Docking Ligands into Flexible and Solvated Macromolecules. 3. Impact of Input Ligand Conformation, Protein Flexibility, and Water Molecules on the Accuracy of Docking Programs

Christopher R. Corbeil and Nicolas Moitessier*

Department of Chemistry, McGill University, 801 Sherbrooke Street W, Montréal, Québec, Canada H3A 2K6

Received November 12, 2008

Several modifications and additions to FITTED1.5 led to the development of FITTED2.6. Among the novel implementations are a matching algorithm-enhanced genetic algorithm and a ring conformational search algorithm. With these various optimizations, we also hoped to remove the biases and to develop a docking program that would provide results (i.e., poses) as independent as possible to the input ligand and protein conformations and used parameters, although keeping the options to provide additional experimental information. These biases were investigated within FITTED2.6 along with FlexX, GOLD, Glide, and Surflex. The input ligand conformation was found to have a major impact on the program accuracy as drops as large as 10−50% were observed with all the programs but FITTED. This comparative study also demonstrates that the accuracy of FITTED is similar to that of other widely used programs. We have also demonstrated that protein flexibility, displaceable water molecules, and ring conformational search algorithms, three of the main FITTED features, significantly increased its accuracy. Finally, we also proposed potential modifications to the available programs to further improve their accuracy in binding mode prediction.

INTRODUCTION

In modern drug design, docking-based virtual screening (VS) methods provide a quick and inexpensive alternative to high-throughput screening. In fact, numerous applications have demonstrated the reasonable level of accuracy of the available methods.[1,2] In parallel, comparative studies evaluated the relative accuracy of previous versions of docking programs in predicting the correct binding modes, typically with Glide and GOLD yielding the best results.[3−9] Many of these studies, which often made use of ligand/protein cocrystal structures, showed that the accuracy of docking the native ligands back to the corresponding protein structures (self-docking) gave reasonable results. However, when examining docking of a ligand to non-native crystal structures of the same protein (cross-docking), the accuracy of most of the programs was significantly lower.[10−12] These failures result in part from the assumption that proteins are rigid objects (the lock-and-key model) even though they are known to be flexible dynamic objects. As a result, this major assumption lead to inaccurate binding pose predictions and low enrichment factors in VS.[13] In fact, implementing protein flexibility has been seen as one major challenge in the development of docking methods.[14−16] Currently, very few programs consider the flexibility of the protein upon docking, although various strategies have been proposed ranging from soft-docking (e.g., smoothed protein structure in AutoDock[11]) to a more exhaustive and therefore time-consuming protein conformational search as seen in Glide when combined with Prime.[17] Docking to conformational ensembles has also been implemented within a few programs such as FlexX-Ensemble,[18] Slide.[19,20] and AutoDock.[21] These various implementations led to significantly improved predictions of binding modes when compared to cross-docking studies.

One of the other challenges in docking and VS is the treatment of key water molecules.[22] In most protein/ligand docking studies, water molecules, if present, are treated on a per protein basis. If the water molecules appear as highly conserved, then they are kept as part of the protein description for the docking run. This approach clearly precludes accurate docking of ligands that would displace these key water molecules upon binding. A commonly described example is HIV-1 protease ligands. A tightly bound water molecule has been observed within the cocrystal structure of HIV-1 protease with KNI-272 and analogues.[23−27] This water molecule may therefore be required for an optimally accurate docking of this set of analogues. In parallel, inhibitors built around a cyclic urea scaffold have been designed to displace this water and would not be properly docked if this water molecule was kept.[28−30] Ideally, water molecules should be displaceable. In fact, a previous report from our group showed that AutoDock gained accuracy when water molecules were made displaceable.[31] Currently only a few docking methods can displace water molecules while docking ligands. For instance, GOLD[32] uses user-defined waters present in the protein input file while FlexX[33] places water molecules within the binding site and keeps the ones interacting strongly. In these two cases, these programs both score with the water present (on) or not (off) and select the best scoring option. A version of FlexX currently in beta testing allows for displaceable waters[34] that were present in the input file. Surprisingly, when the GOLD implementation was reported, it showed no improvement of the docking accuracy. This somewhat unexpected observation questioned

* To whom correspondence should be addressed. Phone: 514-398-8543. Fax: 514-398-3797. E-mail: nicolas.moitessier@mcgill.ca.

10.1021/ci8004176 CCC: $40.75
Published on Web 04/01/2009

the development of functionalities specifically designed to displace water molecules.

Within the past years, we have developed and reported FITTED1.0,[35] followed by FITTED 1.5.[36] FITTED (Flexibility Induced Through Targeted Evolutionary Description) is a docking program that addresses the challenges of protein flexibility and displaceable water molecules. Herein, we describe the development of the next version of this docking program, FITTED2.6, that focuses on accelerating the docking process while keeping similar accuracy. We also focused on reducing the dependence of the accuracy on input parameters and structures. We have previously found that the accuracy of eHiTS was affected by the ligand input structure and further investigation was necessary to evaluate this effect on the accuracy of other programs including ours.[37] To identify its strengths and weaknesses and evaluate these dependencies, we then compared FITTED to some of the most popular docking software available with a specific focus on how changes in input structure and parameters affect docking accuracy.

THEORY AND IMPLEMENTATION

FITTED 1.0 and 1.5, Creating a Virtual Screening Tool out of a Docking Program. Previous reports from our group detailed the development of FITTED versions 1.0[35] and 1.5,[36] and only a brief description is given below. FITTED is a suite of programs that includes FITTED (the docking engine), PROCESS (Protein Conformation Ensemble System Setup, a module for protein file preparation), and SMART (Small Molecule Atom typing and Rotatable Torsion assignment, a module for ligand preparation). Docking a ligand to a protein can be seen as a global optimization problem. The ligand binding mode, protein conformation, water molecule occurrence and locations have to be optimized to provide an optimal free energy of binding. In FITTED, a Lamarckian genetic algorithm addresses the conformational space search. Genetic algorithms are stochastic methods and often start with randomly generated populations, followed by a time-consuming evolution. Because of the large conformational space of the protein/water/ligand complexes, finding the global solution requires a large number of generations and large populations. Short cutting the process is therefore necessary to reduce the CPU time required for a single run. We thought that starting with a population that has already evolved (i.e., lower average energy than random poses) would lead to desired decreases in computational time. Thus using a series of atomic charge constraints and a binding site volume, FITTED1.0 prepared such an initial population intelligently. This approach allows for quicker convergence of the population through evolution. Along with this genetic algorithm, FITTED incorporated a switching function that effectively turns off the water and allows them to be displaced when required. The initial validation of FITTED 1.0 with a small set of protein/ligand complexes showed promises with 76% and 73% success in self-docking and docking to flexible proteins respectively.[35] Although this first version was docking ligands effectively, our eventual goal was to make FITTED a VS tool. Some modifications of the original algorithm were necessary to make it significantly quicker to achieve speeds required by VS tools.

The first step in any virtual screen is the preparation of the virtual database of potential ligands. Since a docking program will attempt to dock any given compound, we first focused on prioritizing "drug-like" molecules for docking. For this purpose, a series of descriptors were implemented into SMART. Bit strings describing the molecular structure generated by SMART could then be exploited by FITTED to filter out compounds with undesired chemical features and/or physical properties. The docking was modified to incorporate a consensus docking approach that enabled FITTED to create and allow the population to evolve in a more intelligent manner than the previous version. This was done by adding pharmacophores and/or automatically generated sets of protein interaction sites (generated by PROCESS) to orient the docking process. When a conformation of a ligand did not match well to the pharmacophore (PharmScore) and/or the interaction sites (MatchScore), the conformation was discarded and a new one was generated. Thus, the inclusion of the interaction sites oriented the docking toward better solutions and, as a result, afforded a 10% increase in accuracy over FITTED 1.0 and a significant decrease in required CPU time.[36] FITTED was then used in the screening of the Maybridge database against HCV polymerase and was successful in identifying two hits in the low micromolar range.[36]

FITTED2.6. Improvements to Remove Dependencies on Input Parameters. When more knowledge is provided to a docking program, the accuracy is expected to increase. For instance, if the ligand in its crystal structure conformation is docked, a program that uses the ligand input structure as an initial guess would most likely outperform any other program in self-docking experiments. However, these experiments would give no information regarding its true accuracy because, in a real drug design scenario, the user does not know the solution. Some other biases, including the selection of parameters and the protocol used to prepare the protein (e.g., protonation state of ionizable residues), can also greatly affect the evaluation of programs. The removal of these dependencies arising from the input parameters has become one of the hot topics in the literature as of late.[38−42]

One of these dependencies is the input conformation of rings. In a VS study, large libraries of compounds are tested *in silico*. These libraries are typically prepared from two-dimensional representations of these molecules, then a 2D to 3D converter such as OMEGA[43] and CORINA[44] is used to generate the 3D coordinates. Most 2D to 3D converters output an esthetically stable state with the option to find a low-in-energy conformation as defined by a force field and a conformational search algorithm. This conformation may not always be the same as the bioactive conformation. If the molecule is acyclic then this poses no problem, since most popular docking software consider the flexibility of acyclic portions of the molecule. Molecules with flexible cyclic structures as in Figure 1 prove to be more challenging. In addition, the conformation of the flexible ring depends on the program used to generate it and is often not fully searched.[38,39,45,46] Our genetic algorithm has therefore been modified to account for ring flexibility as detailed below.

