

## Lessons in Molecular Recognition: The Effects of Ligand and Protein Flexibility on Molecular Docking Accuracy

Jon A. Erickson,<sup>\*,†</sup> Mehran Jalaie,<sup>†,‡</sup> Daniel H. Robertson,<sup>†</sup> Richard A. Lewis,<sup>§</sup> and Michal Vieth<sup>†</sup>

Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, Indiana 46285, and Lilly Research Centre, Sunninghill Road, Windlesham, GU20 6PH, UK

Received April 29, 2003

The key to success for computational tools used in structure-based drug design is the ability to accurately place or “dock” a ligand in the binding pocket of the target of interest. In this report we examine the effect of several factors on docking accuracy, including ligand and protein flexibility. To examine ligand flexibility in an unbiased fashion, a test set of 41 ligand–protein cocomplex X-ray structures were assembled that represent a diversity of size, flexibility, and polarity with respect to the ligands. Four docking algorithms, DOCK, FlexX, GOLD, and CDOCKER, were applied to the test set, and the results were examined in terms of the ability to reproduce X-ray ligand positions within 2.0Å heavy atom root-mean-square deviation. Overall, each method performed well (>50% accuracy) but for all methods it was found that docking accuracy decreased substantially for ligands with eight or more rotatable bonds. Only CDOCKER was able to accurately dock most of those ligands with eight or more rotatable bonds (71% accuracy rate). A second test set of structures was gathered to examine how protein flexibility influences docking accuracy. CDOCKER was applied to X-ray structures of trypsin, thrombin, and HIV-1-protease, using protein structures bound to several ligands and also the unbound (apo) form. Docking experiments of each ligand to one “average” structure and to the apo form were carried out, and the results were compared to docking each ligand back to its originating structure. The results show that docking accuracy falls off dramatically if one uses an average or apo structure. In fact, it is shown that the drop in docking accuracy mirrors the degree to which the protein moves upon ligand binding.

### Introduction

X-ray crystal structures of ligand–protein cocomplexes have been important tools for medicinal chemists in the discovery, design, and optimization of drug candidates.<sup>1–3</sup> These structural data along with the computational analysis tools that have been developed to implement structure-based drug design (SBDD) have proved to be very successful in medicinal chemistry.<sup>4</sup> As a greater number of X-ray crystal structures become available to medicinal chemists, with the advent of structural genomics,<sup>5</sup> computational methods that take advantage of protein–ligand structural data are becoming more critical to the drug design process. In the past 10 years, many robust computational methods have been developed to design, optimize,<sup>6–8</sup> and screen databases for drug lead molecules.<sup>9</sup> Gaining an understanding of the ligand–protein binding event is at the core of these tools and critical to subsequent translation of structure-based information into more potent and drugable ligands. The computational representation of protein–ligand molecular recognition is commonly referred to as “molecular docking”.

X-ray crystal structures of a ligand bound to a protein provide data on the position and thus interactions and potential interactions between the ligand and the

protein.<sup>10</sup> A detailed understanding of the essential interactions making up a protein’s recognition of the ligand can be very useful in terms of structure–activity relationships (SAR), which in turn are important for drug design and optimization. Molecular docking<sup>11–14</sup> methods attempt to place a ligand in the binding pocket of a protein. The many docking methods that have emerged differ mainly with respect to how they treat placement of the ligand, conformational space exploration, representation of protein ligand interactions, and binding affinity estimation (i.e. scoring). These methods are useful tools in drug design, since they help to provide an understanding of the key interactions made by a known ligand or to examine the potentially new interactions of a designed ligand. Due to the importance of understanding and predicting these interactions to the success of virtual screening and structure-based drug design, accurate ligand placement is highly desirable.

How accurate are current docking methods? The only way to verify the accuracy of a ligand binding mode predicted by docking is to compare it to the X-ray crystal structure. Many docking validation studies have been carried out to reproduce the binding modes of protein ligand X-ray structure test sets. The typical procedure involves removal of the ligand from an X-ray structure and then attempting to reproduce the X-ray crystal orientation and conformation with a particular docking method. This is often referred to as the “bound” docking problem.<sup>12</sup> Since it is established that small molecules can and often do change conformations upon binding to a protein,<sup>15</sup> most of these methods incorporate a way

\* Corresponding author. Phone: 317-433-2323. FAX: 317-276-6545. E-mail: jae@lilly.com.

<sup>†</sup> Lilly Research Laboratories.

<sup>‡</sup> Current address: Pfizer Global Research & Development, Ann Arbor, MI.

<sup>§</sup> Lilly Research Centre.

to treat the ligands flexibly during docking. These studies are not exhaustively summarized herein; however, a discussion of a few examples of some commonly used docking routines that represent the current state of docking accuracy puts the current study in context. Although the focus here is on docking accuracy and not scoring, it is very difficult to separate the two cleanly. Many docking algorithms will suggest several low-scoring solutions, but it is difficult to select the "correct" one in the absence of the known solution in a systematic and reproducible fashion. Therefore, only the accuracy of the top-scoring docked orientation is relevant for systematic comparison of methods. In terms of docking accuracy, a threshold of 1.0–3.0 Å root-mean-square deviation (rmsd) between the docked and X-ray pose has been generally considered to be a "successfully" docked structure.<sup>11–14</sup> A value of 2.0 Å was suggested by Gohlke et al.<sup>16</sup> based on the restrictions of crystal structure resolution and will be used here as a measure of docking accuracy success.

One of the first and still widely used docking algorithms is DOCK, developed by Kuntz et al.<sup>17–21</sup> In a recent validation of their treatment of ligand flexibility,<sup>22</sup> they examined docking accuracy on a test set of 12 protein–ligand X-ray structures. This validation set was a good test of their incremental construction method,<sup>20</sup> because it contained ligands with a good range in flexibility (2–13 rotatable bonds). An important detail in docking validation studies is the treatment of the protein and ligand structures. Manipulation of these structures prior to docking can have a dramatic impact on docking accuracy (vide infra). In their work, the starting ligand and protein structures for docking were treated uniformly. The starting ligand conformations were taken from the X-ray crystal structure using atomic charges assigned by the Gasteiger–Marsilli method.<sup>23,24</sup> The protein conformations were also used directly from the X-ray structure; however, care was taken to assign the protonation state of active site histidines. In three cases, the binding pocket was treated with and without waters or cofactors. Their results showed that six of the 12 docked coordinates of the ligands were predicted to be within 2.0 Å rmsd from the X-ray structure. A more comprehensive evaluation of docking accuracy using DOCK was carried out by Tao and Lai.<sup>25</sup> They used a test set of 200 ligand–protein X-ray structures and also found that DOCK was able to successfully dock just over 50% (105 out of 200) of the cases.

FlexX<sup>26</sup> is another widely used docking algorithm in drug design. The initial validation set for FlexX included 19 cocomplex X-ray structures. The starting ligand geometries were taken from the X-ray structure and minimized using SYBYL<sup>27</sup> prior to docking. Preparation of the protein for FlexX requires definition of the binding pocket in terms of "interaction points". In this work the active sites were defined as all atoms within a distance of 6.5–8.0 Å from the bound ligand. The specific distance for each was determined individually for each test case in order to ensure that the active site was enclosed through visual inspection. In addition, active site side chain hydroxyl hydrogen atom torsions were adjusted through manual intervention. The results showed that 14 of the 19 docked ligands achieved a rmsd

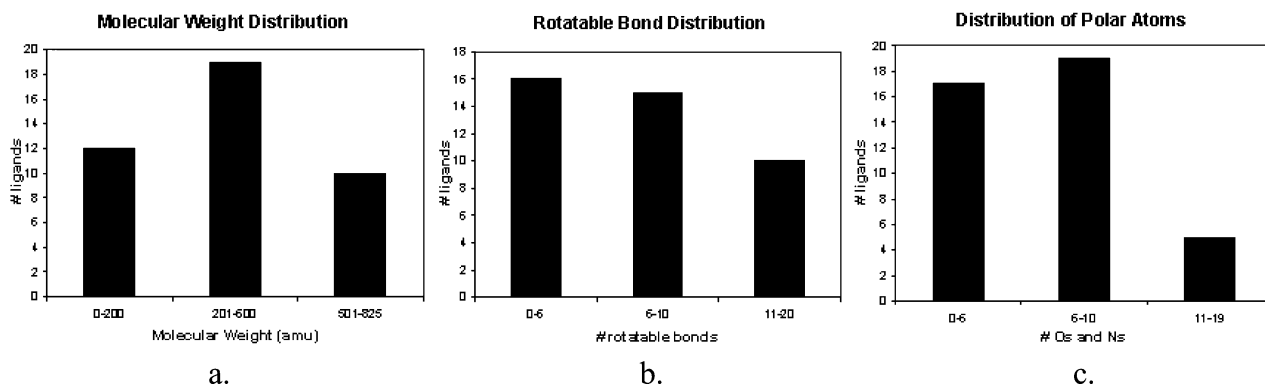
between the X-ray of <2.0 Å, an impressive 74% docking accuracy success rate.

