Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects

Woody Sherman,[‡] Tyler Day,[‡] Matthew P. Jacobson,[§] Richard A. Friesner,[#] and Ramy Farid*,[‡]

Schrödinger, Inc., 120 West 45th Street, 32nd Floor, New York, New York 10036, Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, and Department of Chemistry and Center for Biomolecular Simulation, Columbia University, New York, New York 10027

We present a novel protein−ligand docking method that accurately accounts for both ligand and receptor flexibility by iteratively combining rigid receptor docking (Glide) with protein structure prediction (Prime) techniques. While traditional rigid-receptor docking methods are useful when the receptor structure does not change substantially upon ligand binding, success is limited when the protein must be "induced" into the correct binding conformation for a given ligand. We provide an in-depth description of our novel methodology and present results for 21 pharmaceutically relevant examples. Traditional rigid-receptor docking for these 21 cases yields an average RMSD of 5.5 Å. The average ligand RMSD for docking to a flexible receptor for the 21 pairs is 1.4 Å; the RMSD is ≤1.8 Å for 18 of the cases. For the three cases with RMSDs greater than 1.8 Å, the core of the ligand is properly docked and all key protein/ligand interactions are captured.

Introduction

A key objective of computational structure-based drug design is the prediction of the structure of protein−ligand complexes. If a high resolution structure of the receptor is available, and the receptor structure does not change substantially upon ligand binding, the problem can often be reduced to docking the flexible ligand in the environment of the rigid receptor. Effective algorithms for addressing rigid receptor docking have been developed in the absence of significant steric clashes between receptor and ligand.[1−7] The implications of a rigid receptor model have been studied,[8] and in many cases protein flexibility must be explicitly accounted for by moving outside of the rigid receptor paradigm.[9,10]

When rigid receptor docking fails, one of the simplest approaches is to reduce the van der Waals radii of the protein and/or ligand atoms or delete side chains of residues predicted to be flexible, thus potentially eliminating close contacts.[11,12] However, while this approach may yield the correct ligand binding mode, it may not give insights into specific protein/ligand interactions since the conformation of key residues in the binding site may be inconsistent with the correct ligand structure in the receptor. It is also likely to lead to false positives in virtual screening experiments, as the binding site will effectively increase in size. It is clear that in order to tackle the full protein/ligand structure prediction problem in a robust and accurate manner, it is essential to allow both the structure of the protein and ligand to reorganize. From a computational point of view, this is substantially more challenging than rigid receptor docking, as it involves many more degrees of freedom.

There are a number of ways one could account for both ligand and receptor flexibility. In theory, the correct structure of the protein−ligand complex can be generated by running explicit solvent molecular dynamics simulations, starting from an arbitrary initial guess.[13−15] However, there are no reports in the literature to our knowledge of consistent success on a substantial number of targets.[16] The primary reason for this is that the kinetics of creating the correct protein−ligand complex may be slow (possibly microseconds or longer) due to the need to traverse high energy barriers, particularly if substantial side-chain and/or backbone rearrangement is required. This level of CPU expenditure is generally impractical in a realistic drug discovery environment in which many ligands have to be examined in a relatively short time frame. Furthermore, even if such calculations could be carried out with the available resources, there is no guarantee that the correct answer would be obtained, due to errors in the potential functions, solvation model, boundary conditions, etc.

A number of approaches between the extremes of softened potential docking and full molecular dynamics have been employed in an attempt to model protein flexibility associated with ligand binding (reviewed by Carlson[17]). One such procedure combines various receptor structures (obtained by either experimental or computational means) to form a unified receptor description that can then be targeted using more traditional docking techniques.[18−20] The difficulty with such methods lies in the determination of which receptor conformations to use, how to generate them, and more importantly, how to reasonably combine the properties from the different receptors. A procedure has been proposed to combine various receptor structures using an incompatibility graph to determine valid protein structures to be used for docking;[21] however, the method implies knowledge of existing structures to form the receptor ensemble, which is not always available, and the rate of success at finding poses with RMSDs less than 2 Å from the correct structure as the top ranked solution is approximately 30% for this approach. While this and other approaches for generating unified receptor structures may identify active compounds in a database screen, they do not result in a final protein structure for each ligand and therefore provide minimal insights into the receptor flexibility associated with the specific complex of interest. A second approach involves docking into the members of an "ensemble" independently and aggregating the results.[22−27] While this method shows great promise, it relies on the existence of a sufficiently wide range of representative conformations, which are often not available. In addition, this method cannot

* To whom correspondence should be addressed. Phone: 646-366-9555. Fax: 646-366-9550. E-mail: ramy@schrodinger.com.
‡ Schrödinger, Inc.
# Columbia University.
§ University of California, San Francisco.

identify novel receptor conformations that are not in the initial ensemble. Another strategy involves generating a receptor pharmacophore model based on common elements of a flexible receptor ensemble.[11,28] This approach also relies either on knowledge of an existing receptor ensemble or the generation of an ensemble through other techniques (such as molecular dynamics) and does not reveal precise structural information about the final ligand−receptor complex. Furthermore, this method overlooks structural information not associated with the receptor-derived pharmacophore.

The approach taken here to address receptor flexibility in ligand binding is to combine in an iterative fashion techniques for docking ligands into a rigid receptor with those for modeling receptor conformational changes. We use the docking program Glide[29] to account for ligand flexibility and the Refinement module in the Prime[29] program to account for receptor flexibility. In the work reported here, side-chain degrees of freedom in the receptor are sampled while allowing minor backbone movements through minimization. The key feature of the side-chain prediction algorithm[30,31] is that the sampling is performed in dihedral angle space. That is, although the algorithm uses the same type of force field-based energy functions used in molecular dynamics, the small (femtosecond) Cartesian moves of MD are replaced with large moves in dihedral angle space. Other critical aspects of the protein sampling algorithm, to be described in greater detail below, include (i) rapid elimination of conformations that involve steric clashes, (ii) an efficient minimization algorithm (multi-scale Truncated Newton), (iii) the use of rotamer libraries to sample only energetically reasonable side-chains conformations, and (iv) the use of an accurate energy model (force field + implicit solvent) for scoring the receptor conformations, which is critical for properly treating the receptor strain energy.

The advance reported here is the iterative coupling of Glide and Prime to account for both ligand and protein flexibility, respectively. In principle, one could sample ligand and protein degrees of freedom simultaneously. However, in cases where the receptor induced fit comes primarily from side-chain rearrangements and relatively small backbone conformational changes, an iterative process of sampling ligand and receptor degrees of freedom independently is viable and likely to be more computationally efficient. Our strategy is to first dock ligands into a rigid receptor using a softened energy function such that steric clashes do not prevent at least one ligand pose from assuming a conformation close to the correct one. We then sample the receptor degrees of freedom (i.e., side-chain rotamers) and perform a minimization of the protein/ligand complex for many different ligand poses and attempt to identify low free energy conformations of the receptor−ligand complex. A second round of ligand docking is then performed on the refined protein structures, this time using a hard potential function to further sample ligand conformational space within the refined protein environment. Finally, a composite scoring function is applied to rank the complexes, accounting for the receptor−ligand interaction energy as well as strain and solvation energies. In the present paper, we describe this induced fit docking (IFD) methodology and apply it to a wide range of pharmaceutically relevant examples.

The IFD protocol achieves successful results (ligand RMSD ≤ 1.8 Å) in 18 of the 21 test cases that we have examined to date, which is composed of a diverse range of receptors and ligands. In the three cases with RMSDs over 1.8 Å, the core of the ligand is properly docked and all key ligand/protein interactions are duplicated in the IFD structures. The test cases

include cross-docking between two holo structures in both directions as well as docking into apo structures. In four of the cases, rigid receptor docking and induced fit docking both yield the correct ligand binding pose, demonstrating that the IFD protocol can be applied even in cases were little or no conformational changes in the receptor are required to properly dock the ligand.

Induced Fit Methodology

Overview. Suppose there are crystal structures of two protein−ligand complexes, which we shall refer to as C_A and C_B (the C designating "complex"), which contain the same receptor sequence, but different ligands. The two receptor conformations (R_A and R_B) are adapted to the ligands (L_A and L_B) and hence may be different, possibly significantly different. We shall assume that self-docking (docking of ligand L_A into receptor R_A, or of ligand L_B into receptor R_B) yields good results with the rigid docking program that is available for induced fit studies (in our case, the Glide program, which has been shown to be effective at self-docking[2,32,33]). In this study, both C_A and C_B are available from the PDB; however, in the general drug discovery environment, one might start with only C_B and a 2D structure of L_A in an effort to solve for C_A. A standard "cross docking" test is to dock ligand L_A into receptor R_B. In some cases, cross docking succeeds, indicating that receptor conformations R_A and R_B are fundamentally similar. However, in many cases, cross docking fails due to the incompatibility of R_B with L_A. That is, it is not possible to properly position L_A in receptor conformation R_B, because doing so would lead to significant steric clashes between atoms of L_A and R_B. It is this type of induced fit effect that is the focus of the present paper.

The details of our IFD protocol and the parameters used in this work are given in Materials and Methods; the goal in this section is to outline the basic sequence of calculations and the rationale for adopting it. The discussion will also illuminate the underlying issues that make induced fit a challenging problem and how our solution addresses those problems to achieve accuracy and robustness at an acceptable level of computational effort.

The overall procedure (outlined in Figure 1) has four steps: (1) initial softened-potential docking into a rigid receptor to generate an ensemble of poses; (2) sampling of the protein for each ligand pose generated in the first step; (3) redocking of the ligand into low energy induced-fit structures from the previous step; and (4) scoring by accounting for the docking energy (GlideScore), and receptor strain and solvation terms (Prime energy). The challenge in the first step is to generate at least one reasonably docked pose for the ligand (independent of the score it receives). Without a plausible initial guess for the ligand pose, any attempt to predict reorganization of the protein structure is unlikely to succeed in the context of a limited allotment of CPU time. The principal challenge in the second step is predicting the low energy receptor conformation for a correct ligand pose, starting from the "plausible" initial guess generated in the first step. The primary challenge in the third step is to generate low energy ligand conformations when presented with the correct receptor conformation. The difficulty in the final scoring step lies in properly ranking the complexes such that the top ranked pose correctly predicts the ligand/receptor structure.

The IFD protocol is iterated a second time when the top ranked output structures from step 4 of the first round have nearly identical scores. In this case, the ligand is redocked into
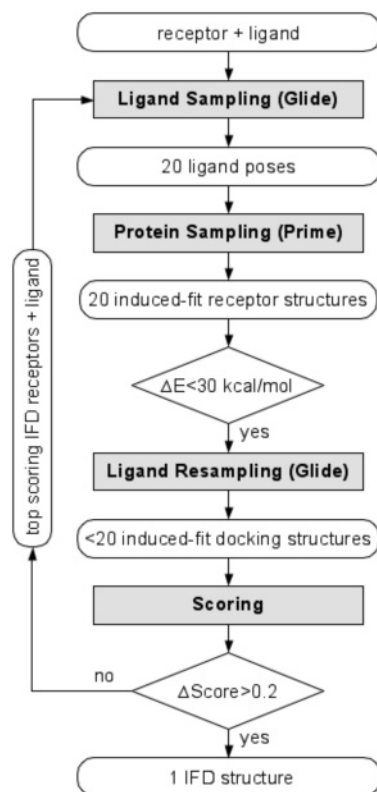
Figure 1. IFD flowchart. ΔE is the energy gap from the lowest energy structure.

top ranked structures from step 4 of the first round. The initial docking step of the second round does not use a softened potential.