Implementation of Ring Flexibility. There are three main strategies to address the issue of flexible ring systems. First a separate tool can create multiple conformations of the ligand to be used as multiple inputs by docking programs. Second, several ring conformations can be exploited during the incremental construction of ligands. Surflex2.1[47] uses

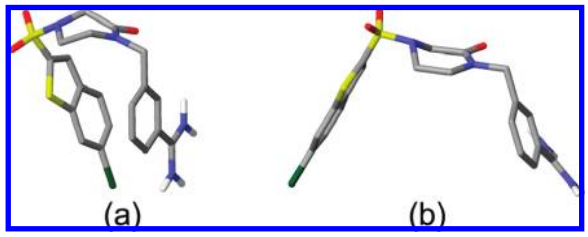

Figure 1. Conformation of 1nfu ligand (a) as observed in the crystal structure and (b) as generated by OMEGA.

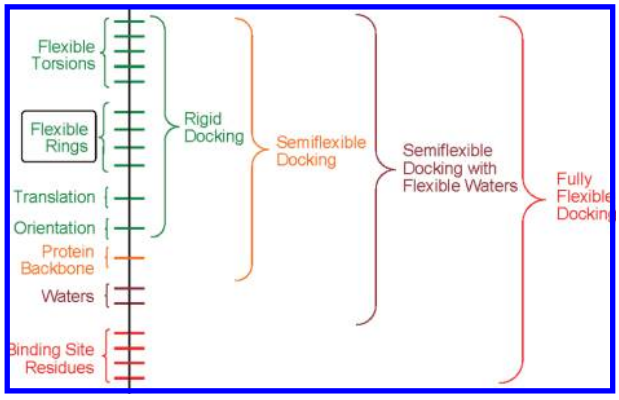

Figure 2. FITTED 1.5 vs FITTED 2.6 chromosome and the various docking modes. Each of the horizontal lines represents a gene (e.g., given conformation of a side chain residue). The box highlights the implementation in FITTED 2.6.

templates of five to seven-membered cycloalkanes to generate multiple input conformations of the rings used by the incremental construction algorithm. Even though the templates are saturated carbocyclic structures, the inclusion of energy minimization steps accounts for the various conformations that may exist for heterocyclic and unsaturated systems. Similarly, Glide,[48] version 5.0, uses the template library from LigPrep to be able to conformationally search larger rings.[49] The third option is to perform the conformational search while docking. GOLD[50−52] exploits the corner flap approach developed by Goto and Osawa,[53] where the atom to be flipped is reflected in the plane formed by the adjacent atoms. The major advantage of this approach is that rings of any size can be searched. The one pitfall is the requirement to have the four adjacent atoms in a plane. In a previous version of Glide, version 4.5, the docking engine used a approach similar to GOLD.[54]

The genetic algorithms rely on the use of chromosomes and the theory of evolution. In the context of docking, the chromosomes are sets of numerical values (genes) that can evolve through genetic operators such as mutations and crossover. These numerical values often define the conformation, orientation and position in space of the ligand (referred to as a pose). The chromosome, as defined in FITTED1.5, included the acyclic flexible torsions, translation and orientation for a given pose of the ligand. Depending on the selected docking mode, the chromosome may also include the protein backbone, water positions and binding site residues (Figure 2).

Since only the acyclic portions of the ligand were included in the chromosome, the conformation of the ring(s) within the ligand remained the same throughout the evolution unless altered by the energy minimization routine. To account for ring flexibility, FITTED2.6 now includes a conformational

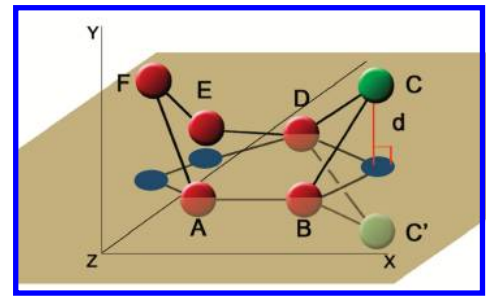

Figure 3. Example of the corner flap approach converting a boat conformation to a chair.

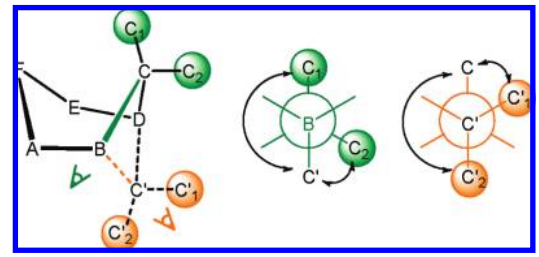

Figure 4. Assumption of torsion equivalencies.

search algorithm for rings during the generation of new conformations (Figures 3 and 4). FITTED2.6 uses a corner flap algorithm similar to that of GOLD but does not impose any criteria to the position in space of A, B, D, and E (Figure 3). This is achieved by creating the plane out of three atoms (A, B, and D) instead of the four atoms required in GOLD. Any distortions of the bond length and angles are next resolved through the energy minimization steps performed by FITTED. To maintain the asymmetry of atom C, GOLD imposes the rotation of two bonds (AB and BC) to position C_1 and C_2 (Figure 4). This approach reinforces the need to have four atoms in a plane. In our current implementation, an assumption is made that the torsion C_1CBC' is equivalent to $C'_2C'BC$ (Figure 4). Thus the Cartesian coordinates of C'_2 can be defined by converting C_2 into C'_2 from its internal coordinates (bonds, angles, and torsions).[55]

Improved Definition of Interaction Sites and a Matching Algorithm. As described previously, the module PROCESS prepares the protein files in the format needed by FITTED. It also probes the binding site and generates additional data for optimal docking. Among this data is a list of potential protein interaction sites (ISs). Geometric rules are applied to find the ideal locations of hydrogen bond donor and acceptor groups referred to as HBD and HBA ISs. In the previous version, ISs were centered on the Ser, Thr, and Tyr hydroxyl oxygen and on metal centers. These points are now placed in the position of the oxygen lone pairs or free metal coordination sites. PROCESS determines these free metal coordination sites by examining the surrounding residues and using the vector bond valence postulate[56] that states that the sum of all the vectors of the coordinated atoms must be equal to 0.

The earlier version of FITTED did not identify the hydrophobic pockets with great accuracy. To resolve this issue, various strategies have been implemented. In the current version, a grid of evenly distributed points is generated and the interaction of a probe atom at each of these grid points with the protein is computed. To be considered hydrophobic (referred to as HYD), the point should not be in close

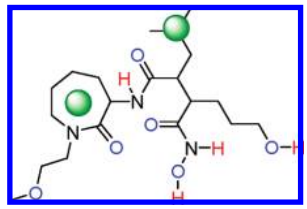

Figure 5. Representation of the ISs found for 1bwi (green = hydrophobic points, blue = hydrogen bond acceptors, red = hydrogen bond donors).

proximity to an HBA or HBD point. Then the van der Waals interaction energy calculated between the probe atom and all the protein carbons should be below a minimum van der Waals energy cutoff value. Applied to a number of proteins, we found this new definition to be more accurate than the previous one.

A weight is then assigned to each point depending on its type. For HBA and HBD, a weight is assigned depending on whether the point was created from a charged or neutral residue then scaled to account for the buriedness of the point.[57] The HYD points are scaled according to the ratio of the van der Waals energy calculated for point over the minimum van der Waals energy cutoff. These weights are then used to compute the MatchScore of each pose as described previously.[36]

Some years ago, we have shown that using pharmacophore oriented docking with a matching algorithm can improve docking accuracy substantially.[58] FITTED 1.5 initiated a move in this direction by orienting the docking using ISs.[36] With the new version, we complete this move with the inclusion of a three-point triangle matching algorithm to orient the ligand instead of the random translation and orientation performed with previous versions. Triangles made of ligand atoms are matched onto triangles made of ISs of identical chemical property (HDB, HBA, or HYD). To optimize the efficiency of this algorithm, only a subset of potential triangle match is used. First, FITTED removes triangles that connect low weight interaction sites. Second, all triangles must contain at least one point of the top 10 ISs as sorted by weights.

The creation of the ligand ISs are based on simple rules. All ligand atoms that are hydrogen bond acceptors (HBA) or donors (HBD) are labeled as HBA or HBD (red and blue atoms in Figure 5). Hydrophobic ligand points (HYD, green spheres in Figure 5) are centroids either placed at the center of hydrophobic rings (rings with a majority of carbons) or at the center of *iso*-propyl, methyl and *tert*-butyl groups (Figure 5). All possible combinations of three-point triangles are then created and stored. The position of these ligand ISs are recalculated with each new conformation.

With the two lists created, a new ligand conformation is randomly generated and a triangle match between the ligand ISs and the protein ISs is sought. The ligand triangle is then superimposed with the protein interaction site triangle. The ligand, that is now oriented within the binding site, proceeds through the consensus docking approach described in FITTED 1.5 and summarized in Figure 6.