Another commonly used docking algorithm in drug design is GOLD.<sup>28</sup> An extensive test set of 100 carefully selected ligand–protein cocomplexes was used to examine the docking accuracy of the GOLD method. The ligands represented variation in size (6–55 heavy atoms) and flexibility (0–30 rotatable bonds). The starting geometries of the ligands were mostly constructed using the SYBYL BUILD<sup>27</sup> module, taking care to assign tautomeric and ionization states according to previous reports. The conformation used as a starting point was not specified. The proteins were also treated on a case-by-case basis, ensuring appropriate tautomeric and ionization states. The definition and size of each active site was individually determined at radii of 5–15.5 Å to ensure the full site was included. Of the 100 ligands in the test set, GOLD was able to dock 66 of them within 2.0 Å of the X-ray crystal determined solution.

In most of these docking methods, the ligand is treated as flexible, but the protein conformation is restricted. It is well-known, however, that protein recognition of a ligand is a dynamic event in which both ligand and protein change conformation in order to maximize total free energy during the association.<sup>29</sup> Despite this, most docking methods employ the rigid protein approximation. This assumption is used mainly due to practical reasons, since the potential search space rapidly becomes impracticable when considering both the ligand and the protein flexibly.<sup>11</sup> Nevertheless, methods have been implemented that allow for partial treatment of protein flexibility.<sup>12,14</sup>

Recently, the effect of protein flexibility on docking accuracy has been accessed using the "unbound" docking problem, that is, docking ligands to a X-ray structure with no or a different ligand bound.<sup>12</sup> Claussen et al.<sup>30</sup> described docking experiments on 10 different proteins each with four or more available X-ray crystal structures with or without ligands bound. They compared a new method (FlexE) for treatment of protein flexibility using a united protein structure constructed from multiple X-ray structures to docking each ligand into all the proteins ("cross-docking") sequentially with FlexX. Results for FlexE and the FlexX cross-docking experiments showed similar accuracies (67% vs 63%), but FlexE was significantly faster. They did not report the comparison of these methods to docking all the ligands to one structure, but it was shown that at least in one case (aldose reductase) using only one structure decreased docking accuracy. In another report, Murray et al.<sup>31</sup> performed cross-docking experiments on the X-ray structures of three proteins (thrombin, thrombolysin, and influenza virus neuraminidase) bound to at least six different ligands. They found a large drop off in docking accuracy when the ligands were docked to the structures other than their own (76–49%).

In this report, we examine the effect of important force field parameters, treatment of ligand flexibility, and the conformation of the protein on the ability of docking algorithms to accurately reproduce X-ray ligand orientations. Specifically, two test sets of protein–ligand complex X-ray crystal structures were assembled. The first is made up of structures with a diverse set of



**Figure 1.** Distribution of molecular properties in test set 1: (a) size (molecular weight), (b) flexibility (number of rotatable bonds), and (c) polarity (number of O's and N's).

ligands with respect to size, polarity, and flexibility. This set was used to examine the variation of ligand atomic charge, dielectric constant, and nonbonded potential on the docking accuracy of three algorithms, DOCK, CDOCKER, and GOLD. This test set of X-ray structures was also used to examine the influence that ligand flexibility has on docking accuracy. First, the DOCK algorithm was used to compare flexible ligand and rigid ligand docking in terms of accuracy. Furthermore, the effect of starting conformation on the docking accuracy of DOCK, CDOCKER, and GOLD was examined by docking ligands with two different starting conformations, one from the experimental structure and the other generated computationally. Finally, the docking accuracy of various flexible ligand treatment methods was compared by applying DOCK, CDOCKER, GOLD, and FlexX to the test set.

The second test set consists of three proteins, each with several ligand–complex X-ray crystal structures. Using this test set, the ability of CDOCKER to reproduce X-ray crystal structure orientations when ligands are docked to various conformations of the protein was examined. Specifically, the ligands for each protein were docked to their original structure, to an “average” structure, and to the nonliganded or apo structure.

## Methods

**(1) Data Sets.** Two sets of X-ray crystal structures were chosen for this analysis from the Protein Data Bank.<sup>32</sup> The first involves 41 protein–ligand complexes with ligands of varying structural diversity designed to test the ability of the docking algorithms to reproduce X-ray orientations (See Table 1). This group of structures was taken primarily from the work of Eldridge et al.<sup>33</sup> and is similar to other sets used<sup>26,28</sup> to evaluate docking accuracy. To provide an objective comparison of methods, all complexes required uniform preparation and treatment. Due to the various abilities of each docking algorithm to handle nonstandard cases such as large polypeptides, oligosaccharides, or metals ions, complexes including these species were not included. An additional goal of this work is to examine the effect of ligand conformation treatment on docking accuracy. To this end, a set of ligands with an even distribution of size, flexibility, and polarity is ideal (See Figure 1). Test set 1 will be used to compare the effects of various force field parameters and ligand starting conformation using DOCK,<sup>17–21</sup> CDOCKER,<sup>34–37</sup> and GOLD.<sup>28</sup> Test set 1 will also be used to compare various treatments of ligand flexibility using these methods along with the FlexX<sup>26</sup> program.

A second validation set of protein–ligand cocomplexes X-ray structures was organized to explore the effect of protein conformation on docking accuracy. For this purpose it is necessary to have access to structures of several ligands bound

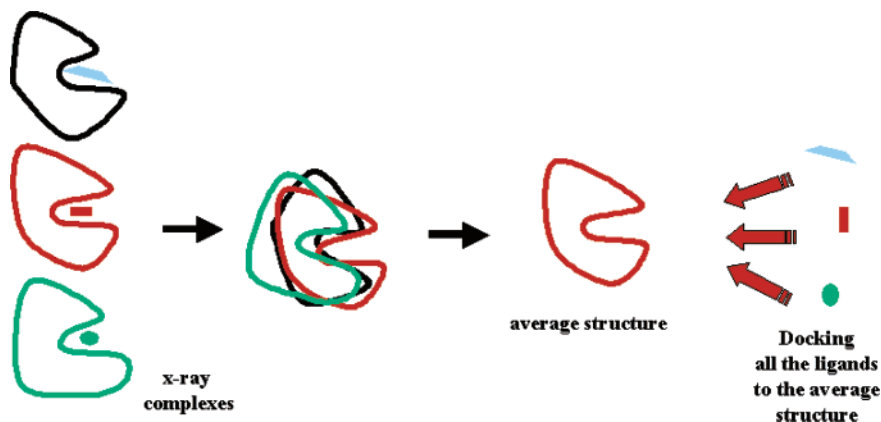
**Table 1.** Composition of X-ray Structure Docking Test Sets 1 and 2

test set	PDB code	protein
1	1abe 1abf 1aoc 1apu 1dbb 1dbj 1dmp 1dog 1dwb 1epo 1ets 1ett 1fax 1hvp 1hsl 1htf 1hvr 1mtw 1nsd 1pgp 1pph 1qbr 1qbt 1qbu 1stp 1tng 1tnh 1ulb 1uvs 1uvt 2cgr 2gbp 2ifb 2phh 2r04 2ypi 3ptb 3tpi 4dfr 4tpi 5abp	assorted
2	1mtw 1az8 1mtu 1tng 3ptb 1bjv 1bju 1tnh 1tni 1tnl	trypsin
2	1a46 1a4w 1abj 1ae8 1afe 1ba8 1bb0 1bcu 1bhx 1bmm 1bmn 1d4p 1dwb 1dwc 1dwd 1dwe 1etr 1ets 1ett 1fpc 1fph 1hai 1tom 1ycp 2hgt 3hat	thrombin
2	1hvp 1hvl 1hvi 1hvs 1hvj 1hvc 1hvk 1dif 1pro 1ajx 1bvg 1htf 1htg 1hbv 1sbg 1hpx 1hsg 1ajv 1hos 1aaq 1hps 2upj 1gnm 1gnn 1gno 1hvr 7hvp 1yth 4hvp 1mtr 9hvp 1cpi 1hiv 1odx 1ytg 1hxb	HIV-1 protease

to the same protein as well as the structure with no ligand bound (apo structure). The data for three proteases, trypsin, thrombin, and the protease of human immunodeficiency virus 1 (HIV1-p), were chosen as a test set for examining the effect of protein conformation (See Table 1) on docking accuracy.

