the hinge axis), the FLIPDock score, especially the interaction between ligand and receptor, can be unrealistic. We are working on incorporating loop generation[32–36] methods to address this problem. Despite this current limitation, we believe that the ability to make numerous side chains flexible in the active site of a receptor is in and of itself valuable as it has been observed that most induced fit motion, especially in the catalytic residues, is accounted for by side chain flexibility.[37,38] FLIPDock is able to tackle such cases by allowing multiple, if not all, side chains near known or potential active sites to be flexible in the context of a general framework allowing much more complex motions to be explored.

Finally, FLIPDock has a component-based architecture: new search engines and additional scoring functions can be added and used interchangeably. This provides a powerful platform for comparing and further understanding these two essential ingredients of docking programs. We are working on the addition of a Larmarckian GA search engine, which is currently the most efficient search technique in AutoDock. We also plan to incorporate alternative search techniques such as particle swarm optimization[39] and additional scoring functions such as the ones recently reviewed in Wang et al.[40] into the FLIPDock framework. Another future direction will be the integration of flexibility prediction methods to assist users in the creation of FT for receptors.

On average, a single energy evaluation of a putative docking in FLIPDock takes about 0.14 seconds in our cross docking experiments on a Linux platform using a single 3.4 GHz Intel Xeon CPU. While the code can be optimized to decrease running time substantially, so far our focus has been on the versatility and generality of the software framework. More information on FLIPDock

DOI 10.1002/prot

is available online at http://www.scripps.edu/~sanner/FLIPDock. The FLIPDock code will be made freely available for academic research.

ACKNOWLEDGMENTS

The authors thank William Lindstrom and Ruth Huey for providing us with the C++ implication of the pairwise AutoDock scoring function, Professors. Arthur Olson and David Goodsell and Dr. Garrett Morris for many helpful discussions.