With earlier versions of FITTED, the minimum MatchScore necessary for a pose to be accepted was manually set and therefore the accuracy of the docking run was heavily dependent on it. This is in fact an appropriate approach in drug design when the user wants to make use of additional information (e.g., a pharmacophore developed from other studies). However, in the case where no information is available, a value for the minimum MatchScore is still requested. With the new additions to the generation and orientation of the ligand, the focus switched to the automatic selection of a MatchScore for a ligand during docking. FITTED2.6 starts with an initial minimum MatchScore, that can either be automatically reduced or increased depending on the ligand. As soon as a new individual is saved, the minimum MatchScore (Min MatchScore in eq 1) is recalculated based on eq 1. In eq 1, MatchScores are all the MatchScore values computed so far while Max MatchScore refers to the maximum value recorded so far. The scaling factors have been empirically defined to orient the docking without affecting the time required for the generation of the initial population.

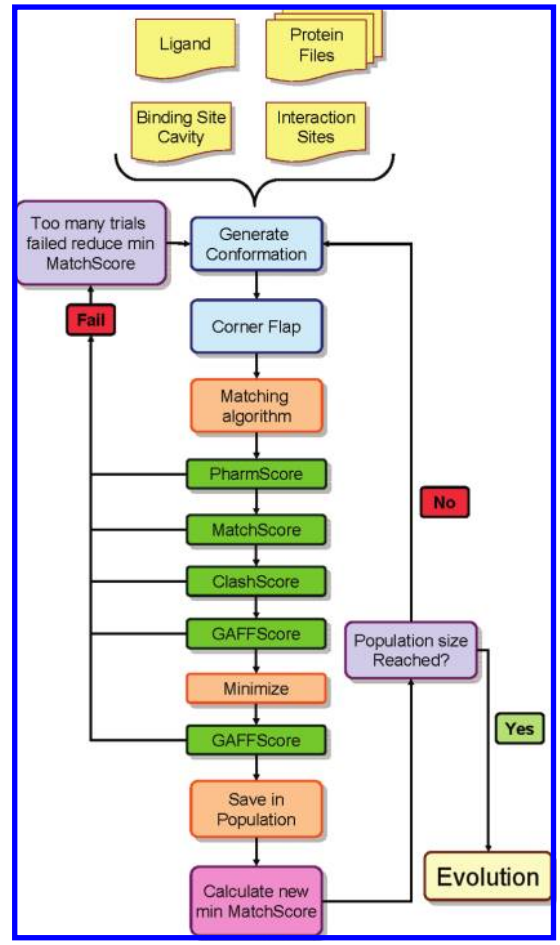

Figure 6. Schematic of the generation of the initial population within FITTED 2.6.

$$\text{Min MatchScore}_{i+1} = \frac{5.0 \times \dfrac{\sum_{0}^{i} \text{MatchScores}}{i} + 2 \times \text{Max MatchScore}}{10.0} \quad (1)$$

Evolution and Convergence. With the previous versions of FITTED, it was necessary to perform multiple runs to find the global minimum. To increase the convergence between various runs, FITTED2.6 incorporates a matching algorithm

Table 1. Testing Set of Ligand/Protein Complexes

protein (abbreviation)	number of structures	include water?	PDB codes
cyclin-dependent kinase 2 (CDK2)	4	yes	1aq1, 1dm2, 1pxn, 1pxn
cyclooxygenase-2 (COX-2)	4	no	1cx2, 1pxx, 3pgh, 4cox
estrogen receptor (ER)	3	yes	1err, 1sj0, 3ert
factor Xa (FXa)	5	yes	1ezq, 1f0r, 1fjs, 1nfu, 1zka
GluK2 Kainate Receptor 6 (GluK2)	5	yes	1s7y, 1s9t, 1sd3, 1tt1, 1yae
HCV polymerase allosteric site (HCV Allo)	9	no	1nhu, 1nhv, 1os5, 2gir, 2hai, 2hwh, 2hwi, 2ilr, 2o5d
HCV polymerase catalytic site (HCV Cat)	7	yes	1yvf, 1z4u, 2fvc, 2gc8, 2giq, 2qe2, 2qe5
HIV-1 protease, mono protonated ASP (HIVP)	5	yes	1b6l, 1eby, 1hpo, 1hpv, 1pro
HIV-1 protease, diprotonated ASP (HIVPD)	5	yes	1ajv, 1ajx, 1hvr, 1hwr, 1qbs
HIV-1 reverse transcriptase (HIVRT)	4	yes	1c1b, 1fk9, 1rt1, 1vrt
mannosidase (Mann)	8	no	1hww, 1hxk, 1ps3, 1r33, 1r34, 1tqt, 2f1a, 2f18
matrix metalloprotease 3 (MMP-3)	4	no	1b8y, 1bwi, 1ciz, 1d8m
P38 map kinase (P38)	5	no	1a9u, 1b17, 1w7h, 1w82, 1w84
thermolysin (Therm)	8	yes	1thl, 1tlp, 1tmn, 3tmn, 4tmn, 5tmn, 6tmn, 8tln
thrombin (Thrn)	5	yes	1dwc, 1etr, 1ets, 1ett, 1tmt
thymidine kinase (TK)	9	yes	1e2k, 1e2p, 1ki3, 1ki4, 1ki7, 1ki8, 1of1, 1qhi, 2ki5
trypsin (Tryp)	5	yes	1f0u, 1o2j, 1o3g, 1o3i, 1qbo
vitamin D receptor (VDR)	5	yes	1db1, 1ie8, 1txi, 2har, 2has

to create the higher quality initial population. Additional modifications were made to the evolution algorithm to better mimic the Lamarckian and Darwinian evolution. We thought to favor the evolution of the best individuals. First, to increase the possibility of the best individuals coupling with each other, we implemented a new evolutionary function called the probability of elitism (pElite operator). This function copies one of the top of individuals, performs a local search on it and passes it on to the next generation. Also a new selection criterion for the next generation called Metropolis evolution was implemented. With this mode, the children replace the parents based on an energy-based Metropolis criterion at a user-defined temperature. With this criterion, higher in energy children have a nonzero probability to survive and be coupled in the next generation. This approach ensures some structural diversity in the population and enable the creation of a population that follows the Boltzmann distribution. This population can next be used for refined scoring.

As a last modification, we moved away from the all atom representation of protein/ligand interactions to the less time-consuming united atom representation. However, the all atom representation is kept to compute the ligand internal energy and preclude any inversion of chiral centers. This hybrid united atom/all atom representation resulted in an increase in speed over the last version of FITTED.

RESULTS AND DISCUSSION

Objectives. To validate this new version we have increased the validation set developed for FITTED 1.0 from 5 proteins (33 crystal structures) to highly diverse 18 proteins and 100 crystal structures (Table 1). In addition to the evaluation of FITTED's accuracy, we decided to investigate the impact of parameters and input structures on accuracy. As discussed above, protein structure, ligand structure (e.g., rings) and selected parameters often have an impact on the docking accuracy. Although some parameters are expected to increase accuracy (i.e., Standard Precision mode vs eXtra Precision mode in Glide or accuracy levels in GOLD), the ligand conformation should not (Table 1). The set described here includes very challenging proteins such as HCV RNA polymerase[9] and metallo enzymes.[37]

Table 2. Comparison of Success Rates of FITTED Versions 1.0, 1.5, and 2.6 Using the "Dock" Docking Mode

	% success		
docking mode	1.0	1.5	2.6
rigid (self-docking)	79	93	79
rigid (cross-docking)	47	75	56
semiflexible	73	84	67
flexible	73	88	67

Table 3. Comparison of Time and Number of Runs Required for Various Versions of FITTED When the "Dock" Docking Mode Is Selected for Rigid Protein Docking

	time (min) per run/number of runs		
protein	1.0	1.5	2.6
TK	63/5	30/3	8.5/1
HIVP	114/5	55/3	22/2

Comparing FITTED Versions 1.0, 1.5, and 2.6. To examine the effect of the various modifications made to FITTED, we compared the current version to the previous ones by using the training set initially developed to test the accuracy of FITTED 1.0.[35] This set is composed of 33 protein−ligand complexes and 5 proteins (HIVP, FXa, Tryp, MMP-3, and TK). Table 2 summarizes the results obtained with the three FITTED versions for self-docking ("Rigid" protein flexibility mode) and docking to flexible proteins using the crystallographic conformation of the ligands. As previously reported, the "SemiFlexible" protein flexibility mode corresponds to docking to a conformational ensemble of protein structures, while flexible docking corresponds to a fully flexible protein.