For both test sets, the protein–ligand cocomplexes were prepared uniformly for docking in order to minimize introduction of bias. For each complex, the ligand, solvent, and any cofactors were removed, leaving only the protein. To eliminate differential effects of the various force fields and minimization schemes in each docking algorithm, no minimization was carried out. Standard charge sets were used for each protein, namely those from CHARMM (the param19/top19 parameter set,<sup>38</sup> polar hydrogens only) for CDOCKER, and AMBER<sup>39</sup> charges for DOCK. Protonation states were assumed to be those most common at pH 7, i.e., lysines, arginines, aspartates, and glutamates are ionized.

**(2) Force Field Parameter Comparison.** Molecular mechanics parameters were examined to query their effect on the ability to reproduce X-ray structure binding modes. The comparison was carried out with DOCK,<sup>17–21</sup> CDOCKER,<sup>34–37</sup> and GOLD<sup>28</sup> to explore how these parameters effect docking accuracy across different methods. The parameters examined included nonbond potential, atomic charge, and dielectric constant. For each complex, the ligand was docked to the protein from which it was removed (“self-docking”) for a range of the particular force field parameter under examination. In each case, all other parameters were held at a constant value. The effect on docking accuracy was compared between four methods of calculating ligand atomic charges, Momany–Rone template charges,<sup>40</sup> CVFF, CVFF91,<sup>41</sup> and Gasteiger–Marsilli charges,<sup>23,24</sup> for the three docking methods. For DOCK and CDOCKER, the dielectric constant was varied between 1 and 4 to examine the effect of electric field potentiation and, in effect, the protein charges. For DOCK, two nonbonded poten-



**Figure 2.** Average protein structure selection.

tials were examined to evaluate the effect on ligand-docking accuracy since the use of “softer” 6–9 nonbonded potential has been useful for in-house virtual screening experiments. To test the generality of this observation, a systematic comparison of docking accuracy between the 6–9 and 6–12 was carried out. Since the CDOCKER method utilizes soft-core potentials as an integral part of the method, this effect was not explored.

**(3) Treatment of Ligand Flexibility.** The effect of ligand flexibility on docking accuracy was examined from two points of view, namely, comparison of the choice of starting conformation and the method of conformational sampling. Two starting conformations were used: the first was taken directly from the cocomplex X-ray crystal structure and the second was generated from Corina,<sup>42</sup> a standard 3D conformation generator commonly used for database construction for docking-based virtual screening. For each of the 41 ligands, Corina was used to generate a 3D conformation and to add hydrogens for use as a starting conformation for docking. The X-ray conformations were extracted from the pdb file and, after adding hydrogens, used as input for docking. The various methods for handling ligand flexibility in the algorithms examined include incremental build-up (DOCK and FlexX), molecular dynamics simulated annealing (CDOCKER), and genetic algorithm (GOLD). When possible (i.e. DOCK), comparison was also made to rigid docking, i.e., no conformational searching used in the docking process, to see the effects of conformational searching during docking.

**(4) Comparison of Protein Conformation.** For each of the proteins in training set 2, three docking experiments were carried out to determine how the protein conformation affects ligand-docking accuracy. The first experiment was to simply redock the ligand back to the protein structure it was extracted from (“self” docking), as was carried out with test set 1. This docking experiment, however, is a best case scenario for ligand-docking accuracy, as any protein conformational changes have already been induced for that particular ligand, but it serves as a useful benchmark for comparison to the other docking experiments. In a standard virtual screening or structure-based design case, a structure with a different or no inhibitor is typically available. How does the use of this type of protein structure affect docking accuracy? The second and third docking experiments were carried out to address this question. To simulate docking to a protein structure with another ligand bound, an “average” protein structure was selected for each protease. For each protein, all the structures were aligned and superimposed as depicted in Figure 2 using QUANTA.<sup>43</sup> The structure that was closest to the mean of the active site coordinates was dubbed the “average structure.” The three structures selected from the sets as the “average structure” were **1mtu** for trypsin, **1hvi** for HIV1-p, and **1ets** for thrombin. Finally, a comparison of docking to the apo structures for each protein was included. The apo structures are defined as those crystallized without an inhibitor. The structures chosen for the apo experiments were **1aks** for trypsin, **1hgt** for thrombin, and **3phv** for HIV1-p. The degree of conformational change that occurs upon ligand binding to these three proteins

**Table 2.** Comparison of Ligand and Protein Flexibility in Test Set 2 X-ray Structures

protein	av no. of ligand rotatable bonds	average rmsd (Å) <sup>a</sup>	
		cocomplexes	apo
trypsin	5	0.15	1.6
thrombin	13	0.31	1.0
HIV-1p	19	0.73	2.0

<sup>a</sup> Average root-mean-square deviation (rmsd) between binding pockets of ligand–protein cocomplexes themselves and the apo structure (see text for detailed explanation).

was compared by calculating the root-mean-square deviation (rmsd) of the aligned active sites. Table 2 shows that the three proteins display a range of flexibility in response to ligand binding demonstrated by the range of rmsd values between the active site coordinates of the ligand-complexed structures and the apo structure. This allows an examination of the effect of subtle active site residue movement on the accuracy of ligand docking.

**(5) Docking Methods.** To help generalize the effects of ligand and protein conformation on docking accuracy, it is necessary to compare the effects using a variety docking algorithms, all of which use a rigid protein structure. This constraint is a known weakness of most common docking algorithms; however, it is commonly employed in order to reduce conformational search space. In addition to DOCK,<sup>17–21</sup> CDOCKER,<sup>34–37</sup> and GOLD,<sup>28</sup> FlexX<sup>26</sup> was applied to test set 1. Each algorithm employs alternative ways of scoring and treating ligand flexibility. Performance of these methods on docking accuracy may help to generalize conclusions regarding the role of ligand complexity on docking accuracy. Unlike DOCK, CDOCKER, and GOLD, the various docking parameters were not explored for FlexX as they have been described elsewhere.<sup>26</sup> For each docking program, a brief description of the method and the parameters used in this work are given below.

**DOCK 4.0.1.** DOCK<sup>17–21</sup> utilizes a sphere-matching algorithm to fit ligand atoms to spheres in the binding pocket. These spheres are complementary to the receptor molecular surface, i.e., a negative image of the binding pocket. This study utilized the SPHGEN<sup>44</sup> algorithm to create spheres for each of the 41 complexes in test set 1. All the spheres within a 7 Å radius of the ligand centroid were used in this work. DOCK employs an incremental build algorithm<sup>20</sup> (the “anchor-first method”) to treat ligand flexibility. DOCK utilizes a force field scoring method that is grid-based for increased speed. Grids were calculated for each protein using a 0.3 Å spacing that fully enclosed the spheres. An all-atom form of the protein and AMBER<sup>39</sup> atomic charges were used. Preliminary comparisons show very little difference between AMBER and CHARMM on docking accuracy with test set 1. Except for those variables examined here (ligand charge, dielectric, nonbond potential) the DOCK matching, flexibility, force field, and minimization options were developed to provide optimal docking accuracy

**Table 3.** DOCK Parameter Options Used for Docking Experiments on Test Set 1

type <sup>a</sup>	parameter	value
matching	maximum no. of orientations	500
matching	minimum nodes	3
matching	maximum nodes	6
flexibility	configurations per cycle	100
flexibility	peripheral search	yes
flexibility	torsion drive	yes
flexibility	torsion minimize	yes
flexibility	reminimize anchor	yes
flexibility	minimize anchor	yes
flexibility	reminimize ligand	yes
scoring	energy cutoff distance	10
scoring	bump filter	yes
scoring	bump maximum	8
scoring	distance dielectric	yes
scoring	atom model	all

<sup>a</sup> Flexibility options apply only to flexible-ligand-docking results.

on test 1, as shown in Table 3. All other variable parameters were set at the default values.