Initial Ligand Sampling. The key challenge in the initial ligand docking step is minimizing the protein−ligand steric clashes that are manifested when docking into the unmodified binding site, while retaining sufficient structure in the modified binding site to avoid generating a large number of infeasible poses. The approach taken here to eliminate or at least significantly reduce blocking steric clashes is to scale the van der Waals radii of ligand and receptor atoms by 50%, and to temporarily replace residues predicted to be highly flexible with alanine. Van der Waals scaling is an automated, unbiased procedure that is easily applied to an arbitrary induced fit problem. However, in cases where a particular residue is directly blocking access to the binding site pocket, simply scaling the van der Waals radii of the receptor and ligand will not work. In this type of situation, removal of the side chain of the critical blocking residue (i.e., temporary mutation to Ala) is essential to obtain a low RMSD pose, suitable for serving as an initial guess in the protein sampling step. It is expected that residues with a high degree of structural flexibility are the ones most likely to block binding sites and therefore require mutation to Ala prior to the initial docking step. We have implemented a simple, unbiased procedure that identifies highly flexible residues that should be temporarily mutated to Ala, described in Materials and Methods. With implementation of this procedure, it is sufficient in the initial docking step to retain the top 20 poses as ranked by the GlideScore.

Receptor Sampling. In the second stage of the IFD protocol, the 20 poses retained from the first stage are used as a basis for protein structure prediction via the Refinement module of the Prime program.[30,31,34] Any residues that were replaced with alanine in the first stage of docking are restored to their original residue type, and then side-chain prediction and minimization

are performed for all 20 ligand/protein complexes. The backbone and ligand are also energy minimized, but not otherwise sampled. In all cases presented here, only residues having at least one atom within 5 Å of any of the 20 ligand poses are sampled; all other residues are held fixed. We have found that using a shell of 3 Å does not allow for sufficient protein flexibility and that a shell of 10 Å introduces noise as a result of sampling problems associated with the significantly larger number of movable residues. Although we have not exhaustively investigated the effect on shell size, it appears that a shell of 5 Å provides sufficient flexibility, while not overwhelming the computation with noise.

The output from this stage is 20 new receptor conformations optimized around each of the ligand poses from the first stage. Each structure is ranked by the total Prime energy (molecular mechanics plus implicit solvation) of the system, which includes the strain energy of the protein and ligand as well as the protein/ligand interaction energy. Of the 20 test cases described below that have a good pose from the initial ligand sampling stage, there is at least one complex within 20 kcal/mol of the lowest energy structure in which the ligand pose is within 2 Å of the correct answer. However, it is not always possible to identify the correct complex structure at this stage due to inaccuracies in the force field and solvent model, and incomplete sampling of both the ligand and receptor.

Ligand Resampling. In this stage, the ligand is redocked into the induced-fit structures from the previous stage that are within 30 kcal/mol of the lowest energy structure. This 30 kcal/mol window was set prior to running the test cases reported here, so although in retrospect, it appears a 20 kcal/mol window would have been sufficiently large to capture at least one good pose from the initial docking step, the larger 30 kcal/mol window was retained for these studies. Glide docking parameters at this stage are returned to their default values, thereby providing a hard potential function that will enhance sampling around the poses generated in the previous step. The ability of Glide to correctly predict the binding mode of a protein−ligand complex when presented with a receptor structure in a conformation that can accommodate the ligand has been tested on a substantial test suite of 263 complexes and is described elsewhere.[2]

Final Scoring. We have found that a high degree of robustness in selecting the correct binding mode can be achieved by combining the Prime energy and GlideScore in suitable proportions. Here, we briefly consider the strengths and weaknesses of the individual energy-based terms, and why combining the two provides a robust solution.

The energy model in Prime has been extensively tested in its ability to predict loop and side chain structures, via benchmark comparisons with PDB structural data.[30,31,34,35] While the great majority of loop structures are accurately predicted (median RMSDs are less than 1 Å for loops up to 10 residues), there are occasional outliers in which an alternative structure is selected in preference to the crystal structure, indicating difficulties with the force field or solvation terms, incorrect assignment of protonation states, or neglect of entropy effects. Force fields and solvation parameters for ligands are likely not as accurate as those for the protein residues, due to the wide range of chemistry that must be covered. Perhaps more importantly, the sampling performed in our IFD protocol described above represents a tradeoff between efficiency and thoroughness; there will be situations in which the lowest free energy nativelike pose is not located due to trapping in a nearby metastable local minimum, and this may enable an alternative

basin of attraction to appear lower in free energy. For these reasons, although the Prime energy succeeds in selecting the correct protein−ligand complex in many cases, it fails in others.

We have found from these studies that the Glide energy used to predict binding affinity, GlideScore, obtained from redocking into the hard potential, provides a greatly improved heuristic measure of the correctness of the binding mode of the complex. This result can be understood from the observation that ligand binding affinity is primarily driven by the hydrophobic effect, i.e., displacement of explicit waters into bulk solution coupled with burial of hydrophobic atoms of the ligand in lipophilic regions of the binding site cavity. The continuum solvation model used in Prime has its greatest difficultly in dealing with the free energies of bound waters in the binding site cavity; the empirically calibrated Glide binding affinity estimator has been optimized to provide accurate estimation of these quantities. Furthermore, the GlideScore function is "softer" than the molecular mechanics potential functions used in Prime and hence is less susceptible to random noise introduced by minor steric clashes that could in fact be relieved by better sampling. As long as the protein molecular mechanics energy is not too high, the correct docking pose will be dominated by optimization of the protein−ligand interaction energy, which is captured by GlideScore reasonably well.

The composite score used for final ranking of compounds is GlideScore + 0.05 × PrimeEnergy. This implies that in most instances, the GlideScore term is dominant, in line with the arguments made above. However, the small (5%) admixture of Prime energy is sufficient to eliminate predicted protein structures for which the energy gap (as compared to the lowest Prime energy) is large enough to rise above the sources of noise discussed above. Additional test cases will be required to fully validate the particular admixture used here; furthermore, the optimal admixture may change as the Glide and Prime energy models are improved over time. For the present, robust performance on the test suite presented below provides more than anecdotal evidence that the combination (a one parameter fit) is highly effective.

If the gap in composite scores between top ranked structures is below 0.2, indicating nearly isoenergetic solutions, the entire IFD protocol is repeated for the top ranked receptor structures using the results from the first round of IFD as a starting point. The only difference in this second round of IFD is that the default Glide potential function is used in the initial ligand docking stage instead of the softened potential described above. This ensures that ligand poses similar to ones that survived the first IFD round are sampled and scored in subsequent steps.

Graphical User Interface. The IFD protocol described here can be run from a graphical user interface accessible from within the program Maestro.[29] While it is possible to run the entire procedure in a fully automated mode, the interface supports modification of all parameters described above.

Materials and Methods

Structure Preparation. The coordinates for all proteins were obtained from the RCSB Protein Data Bank (PDB). Structures were prepared using the Maestro[29] software package and aligned using the Protein Structure Alignment module in Prime.[36] Hydrogen atoms were added, and the proteins were inspected visually for accuracy in the χ_2 dihedral angle of Asn and His residues and the χ_3 angle of Gln, and rotated by 180° when needed to maximize hydrogen bonding. The proper His tautomer was also manually selected to maximize hydrogen bonding. All Asp, Glu, Arg, and Lys residues were left in their charged state except for Glu143 of both thermolysin structures, which were protonated at the epsilon oxygen

to interact with the negatively charged carboxylate group of the ligand. Also, His231 of both thermolysin structures was prepared in its positively charged form to maximize hydrogen bonds with the protein and cocrystallized ligand.

If a PDB structure was missing side-chain atoms, Prime[29] was used to predict their location. Water molecules in all structures were removed. A brief relaxation was performed on each starting structure using the *Protein Preparation* module in Maestro with the "Refinement Only" option. This is a two-part procedure that consists of optimizing hydroxyl and thiol torsions in the first stage followed by an all-atom constrained minimization in the second stage to relieve clashes. The minimization was terminated when the RMSD reached a maximum value of 0.18 Å.

Ligands were obtained from the PDB crystal structures, and the appropriate bond orders and formal charges were manually adjusted in Maestro. Each ligand was subjected to a full minimization in the gas phase using the Impact program with the OPLS force field to eliminate bond length and bond angle biases from the crystal structure. A torsional constraint in the minimization was added to the hymenialdisine ligand from CDK-2 entry 1dm2 to keep the five-membered and seven-membered rings planar. This was due to inadequacies of the OPLS force field to properly account for the conjugated π-system. Additionally, the 1dm2 ligand had two tautomers with approximately equal energies in implicit solvent calculations. For this case, induced fit calculations were performed on both forms of the ligand and the lowest composite score at the end of the induced fit procedure is reported. In all cases, the minimized ligands were the starting point for the initial softened-potential docking step.

Temporary replacement of at least one flexible residue with alanine prior to the initial ligand docking step was performed for 14 of the 21 complexes. The rules used to select alanine mutations are presented below. The rules are followed in the order presented until a maximum of three residues have been identified. Table 1 shows the residues for each receptor that satisfy these rules.

1. If the X-ray structure of the receptor does not contain a bound ligand (apo protein), it is superimposed on existing holo structures of the same protein, if available. Any residues in the binding site with side-chain atoms that deviate more than 2.5 Å from the closest atom in the same residue of the holo protein are mutated to Ala. Note, this is not a comparison to identical atom types on the different proteins, but rather a comparison to the closest atom, which may or may not be the same atom type. This is done in order to distinguish residues that move significantly in space from those that undergo motions that place the overall side chain density in approximately the same place by interchanging positions of some atoms. In these studies, the protein that contained the ligand being docked was not included in this step.

2. Residues with side chains having multiple occupancy or missing density that are within 5 Å of the cocrystallized ligand are mutated to Ala.

3. If there are multiple independently refined structures in the unit cell, they are superimposed, and residues with side chains within 5 Å of the cocrystallized ligand having atoms that deviate more than 1.5 Å from the closest atom in the same residue are mutated to Ala. See rule 1 above for a description of how the deviation is calculated.

4. Residues with side chains atoms having *B*-factors greater than 40 Å^2 within 5 Å from the cocrystallized ligand are ranked by the atom with the highest *B*-factor on each side chain. If less than three residues have been mutated to Ala in steps 1−3, residues are taken from this list starting with the highest *B*-factors until a maximum of three residues (including those chosen in steps 1−3) have been mutated to Ala. We have found that mutating more than three residues in the active site to alanine creates a pocket that is excessively large, making it difficult to identify reasonable poses in the initial softened potential docking step.