Overall accuracy has declined between versions 1.5 and 2.6 (Table 2), although this set is not large enough to provide statistically relevant evaluation. Gratifyingly, an overall increase in speed and convergence between multiple runs was also recorded (Table 3). This drop in accuracy is attributed to the manual selection of the minimum Match-Score in version 1.5 that is now automatically determined during the generation of the initial population. As a result, docking to FXa, which was fairly successful with the previous versions shows very poor accuracy with the current version (i.e., one out of 5 ligands is docked accurately) while

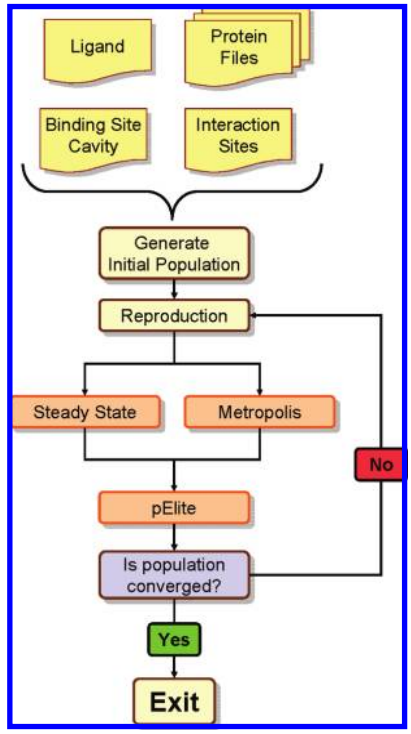

Figure 7. Schematic of the evolution cycle of FITTED 2.6.

docking to the other proteins demonstrated accuracy of 60% (Tryp), 75% (MMP-3) and even 100% (HIVPD, HIVP, TK). Manual selection of high Min MatchScore values with version 2.6 forces the key ionic interactions between the ligands and binding site Asp of FXa and Tryp and restores an accuracy similar to that of version 1.5. This automatic determination of the minimum MatchScore is important as in a blind docking study no information is given and determination of this minimum MatchScore value would be difficult.

A 3-fold increase in speed is seen with the newer version of FITTED. This increase can be in part attributed to the introduction of a matching algorithm to orient the ligand when generating the initial population and to the use of the hybrid united atom/all atom representation described above. In the current work, both an extensive conformational search is carried out using the "Dock" docking mode along with a significantly quicker (i.e., less generations) "VS" docking mode. The increase in the convergence of the runs (a single run is often enough with the current version) can be attributed in part to the inclusion of the new pElite evolutionary operator and to the matching algorithm. It should be stressed that in VS mode the time can be significantly reduced (down to 3−4 min for TK inhibitors) and that code optimization is ongoing to further improve the necessary CPU time for a single run.

Comparing Dependencies on Input Structures. We next turned our attention to the impact of the input structure of both the ligand and the protein on the pose prediction accuracy of FITTED. We also investigated how other docking programs perform under the same conditions. For this comparison, we decided to focus more specifically on three important features that can affect the accuracy of docking programs, the ligand input structure, protein input structures and the inclusion of bridging water molecules.

Discussions with the developers and/or technical support of each program allowed a fair comparative study and an optimal use of these programs. In fact, following recommendations, many conditions (set of parameters) were tried. In addition, all the major scoring functions were tried if more than one was available (in FlexX and GOLD). Some representative data is shown in Figure 8. Different levels of accuracy were also tried with some of the programs as described in the legend of Figure 8.

Prior to the description of the experiments and results, we thought this study should be put in context. In this work, we wish to evaluate the docking ability of programs without any information other than the crystal structure of the protein. Obviously in the context of drug design and screening, any relevant information should be given to the docking program. However, this would bring too many variables to this study as ISs in FITTED or pharmacophores in FlexX (FlexX-Pharm) can be trained and would significantly increase their respective accuracy.

We also wish to stress that the primary goal of this study is not to compare programs but to evaluate the impact of the input parameters on their pose prediction accuracy. In addition, as in any comparative study, the data collected in this work should be considered with care as the set is still not large enough to draw conclusions on their respective accuracy, and some hidden biases may remain as discussed below. In addition, we have used the rmsd between ligand crystal structures and docked poses to measure the docking accuracy. This criterion is believed to be appropriate to evaluate the impact of input parameters (the relative accuracy under two different conditions) but not to compare programs.

Ligand Input Conformations. We first looked at the impact of the ligand structure. Previously reported comparative studies typically have only used one conformation of the ligand either the crystal[6] or a noncrystal[5,9,59] conformation of ligands. In a real drug design scenario, the bioactive conformation is unknown and therefore the noncrystal conformation represents a more realistic scenario. To evaluate the bias when using the crystallographic conformation of the ligand, we compared the accuracy of the docking programs when the crystal ligand or OMEGA[43] generated structures were used alternatively as input (Figure 8). In this work, the OMEGA generated structure were used to assess the programs ability to dock the noncrystal conformation of the ligand. To our knowledge, none of the docking programs assessed in this work have been trained with OMEGA-generated structures. However, these ligand structures cannot be considered as completely unbiased as these conformations may be preferably docked by one of these programs, a bias that we have not evaluated.

In this work, the docked pose is assumed to be accurately predicted if it is within 2.0 Å of the crystal binding mode when performing self-docking experiments. Even though the use of rmsd values is known to be misleading, we believe that it will clearly reveal drops or increases in accuracy induced by specific parameters or conformations. When evaluating the accuracy in cross-docking experiments, the additional error introduced when superposing the protein structures should be considered. Thus, an arbitrary rmsd of 2.25 Å in cross-docking experiment was selected as a criterion of success.

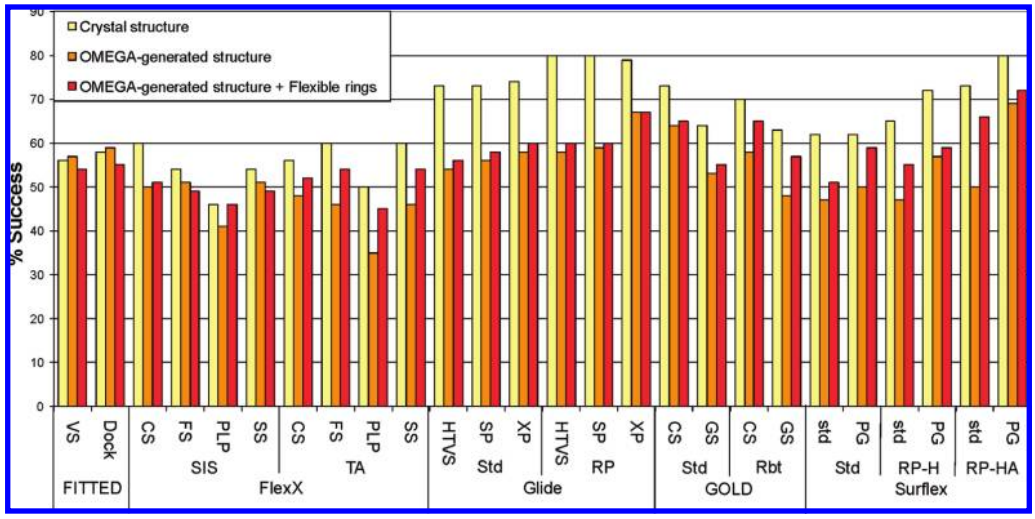

Figure 8. Accuracy vs ligand and protein conformations. For legend see Table 4.

Table 4. Abbreviations used in Figure 8

program	abbreviations	
FITTED	VS	virtual screening mode
	dock	docking mode
FlexX	SIS	single interaction scan (matching algorithm)
	TA	triangle algorithm
	CS	ChemScore scoring function
	FS	FlexXScore scoring function
	PLP	PLP Scoring scoring function
	SS	ScreenScore scoring function
Glide	RP	refined protein (optimized protein structure)
	Std	nonrefined protein
	HTVS	high-throughput virtual screening mode
	SP	standard precision mode
	XP	extra precision mode
GOLD	Std	standard (automatic selection of parameters)
	Rbt	robust
	GS	GoldScore scoring function
	CS	ChemScore scoring function
Surflex	Std	standard docking
	PG	pgeom
	Std	nonrefined protein
	RP-H	protein structure with hydrogen atom positions refined
	RP-HA	refined protein structure with a constrained optimization of the heavy atoms

As can be seen in Figure 8, the accuracy of all the programs drops by 10 to 20% when moving from crystal ligand structures (yellow bar) to OMEGA-generated structures (orange bars), except for FITTED for which no change in accuracy was observed. This first piece of data confirmed that a docking program should not be evaluated by using ligand crystal structures as input. We then used the ring conformational search features when available (red bars in Figure 8). The overall accuracy increases although not reaching the one observed with crystal structures. The ring conformational search engine used by Surflex, which covers a wider range of rings, are clearly more efficient than the ones used by Glide4.5, GOLD, and FITTED. While this manuscript was in preparation, a new version of Glide that features a new ring search method was released but was not used in this work. This most recent version of Glide (v5.0) uses a ring library that accounts for both larger rings and more ring systems and includes small heterocyclic rings previously not included. These additional features may increase its accuracy. As no drop was observed between the use of crystal ligand structures and OMEGA-generated ligand conformations, the FITTED ring search algorithm was not expected to improve the binding mode prediction significantly. A closer look at the data for FITTED reveals that the ring conformation is often searched even when the specific feature is turned off. This can be rationalized by the generation of highly distorted structures and their optimization through energy minimization. As a first conclusion, all the programs but FITTED are very sensitive to the input ligand conformations and the implementation of a ring conformational search engine can reduce this dependency.