**CDOCKER.** CDOCKER is an in-house docking method that has been described previously.<sup>34–37</sup> In general, CDOCKER generates ligand “seeds”<sup>50</sup> to populate the binding pocket. Each seed is then subjected to high temperature MD using a modified version of CHARMM. For the initial stage MD, a soft-core potential is used. Each of the structures from the MD run are then located and fully minimized. The solutions are then clustered according to position and conformation and ranked by energy. CHARMM charges are used for the protein structure, i.e., the param19/toph19 parameter set<sup>38</sup> using only polar hydrogens. Initial docking experiments on test set 1 showed little difference when AMBER charges were used. CDOCKER only allows for flexible ligand treatment.

**FlexX (version 1.10.1 L).** FlexX<sup>26</sup> relies on a “triples” matching algorithm between interaction sites on the ligand and the protein. The interaction sites are defined as hydrogen-bond-donor and -acceptor groups as well as hydrophobic contacts such as aromatic rings and methyl groups. Interactions take place for complimentary matches between protein and ligand sites. Ligand flexibility is treated by an incremental build-up method that uses the MINUMBA<sup>45</sup> conformer library to grow ligands during the docking process. Here the FlexX method as implemented in SYBYL<sup>27</sup> was applied to test set 1 using most of the default parameters. A radius of 6.5 Å was used to define the active site interaction points. Formal charges were applied to the ligands, since it improved docking accuracy on test set 1. No other exploration of parameters was carried out.

**GOLD (version 1.1).** Inclusion of ligand flexibility in molecular docking using a genetic algorithm (GA) is featured by the GOLD method.<sup>28</sup> The binding site is defined manually or by a cavity detection program, and all hydrogen-bonding sites within the site are defined. This information is used to encode a “chromosome” with all the potential hydrogen bond interactions along with a representation of the ligand’s flexibility. The GA then samples the ligand conformation, orientation, and position in the active site and uses a fitness function comprised of hydrogen bonding, pairwise force field interaction, and ligand internal energy terms to select poses in the docking process. Typically, several GA runs are required for binding mode identification. In this work, 10 GA runs with 1000 operations were used. Like DOCK and CDOCKER experiments, four ligand atomic charge methods were compared to determine the optimal one for docking accuracy. The location of the active site was set using the center of mass of the ligand taken from the X-ray protein–ligand complex. The size of the site was set to 22 Å, and the flood-fill cavity detection program was used, to provide a fair comparison to the other docking protocols.

## Results

To compare the results in docking accuracy when varying force field parameters, conformational flexibility, and docking method, an accuracy metric is needed. Since this work is focused on the ability to reproduce X-ray crystallographically determined ligand coordinates, the root-mean-square deviation (rmsd) of the heavy atom coordinates between the docked ligand and its X-ray orientation is used. Acceptable deviations have ranged from 1.0 to 3.0 Å in other docking comparison and evaluation work.<sup>11–14</sup> Here a value of 2.0 Å was chosen on the basis of the analysis of Gohlke et al. in considering the errors in crystallographic data.<sup>16</sup> To maintain consistency, a single in-house program was used to calculate the rmsd between the X-ray orientation and the docked orientation in question. The internal program considers all possible symmetry equivalent poses and gives identical results to rmsd values calculated by SYBYL<sup>27</sup> and other commercially available packages. For each docking run, only the top scoring pose, selected by the algorithm’s docking function, is considered.

**(1) Test Set 1 Docking Results.** The results of several docking accuracy comparisons carried out using DOCK, CDOCKER, and GOLD are shown in Table 4. Since there is a degree of randomness in the matching and minimization routines of DOCK and CDOCKER, each docking experiment was repeated three times using a different random seed. The number of docked orientations with a rmsd value to the X-ray structure of less than 2.0 Å for each docking experiment was recorded as “successful.” The average percent success over the test set and the standard deviation (SD) of each run are shown. To comment on the differences observed between docking trials, pairwise *t*-tests were also carried out. The resulting *P*-values (i.e. those <0.05) reveal statistically significant differences and are available as Supporting Information (Tables S-3, S-4, and S-5). A summary of the comparison of the flexible ligand-docking algorithms examined here is displayed in Table 5 (A full table of each rmsd for each complex for each method is available as Supporting Information, Table S-6). For these results, the best single run for each method is compared using the Corina ligand geometries as the starting conformation.

**(2) Test Set 2 Docking Results.** In this work the effects of protein conformation on ligand-docking accuracy was addressed by three docking experiments to three proteins as described in the Methods section. Each ligand was docked back to its own cocrystal structure, to the “average structure”, and to the apo structure. As before, the success rate is defined as the percentage of docked ligand orientations whose rmsd to the X-ray orientation is <2.0 Å. From the results on training set 1, it is clear that all four methods perform equitably in the “self-docking” paradigm, so further comparison here is unwarranted. Furthermore, timing issues are not critical to conclusions drawn with regard to the effects of protein conformation on ligand-docking accuracy. In light of this, the most accurate method, CDOCKER, was applied to test set 2. Table 6 shows a summary of the docking success rate in the self, average, and apo experiments. (The single best rmsd result for each

**Table 4.** Docking Success Rate Comparison for Variations in Starting Ligand Conformation, Atomic Charge, Dielectric Constant, Nonbonded Potential, and Docking Method for Test Set 1

starting conformer	ligand charge <sup>b</sup>	dielectric	potential <sup>c</sup>	% success rate (SD) <sup>a</sup>			
				DOCK		CDOCKER	GOLD
				rigid	flexible		
X-ray	Gasteiger	3	6–9	84.6 (2.8)	53.7 (0.0)	79.7 (3.0)	63.4 (NA)
X-ray	MR	3	6–9	86.2 (2.8)	52.8 (1.4)	79.0 (3.0)	60.9 (NA)
X-ray	CVFF	3	6–9	81.3 (1.4)	52.8 (1.4)	73.4 (3.0)	58.5 (NA)
X-ray	CVFF91	3	6–9	80.5 (2.4)	49.6 (1.4)	73.2 (0.0)	56.2 (NA)
X-ray	Gasteiger/MR	1	6–9	82.1 (1.4)	52.8 (1.4)	63.4 (2.0)	
X-ray	Gasteiger/MR	2	6–9	84.6 (1.4)	54.5 (1.4)	78.0 (2.0)	
X-ray	Gasteiger/MR	3	6–9	84.6 (2.8)	53.7 (0.0)	79.0 (3.0)	
X-ray	Gasteiger/MR	4	6–9	86.2 (1.4)	56.9 (3.7)	78.0 (3.5)	
X-ray	Gasteiger	3	6–9	84.6 (2.8)	53.7 (0.0)		
X-ray	Gasteiger	3	6–12	76.4 (3.7)	48.0 (1.4)		
Corina <sup>d</sup>			-	63.4 (NA)	63.4 (NA)	63.4 (NA)	63.4 (NA)
Corina	Gasteiger	3	6–9	43.9 (0.0)	49.6 (1.4)	73.2 (4.9)	46.3 (NA)
Corina	MR	3	6–9	43.9 (2.4)	48.0 (1.4)	75.6 (6.4)	46.3 (NA)
Corina	CVFF	3	6–9	42.3 (1.4)	48.8 (2.4)	64.2 (7.0)	46.3 (NA)
Corina	CVFF91	3	6–9	42.3 (1.4)	47.2 (1.4)	66.7 (3.7)	43.9 (NA)
Corina	Gasteiger/MR	1	6–9	43.9 (2.4)	48.8 (2.4)	56.1 (0.0)	
Corina	Gasteiger/MR	2	6–9	43.1 (1.4)	50.4 (1.4)	65.0 (4.1)	
Corina	Gasteiger/MR	3	6–9	43.9 (0.0)	49.6 (1.4)	75.6 (6.4)	
Corina	Gasteiger/MR	4	6–9	43.9 (0.0)	48.8 (0.0)	72.4 (2.3)	
Corina	Gasteiger	3	6–9	43.9 (0.0)	49.6 (1.4)		
Corina	Gasteiger	3	6–12	31.7 (2.4)	44.7 (1.4)		

<sup>a</sup> The percentage of the number of successfully docked ligands (i.e. ligand orientations with rmsd < 2.0 Å from the X-ray position) in test set 1. The average over three independent docking experiments for each of the 41 complexes are shown for the DOCK and CDOCKER methods along with the standard deviations (SD). Only one docking experiment for each of the 41 complexes was carried out using GOLD.