Glide Docking. Glide 3.5[2,3] was used for all docking calculations. In the first stage of the IFD protocol, softened-potential docking is performed to generate 20 initial poses. The softened-potential docking consisted of scaling the van der Waals radii by 0.5 (for

Table 1. Potentially Flexible Residues That Were Replaced by Alanine Prior to the Initial Ligand Docking Step with an Explanation for How These Residues Were Identified[a]

| target | receptor (R_B) PDB ID | rules | | | | final Ala replacement |
		1. apo with difference compared to holo structure	2. partial occupancy or missing density	3. difference between structures in same crystal	4. *B*-factor >40 Å^2	
aldose reductase	2acr	_:Phe122, _:Leu300, _:Ser302	−	N/A	−	_:Phe122, _:Leu300, _:Ser302
antibody DB3	1dba	H:Trp100	−	N/A	H:Trp100	H:Trp100
CDK2	1buh	A:Phe80	−	N/A	−	A:Phe80
CDK2	1dm2	N/A	A:Asn132	N/A	A:Asp145	A:Asn132, A:Asp145
CDK2	1aq1	N/A	−	N/A	−	−
COX-2	3pgh	N/A	−	−	−	−
COX-2	1cx2	N/A	−	−	−	−
estrogen receptor	1err	N/A	−	−	A:His524, A:Leu536, A:Leu539	A:His524, A:Leu536, A:Leu539
estrogen receptor	3ert	N/A	−	N/A	A:Glu419, A:His524	A:Glu419, A:His524
FXa	1ksn	N/A	−	N/A	A:Glu147, A:Gln192, A:Arg222	A:Glu147, Gln192, Arg222
FXa	1xka	N/A	−	N/A	−	−
HIV-RT	1rth	N/A	−	N/A	A:Glu138, A:Trp229, A:Leu234	A:Glu138, A:Trp229, A:Leu234
HIV-RT	1c1c	N/A	−	N/A	A:Lys102, A:Phe227, A:Leu234	A:Lys102, A:Phe227, A:Leu234
neuraminidase	1a4q	N/A	−	−	−	−
neuraminidase	1nsc	N/A	−	−	−	−
PPARγ	1fm9	N/A	−	N/A	D:Phe282, D:Gln286, D:Phe363	D:Phe282, D:Gln286, D:Phe363
PPARγ	2prg	N/A	−	A:Phe282, A:Gln286	A:Phe363	A:Phe282, A:Gln286, A:Phe363
thermolysin	1kjo	N/A	−	N/A	−	−
thermolysin	1kr6	N/A	A:Leu133	N/A	−	A:Leu133
thymidine kinase	1ki4	N/A	−	−	−	−
thymidine kinase	1kim	N/A	−	A:Tyr132	−	A:Tyr132

[a] See text for a detailed description of the four rules.

receptor atoms with partial atomic charge $q \leq 0.25e$ and for ligand atoms with $q \leq 0.15e$) except in the event when alanine substitutions (described above) were introduced, in which case the receptor scaling was set to 0.7. This was done to partially offset the increased cavity size created by the alanine mutation. To enhance the hit rate of poses in the initial docking that are close to the correct answer, the Glide hydrogen bond energy cutoff filter (normally set to 0 kcal/mol) was decreased to −0.05 kcal/mol. This ensures that all retained poses contain at the very least a weak hydrogen bond with the receptor. Second, the Glide Coulomb−vdW energy cutoff filter is increased to 100 kcal/mol, enabling toleration of more steric clashes than in a normal docking run. In the case of HIV−RT with the 1rth/1c1c ligand/receptor pair, the calculations were performed with the hydrogen bond filter set to the default value of 0.0 kcal/mol. This was because with the default hydrogen bond filter of −0.05 only 1 pose was scored in the final stage, while all others were eliminated by the hydrogen bond filter, revealing that the binding site is highly nonpolar and that the hydrogen bond filter was eliminating too many poses. The 1rth ligand does not make any hydrogen bonds with the receptor in the crystal structure.

Poses with an RMSD of less than 0.5 Å and a maximum atomic displacement of less than 1.2 Å were eliminated as redundant in order to increase diversity in the retained ligand poses. An inner grid box of 10 Å was used to fit the ligand center while the outer grid box was defined by the longest ligand dimension in the starting protein. In the case when no ligand was present in the starting protein, an outer box size of 26 Å was used. The center of the grid box was defined by the center of the bound ligand when present and by a manually selected set of residues that surrounded the binding site for the apo proteins (2acr and 1buh). All docking calculations were run in the "Standard Precision" (SP) mode of Glide.

Prime Refinement. For each of the top 20 poses (with respect to GlideScore) from the initial softened-potential docking step, a full cycle of protein refinement was performed. Prime uses the OPLS parameter set[30,37,38] and a surface Generalized Born implicit solvent model.[39,40] First, a list is generated consisting of all residues

having at least one atom within 5 Å of an atom in any of the 20 ligand poses. Residues directly bound to metal ions in the binding site (such as Glu166, His146, and His142 bound to Zn in thermolysin) were manually omitted from the list. This unified list was used in the prediction of all structures, therefore allowing for a fair comparison in energies between the different refinement calculations. All side chains in the list underwent a conformational search and minimization.[31] Residues that were mutated to alanine in the initial docking stage were returned to their original identity prior to the search. After convergence to a low-energy solution, an additional minimization was performed allowing all residues in the list (backbone and side chain) and the ligand to be relaxed. The complexes were ranked by Prime energy (molecular mechanics plus solvation) and those within 30 kcal/mol of the minimum energy structure were passed through for a final round of Glide docking and scoring.

Ligand Resampling and Final Scoring. The minimized ligand used in the first docking step is redocked using Glide with default settings into each of the 20 receptor structures produced in protein refinement step. A composite score that accounts for the protein/ ligand interaction energy (GlideScore) and the total energy of the system (Prime energy) is calculated using the following equation: GlideScore + 0.05 × PrimeEnergy. In the case of Factor Xa with receptor 1xka a second round of IFD was performed using the results from the first round because the top three structures from the first round had nearly identical composite scores and different RMSDs, making it impossible to select a top scoring structure. The only difference in the second round is that default Glide settings (i.e., 1.0 and 0.8 for the van der Waals scaling for receptor and ligand atoms, respectively, and 0.0 for both the Coulomb-vdW and hydrogen bond energy cutoffs) were used for the initial docking step.

Results

Table 2 presents results for the induced fit cases included in our initial test suite. The test cases are formulated as described

Table 2. Ligand RMSDs (excluding hydrogens) for Rigid Receptor Docking and Induced Fit Docking of Ligand L_A from the Indicated PDB Structure into PDB Receptor Structure R_B

target	receptor (R_B)	ligand (L_A) from:	R_B residues that clash with L_A	ligand RMSD (Å)	
				rigid receptor docking	induced fit docking
aldose reductase	2acr:_	1ah3	_:W111, _:F122, _:L300	6.5	0.9
antibody DB3	1dba:H	1dbb	H:Y97, H:W100	7.6	0.3
CDK2	1buh:A	1dm2	A:K33, A:F80	6.4	1.1
CDK2	1dm2:A	1aq1	A:I10[b], A:G13, A:K33, A:L83[b], A:H84[b]	6.2	0.8
CDK2	1aq1:_	1dm2	−	0.6	0.8
COX-2	3pgh:A	1cx2	A:Y355, A:V523	11.1	1.0
COX-2	1cx2:A	3pgh	A:R120	6.6	1.0 (0.5[c])
estrogen receptor	1err:A	3ert	A:F404, A:M421, A:H524, A:L525	5.3	1.0
estrogen receptor	3ert:A	1err	A:L346, A:G420[b], A:M421, A:I424	2.3	1.4 (1.0[c])
Factor Xa	1ksn:A	1xka	A:Y99	9.3	1.5
Factor Xa[d]	1xka:C	1ksn	C:F174, C:Q192	5.3	1.5
HIV-RT	1rth:A	1c1c	A:W229, A:P236	2.5	1.3
HIV-RT	1c1c:A	1rth	A:P95, A:L100, A:K101[b], A:F227, B:E138	12.0	2.5
neuraminidase	1nsc:A	1a4q	A:I220, A:A244[b], A:E274	3.9	0.8
neuraminidase	1a4q:A	1nsc	−	1.0	1.7
PPARγ	1fm9:D	2prg	D:G284, D:S289	9.1	1.8 (0.4[e])
PPARγ	2prg:A	1fm9	A:F282, A:Q286, A:F360, A:F363, A:Y473	9.8	3.0 (1.5[f])
thermolysin	1kr6:A	1kjo	−	1.1	1.3
thermolysin	1kjo:A	1kr6	A:N112	3.5	3.2 (1.6[g])
thymidine kinase	1kim:A	1ki4	A:Y132, A:A168	4.7	0.4
thymidine kinase	1ki4:A	1kim	−	0.5	1.2

[a] Induced fit docking results are for the top ranked structure for each receptor/ligand pair. [b] Backbone clash. [c] RMSD of 2nd ranked IFD structure, which has nearly identical composite score as the top ranked structure; see text for details. [d] Second round of IFD was performed because nearly isoenergetic structures were returned from the first round. [e] RMSD excluding 10 atoms in the partially solvent exposed methyl-2-pyridinylamino tail of the ligand that has atoms with very high *B*-factors (>60 Å²). [f] RMSD excluding 13 atoms in the partially solvent exposed methylphenyloxazole tail of the ligand. [g] RMSD excluding 6 atoms in the symmetric di-carboxylate that are flipped 180° in the IFD structure.

in the Induced Fit Methodology section above; the ligand L_A from complex C_A is docked into the receptor conformation R_B from complex C_B. In four of the 21 test cases, rigid receptor docking yielded the correct binding mode (RMSD ≤ 1.8 Å), implying that R_B has a similar conformation to R_A in the binding site such that L_A binds in the correct mode without substantial steric clashes with the receptor. These four examples serve to demonstrate that even when induced fit effects are minimal, the IFD protocol can still yield the correct answer. This is important as it is not always obvious whether a receptor structure will undergo induced fit conformational changes upon binding of a ligand other than the one that it was cocrystallized with. The remaining 17 test cases required induced fit docking as evidenced by the large ligand RMSD values obtained from rigid receptor docking. Visual inspection of these examples confirmed that placing L_A in the correct binding mode into receptor R_B would result in steric clashes with the receptor. Table 2 includes a list of residues in R_B that clash with L_A when it is transplanted from C_A. In most cases, the ligands clash with side-chain atoms of the receptor, but in a few cases the ligand clashes with backbone atoms as well. The IFD protocol described above is capable of sampling dramatic side-chain conformational changes as well as minor changes in the backbone structure.