When comparing programs, the accuracy does not change much between programs but is led by Surflex (68% with the fully relaxed protein structures), followed by Glide (66% in XP mode), GOLD (65% with ChemScore and flexible rings), FITTED (59%), and FlexX (54% with FlexScore). It should be stressed that this study is carried out with a very difficult testing set including some of the most challenging proteins such as HCV RNA polymerase. From now on, we will only present the data for the best set of conditions for each program unless there is a significant deviation or interesting point of discussion. In fact, the best conditions described in Figure 8 were found to be the best conditions for most of the following studies. In addition, only the OMEGA-generated structures will be discussed as we believe that the data obtained with these represents the true accuracy of the docking programs.

Protein Input Conformations. We next looked at the effect of the protein conformation. The Glide and Surflex developers recommend relaxing the protein structure prior to docking ligands. This procedure (referred to as "refined proteins" in this manuscript, see Table 4) aims at removing any inaccuracies in the crystal structure. However it is often carried out keeping the cocrystallized ligand in place and can be seen as a bias for self-docking experiments. As can be seen in Figure 8, these procedures appeared to have moderate impact on the accuracy for Glide (increase of 4%) but a significant impact with Surflex when the fully refined protein is used. In this later case, an increase of 15% is observed in the standard docking mode and 12% is the advanced docking procedure (pgeom) is used.

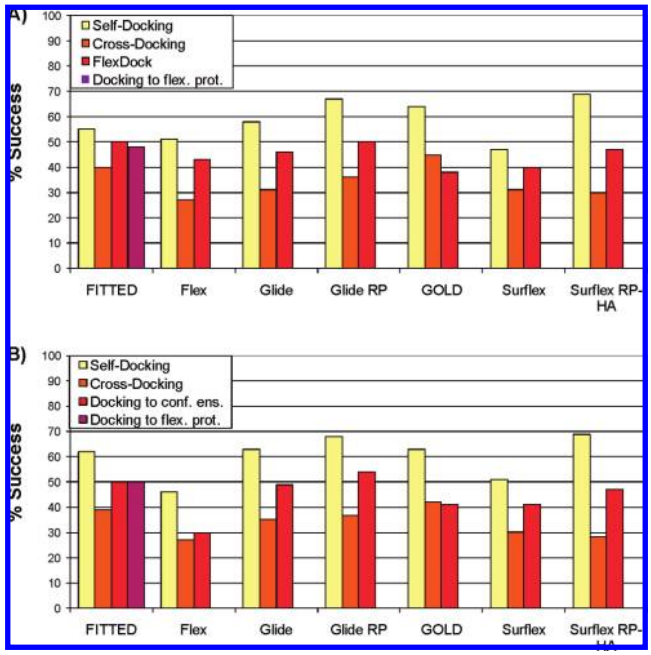

Figure 9. Self-docking vs cross-docking for protein (A) with no waters and (B) with key water molecules.

The study described above was carried out using a set of self-docking experiments (i.e., the protein structure with its native ligand). We next turned our attention to cross-docking experiments because we believe these experiments would be more representative of the true accuracy of the docking programs when performing a virtual screen. As expected, all the programs demonstrated a much poorer accuracy in this set of experiments, with GOLD being the most accurate, although not significantly (Figure 9A). Drops in the cross-docking success rate relative to the self-docking rate of as large as nearly 40% were recorded. The largest drops were attributed to Surflex and Glide when the refined protein conformations were used. Nevertheless, cross-docking to the refined proteins remains slightly more accurate than to the crystal structures with Glide. This observation confirms the developers' recommendation but also demonstrates that this is a clear bias when comparing programs running only self-docking experiments. Such large drops in accuracy between self- and cross-docking have often been observed.[10−13,60,61] When the proteins are considered flexible (by selecting the best scoring poses of the cross-docking experiments, also known as docking to conformational ensembles), the accuracy of all the programs but GOLD is significantly improved (Figure 9).

Within FITTED, the protein can be made flexible without having recourse to multiple runs with run times similar to rigid protein docking. In our previous report, the protein conformational ensembles used to evaluate this feature included the cognate protein structures (i.e., protein conformation when cocrystallized with the ligand to be docked) together with other protein conformations.[35,36] This approach, incorporating both self- and cross-docking experiments in a single run, allowed us to demonstrate that FITTED was able to identify the best protein conformation for a given ligand. However, when evaluating the docking accuracy, we believe that this specific protein conformation should not be included. Thus, in this work, only the non-native protein structures were included. This feature makes FITTED slightly more

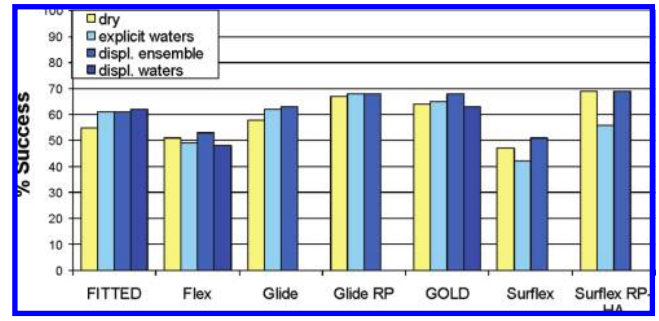

Figure 10. Accuracy and water molecules in self-docking experiments.

accurate than Glide, GOLD, Surflex and FlexX, while FITTED was found to be less accurate in self-docking experiment.

Water Molecules. The three major features of FITTED are protein flexibility, ring conformational search, and displaceable water molecules. At this stage, the importance of the first two had been investigated. But do water molecules significantly affect the accuracy as well? To investigate the role of key water molecules, we carried out self-docking experiments and looked at 4 distinct cases: (i) all waters were removed from the protein crystal structures (no waters), (ii) all key waters were kept (explicit waters), (iii) the best scoring of the no waters and explicit waters experiments were kept, simulating a displaceable ensemble of water molecules (displ. ensemble), and (iv) waters were made displaceable whenever the feature was available (displ. waters). The collected data is shown in Figure 10. A small increase was observed for all the programs when the key waters were kept. When the waters were made displaceable such as for the displ. ensemble and displ. waters experiments, the accuracy further increased. As previously observed by the GOLD developers, the strategy implemented in GOLD did not improve the docking ability of this program.[32] The same observation was made with the FlexX program. In contrast, the displaceable water approach implemented in FITTED was found to be more accurate than the explicit water and even than the displ. ensemble. In fact, displ. ensemble simulates either all the waters on or none while FITTED can displace each water independently. In addition, as the various optimizations of FITTED have been carried out with this feature, the best results are obtained when the displaceable water option is used.

We then investigated the combination of protein flexibility and water molecules. Figure 9B summarizes the data for self- and cross-docking experiments in presence of displaceable waters if implemented and displ. ensemble if not. Once more, a slight improvement is observed with most of the programs indicating that considering both features leads to at least similar or improved results (Figure 9) and should be considered by other developers.

The presence of hydrogen bond donors or acceptors in the protein binding site is expected to help finding the proper orientation of the ligand. In contrast, nondirectional hydrophobic interactions are directly related to the nature of the compound and protein binding site and their respective solvation energies more than to any "real" hydrophobic interactions between protein and ligands. These interactions are therefore expected to be more difficult to identify in a computationally tractable manner. The proteins of our set were classified as polar (CDK2, COX-2, ER, FXa, GluK2,

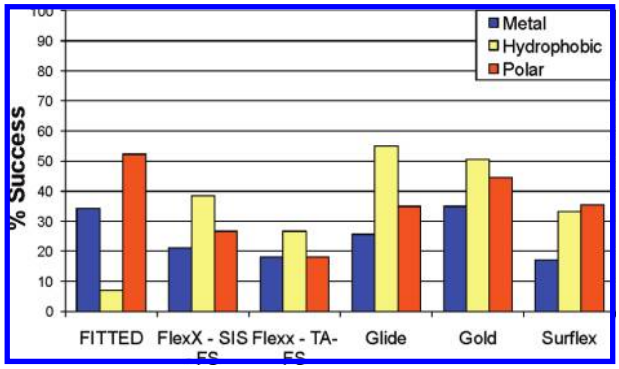

Figure 11. Protein class and accuracy on cross-docking experiments.

HIVP, HIVPD, HIVRT, Thrn, TK, Tryp, VDR), hydrophobic (HCV Allo, HCV Cat, P38), and metal-containing enzymes (Mann, MMP3, Therm), based on the main ISs identified by PROCESS. The cross-docking data was reorganized to account for this factor and the results are summarized in Figure 11. GOLD appears to be fairly insensitive to the protein type, while Surflex and FlexX were much less accurate with metalloenzymes. The automatic metal parameters within GOLD and FITTED may explain the good accuracy with metalloenzymes. More striking is the much greater accuracy of FITTED with hydrophilic enzymes than with hydrophobic enzymes, while Glide and FlexX are significantly more accurate with hydrophobic proteins than with polar proteins. Interestingly, the SIS algorithm in FlexX, developed to improve the accuracy with hydrophobic proteins, lead to increase in accuracy with this class of proteins when compared to the traditional FlexX algorithm. Once more this data indicates that the set used for any comparative study would have a significant effect on the relative accuracies of programs. For instance, FITTED would be the second best program if hydrophilic proteins were selected while being the worst if only hydrophobic proteins were selected.