<sup>b</sup> Gasteiger charges were used for DOCK while Momany–Rome (MR) charges were used with CDOCKER to compare the effect of dielectric constant on docking accuracy. <sup>c</sup> Nonbonded potential was varied for DOCK experiments only. <sup>d</sup> Nondocked best rigid rms fit of the Corina to the X-ray geometry (comparison of conformations, i.e., similarity of X-ray and Corina conformation) for the 41 complexes.

**Table 5.** Docking Success Rate Comparison for DOCK, CDOCKER, FlexX, and GOLD

docking algorithm	success rate <sup>a</sup> (%)	av/median CPU times (s) <sup>b</sup>
DOCK	51.2	200/114
CDocker	82.9	2012/1740
FlexX	53.7	65/35
GOLD	46.3	824/708

<sup>a</sup> Success rates of the four algorithms represent the best single docking experiment obtained using a Corina-generated conformer as a starting point. <sup>b</sup> Times for a Silicon Graphics R12000 CPU.

ligand for each docking experiment is available as Supporting Information in Table S-7).

## Discussion

**Rigid Ligand Docking.** In examining the influence of various force field parameters, ligand flexibility, and docking algorithms, a systematic build-up in complexity is useful to quantify the effects they have on docking accuracy. A minimal test of a docking program, in this case DOCK, is to reproduce the X-ray orientation, given the correct ligand conformation. Here the search space

is reduced significantly, since no internal conformational flexibility is permitted. This experiment allows the examination of the effect variation of the force field parameters has on docking accuracy without the confounding effects of ligand conformation or flexibility. In the best case, DOCK was able to correctly place the X-ray conformation in of 35 of the 41 complexes on average in test set 1, yielding a docking accuracy of better than 85%. This shows that DOCK is very reasonable search method in terms of the six rotational and translational degrees of freedom to be explored in ligand docking (see Table 4). This success rate sets a baseline for the subsequent docking experiments.

Examination of the influence force field parameter variation has on ligand docking reveals that there is a slight effect for DOCK (Table 4). For instance, the range in average docking accuracy success is 2.3 complexes (33 for CVFF to 35.3 for Momany–Rone, of the 41) when the ligand atomic charge is varied, a value just above the standard deviation (1.2) of these experiments. Similarly, variation of dielectric constant shows a docking success range of less than 2 complexes on average.

**Table 6.** Docking Success Rate Results for Test Set 2 Examining the Effects of Protein Flexibility on Docking Accuracy

protein	Success Rate <sup>a</sup> (%)					
	self <sup>b</sup>	SD <sup>c</sup>	average <sup>d</sup>	SD	apo <sup>e</sup>	SD
trypsin	66.7 (53.3)	4.7 (4.7)	60 (56.7)	0 (4.7)	36.7 (36.7)	4.7 (4.7)
thrombin	35.9 (32.1)	3.6 (9.1)	26.9 (12.8)	3.1 (5.1)	9.0 (5.1)	4.8 (3.6)
HIV-1p	50.0 (28.7)	3.9 (6.7)	35.2 (18.5)	5.7 (5.7)	3.7 (0.9)	3.5 (1.3)

<sup>a</sup> Results for docking experiments used conformations derived directly from the X-ray structures as starting points, and the results in parentheses used a Corina conformer as a starting point for docking. The success rate is the average for three independent docking experiments. <sup>b</sup> Results of ligand docking to the protein structure taken from the exact complex (i.e. benzamide to **3ptb** structure etc.). <sup>c</sup> Standard deviation based on three independent docking experiments. <sup>d</sup> Results of docking all ligands to one selected complex structure. For trypsin, **1mtu** was chosen to represent the average structure, while choices for thrombin and HIV1-p were **1ets** and **1hvi**, respectively. <sup>e</sup> Results of docking all ligands to the protein structure crystallized without inhibitor present in the binding site. For trypsin, **1aks** was chosen to represent the apo structure, while choices for thrombin and HIV1-p were **1hgt** and **3phv**, respectively.

Most of these changes are not statistically significant according to pairwise *t*-tests. It is interesting that the CVFF91 charges are the only set that shows a significant decline in docking success rate. On the other hand, the change from standard 6–12 potential to the softer 6–9 had a more significant effect on rigid-ligand-docking accuracy (Table 4). The average docking success jumps 3.5 complexes (~12%) by going to the softer potential. A similar influence of charge parameter variation is seen in rigid docking beginning with a Corina-generated conformation instead of the ligand conformation from the X-ray structure. The docking success accuracy ranges are less than 1 complex for ligand atomic charge and dielectric constant. Again, a more significant effect on docking accuracy is observed when the nonbonded potential was softened. Docking success accuracy increased by 5 complexes when moving to the 6–9 from the standard 6–12 and was shown to be statistically significant as seen in Table 4.

The starting conformation of the ligand has a much larger effect for nonflexible docking than force field scoring parameters. Docking accuracy dropped by almost half (from 86% to 44%) when starting from the non-X-ray conformation. This emphasizes the importance of the ability to treat ligands flexibly (*vide infra*). The proper baseline for comparison in this case, however, is not the rigid-ligand X-ray results, but rather the number of Corina-generated conformers that are within 2.0 Å rmsd of the X-ray conformation. To examine this, the Corina conformers were overlaid on the X-ray conformer using a rigid atom–atom least-squares fitting procedure. Surprisingly, 63.4% (26 of 41) of the Corina conformers had rmsds of <2.0 Å to the X-ray conformer. This suggests that, for some ligands, automated conformer generators, such as Corina, produce X-ray-like conformations for this test set, which is consistent with several studies on conformer generators. Using these 26 as a comparison, DOCK was able to position them with ~70% accuracy rate. The maximum fitted rmsd between the Corina and X-ray conformers that DOCK failed to successfully dock (i.e. within 2.0 Å) was 1.3 Å, suggesting a threshold for the use of rigid docking with DOCK. Moreover, those molecules whose fitted rmsd to the X-ray that are less than 2.0 Å are, on average, small and rigid with a mean molecular weight and number of rotatable bond value of 250.5 amu and 4.5, respectively. The other 15 ligands in test set 1 had an average molecular weight of 537.2 amu and 11.2 rotatable bonds. These observations may suggest cutoffs for the size and flexibility of rigid docking with DOCK using Corina conformers for virtual screening.

**Flexible Ligand Docking.** The previously discussed experiments using the rigid-ligand approximation highlight the importance of allowing ligands to fully relax during docking. Here several aspects of flexible ligand docking were examined. First a comparison between rigid and flexible ligand docking using DOCK is made (Table 4). Since DOCK employs an incremental build-up algorithm, the starting conformation used in docking does not have as significant an effect as was seen in rigid docking, because the ligand structure is fragmented and then reconstructed during docking. The docking success rate does increase, however, by ~6%

(i.e. an average of 2.5 more ligands docked successfully) starting with the X-ray conformation instead of a Corina conformer. This suggests that the incremental build-up method suffers from incomplete sampling (possibly due to limitations of the rotamer library), noninclusion of ring flexibility, and/or the lack of proper treatment of intraligand forces. Overall, considering ligand flexibility does significantly (*P*-value = 0.0022) improve docking success rates. Starting from the X-ray conformer, flexible docking success rates are lower than rigid docking, but this is expected, since the correct conformation is provided in the rigid docking case, while the algorithm is forced to find the correct conformation in the flexible case. Similarly, the results starting with the Corina conformation improve by ~2.5 complexes docked correctly, on average, in test set 1 (~6% improvement). This small improvement may justify the use of fully flexible ligand docking in virtual screening of large databases if docking times are reasonable especially for larger ligands (*vide supra*). It should be noted that docking time was not optimized for these experiments. DOCK parameters can be “tuned” to individual proteins to significantly reduce average flexible-ligand-docking times. In fact, successful in-house virtual screening experiments have attained times of ~5 s per molecule per CPU (SGI R10000).