REFERENCES

1. Bonvin A. Flexible protein–protein docking. Curr Opin Struct Biol 2006;16:194–200.
2. Teague SJ. Implications of protein flexibility for drug discovery. Nat Rev Drug Discov 2003;2:527–541.
3. Ma B, Shatsky M, Wolfson HJ, Nussinov R. Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations. Protein Sci 2002;11:184–197.
4. Barril X, Morley SD. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. J Med Chem 2005;48:4432–4443.
5. Carlson HA. Protein flexibility and drug design: how to hit a moving target. Curr Opin Chem Biol 2002;6:447–452.
6. Claussen H, Buning C, Rarey M, Lengauer T. FlexE: efficient molecular docking considering protein structure variations. J Mol Biol 2001;308:377–395.
7. Lin J-H. ALPJRSJAM. The relaxed complex method: Accommodating receptor flexibility for drug design with an improved scoring scheme. Biopolymers 2003;68:47–62.
8. Wong CF, Kua J, Zhang Y, Straatsma TP, McCammon JA. Molecular docking of balanol to dynamics snapshots of protein kinase A. Proteins 2005;61:850–858.
9. Carlson HA, McCammon JA. Accommodating protein flexibility in computational drug design. Mol Pharmacol 2000;57:213–218.
10. Wong CF, Kua J, Zhang Y, Straatsma TP, McCammon JA. Molecular docking of balanol to dynamics snapshots of protein kinase A. Proteins: Struct FunctBioinform 2005;61:850–858.
11. Österberg F, Morris G, Sanner M, Olson A, Goodsell D. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. Proteins: Struct Funct Genetics 2002;46:34–40.
12. Cavasotto CN, Abagyan RA. Protein flexibility in ligand docking and virtual screening to protein kinases. J Mol Biol 2004;337:209–225.
13. Cavasotto CN, Kovacs JA, Abagyan RA. Representing receptor flexibility in ligand docking through relevant normal modes. J Am Chem Soc 2005;127:9632–9640.
14. Sherman W, Day T, Jacobson M, Friesner R, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. J Med Chem 2006;49:534–553.
15. Mizutani MY, Takamatsu Y, Ichinose T, Nakamura K, Itai A. Effective handling of induced-fit motion in flexible docking. Proteins 2006;63:878–891.
16. Schrodinger Suite 2006. Available at biowulf.nih.gov/apps/schrodinger/docs/inducedfit/user_manual/ifd06_user_manual.pdf.
17. Alonso H, Bliznyuk AA, Gready JE. Combining docking and molecular dynamic simulations in drug design. Med Res Rev 2006.
18. Zhao Y, Stoffler D, Sanner M. Hierarchical and multi-resolution representation of protein flexibility. Bioinformatics 2006;22:2768–2774.
19. Brooijmans N, Kuntz I. Molecular recognition and docking algorithms. Annu Rev Biophys Biomol Struct 2003;32:335–373.
20. Morris G, Goodsell D, Halliday R, Huey R, Hart W, Belew R, Olson A. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1999;19:1639–1662.
21. Lutz M, Asher D. Learning python. Sebastopol, CA: O'Reilly & Associates; 1999.
22. Sanner MF. A component-based software environment for visualizing large macromolecular assemblies. Structure 2005;13:447–462.
23. Sanner MF, Stoffler D, Olson AJ. ViPEr, a visual programming environment for Python. In: Proceedings of the 10th International Python conference. Alexandria, VA; February 4–7, 2002; pp 103–115.
24. David S. Goodsell, Personal communication.
25. Dunbrack RL. Rotamer libraries in the 21st century. Curr Opin Struct Biol 2002;12:431–440.
26. Case DA, Cheatham TE, III, Darden T, Gohlke H, Luo R, Merz KM, Jr., Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. J Comput Chem 2005;26:1668–1688.
27. de Groot BL, van Aalten DM, Scheek RM, Amadei A, Vriend G, Berendsen HJ. Prediction of protein conformational freedom from distance constraints. Proteins 1997;29:240–251.
28. Ala PJ, DeLoskey RJ, Huston EE, Jadhav PK, Lam PY, Eyermann CJ, Hodge CN, Schadt MC, Lewandowski FA, Weber PC, McCabe DD, Duke JL, Chang CH. Molecular recognition of cyclic urea HIV-1 protease inhibitors. J Biol Chem 1998;273:12325–12331.
29. Tong L, Pav S, Mui S, Lamarre D, Yoakim C, Beaulieu P, Anderson PC. Crystal structures of HIV-2 protease in complex with inhibitors containing the hydroxyethylamine dipeptide isostere. Structure 1995;3:33–40.
30. Melnick M, Reich SH, Lewis KK, Mitchell LJ, Jr., Nguyen D, Trippe AJ, Dawson H, Davies JF, II, Appelt K, Wu BW, Musick L, Gehlhaar DK, Webber S, Shetty B, Kosa M, Kahil D, Andrada D. Bis-tertiary amide inhibitors of the HIV-1 protease generated via protein structure-based iterative design. J Med Chem 1996;39:2795–2811.
31. Hetényi C, van der Spoel D. Efficient docking of peptides to proteins without prior knowledge of the binding site. Protein Sci 2002;11:1729–1737.
32. Crivelli S, Kreylos O, Hamann B, Max N, Bethel W. ProteinShop: a tool for interactive protein manipulation and steering. J Comput-Aid Molec Des 2004;18:271–285.
33. Canutescu AA, Dunbrack RL, Jr. Cyclic coordinate descent: a robotics algorithm for protein loop closure. Protein Sci 2003;12:963–972.
34. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. Proteins 2004;55:351–367.
35. Coutsias EA, Seok C, Jacobson MP, Dill KA. A kinematic view of loop closure. J Comput Chem 2004;25:510–528.
36. Amarda Shehu, Cecilia Clementi, Lydia E, Kavraki. Modeling protein conformational ensembles: from missing loops to equilibrium fluctuations. Proteins: Struct Funct Bioinform 2006;65:164–179.
37. Gutteridge A, Thornton J. Conformational changes observed in enzyme crystal structures upon substrate binding. J Mol Biol 2005;346:21–28.
38. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. Proteins 2000;39:261–268.
39. Clerc M, Kennedy J. The particle swarm—explosion, stability, and convergence in a multidimensional complex space. IEEE Trans Evolut Computat 2002;6:58–73.
40. Wang R, Lu Y, Fang X, Wang S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. J Chem Inf Comput Sci 2004;44:2114–2125.