As a summary, the accuracy of each of the assessed programs using the optimal conditions is shown in Figure 12 for self-docking and cross-docking experiments as well as docking to flexible proteins when available. Overall, the levels of accuracy given here are significantly lower than the ones provided in other comparative studies.[5,6,9] In fact, we found our testing set to be much more challenging than the one we used previously. In addition, part of this drop (10−20%) is directly attributed to the use of OMEGA-generated structures and another part (10−30%) to the use of cross-docking experiments in place of self-docking experiments used elsewhere.

To further assess the impact of such a protein-specific training and the novel placement of ISs implemented in the current version of PROCESS, we have carried out additional experiments. When protein-specific information (e.g., ISs derived from known ligand/protein complexes) is manually given to FITTED, the accuracy increases significantly (data not shown). This clearly demonstrates that this manual placement of ISs-and more specifically hydrophobic sites-for docking with FITTED remains better than the automated placement which should be further improved.

Overall, FITTED, Glide, GOLD, and Surflex show very similar accuracies on our testing set for self-docking (i.e., rigid protein). When protein flexibility is considered (ligands

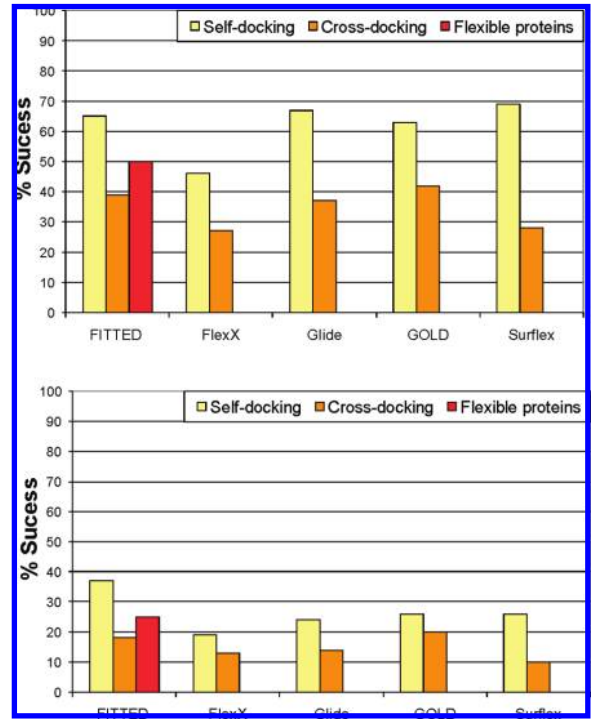

Figure 12. Accuracy of program with OMEGA-generated structures: FITTED, Dock mode; FlexX, ScreenScore and SIS used for the incremental construction; Glide, XP and refined protein; GOLD, ChemScore; Surflex, pgeom and protein with refined hydrogen positions. For the flexible protein, the "fully flexible" protein mode is used with FITTED implementation. (a) Success criterion: rmsd ≤ 2.0 Å for self-docking and 2.25 Å for cross docking; (b) rmsd ≤1.0 Å for self-docking and 1.25 Å for cross docking.

docked to all non-native protein structures in multiple runs and in a single run with FITTED), FITTED is slightly more accurate (the only one featuring displaceable waters and protein flexibility simultaneously), followed by GOLD and Glide. It is worth recalling that FITTED was outperformed by the other programs in the self-docking experiments. It is clear from Figures 9 and 12 that implementing protein flexibility would significantly improve Surflex, Glide, and FlexX accuracy, while no significant improvements are expected for GOLD, which already uses a soft protein representation (Lennard-Jones 8−4).

The following numbers are given as rough estimates as these programs were run on various computers and supercomputers with varying processor speed, and some programs (Surflex, FlexX) do not output the CPU time. In extreme cases, FlexX docks a compound every 30 s, while FITTED is the slowest by a factor of 10 to 15. When the criterion of success is made more stringent (rmsd ≤1.0 Å for self-docking and ≤1.25 Å for cross docking, Figure 12b), FITTED slightly outperforms the other programs for self-docking, and all programs show similar accuracy in cross-docking with this stricter criterion. It also shows that the arbitrary limit of 2 Å does not affect much the ranking of programs by their rmsd-derived accuracy.

From this comparative study, we confirmed that a few guidelines should be considered to perform a proper evaluation: (1) the ligand should be in a conformation other than the crystal structure, (2) both cross-docking and self-docking experiments should be carried out, and (3) refining the protein

structure using the cocrystallized ligand may bias the self-docking accuracy but does not affect the cross-docking accuracy.

CONCLUSION

We have further modified our docking program FITTED and implemented a ring search method into the genetic algorithm as well as a matching algorithm to produce the initial population. This advanced version was tested against major docking programs. It should be stressed that this work was not intended to rank programs as the ranking varies from one set of protein/ligand complexes to another. In fact, this work demonstrated that ranking can significantly vary depending on the protein/ligand set considered (e.g., hydrophilic, hydrophobic) as well as the input ligand and protein conformations (e.g., crystal structures or OMEGA-generated, self- or cross-docking, with or without water molecules). With this study, we demonstrated the impact of protein and ligand conformations as well as protein flexibility and water molecules on the accuracy of docking programs. We have been working on these last two properties for the past few years and have shown herein that these two features significantly improve the accuracy of our docking program FITTED. The placement of hydrophobic interaction sites has been identified as a remaining issue and more work are currently ongoing to better understand and identify hydrophobic pockets. This work may also serve the developers to better understand the weaknesses and strengths of their respective programs.

EXPERIMENTAL SECTION

Preparation of the Docking Set. Structures were downloaded from the PDB[62] and selected on the basis of the diversity of the ligands, presence of water molecules, flexibility of the protein, and resolutions of the crystal structure below 2.5 Å. In some cases, crystal structures with resolutions higher than 2.5 Å were kept to increases the diversity of the conformations seen within a protein. All structures were prepared using Maestro,[63] the graphical interface to the Schrödinger Suite of programs. Structures of the same protein were then superimposed using the protein structure alignment option within Maestro. The protein sequences where then homogenized by mutating and deleting missing residues when at least 10 Å from the binding cavity. If a missing residue was closer than the minimum distance the structure was removed from the set. Hydrogen atoms where added using Maestro and energy-minimized using the OPLS_2005 forcefield. All nonconserved waters were removed from all the structures. Conserved or key water molecules were defined as water molecules that make at least 2 hydrogen bonds with the protein and one with the ligand. The protein and ligand structure were then separated. The ligand crystal structure was used as input into OMEGA[64] to generate new starting conformations. For this study we had OMEGA only output the most thermodynamically stable conformation using all standard default values.

Docking Programs Methodology. A recent review by our group[65] found over 60 docking programs that have been published. It is becoming ever harder to distinguish which program is best for a specific protein or in general. To assess how well FITTED performs compared to other docking programs a small comparative study was undertaken using FlexX, Glide, GOLD, and Surflex. It is worth noting that even though AutoDock[11] or a combination of Glide and Prime[17] can allow for protein flexibility they were not used due to time constraints. Also FlexX does have a module for protein flexibility (FlexX-Ensemble) but the version used in this study was incompatible with FlexX-Ensemble module.

For all docking runs the OMEGA generated ligand conformation was used for self-docking, cross-docking and flexible-protein docking. The crystallized ligand structure was run separately but only the self-docking data is shown. All the docking experiments were performed using dry proteins (no waters present) unless otherwise stated. When proteins structures contain a key water molecule(s) additional docking experiments were preformed to the wet protein (only the key water molecules for that protein crystal structure are present). If the docking program has the ability to dock with displaceable crystallographic waters additional sets of docking experiments were performed to a displaceable water−protein structure (all possible water positions occur). When defining the active site for all the proteins, the largest ligand of the set for a particular protein was used.

In all cases, docking success for docking was measured using the standard rmsd criterion (rmsd between the heavy atoms of the docked posed and the reported crystal structure). During self-docking run we used a criterion of less than 2.0 Å but increased this to 2.25 Å for cross docking to account for the error resulting from the superposition of the proteins. Only the top scoring pose was used for accuracy measures as it would be the one picked in a VS experiment whether the docking run was successful. The rmsd was calculated using the tool provided by the program. One exception was in the case of FlexX and Glide, where the GOLD rmsd script was also used for the calculation of the RMSDs for HIV-1 proteases ligands. Because of the C_2 symmetric nature of the HIV-1 protease binding site it was necessary to calculate the rmsd on 2 orientations of the ligand, the original and the ligand rotated 180°. This second rmsd could not be done since these programs output the rmsd within the output file of the run and could not be recomputed. This second rotated ligand was done by rotation of a duplicated copy of the protein/ligand complex in space and resuperimposition using InsightII. With all programs the rmsd was calculated using both orientations for the HIV-1 protease ligands with the lowest rmsd being kept.