The effect of force field parameter variation on flexible ligand docking using DOCK is also displayed in Table 4. It can be seen that variation of ligand atomic charge and dielectric constant has a very small effect on docking accuracy success. There is a range of less than 2 ligands docked successfully when ligand charge or dielectric constant is changed, making flexible ligand docking less sensitive than was seen with rigid ligand docking. The only significant effect is observed when using the CVFF91 charges. The other three charge sets showed no significant change in success rates. Similar to the rigid ligand case, the softer 6–9 potential showed significant improvement in flexible-ligand-docking accuracy over the standard 6–12 but to a lesser extent. Docking success improved ~5% (i.e. 2 complexes) on average.

On the other hand, atomic charge and dielectric constant had a slightly larger effect on docking accuracy for the CDOCKER method. The docking accuracy success range between the atomic charge sets was 2.7 and 4.7 ligands on average (6.5 to 11.4%) starting with the X-ray conformation and Corina conformation, respectively. The differences between the Momany–Rone charge set and the other three were not significant, however, as indicated by pairwise *t*-test results. A wider range was observed for dielectric constant variation where the docking accuracy success rate range was 15.6% (6.4 ligands) for the X-ray conformation and 19.5% (8 ligands) when a Corina conformer is docked. This was due mainly to the drop off in accuracy seen in using a dielectric constant of 1, which showed a significant difference between the other dielectrics. The fact that CDOCKER is more sensitive to atomic charge and dielectric constant than DOCK is not too surprising, given the differences in the algorithms. DOCK uses a nonenergetic method to place the ligand in the binding pocket followed by a ranking of poses by interaction energy, while CDOCKER's approach involves explora-

**Table 7.** Correlation of Ligand Properties with the rmsd between Docked and X-ray Orientations

property	correlation ( $r^2$ ) <sup>a</sup>			
	DOCK	FlexX	CDOCKER	GOLD
molecular weight	0.65	0.35	0.09	0.36
no. of rotatable bonds	0.67	0.49	0.07	0.28
no. of polar atoms	0.37	0.15	0.25	0.07

<sup>a</sup> Correlation coefficient squared between the rmsd and molecular weight, number of rotatable bonds and the number of O's and Ns for each docking method. The rmsd was taken from the best single docking experiment for the 41 complexes for each method.

tion of an energetic landscape for both. It is interesting that accuracy tends to improve for both methods using the Momany–Rone or Gasteiger atomic charges and a dielectric constant of 3. These charge sets also provide the best results for the GOLD algorithm.

The marked difference in flexible-ligand-docking accuracy shown between CDOCKER, DOCK, and GOLD emphasizes the importance of docking methodology and treatment of energy landscape. Table 5 expands the comparison to include FlexX,<sup>26</sup> another well-established and widely used docking program. On average, all four methods perform well and are able to reproduce X-ray orientations within 2.0 Å rmsd in 50% or more of the cases, even though widely different methods for ligand positioning and flexibility are used in each. This result is in line with other comparisons of flexible-ligand-docking methods.<sup>20,26</sup> The two most similar methods, DOCK and FlexX, both use a geometric matching routine for positioning and incremental build-up for ligand flexibility, but they differ substantially in the scoring method. Nevertheless, both are able to flexibly dock a Corina conformer correctly in just over 50% of the cases. The same is true for GOLD, which uses a genetic algorithm and an empirically derived fitness function. CDOCKER's simulated annealing molecular dynamics method, on the other hand, is able to reproduce X-ray ligand structures in significantly more cases (>80%) than the other methods, but at a much higher cost (10 times or more slower due to the explicit treatment of all interactions; a grid treatment of the protein–ligand interaction shows a mean speed-up of 3 times<sup>34</sup>).

To help understand the key determinants of ligand-docking accuracy, the test set ligands were examined more closely with respect to docking accuracy and docking method. How do molecular properties of the ligands such as the size (molecular weight), flexibility (number of rotatable bonds), and polarity (number of polar atoms, i.e., oxygens and nitrogens) affect the ability of a docking method to accurately reproduce the bound conformation and position (see Figure 1)? Table 7 displays the correlation between the rmsd between the docked orientation and the X-ray orientation and the properties representing ligand size, flexibility, and polarity. There is a striking dependence of docking accuracy on the size and flexibility of the ligand in three of the four docking algorithms. [It should be noted that molecular weight and the number of rotatable bonds are slightly correlated ( $r^2 = 0.65$ ).] The docking accuracy of DOCK and FlexX (and to a lesser extent, GOLD) have a strong dependence on the number of rotatable bonds of the ligand. In fact, these three methods perform very well when the ligand has fewer than eight rotatable

**Table 8.** Docking Accuracy Compared to Ligand Flexibility for Test Set 1

no. of rotatable bonds	total no. of ligands <sup>b</sup>	no. of ligands docked correctly <sup>a</sup>			
		DOCK	FlexX	GOLD	CDOcker
<8	20	18	16	14	19
≥8	21	3	6	4	15

<sup>a</sup> Number of complexes with docked vs X-ray rmsd < 2.0Å.

<sup>b</sup> Total number of complexes with the specified number of rotatable bonds in test set 1.

bonds. Docking accuracy success rates increase from just over 50% to 90%, 80%, and 70% for DOCK, FlexX, and GOLD, respectively, for ligands that have less than eight rotatable bonds (see Table 8). This dependence can be rationalized by the lack of proper sampling of rotatable bond space during the docking process. Even though these methods do a reasonable survey of the conformational space of the ligands, exhaustive searching is not practicable using current technology. Incomplete sampling can be established as the source of inaccurate docking if the scores of the best docked solutions are higher (less favorable) than the X-ray orientation. In the DOCK results, for example, when comparing the docking energy scores (i.e. ligand–protein interaction energies) of the docked orientations to the scores of the minimized X-ray orientations, it is revealed that the scores of the X-ray orientation are lower in 17 out of 20 “misdocked” ligands (i.e., those with rmsd > 2.0 Å). On the other hand, all seven of the “misdocked” ligands from the CDOCKER results, whose rmsd of docking poses vs X-ray was not correlated with ligand flexibility, had lower scores than the minimized X-ray orientation. This suggests a docking function inadequacy rather than incomplete conformational analysis. The higher success in those ligands with less than eight rotatable bonds (95%) compared to those with eight or more rotatable bonds (71%) for CDOCKER, however, implies that the docking functions' ability to distinguish between correct and incorrect conformations may decrease for ligands with large numbers of viable low-energy conformations.

Examination of the seven complexes that CDOCKER failed to reproduce correctly may support the conclusion that failures probably arise from scoring rather than sampling. CDOCKER fails on only one ligand with fewer than eight rotatable bonds (PDB code **1dog**). Interestingly, all of the algorithms fail to correctly dock 1-deoxyojicimycin to glucoamylase. (Another docking algorithm, PRO-SELECT also failed to dock these ligand accurately.<sup>33</sup>) This structure has two 1-deoxyojicimycin molecules bound in close proximity in the binding pocket. Here we are attempting to reproduce the molecule with the “strong electron density”, but the possibility of the other binding mode may complicate the docking process. The other six “misdocked” ligands have a range of flexibility, namely, two ligands with eight rotatable bonds and one each with nine, 12, 14, and 15 rotatable bonds. Three of the six failures contain a sulfonamide linker (**1pph**, **1uvt**, and **1uvs**). This group may be problematic for the CHARMM force field, leading to an incorrect docking result.<sup>46,47</sup> FlexX, on the other hand, is able to correctly dock two of three of these complexes, **1pph** and **1uvt**, and has a smaller rmsd on the third, **1uvs**, to the X-ray pose than the CDOCKER



result. The extensive MINUMBA<sup>45</sup> conformer library used by FlexX to explore conformational space during the incremental build-up procedure may better represent the sulfonamide group. This failure is not inherent in the MD sampling method of CDOCKER, but rather, these results point to a systematic error in the underlying force field for this particular group. Overall, the direct exploration of the energetic hyperspace seems to provide a significant advantage in docking accuracy, especially that of the more flexible ligands. We can speculate that in the face of many conformers with low energies, those found by the molecular dynamics sampling pathway may reflect those actually sampled in the ligand binding event more so than the nonenergetically driven methods of the other three algorithms.