The average ligand RMSD for the IFD structures presented in Table 2 is 1.4 Å. The corresponding average RMSD for rigid receptor cross docking is 5.5 Å. The ligand RMSD for over 85% (18/21) of the cases is less than or equal to 1.8 Å. Excluding the three cases with RMSD over 1.8 Å, the average RMSD for the IFD structures is 1.1 Å, and 5.0 Å for rigid receptor cross docking. In the three cases with RMSDs over 1.8 Å, the core of the ligand is properly docked and all the key ligand/protein interactions are duplicated in the IFD structures. What follows is a detailed analysis of the 21 test cases.

Aldose Reductase. Rigid receptor cross docking of ligand **1** (tolrestat) into the 2acr structure yields a ligand pose with an RMSD of 6.5 Å when compared to the structure of **1** in 1ah3. The primary reason for this is that 2acr was cocrystallized with

the herbicide cacodylic acid (dimethylarsinic acid), which is a very small molecule that binds exclusively in the rigid catalytic binding site (near the NADP⁺ cofactor) and does not occupy the more flexible specificity pocket. Therefore, three residues, Trp111, Leu300, and Phe122, protrude into the specificity pocket, thus blocking binding of **1** in the correct pose. The initial ligand sampling step of the IFD protocol yielded 5 poses with ligand RMSDs below 2 Å. The lowest energy structure from the protein sampling step has a ligand RMSD of 1.4 Å. The second ranked structure (1.5 kcal/mol higher in energy) has an RMSD of 0.8 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 0.8 Å. Hydrogen bonds to Trp111, Tyr 48, and His110 in the 1ah3 crystal structure are duplicated in the top ranked IFD structure (Figure 2). The residue that undergoes the largest induced fit effect is Phe122, which is accurately predicted in the IFD structure. The χ_1 and χ_2 values in the starting 2acr structure are −76.2 and 78.0, respectively, and −57.1° and 103.2° in the target 1ah3 structure. The corresponding values in the best scoring IFD structure are −59.5° and 98.8°, which match well with the target values. In addition, Leu300 shifts out of the binding pocket, thus allowing **1** to bind in the correct pose. This shift is accurately predicted in the IFD structure. All other side-chains conformations in the IFD structure matched very well with the 1ah3 structure.

Anti-Progesterone Antibody DB3. Rigid receptor cross docking of ligand **2** (progesterone) into the 1dba (apo) structure yields a ligand pose with an RMSD of 7.6 Å when compared to the structure of **2** in 1dbb. The primary reason for this is that 1dba is an apo structure, which allows the indole side chain of TrpH100 and to a lesser extent the side chain of TyrH97 to protrude into the binding site. This conformation of the binding site, referred to as the "closed" form,[41] precludes binding of **2** in the correct pose to 1dba. The initial ligand sampling step of the IFD protocol yielded three poses with ligand RMSDs below 2 Å (ranging from 0.5 to 1.9 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 7.4 Å.

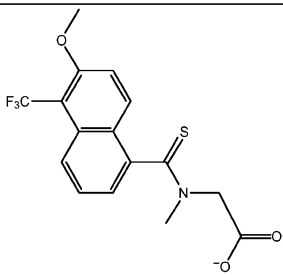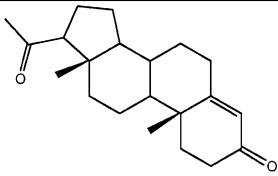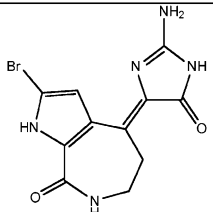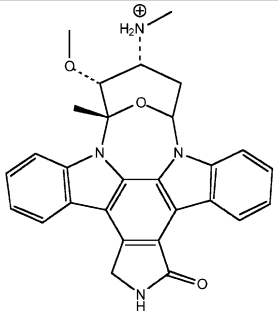**Table 3.** Ligand Structures and Names
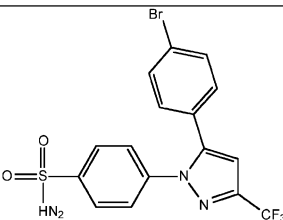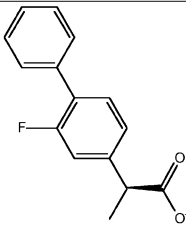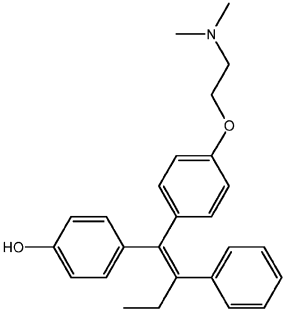
Structure	PDB code	Symbol used in the text	Name
	1ah3	1	tolrestat
	1dbb	2	progesterone
	1dm2	3	hymenialdisine
	1aq1	4	staurosporine
	1cx2	5	SC-558
	3pgh	6	flurbiprofen
	3ert	7	4-hydroxytamoxifen

Table 3. (Continued)
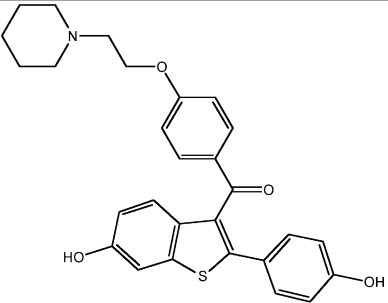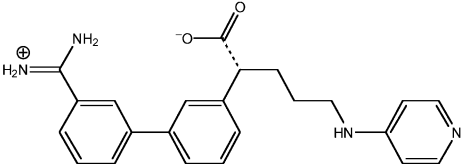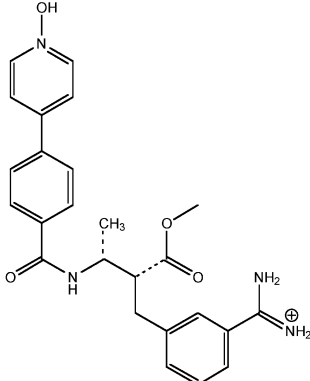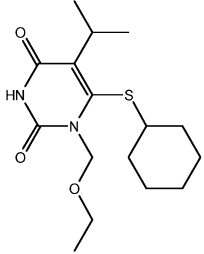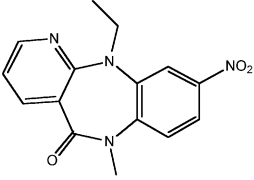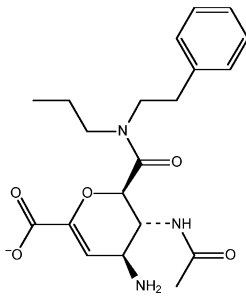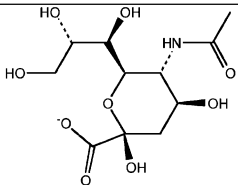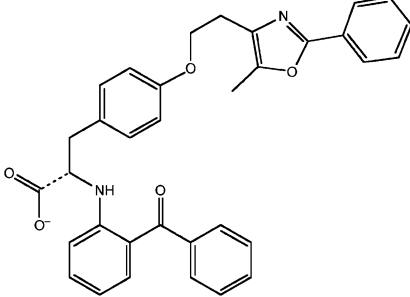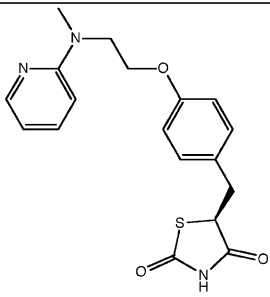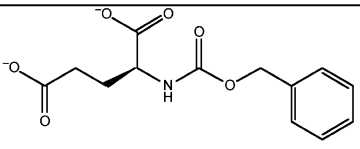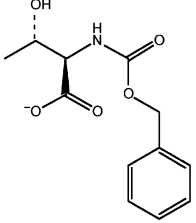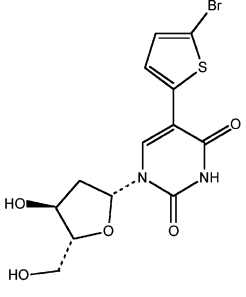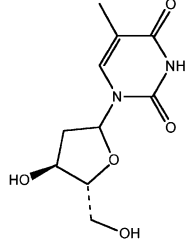
Structure	PDB code	Symbol used in the text	Name
	1err	**8**	raloxifene
	1xka	**9**	FX-2212A
	1ksn	**10**	FXV673
	1c1c	**11**	TNK-6123
	1rth	**12**	1051U91
	1a4q	**13**	dihydropyran-phenethylpropylcarboxamide

Table 3. (Continued)