In all cases, we have been in contact with either the developers themselves or the technical support of the programs discussed herein to determine the best conditions for our comparative study. Where they were uncertain, we ran all possibilities.

FlexX 3.1.0.[33,60,66,67] FlexX uses an incremental construction algorithm to build up the ligand within the active site. To determine the placement of the fragments FlexX uses a set of interaction sites then uses a matching algorithm to find the best match between the fragment and the interactions. FlexX can account for displaceable water molecules using the particle water concept where all possible combinations of the water being present or not present are tried and the best scoring combination is kept. We used the FlexX3.1.0 interface to construct all the project files for each individual crystal structure. For each protein the binding site hydrogen positions for the protein were manually oriented to create

optimal hydrogen bond with either the protein and/or native ligand. For each structure where water molecules occur, a project file was created for the dry protein (no waters), the wet protein (waters are treated as spheres) and the displaceable waters protein (waters are considered as spheres and allowed to be displaced). Four settings.pxx files were created so that we could run FlexX through command line interface using FlexXScore, ChemScore, ScreenScore, and PLP scoring functions. Within the. bat file, we would turn on the ring search using the corina_f executable[68] provided by Molecular Networks by using the keyword SET RING_MODE to 1 and/or turn on the SIS docking algorithm by using the PLACEBAS 1 keyword. At the time of this publication, FlexX-Ensemble was not available for FlexX3.1.0 and was deemed not ready for a comparative study by the developers.

Glide4.5.[48,69,70] Glide uses a funnel approach to docking by initially creating a series of ligand conformations then removing the unfavorable ones. With this done a refinement is performed by doing an energy minimization followed by a restricted Monte Carlo search on the lowest energy conformations. This Monte Carlo search is used to refine the initial structure.

The protein structures in mol2 format were prepared using the protein preparation wizard with default values (the proteins in future referred to as refined proteins). Grids were prepared for the initial prepared protein as well as the refined proteins using a 30 Å box with the center of the grid being defined by using the largest ligand of the protein in our set. Default parameters were used to dock with Glide for HTVS, SP, and XP docking modes. With each docking mode, both grids were used individually. With Glide the default is to allow for flexible rings and therefore to study Glide with rigid rings this functionality was turned off.

GOLD3.2.[71] GOLD performs a conformational search of the ligand by using a genetic algorithm. When dealing with displaceable waters GOLD considers all possible combinations of the water being present or not present keeping the best scoring combination. The prepared ligand and proteins were used in mol2 format for GOLD. The automatic settings with the default parameters were used. When docking to the wet protein (proteins with key water molecules), the orientation of the waters is optimized but the waters are not displaced. When docking to displaceable water−protein structures, the waters are set to displaceable and their orientation is optimized. To examine the corner approach the flip_free_corners was set to 1 in the.conf file. Upon discussions with the developers it was suggested to try a more robust search. This was done by using the keywords autoscale = 1.5 and autoscale_nopt_min = 15000. It was also suggested to set early_termination to 0. Both additions were tried and are referred to as the robust search in the results.

Surflex2.3.[72] The Surflex docking algorithm combines a shape matching algorithm with the matching of a protomol that is similar to a pharmacophore. Surflex uses an incremental construction algorithm with relinking of the fragmented ligand. No interface was provided with Surflex which was therefore used in command line. For Surflex the prepared ligands and proteins were used in mol2 format. The protomol was initially generated using the largest ligand of the set for that protein. Surflex was then used to dock using all the default values. To perform the conformational search of rings the +rings command was used. Upon discussion with the developers, it was suggested trying the -pgeom command that is meant to increase docking accuracy and use their program to optimize the hydrogens of the proteins. Docking runs using these suggested conditions were performed.

Table 5. List of Ligands Used to Define Protein Binding Sites

protein	PDB code
CDK2	1aq1
COX-2	4cox
estrogen receptor	1sj0
factor Xa	1nfu
kainate glutamate receptor	1yae
HCV polymerase allosteric pocket	2o5d
HCV polymerase catalytic pocket	2fvc
HIV-1 protease	1pro
HIV-1 protease diols	1hvr
HIV reverse transciptase	1vrt
mannosidase	2f18
MMP-III	1d8m
p38 map kinase	1w82
thermolysin	3tmn
thymidine kinase	1tmt
thrombin	1qhi
trypsin	1qbo
vitamin D receptor	2has

FITTED 2.6.[73] The files describing the proteins, interaction sites and cavity sites were prepared using the PROCESS module while the ligands were prepared using SMART. The created files were next used by FITTED. The default parameters were used with each of these three programs. FITTED is now available at www.Fitted.ca.

ACKNOWLEDGMENT

We thank CIHR and Virochem Pharma for financial support, as well as the Canadian Foundation for Innovation for financial support through the New Opportunities Fund program. C.R.C. held a CIHR-funded Chemical Biology Scholarship during a portion of this study. We are thankful to the RQCHP for allocation of computer resources for this study. We would also like to thank the following people for the input and suggestions on how to improve the docking accuracy of their programs: Ajay Jain, UCSF (SurFlex) and the support departments of both CCDC (GOLD) and Schrödinger (Glide).

Supporting Information Available: Detailed listing of results and the comparative study set. This material is available free of charge via the Internet at http://pubs.acs.org.

Note Added in Proof. In the initial study, eHiTS was included but had to be removed after discussions with the developers who informed us of a problem related to the report of the rmsd values. This problem led to a significant overestimation of the pose prediction accuracy of eHiTS. This problem has been corrected in the most recent version.

REFERENCES AND NOTES

(1) Cozza, G.; Bonvini, P.; Zorzi, E.; Poletto, G.; Pagano, M. A.; Sarno, S.; Donella-Deana, A.; Zagotto, G.; Rosolen, A.; Pinna, L. A.; Meggio, F.; Moro, S. Identification of ellagic acid as potent inhibitor of protein kinase CK2: A successful example of a virtual screening application. *J. Med. Chem.* **2006**, *49* (8), 2363–2366.

(2) De Graaf, C.; Oostenbrink, C.; Keizers, P. H. J.; Van Der Wijst, T.; Jongejan, A.; Vermeulen, N. P. E. Catalytic site prediction and virtual screening of cytochrome P450 2D6 substrates by consideration of water and rescoring in automated docking. *J. Med. Chem.* **2006**, *49* (8), 2417–2430.

(3) Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem.* **2000**, *43* (25), 4759–4767.

(4) Bursulaya, B. D.; Totrov, M.; Abagyan, R.; Brooks Iii, C. L. Comparative study of several algorithms for flexible ligand docking. *J. Comput.-Aided Mol. Des.* **2003**, *17* (11), 755–763.

(5) Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: Comparative data on docking algorithms. *J. Med. Chem.* **2004**, *47* (3), 558–565.

(6) Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins* **2004**, *56* (2), 235–249.

(7) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins* **2004**, *57* (2), 225–242.

(8) Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.* **2005**, *48* (4), 962–976.

(9) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49* (20), 5912–5931.

(10) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337* (1), 209–225.

(11) Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in autodock. *Proteins* **2002**, *46* (1), 34–40.

(12) Murray, C. W.; Baxter, C. A.; Frenkel, A. D. 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* (6), 547–562.

(13) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: The effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47* (1), 45–55.

(14) Cavasotto, C. N.; J.W. Orry, A.; Abagyan, R. A. The challenge of considering receptor flexibility in ligand docking and virtual screening. *Curr. Comput.-Aided Drug Des.* **2005**, *1*, 423–440.

(15) Klebe, G. Virtual ligand screening: Strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11* (13−14), 580–594.

(16) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein−ligand docking: Current status and future challenges. *Proteins* **2006**, *65* (1), 15–26.

(17) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49* (2), 534–553.

(18) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FLEXE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308* (2), 377–395.

(19) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Perspect. Drug Discovery* **2000**, *20*, 171–190.

(20) Zavodszky, M. I.; Lei, M.; Thorpe, M. F.; Day, A. R.; Kuhn, L. A. Modeling correlated main-chain motions in proteins for flexible molecular recognition. *Proteins* **2004**, *57* (2), 243–261.

(21) Sotriffer, C. A.; Dramburg, I. "In situ cross-docking" to simultaneously address multiple targets. *J. Med. Chem.* **2005**, *48* (9), 3122–3125.

(22) Li, Z.; Lazaridis, T. Water at biomolecular binding interfaces. *Phys. Chem. Chem. Phys.* **2007**, *9* (5), 573–581.

(23) Baldwin, E. T.; Bhat, T. N.; Gulnik, S.; Liu, B.; Topol, I. A.; Kiso, Y.; Mimoto, T.; Mitsuya, H.; Erickson, J. W. Structure of HIV-1 protease with KNI-272, a tight-binding transition-state analog containing allophenylnorstatine. *Structure* **1995**, *3* (6), 581–590.

(24) Wang, Y. X.; Freedberg, D. I.; Wingfield, P. T.; Stahl, S. J.; Kaufman, J. D.; Kiso, Y.; Bhat, T. N.; Erickson, J. W.; Torchia, D. A. Bound water molecules at the interface between the HIV-1 protease and a potent inhibitor, KNI-272, determined by NMR. *J. Am. Chem. Soc.* **1996**, *118* (49), 12287–12290.