**Protein Flexibility.** All of the docking algorithms examined here, as well as those most frequently employed in structure-based drug design and virtual screening, treat the protein as rigid. This assumption is made for practical considerations, given the large increase in the already large search space when the protein conformation is included, even though it is well-established that the ligand binding event is dynamic, often involving large protein movement.<sup>29</sup> Here the effect of applying the rigid protein constraint is examined utilizing the X-ray structures of three proteins both bound to several different molecules and with no ligand in the binding pocket.

Three docking experiments were carried out for each ligand using the CDOCKER method. The first involved docking the ligand back to the structure it was taken from or “self” docking, as was carried out with test set 1. This provides a baseline for comparison, since these structures have the experimentally determined induced protein movement incorporated. In the second two docking experiments, all of the ligands, within each protein class, were docked to a single structure. The ligands were docked to an “average” structure that was determined by selecting a structure whose binding site coordinates were the closest to the average position with respect to the rmsd of all the structures. They were also docked to the “apo” structure, that is, the structure without a bound inhibitor. These experiments show how docking accuracy, determined as the number of docked solutions whose heavy-atom coordinates are less than 2.0 Å rmsd from the X-ray position, is effected by small protein movements.

Table 2 characterizes the degree of protein movement caused by ligand binding. There is an interesting range of protein movement represented by these three proteases. The largest difference is seen between the structures with an empty binding pocket to those with a bound ligand. The average rmsd values between the coordinates of the structures with a bound ligand and the apo structure are quite large, suggesting the importance of using a ligand-bound structure if available for structure-based design or virtual screening. The protein with the largest movement that occurs upon ligand binding is HIV-1 protease, where a 2.0 Å movement is seen on average. As expected, it is very difficult to reproduce X-ray orientations within a 2.0 Å rmsd when the change in the binding site is on the same order. Table 6 highlights this difficulty, showing that docking accuracy rates fall off greatly when the apo

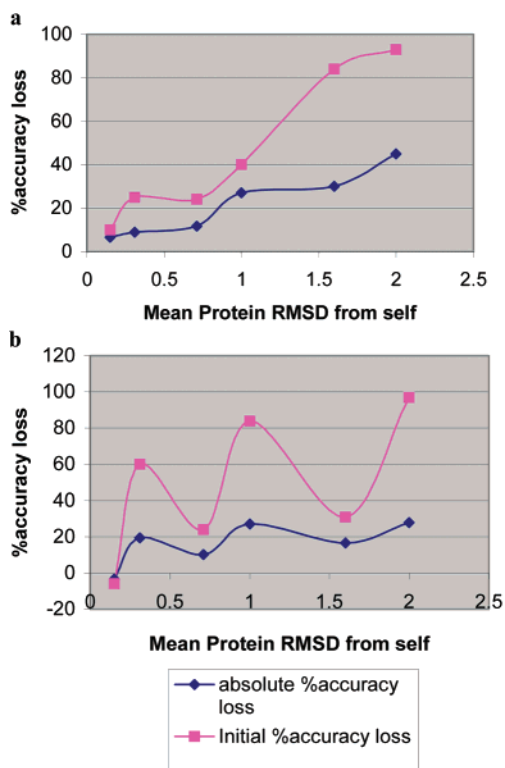
structure is used. Accuracy rates drop 45%, 75%, and 93% for trypsin, thrombin, and HIV-1 protease, respectively, when the apo structure is used instead of docking the ligand back to its own structure.

The effect that different ligands have on the protein structure varies in this set of proteases. The movement as measured by the average rmsd between the binding pockets of the structures increases from trypsin to thrombin to HIV-1 protease (see Table 2). Table 6 shows that this degree of movement is reflected in docking accuracy rates. The protein with the least amount of movement, trypsin, also shows the highest docking accuracy to the average structure. In fact, docking accuracy success has an inverse correlation with the degree of protein movement, i.e., docking accuracy success increases as protein movement decreases going from HIV-1 protease to thrombin to trypsin.

These results are confounded, however, with the average flexibility of the respective ligands that exhibit the same trend. As the average number of rotatable bonds increases for the ligands of each protein, the docking success rates decrease. Although there was not a strong relationship seen between ligand flexibility as measured by the number of rotatable bonds and the rmsd between the docked and the X-ray orientations for test set 1 using CDOCKER, it was clear that docking success rates did decrease for highly flexible ligands (those with more than eight rotatable bonds; see Table 7). A way to examine the effect of protein movement on docking accuracy without the confounding effect of ligand flexibility is to look at the decrease in docking accuracy between the self-docking experiments and those utilizing the average structure. The decrease in docking accuracy shows the trend trypsin < thrombin < HIV-1 protease, exhibiting 6.7%, 9.0%, and 11.7% drops in accuracy rates, respectively. This trend, although not as strong, suggests that docking accuracy depends on the amount the protein moves upon ligand binding. This observation is strongly supported by Figure 3a, which shows the loss of docking accuracy as a function of mean protein rmsd from the initial complex. It is worth noting that almost 90% of the initial docking accuracy is lost if the mean protein rearrangement is greater than 1.5 Å. This observation has important implications for homology modeling and indirectly suggests that prediction of ligand binding geometries would be reliable only for the most refined homology models.

## Conclusions

An attempt was made in this report to explore ligand-docking methodologies in order to make some comments on the determinants of the successful prediction of protein–ligand binding modes. Thus, the focus of this work was on the ability of standard docking methods to accurately reproduce X-ray crystallographically determined orientations rather than on scoring (i.e. reproducing measured binding energies). To carry out the docking experiments, two carefully chosen test sets were created. Care was taken to treat each system in a uniform and consistent manner to avoid introduction of bias. For test set 1, the ligands provide a range in molecular size, polarity, and flexibility (See Figure 1). This set was used to examine the effect of ligand charge,



**Figure 3.** Loss of docking accuracy as a function of protein structure rearrangement from the original complex. Data from Tables 2 and 6 are used. Mean rms deviation from the original complex for average and apo structures for trypsin, HIV1-p, and thrombin are plotted on the X axis. Corresponding absolute loss of docking accuracy is shown in diamonds. Filled squares represent the relative loss of docking accuracy (absolute loss divided by the initial docking accuracy). The figure demonstrates that protein active site rearrangement greater than 1.5 Å leads to almost complete lack of reproduction of the “true” binding mode. (a) Starting geometries for ligands derived directly from the X-ray structures. (b) Starting geometries for ligands derived from a Corina-generated conformer.

size, and flexibility on docking accuracy. For test set 2, the three proteases displaying a range of protein flexibility in response to ligand binding were selected to examine the effects of protein flexibility on docking accuracy (see Table 2). Using these training sets and four docking algorithms, we have tried to investigate the key determinants of docking accuracy.