Structure	PDB code	Symbol used in the text	Name
	1nsc	**14**	sialic acid
	1fm9	**15**	GI262570
	2prg	**16**	rosiglitazone, BRL49653
	1kr6	**17**	benzyloxycarbonyl-D-glutamic, Z-D-glutamic acid
	1kjo	**18**	N-benzyloxycarbonyl-L-threonine, Z-L-threonine
	1ki4	**19**	5-bromothienyldeoxyuridine
	1kim	**20**	deoxythymidine
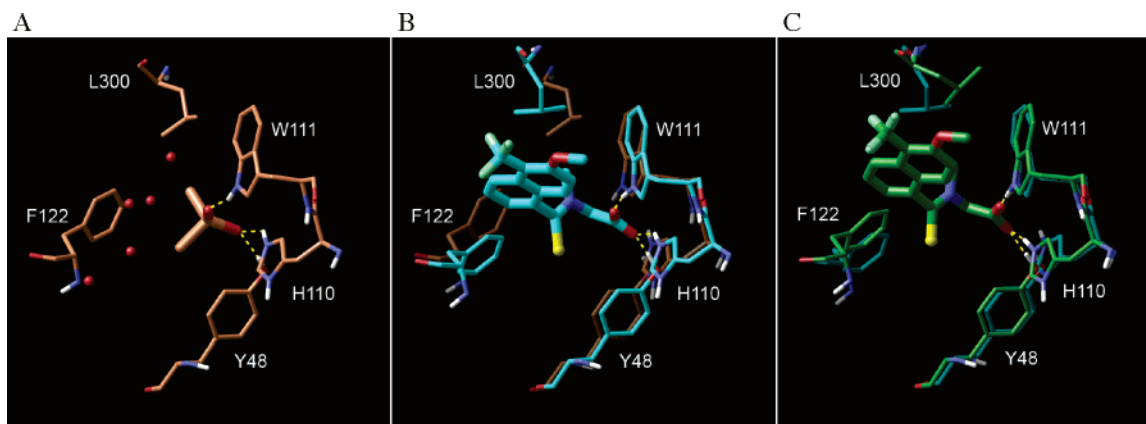
Figure 2. (A) Crystal structure of the binding site of 2acr with the ligand cacodylic acid and bound water molecules; (B) crystal structure of the binding site of 1ah3 with the ligand **1** (tolrestat) superimposed on the 2acr structure; (C) top ranked IFD structure in which **1** was docked into 2acr. The binding site of 1ah3 is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
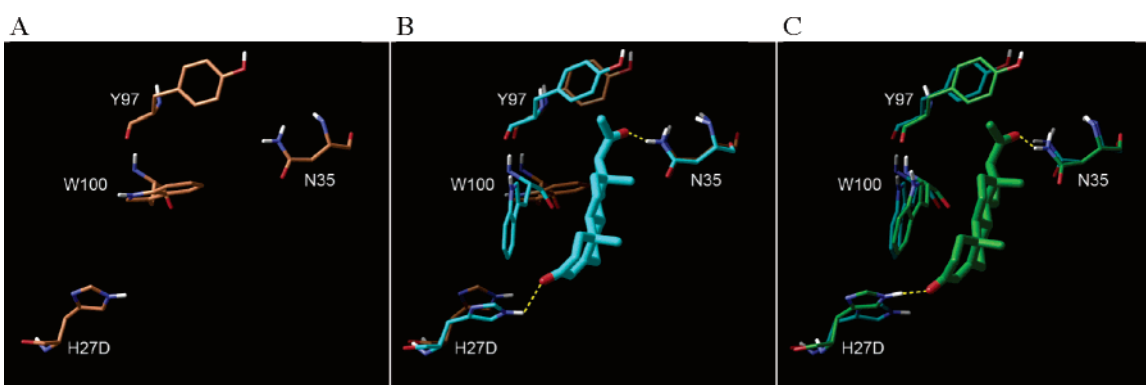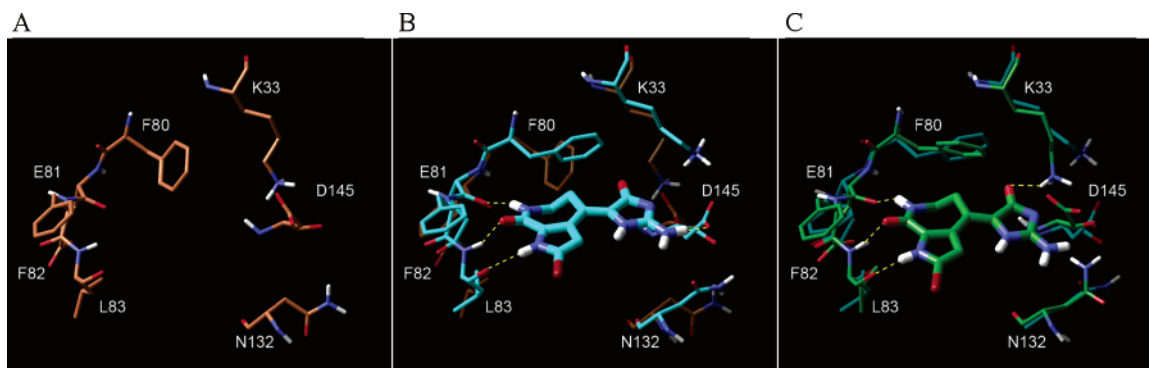

Figure 3. (A) Crystal structure of the binding site of the apo protein 1dba; (B) crystal structure of the binding site of 1dbb with **2** (progesterone) superimposed on the 1dba structure; (C) top ranked IFD structure in which **2** was docked into 1dba. The binding of site of 1dbb is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.

The second ranked structure (8.7 kcal/mol higher in energy) has an RMSD of 0.4 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 0.3 Å. The second and third ranked IFD structures also have ligand RMSDs of 0.3 Å. In the top ranked IFD structure, the side chain of TrpH100 swings out of the pocket, producing the so-called "open" form[41] of the binding site, thus allowing **2** to bind in the correct pose (Figure 3). The side chain of Tyr97 also rotates slightly about χ_2 to a conformation that more closely matches that of 1dbb. The χ_2 angle of Tyr97 in the starting 1dba structure is 90.9°, and 119.3° in 1dbb; in the IFD structure, χ_2 is 106.6°.

Cyclin-Dependent Kinase 2 (CDK-2). Rigid receptor cross docking of ligand **3** (hymenialdisine) into the 1buh structure yields a ligand pose with an RMSD of 6.4 Å when compared to the pose of **3** in 1dm2. The primary reason for this is that 1buh was crystallized without a ligand, which allows Phe80 and Lys33 to protrude into the binding site, thus blocking binding of **3** in the correct pose. The initial ligand sampling step of the IFD protocol yielded five poses with ligand RMSDs below 2 Å (ranging from 0.7 to 1.5 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 5.8 Å. However, a structure 0.3 kcal/mol higher in energy has an RMSD of 1.1 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 1.1 Å. The IFD structure captures side-chain rearrangements of Lys33 and Phe80 needed to correctly dock **3** (Figure 4). In the 1dm2 structure there are a total of four hydrogen bonds between the ligand and the receptor. The top IFD structure captures the three critical hydrogen bonds with the backbone of Glu81,

Phe82, and Leu83. However, interaction of the amine on the imidazol-4-one group is not precisely duplicated in the IFD structure. In 1dm2, the amine group is hydrogen bonded to Asp145, which in turn is hydrogen bonded to Lys33. In the IFD structure, Asp145 is also hydrogen bonded to Lys33, but is 4.5 Å from the ligand amino group. The reason is that the backbone of Asp145 is shifted by 2.2 Å away from the pocket in 1buh compared to 1dm2, and this dramatic movement of the backbone is not sampled in the current IFD protocol. Instead Asp145 is hydrogen bonded to the amino group of Asn132.

Rigid receptor cross docking of ligand **4** (staurosporine) into the 1dm2 structure yields a ligand pose with an RMSD of 6.2 Å when compared to the structure of **4** in 1aq1. The primary reason for this is that 1dm2 was cocrystallized with a smaller ligand (**3**), which allows the side-chain and backbone atoms of a number of residues (Ile10, Gly13, Lys33, Leu83, and His84) to protrude into the binding site to differing degrees. This blocks **4** from binding to 1dm2 in the correct pose. The initial ligand sampling step of the IFD protocol yielded 4 poses with ligand RMSDs below 2 Å (ranging from 0.5 to 1.4 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 4.0 Å. The second ranked structure (6.5 kcal/mol higher in energy) has an RMSD of 0.7 Å. The top ranked structure from the ligand resampling and scoring step has a ligand RMSD of 0.8 Å. The second ranked IFD structure has a ligand RMSD of 1.1 Å. The backbone carbonyls of Ile10, Leu83, and His84 in 1dm2 all block binding of **4** in the correct pose. In the top ranked IFD structure, these three carbonyls retreat from the binding site to locations nearly identical to those of

Figure 4. (A) Crystal structure of the binding site of the apo protein 1buh; (B) crystal structure of the binding site of 1dm2 with **3** (hymenialdisine) superimposed on the 1buh structure; (C) top ranked IFD structure in which **3** was docked into 1buh. The binding site of 1dm2 is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
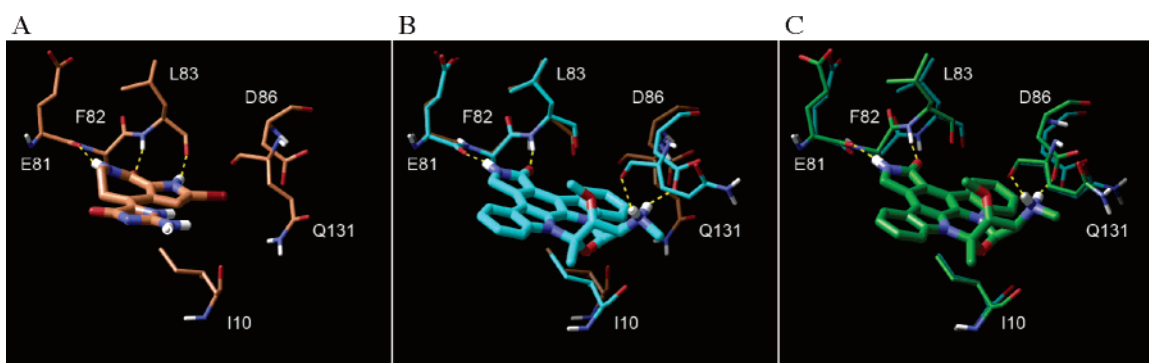
Figure 5. (A) Crystal structure of the binding site of 1dm2 with **3** (hymenialdisine); (B) crystal structure of the binding site of 1aq1 with **4** (staurosporine) superimposed on the 1dm2 structure; (C) top ranked IFD structure in which **4** was docked into 1dm2. The binding site of 1aq1 is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown. The side chain of Phe82 has been removed for clarity as it is largely obscured by the ligand and has nearly identical conformation in all three structures.

the 1aq1 structure. Gly13 and the side chain of Lys33 also move slightly out of the pocket (by 0.3 and 0.9 Å, respectively) to allow **4** to bind in the correct binding mode (not shown). The four hydrogen bonds between **4** and the 1aq1 receptor are precisely duplicated in the top ranked IFD structure (Figure 5).

The third CDK2 example involves cross docking of the smaller **3** ligand into the large binding site of 1aq1. Rigid receptor docking yields a ligand pose with an RMSD of 0.6 Å. Induced fit docking also yields a good pose with an RMSD of 0.8 Å.

Cyclooxygenase-2 (COX-2). Rigid receptor cross docking of inhibitor **5** into the 3pgh structure yields a ligand pose with an RMSD of 11.1 Å when compared to the structure of **5** in 1cx2. The primary reason for this is that 3pgh was cocrystallized with a much smaller ligand (the nonselective inhibitor **6**), which allows the side chains of Tyr355 and Val532 to protrude into the binding site. This blocks **5** from binding to 3pgh in the correct pose. The initial ligand sampling step of the IFD protocol yielded 10 poses with ligand RMSDs below 2 Å (ranging from 1.0 to 1.6 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 1.3 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 1.0 Å. The second and third ranked IFD structures have ligand RMSDs of 1.0 and 1.1 Å, respectively. In the top ranked IFD structure, the backbone and side-chain atoms of Val523 (which blocks binding of **5** in the correct pose) move "up" away from the binding site to a location nearly identical to that of the 3pgh structure (Figure 6C). Tyr355, which also blocks binding of **5**, moves "down" away from the binding site to a location that allows **5** to bind in the correct pose. There are three hydrogen bonds between **5** and the receptor in the

1cx2 structure: Arg513 to the ligand sulfonamide group, Arg120 to the ligand trifluoromethyl group, and the ligand sulfonamide group to the backbone carbonyl of Ser353. All three hydrogen bonds are precisely duplicated in the top ranked IFD structure (Figure 6C). Ser353, which does not move during the induced fit refinement, is not displayed in the Figure for clarity.