(25) Kervinen, J.; Thanki, N.; Zdanov, A.; Tino, J.; Barrish, J.; Lin, P. F.; Colonno, R.; Riccardi, K.; Samanta, H.; Wlodawer, A. Structural analysis of the native and drug-resistant HIV-1 proteinases complexed with an aminodiol inhibitor. *Protein Pept. Lett.* **1996**, *3* (6), 399–406.

(26) Hong, L.; Zhang, X. J.; Foundling, S.; Hartsuck, J. A.; Tang, J. Structure of a G48H mutant of HIV-1 protease explains how glycine-48 replacements produce mutants resistant to inhibitor drugs. *FEBS Lett.* **1997**, *420* (1), 11–16.

(27) Louis, J. M.; Dyda, F.; Nashed, N. T.; Kimmel, A. R.; Davies, D. R. Hydrophilic peptides derived from the transframe region of Gag-Pol inhibit the HIV-1 protease. *Biochemistry* **1998**, *37* (8), 2105–2110.

(28) Lam, P. Y. S.; Jadhav, P. K.; Eyermann, C. J.; Hodge, C. N.; Ru, Y.; Bacheler, L. T.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Wong, Y. N.; Chang, C. H.; Weber, P. C.; Jackson, D. A.; Sharpe, T. R.; Erickson-Viitanen, S. Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors. *Science* **1994**, *263* (5145), 380–384.

(29) Grzesiek, S.; Bax, A.; Nicholson, L. K.; Yamazaki, T.; Wingfield, P.; Stahl, S. J.; Eyermann, C. J.; Torchia, D. A.; Nicholas Hodge, C.; Lam, P. Y. S.; Jadhav, P. K.; Chang, C. H. NMR evidence for the displacement of a conserved interior water molecule in HIV protease by a non-peptide cyclic urea-based inhibitor. *J. Am. Chem. Soc.* **1994**, *116* (4), 1581–1582.

(30) Hodge, C. N.; Aldrich, P. E.; Bacheler, L. T.; Chang, C. H.; Eyermann, C. J.; Garber, S.; Grubb, M.; Jackson, D. A.; Jadhav, P. K.; Korant, B.; Lam, P. Y. S.; Maurin, M. B.; Meek, J. L.; Otto, M. J.; Rayner, M. M.; Reid, C.; Sharpe, T. R.; Shum, L.; Winslow, D. L.; Erickson-Viitanen, S. Improved cyclic urea inhibitors of the HIV-1 protease: Synthesis, potency, resistance profile, human pharmacokinetics and X-ray crystal structure of DMP 450. *Chem. Biol.* **1996**, *3* (4), 301–314.

(31) Moitessier, N.; Westhof, E.; Hanessian, S. Docking of aminoglycosides to hydrated and flexible RNA. *J. Med. Chem.* **2006**, *49* (3), 1023–1033.

(32) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W. M.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.* **2005**, *48* (20), 6504–6515.

(33) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: Placing discrete water molecules during protein- ligand docking predictions. *Proteins* **1999**, *34* (1), 17–28.

(34) 14.3.5 Mapping amino acids to templates. In *FlexX Release 3 with GUI User Guide and Technical Reference*; BiosolveIT GmbH: Sankt Augustin, Germany, 2007; p 310.

(35) Corbeil, C. R.; Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47* (2), 435–449.

(36) Corbeil, C. R.; Englebienne, P.; Yannopoulos, C. G.; Chan, L.; Das, S. K.; Bilimoria, D.; Heureux, L.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 2. Development and application of FITTED 1.5 to the virtual screening of potential HCV polymerase inhibitors. *J. Chem. Inf. Model.* **2008**, *48* (4), 902–909.

(37) Englebienne, P.; Fiaux, H.; Kuntz, D. A.; Corbeil, C. R.; Gerber-Lemaire, S.; Rose, D. R.; Moitessier, N. Evaluation of docking programs for predicting binding of Golgi α-mannosidase II inhibitors: A comparison with crystallography. *Proteins* **2007**, *69* (1), 160–176.

(38) Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection−What can we learn from earlier mistakes. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3), 213–228.

(39) Good, A.; Oprea, T. Optimization of CAMD techniques 3. Virtual screening enrichment studies: A help or hindrance in tool selection. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3), 169–178.

(40) Jain, A. N. Bias, reporting, and sharing: Computational evaluations of docking methods. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3−4), 201–212.

(41) Jain, A. N. Bias, reporting, and sharing: Computational evaluations of docking methods. *J. Comput.-Aided Mol. Des.* **2007**, 1–12.

(42) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* **2007**, *50* (4), 726–741.

(43) Boström, J.; Greenwood, J. R.; Gottfries, J. Assessing the performance of OMEGA with respect to retrieving bioactive conformations. *J. Mol. Graph. Modell.* **2003**, *21* (5), 449–462.

(44) Gasteiger, J.; Sadowski, J.; Schuur, J.; Selzer, P.; Steinhauer, L.; Steinhauer, V. Chemical information in 3D Space. *J. Chem. Inf. Comput. Sci.* **1996**, *36* (5), 1030–1037.

(45) Boström, J. Reproducing the conformations of protein-bound ligands: A critical evaluation of several popular conformational searching tools. *J. Comput.-Aided Mol. Des.* **2001**, *15* (12), 1137–1152.

(46) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22* (3−4), 133–139.

(47) Jain, A. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J. Comput.-Aided Mol. Des.* **2007**, *21* (5), 281–306.

(48) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry,

J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47* (7), 1739–1749.
(49) 5.2.3 Setting Docking Options. In *Glide 5.0 User Manual*, Schrödinger, LLC.: Portland, Oregon, 2008; p 44.
(50) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245* (1), 43–53.
(51) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
(52) Payne, A. W. R.; Glen, R. C. Molecular recognition using a binary genetic search algorithm. *J. Mol. Graph* **1993**, *11* (2), 7491 + 121.
(53) Goto, H.; Osawa, E. Corner flapping: A simple and fast algorithm for exhaustive generation of ring conformations. *J. Am. Chem. Soc.* **1989**, *111* (24), 8950–8951.
(54) 5.2.3 Setting Docking Options. In *Glide 4.5 User Manual*, Schrödiger, LLC.: 2007; p 42.
(55) Thompson, H. B. Calculation of cartesian coordinates and their derivatives from internal molecular coordinates. *J. Chem. Phys.* **1967**, *47* (9), 3407–3410.
(56) Harvey, M. A.; Baggio, S.; Baggio, R. A new simplifying approach to molecular geometry description: The vectorial bond-valence model. *Acta Crystallogr. B* **2006**, *62* (6), 1038–1042.
(57) O'Boyle, N. M.; Brewerton, S. C.; Taylor, R. Using buriedness to improve discrimination between actives and inactives in docking. *J. Chem. Inf. Model.* **2008**, *48* (6), 1269–1278.
(58) Moitessier, N.; Henry, C.; Maigret, B.; Chapleur, Y. Combining pharmacophore search, automated docking, and molecular dynamics simulations as a novel strategy for flexible docking. Proof of concept: Docking of arginine-glycine-aspartic acid-like compounds into the $\alpha_v\beta_3$ binding site. *J. Med. Chem.* **2004**, *47* (17), 4178–4187.
(59) Chen, H.; Lyne, P. D.; Giordanetto, F.; Lovell, T.; Li, J. On evaluating molecular-docking methods for pose prediction and enrichment factors. *J. Chem. Inf. Model.* **2006**, *46* (1), 401–415.
(60) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, *37* (2), 228–241.
(61) Birch, L.; Murray, C. W.; Hartshorn, M. J.; Tickle, I. J.; Verdonk, M. L. Sensitivity of molecular docking to induced fit effects in influenza virus neuraminidase. *J. Comput.-Aided Mol. Des.* **2002**, *16* (12), 855–869.
(62) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J. B. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112* (3), 535–542.
(63) *Maestro, 8.0*; Schrödinger, LLC.: Portland, OR, 2007.
(64) *OMEGA, 2.2.1*; Open Eye Scientific Software: Sante Fe, NM, 2007.
(65) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: A long way to go. *Br. J. Pharmacol.* **2008**, *153*, S7–S26, SUPPL. 1.
(66) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261* (3), 470–489.
(67) *FlexX, 3.1.0*; BioSolveIT: Sankt Augustin, Germany, 2008.
(68) *Corina_F, 3.4*; Molecular Networks: Erlangen, Germany, 2008.
(69) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47* (7), 1750–1759.
(70) *Glide, 4.5*; Schrödinger, LLC.: Portland, OR, 2007.
(71) *GOLD, 3.2*; Cambridge Crystallographic Data Centre: Cambridge, U.K., 2007.
(72) *Surflex, 2.3*; BioPharmics, LLC: San Fransico, CA, 2008.
(73) Corbeil, C. R.; Englebienne, P.; Moitessier, N. FITTED, version 2.6; McGill University: Montreal, Quebec, Canada, 2008.

CI8004176