Do small changes in the force field parameters used for docking and scoring have a significant effect on docking accuracy? In general, the methods used for ligand atomic charge assignment were found to produce small effects on docking accuracy. The largest change in docking accuracy when ligand charges were varied over four different charge methods and three docking algorithms was a ~10% drop. This occurred when using CVFF charges with CDOCKER. For DOCK and GOLD, docking accuracy was reduced by only ~2% when changing charge set. This is consistent with GOLD docking function. Similarly, varying the effective charge of the protein by changing the dielectric constant from 4 to 1 had a small effect on docking accuracy for DOCK (the accuracy dropped less than 2%), while the drop in docking accuracy for CDOCKER was larger (the accuracy dropped from 75% to 56%). Finally, softening the nonbonded potential (from a 6–12 to a 6–9) in DOCK increased docking accuracy by about 5%. These results

are in line with the success of using the soft-core potentials in CDOCKER.<sup>34–37</sup>

How do the molecular properties of the ligand affect the ability to accurately dock it? The link between docking accuracy and ligand flexibility has been established here in a systematic study using four different docking algorithms. For the docking algorithms that employ an incremental build-up method to treat ligand flexibility (FlexX, DOCK), there is a significant correlation with ligand flexibility. For GOLD and CDOCKER, the correlation is not as strong, but docking accuracy clearly decreases for ligands with eight or more rotatable bonds. Extending the analysis to include all 113 self-docking cases from test set 1 and 2 using CDOCKER further supports the link between ligand flexibility and docking success accuracy. CDOCKER was able to successfully dock 54% of the 113 complexes, even though the ligands are very flexible (average number of rotatable bonds = 12), however, docking success rates dropped from 76.7% for ligands with less than eight rotatable bonds to 45.8% for ligands with eight or more. This supports the link between ligand flexibility and docking accuracy. Overall, ligand flexibility appears to be a major determinant in the ability to dock a ligand accurately across all four methods examined here. Moreover, the performance of CDOCKER on a large range of ligands supports the use of molecular dynamics as an attractive search method for molecular docking.

How does protein flexibility affect docking accuracy? This work has established the dramatic effect that protein conformation has on the ability to accurately dock ligands. Looking at three proteins with varying degree of conformational change upon ligand binding demonstrated that not only was there was a clear drop in accuracy when the ligand was docked to an X-ray structure other than its own but also that this drop off is correlated with the degree of the protein movement in the active site. These results are in agreement with other unbound ligand studies<sup>30,31</sup> and have major implications in structure-based drug design and virtual screening. The drop off in docking accuracy observed here for using protein structures with no inhibitor bound or those in complex to very different ligands should be factored into structure-based design efforts. This problem is especially magnified when using homology models constructed from templates with low sequence similarity to the target. Studies are underway to expand this work to other protein classes in addition to proteases to further explore the effect of protein flexibility on molecular docking.

**Acknowledgment.** We thank Dr. Matthias Rarey (GMD) and Dr. David Lewis (Tripos) for their help in FlexX evaluation and the members of Eli Lilly computational group for helpful discussions. This work was supported by Eli Lilly Postdoctoral Program.

**Supporting Information Available:** Tables containing the set test ligand structures (Tables S1 and S2), the pairwise *t*-test results for DOCK and CDOCKER experiments (Tables S3, S4, and S-5), and the best rmsd values between docked and X-ray poses for each docking experiment (Tables S-6 and S-7). This information is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Babine, R. E.; Bender, S. L. Molecular recognition of protein–ligand complexes: Applications to drug design. *Chem. Rev.* **1997**, *97*, 1359–1472.
- (2) Kuntz, I. D. Structure-based strategies for drug design and discovery. *Science* **1992**, *257*, 1078–1082.
- (3) Greer, J.; Erickson, J. W.; Baldwin, J. J.; Varney, M. D. Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *J. Med. Chem.* **1994**, *37*, 1035–1054.
- (4) Charifson, P. S.; Kuntz, I. D. Recent successes and continuing limitations in computer-aided drug design. In *Practical Application of Computer-Aided Drug Design*; Charifson, P. S. Ed.; Dekker: New York, 1997; pp 1–37.
- (5) Stevens, R. C.; Yokoyama, S.; Wilson, L. A. Global efforts in structural genomics. *Science* **2001**, *294*, 89–92.
- (6) Gillet, V. J. De novo molecular design. *Methods Principles Med. Chem.* **2000**, *8*, 49–69.
- (7) Waszkowycz, B. New methods for structure-based de novo drug design. *Adv. Drug Discovery Techniques* **1998**, 143–164.
- (8) Clark, D. E.; Murray, C. W.; Li, J. Current issues in de novo molecular design. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 1997; pp 67–125.
- (9) Schneider, G.; Bohm, H.-J. Virtual Screening for Bioactive Molecules. In *Methods and Principles in Medicinal Chemistry*; Mannhold, R., Kubinyi, H., Timmerman, H., Eds.; Wiley-VCH: Weinheim, 2000.
- (10) Whittle, P. J.; Blundell, T. L. Protein structure-based drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 349–375.
- (11) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein–small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151–166.
- (12) Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 409–443.
- (13) Muegge, I.; Rarey, M. Small molecule docking and scoring. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH: New York, 2001.
- (14) Abagyan, R.; Totrov, M. High-throughput docking for lead generation. *Curr. Opin. Chem. Biol.* **2001**, *5*, 375–382.
- (15) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. A. Conformational changes of small molecules binding to proteins. *Bioorg. Med. Chem.* **1995**, *295*, 411–428.
- (16) Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein–ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
- (17) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
- (18) Meng, E. C.; Shoichet, B. K.; Kuntz, I. D. Automated docking with grid-based energy evaluation. *J. Comput. Chem.* **1992**, *13*, 505–524.
- (19) Gschwend, D. A.; Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: On-the-fly optimization and degeneracy removal. *J. Comput.-Aided Mol. Des.* **1996**, *10*, 123–132.
- (20) Makino, S.; Kuntz, I. D. Automated flexible ligand docking method and its application for database search. *J. Comput. Chem.* **1997**, *18*, 1812–1825.
- (21) Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **1997**, *18*, 1175–1189.
- (22) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–428.
- (23) Gasteiger, J.; Marsili, M. A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* **1978**, *34*, 3181–3184.
- (24) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity: A rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219–3222.
- (25) Tao, P.; Lai, L. Protein ligand docking based on empirical method for binding affinity estimation. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 429–446.
- (26) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- (27) Tripos SYBYL Molecular Modeling Software, 6.x ed.; Tripos Inc., St. Louis, MO.
- (28) 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*, 727–748.
- (29) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. Complexity and simplicity of ligand–macromolecule interactions: The energy landscape perspective. *Curr. Opin. Struct. Biol.* **2002**, *12*, 197–203.
- (30) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377–395.
- (31) 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*, 547–562.
- (32) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (33) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425–445.
- (34) Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. A detailed analysis of grid-based molecular docking. A case study of CDOCKER—A CHARMM based MD docking algorithm. *J. Comput. Chem.* **2003**, *24*, 1549–1562.
- (35) Vieth, M.; Hirst, J. D.; Kolinski, A.; Brooks, C. L., III Assessing energy functions for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1612–1622.
- (36) Vieth, M.; Hirst, J. D.; Dominy, B. N.; Daigler, H.; Brooks, C. L., III Assessing search strategies for flexible docking. *J. Comput. Chem.* **1998**, *19*, 1623–1631.
- (37) Vieth, M.; Cummins, D. J. DoMCoSAR: A Novel Approach for Establishing the Docking Mode That Is Consistent with the Structure–Activity Relationship. Application to HIV-1 Protease Inhibitors and VEGF Receptor Tyrosine Kinase Inhibitors. *J. Med. Chem.* **2000**, *43*, 3020–3032.
- (38) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization and dynamics calculation. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (39) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Wang, J.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, R. J.; Duan, Y.; Pitera, J.; Massova, I.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER*; 7 ed.; University of California, San Francisco.
- (40) Momany, F. A.; Rone, R. Validation of the General Purpose QUANTA 3.2/CHARMM Force Field. *J. Comput. Chem.* **1992**, *13*, 888–900.
- (41) *Insight, Release 2000*; Accelrys, San Diego, CA, 2002.
- (42) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537–547.
- (43) MSI QUANTA; 4.6 ed.; Molecular Simulations Inc., San Diego, CA, 1997.
- (44) DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.* **1988**, *31*, 722–729.
- (45) Klebe, G.; Mietzner, T. A fast and efficient method to generate biologically relevant conformations. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 583–606.
- (46) Bindal, R. D.; Golab, J. T.; Katzenellenbogen, J. A. Ab initio calculations on *N*-methylmethanesulfonamide and methyl methanesulfonate for the development of force field torsional parameters and their use in the conformational analysis of some novel estrogens. *J. Am. Chem. Soc.* **1990**, *112*, 7861–7868.
- (47) Nicholas, J. B.; Vance, R.; Martin, E.; Burke, B. J.; Hopfinger, A. J. A molecular mechanics valence force field for sulfonamides derived by ab initio methods. *J. Phys. Chem.* **1991**, *95*, 9803–9811.

JM030209Y