Rigid receptor cross docking of **6** (flurbiprofen) into the 1cx2 structure yields a ligand pose with an RMSD of 6.6 Å when compared to the structure of **6** in 3pgh. The primary reason for this is that 1cx2 was cocrystallized with a ligand (**5**) that contains an extra Br atom on the terminal phenyl ring and a trifluoromethyl group in place of the larger propionate group of 1cx2. This has the effect of forcing the side chain of Ser530 out of the binding site by about 0.8 Å. In addition, the smaller trifluoromethyl group allows the side chain of Arg120 to protrude into the binding site by about 1.5 Å relative to 3pgh (Figure 7A and 6B). The end result for rigid cross docking of **6** into 1cx2 is that the ligand binds in an orientation that is 180° flipped from that of the 3pgh crystal structure. The carboxylic group, rather than being hydrogen bonded to Tyr355 and Arg120 is hydrogen bonded to Tyr385 and Ser530 on the opposite side of the binding site. The initial ligand sampling step of the IFD protocol yielded 12 poses with ligand RMSDs below 2.0 Å (ranging from 1.3 to 1.9 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 1.7 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 1.0 Å. The second ranked IFD structure is nearly isoenergetic with respect to the composite score and has an RMSD of 0.5 Å. The two hydrogen bonds between **6** and Tyr355 and Arg120 in the 3pgh structure are duplicated in the IFD structure (Figure 7C). Both residues
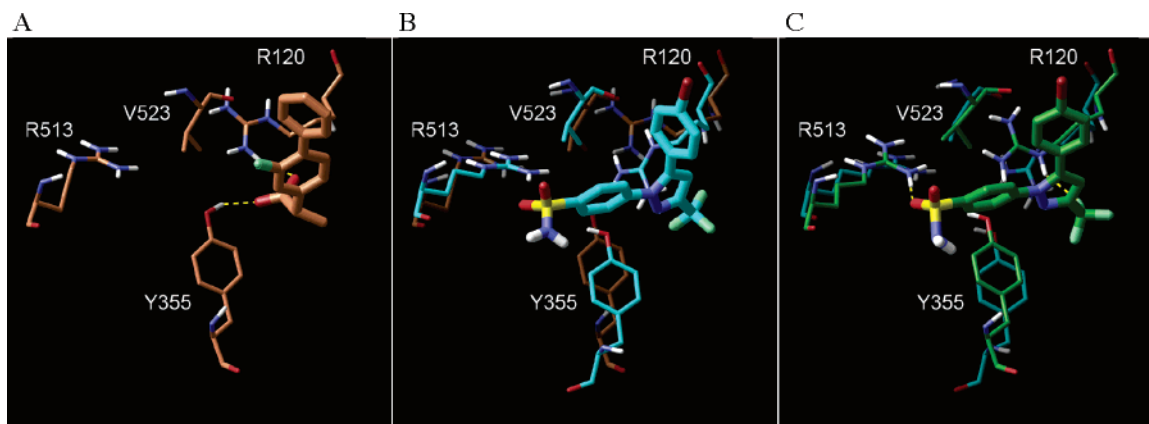
Figure 6. (A) Crystal structure of the binding site of 3pgh with **6** (flurbiprofen); (B) crystal structure of the binding site of 1cx2 with **5** superimposed on the 3pgh structure; (C) top ranked IFD structure in which **5** was docked into 3pgh. The binding site of 1cx2 is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
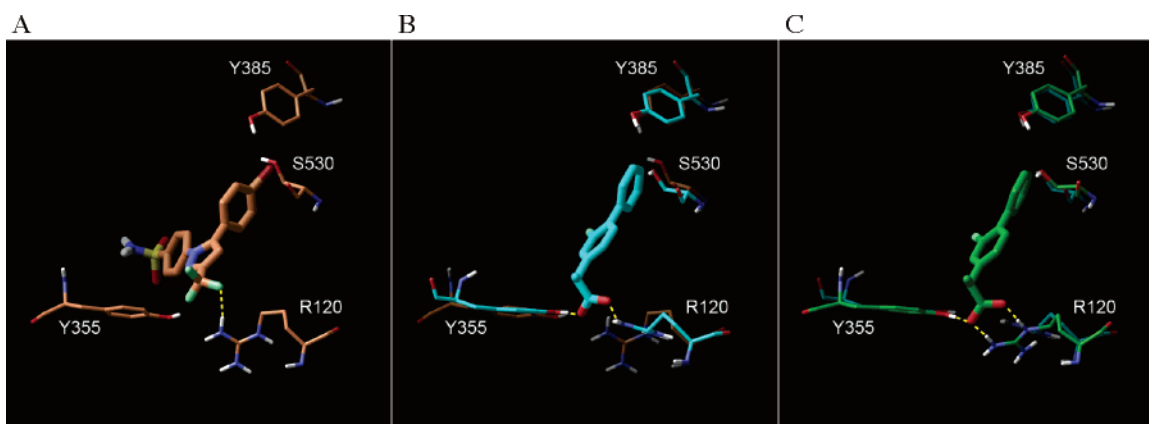
Figure 7. (A) Crystal structure of the binding site of 1cx2 with **5**; (B) crystal structure of the binding site of 3pgh with **6** (flurbiprofen) superimposed on the 1cx2 structure; (C) top ranked IFD structure in which **6** was docked into 1cx2. The binding site of 3pgh is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
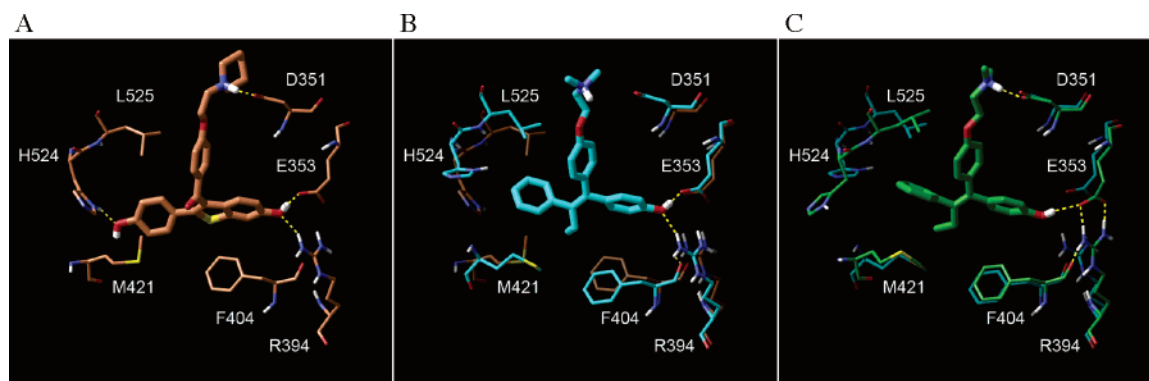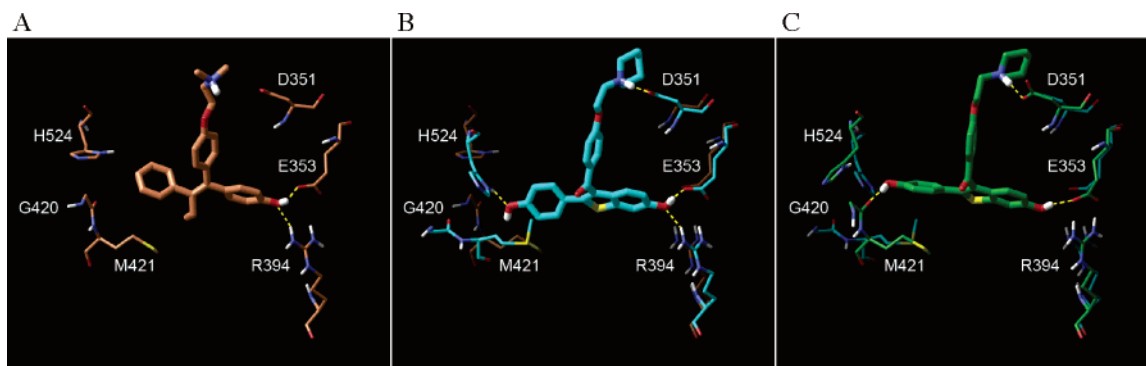

Figure 8. (A) Crystal structure of the binding site of 1err with **8** (raloxifene); (B) crystal structure of the binding site of 3ert with **7** (4-hydroxytamoxifen) superimposed on the 1err structure; (C) top ranked IFD structure in which **7** was docked into 1err. The binding site of 3ert is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.

adopt conformations that more closely match the 3pgh structure. In addition, the conformation of Ser530 changes slightly in the IFD structure and is nearly superimposable with that of the 3pgh crystal structure (Figure 7C).

Estrogen Receptor. Rigid receptor cross docking of ligand **7** (4-hydroxytamoxifen) into the 1err structure yields a ligand pose with an RMSD of 5.3 Å when compared to the structure of **7** in 3ert. There are a number of residues that adopt conformations in 1err that preclude binding of **7** in the correct pose. The most significant difference between 1err and 3ert is the side-chain conformation of His524; in 1err, His524 is hydrogen bonded to the phenoxy group of ligand **8** (raloxifene),

while in 3ert there is no analogous hydrogen bond acceptor on the ligand. This places His524 into a position that precludes binding of **7** in the correct pose (Figure 8B). In addition, Phe404, Met421, and Leu525 preclude proper binding of **7** to the rigid 1err structure. The initial ligand sampling step of the IFD protocol yielded 10 poses with ligand RMSDs below 2.0 Å (ranging from 1.1 to 1.6 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 1.2 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 0.95 Å. In addition to the low ligand RMSD, there is remarkable similarity between the 3ert crystal structure and the top ranked IFD structure. The conformation

Figure 9. (A) Crystal structure of the binding site of 3ert with **7** (4-hydroxytamoxifen); (B) crystal structure of the binding site of 1err with **8** (raloxifene) superimposed on the 3ert structure; (C) top ranked IFD structure in which **8** was docked into 3ert. The binding site of 1err is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
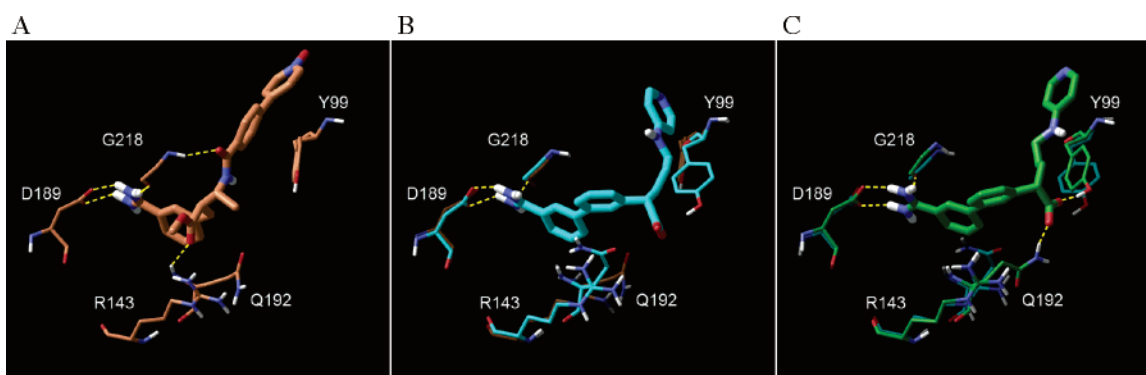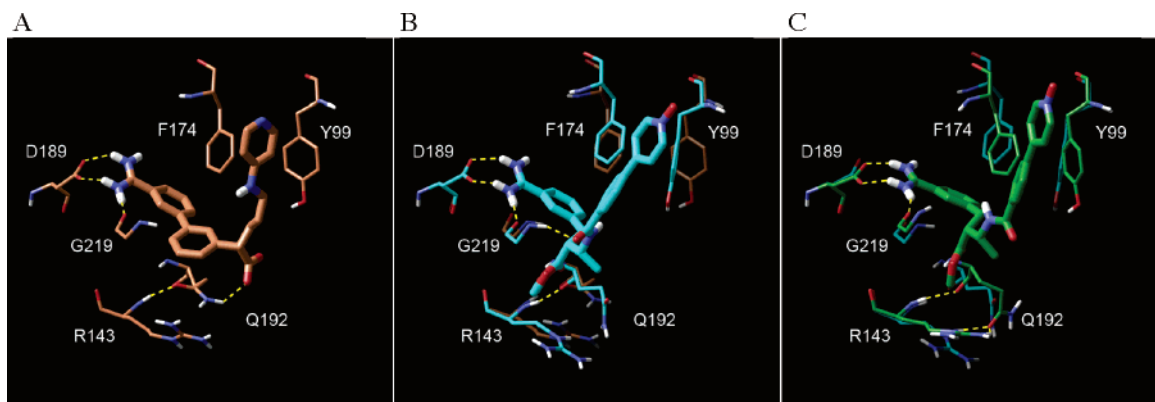

Figure 10. (A) Crystal structure of the binding site of 1ksn with **10**; (B) crystal structure of the binding site of 1xka with **9** superimposed on the 1ksn structure; (C) top ranked IFD structure in which **9** was docked into 1ksn. The binding site of 1xka is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.

of Phe404, Met421, and Leu525, which in 1err preclude binding of **7** in the correct pose, are nearly superimposable in both structures (Figure 8C). His524 rotates to a conformation that allows 4 **7** to bind in the correct pose, although its conformation differs somewhat from the 3ert structures. Interestingly, the *B*-factors for His524 are the highest of any residue in the binding site (greater than 50 Å^2), suggesting that this residue in 3ert is somewhat flexible.

Rigid receptor cross docking of ligand **8** (raloxifene) into the 3ert structure yields a ligand pose with an RMSD of 2.3 Å when compared to the structure of **8** in 1err. The primary reason for this is that 3ert was cocrystallized with a smaller ligand (4-hydroxytamoxifen) that does not contain the hydroxyl group on the phenyl ring or the ketocarbonyl of **8** (Figure 9A). The absence of the hydroxyl group, which is hydrogen bonded to His524 in the 1err crystal structure (Figure 9B), causes the side chain of His524 to swing out of the binding site. The absence of the ketocarbonyl in 4-hydroxytamoxifen allows Leu346 to protrude into the binding site by about 0.8 Å relative to the location of Leu346 in 3ert, thus blocking binding of the larger ligand, **8** (not shown). The initial ligand sampling step of the IFD protocol yielded 12 poses with ligand RMSDs below 2 Å (ranging from 1.0 to 1.9 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 1.1 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 1.4 Å. The second ranked IFD structure, which is nearly isoenergetic with the top ranked structure with respect to the composite score (0.09 higher) has a ligand RMSD of 1.0 Å. In the top ranked IFD structure, Leu346 and Ile424 adopt conformations very similar to those observed in 1err (not shown). In addition, the hydrogen bonds to Asp351 and Glu353 are duplicated in the IFD structure

(Figure 9C). The side chain of Arg394 is hydrogen bonded to a hydroxyl group of **8** in 1err; however, this hydrogen bonded is absent in the IFD structure, in which Arg394 forms a tighter salt bridge with Glu353. Although His524 in the IFD structure moves toward the ligand to a position that more closely matches that of 1err, it does not form the hydrogen bond to the hydroxyl group of **8** observed in 1err. Instead, a hydrogen bond to the backbone of Gly420 is observed in the IFD structure.

Factor Xa. Rigid receptor cross docking of ligand **9** into the 1ksn structure yields a ligand pose with an RMSD of 9.3 Å when compared to the structure of **9** in 1xka. The primary reason for this is that 1ksn was cocrystallized with a rigid, linear ligand (**10**), which allows the side chain of Tyr99 to protrude into the binding site in such a way to preclude binding of **9** in the correct pose. The initial ligand sampling step of the IFD protocol yielded two poses with ligand RMSDs below 2.0 Å (ranging from 1.5 to 1.6 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 2.8 Å. The second ranked structure (1.9 kcal/mol higher in energy) also has an RMSD of 3.1 Å and the third ranked structure (2.0 kcal/mol higher in energy than the lowest energy structure) has an RMSD of 1.5 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 1.5 Å. The bidentate hydrogen bond between the guanine group of **9** and Asp189 is duplicated in the IFD structure (Figure 10C). In addition, the side chain of Tyr99 rotates to a conformation that is similar to the 1xka crystal structure, allowing **9** to dock in the correct pose. Interestingly, Tyr99 in the IFD structure forms a hydrogen bond to the carboxylate group of the ligand, which is not observed in 1xka. However, analysis of other Factor Xa crystal structures reveals that Tyr99 does indeed hydrogen bond to ligands highly analogous to **9**. For example, the cocrystallized ligand in 1iqi

Figure 11. (A) Crystal structure of the binding site of 1xka with **9**; (B) crystal structure of the binding site of 1ksn with **10** superimposed on the 1xka structure; (C) top ranked IFD structure in which **10** was docked into 1xka. The binding site of 1ksn is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
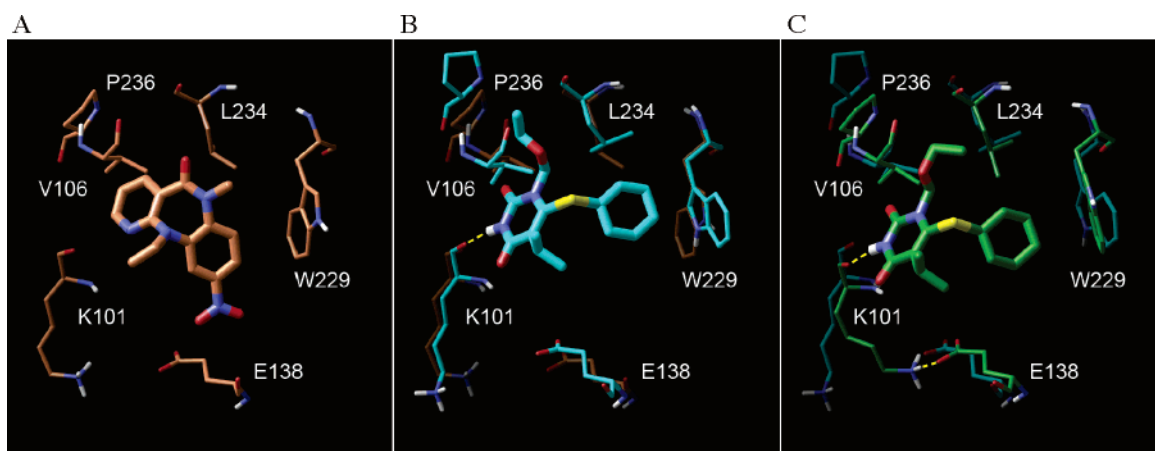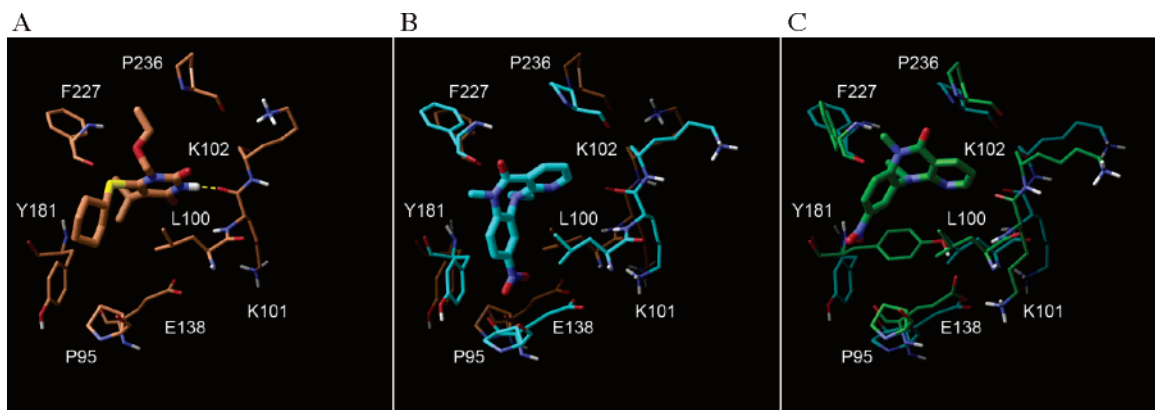

Figure 12. (A) Crystal structure of the binding site of 1rth with **12**; (B) crystal structure of the binding site of 1c1c with **11** superimposed on the 1rth structure; (C) top ranked IFD structure in which **11** was docked into 1rth. The binding site of 1c1c is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.

(M55125) contains a carboxylate group analogous to the one in **9** and is hydrogen bonded to Tyr99. Another interesting difference between the 1xka crystal structure and the IFD structure is a hydrogen bond between Gln192 and the carboxylate group of **9** that is only observed in the IFD structure. Again, analysis of other Factor Xa crystal structures reveals that a similar hydrogen bond is observed for ligands analogous to **9**. For example, the cocrystallized ligand in 1fax (DX-9065a) contains a carboxylic group analogous to the one in **9** and is hydrogen bonded to bonded to Gln192. The amide group of Gln192 in 1fax is flipped 180° relative to the orientation in 1xka. It is not clear from the 1xka structure whether the conformation of the side-chain amide of Gln192 is correct; energy calculations show that a flip of 180° in the χ_3 angle produces a structure that is about 0.1 kcal/mol lower in energy.

Rigid receptor cross docking of ligand **10** into the 1xka structure yields a ligand pose with an RMSD of 5.3 Å when compared to the structure of **10** in 1ksn. The primary reason for this is that 1xka was cocrystallized with a ligand (**9**) that occupies a slightly different region of the binding site in one area, which allows the side-chain atoms of residues Arg143 and Gln192 to protrude into the binding site. This blocks **10** from binding to 1ksn in the correct pose. The initial ligand sampling step of the IFD protocol did not yield any poses with ligand RMSD below 2 Å. However, two reasonably good poses were produced, both with RMSDs of 3.1 Å. Three nearly isoenergetic low energy structures (separated by 0.6 and 0.3 kcal/mol) were produced by the protein sampling step; the ligand RMSDs are

9.7, 2.9, and 2.3 Å. The top three structures from the ligand resampling and scoring step have RMSDs of 3.8, 2.1 and 1.6 Å and nearly identical composite scores (the top ranked and third ranked are with 0.2 of another). Since it is not possible to unambiguously identify the correct pose among these three nearly isoenergetic structures, a second round of IFD was performed, as described in Materials and Methods, to select the top pose. The final ligand resampling and scoring step of the second round of IFD produced six nearly isoenergetic poses with RMSDs ranging from 1.5 to 1.8 Å. In the top IFD structures the side chains of Arg143 and Gln192 (both of which block binding of **10** in the correct pose) move away from the binding site to positions similar to those of the 1ksn structure. Additionally, there are shifts in the side-chain positions of Tyr99 and Phe174 that help with the packing of the **10** ligand in the correct binding mode. Four out of the five hydrogen bonds between **10** and the 1ksn receptor are replicated in the top ranked IFD structure (Figure 11), while the hydrogen bond to Gly219 is lost in exchange for a new hydrogen bond to the side chain of Arg143.

HIV Reverse Transcriptase (HIV-RT). Rigid receptor cross docking of ligand **11** into the 1rth structure yields a ligand pose with an RMSD of 2.5 Å when compared to the structure of **11** in 1c1c. The reason is that Pro236 in 1rth, which is in a short four residue loop, is directed into the binding site relative to its conformation in 1c1c (Figure 12B). This blocks binding of **11** through a clash with the ethoxy group of the ligand, which is not present in the 1rth ligand (**12**). In addition, Trp229 to a

Figure 13. (A) Crystal structure of the binding site of 1c1c with **11**; (B) crystal structure of the binding site of 1rth with **12** superimposed on the 1c1c structure; (C) top ranked IFD structure in which **12** was docked into 1c1c. The binding site of 1rth is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.

lesser extent protudes into the binding site, causing a clash with the phenyl ring of the ligand (Figure 12B). The initial ligand sampling step of the IFD protocol yielded 13 poses with ligand RMSDs below 2 Å (ranging from 1.1 to 1.7 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 1.2 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 1.3 Å. The second ranked IFD structure, which is nearly isoenergetic with respect to the composite score (0.2 higher), has a ligand RMSD of 0.9 Å. Trp229 in the IFD structure moves by about 0.8 Å out of the binding site to accommodate the phenyl ring of **11**. By comparison, Trp229 in 1c1c is shifted out of the binding site by about 1.0 Å relative to 1rth. No movement of Pro236 is observed; however, the ligand still binds in nearly the correct binding mode. The reason is that the flexible ethoxy group of **11** swings away from the Pro236 to a hydrophobic region of the protein defined by Leu234. The single hydrogen bond between the ligand and the backbone carbonyl of Lys101 is duplicated in the IFD structure (Figure 12C).

Rigid receptor cross docking of ligand **12** into the 1c1c structure yields a ligand pose with an RMSD of 12.0 Å when compared to the structure of **12** in 1rth. There are a number of structural differences between the binding sites of 1c1c and 1rth that lead to this large RMSD. First, the ligand 1c1c (**11**) is hydrogen bonded to the backbone carbonyl of Lys101 (Figure 13A), which pulls this residue and Leu100 into the binding site by 0.6−1.4 Å, blocking binding of **12** into the rigid 1c1c structure. Second, the smaller cyclohexyl group of **11**, compared to the nitrophenyl group of **12**, allows Pro95 and Glu138 to shift into the binding site (Figure 13B). This blocks **12** from binding in the correct pose. Finally, Phe227 swings into the binding site by about 1.0 Å, also blocking binding of **12** in the correct pose (Figure 13B). The initial ligand sampling step of the IFD protocol does not produce any poses with RMSD below 2 Å; the pose closest to **12** in 1rth has an RMSD of 2.1 Å. The lowest energy structure from the protein sampling step has a ligand RMSD of 4.2 Å. The second ranked structure (4.2 kcal/mol higher in energy) has an RMSD of 2.6 Å and the third ranked structure (4.7 kcal/mol higher in energy than the lowest energy structure) has an RMSD of 2.1 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 2.5 Å (Figure 13C). Despite the greater than 2.0 Å RMSD, the top ranked IFD structure exhibits all the correct ligand/protein interactions observed in the 1rth crystal structure, with the exception of Tyr181. In the crystal structure, Tyr181 forms a "shifted parallel" $\pi-\pi$ stacking interaction with the nitrophenyl group of the ligand, while in the IFD structure,

the ligand forms a T-shaped $\pi-\pi$ stacking interaction with Tyr181. **12** is rotated in the binding site by about 40° relative to its position in 1rth. The side chains of Leu100, Lys102, and Glu138, and the backbones of Pro95, Lys101, and Pro236 all shift to locations in the IFD structure that are very similar to the 1rth structure.

Neuraminidase. Rigid receptor cross docking of the dihydropyrancarboxamide inhibitor in 1a4q (dihydropyran-phenethylpropylcarboxamide, **13**) into the 1nsc structure yields a ligand pose with an RMSD of 3.9 Å when compared to the 1a4q crystal structure. The primary reason for this is that 1nsc was cocrystallized with a smaller ligand (sialic acid), which allows the side-chains of Ala244 and Ile220 to collapse into the pocket that in 1a4q accommodates the phenethyl group of **13**. In addition, Glu274, which is hydrogen bonded to the two hydroxyl groups of sialic acid, protrudes into the binding pocket, thus blocking binding of the propyl group of **13**. The initial ligand sampling step of the IFD protocol yielded three poses with ligand RMSDs below 2.0 Å (ranging from 0.9 to 1.9 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 3.8 Å. The fourth ranked structure (8.1 kcal/mol higher in energy) has an RMSD of 1.3 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 0.8 Å. The pocket that accommodates the phenethyl group of **13** is enlarged by about 0.7 Å relative to the starting 1nsc structure. In 1nsc, the distance from the CB of Ala244 to CD1 of Ile220 is 5.99 Å, while in 1a4q this distance is 6.76 Å. In the IFD structure, this separation distance is 6.72 Å. Additionally, the side chain of Glu274, which in 1nsc blocks binding of **13**, adopts a conformation in the IFD structure that is very similar to that of the 1a4q crystal structure. All the hydrogen bonds to the ligand in 1a4q are duplicated in the IFD structure (Figure 14C).

Rigid receptor cross docking of ligand **14** (sialic acid) into the 1a4q structure yields a ligand pose with an RMSD of 1.0 Å when compared to the structure of **14** in 1nsc. IFD also yields the correct pose with slightly higher RMSD of 1.7 Å.

Peroxisome Proliferator Activated Receptor Gamma (PPARγ). Rigid receptor cross docking of ligand **15** into the 2prg structure yields a ligand pose with an RMSD of 9.8 Å when compared to the structure of **15** in 1fm9. The primary reason for this is that 2prg was cocrystallized with a much smaller ligand (**16**), which allows the side-chain atoms of a number of residues to protrude into the binding site to differing degrees, thus blocking binding of the larger **15** ligand in the correct pose. The most striking difference between the two structures is Phe363, which in 2prg is rotated to a conformation
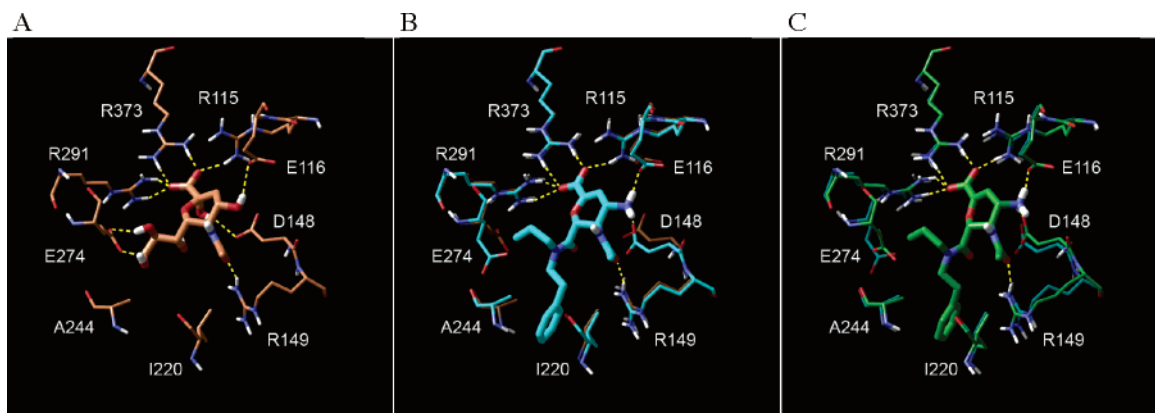
Figure 14. (A) Crystal structure of the binding site of 1nsc with **14** (sialic acid); (B) crystal structure of the binding site of 1a4q with **13** (dihydropyran-phenethylpropylcarboxamide) superimposed on the 1nsc structure; (C) top ranked IFD structure in which **13** was docked into 1nsc. The binding site of 1a4q is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
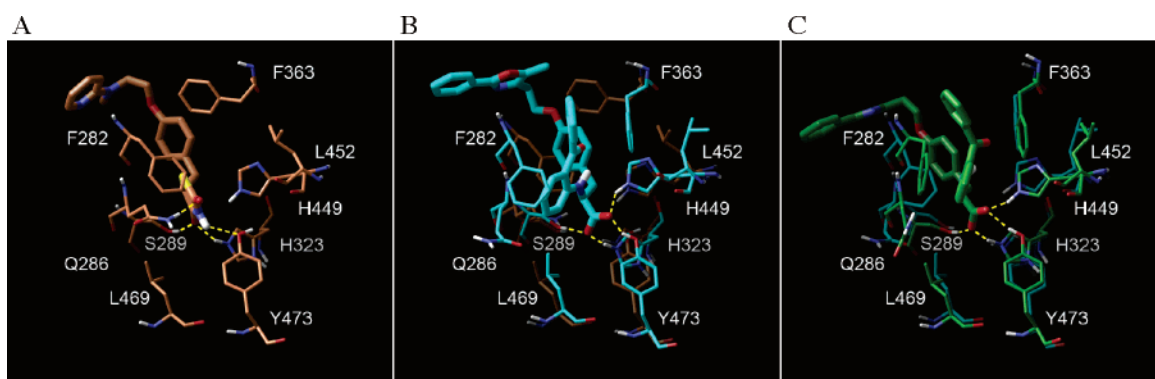
Figure 15. (A) Crystal structure of the binding site of 2prg with **16**; (B) crystal structure of the binding site of 1fm9 with **15** superimposed on the 2prg structure; (C) top ranked IFD structure in which **15** was docked into 2prg. The binding site of 1fm9 is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.

that in rigid docking would block one of the terminal phenyl groups of **15** (Figure 15B). The side chains of three other aromatic residues (Phe282, Phe360, and Tyr473) also block binding of the larger **15** ligand by virtue of being rotated into the binding site toward the smaller **16** ligand. In addition, Gln286 in 2prg, which is hydrogen bonded to one of the carbonyls of the thiazolidinedione group of **16**, is directed into the binding site, thus blocking **15**. The initial ligand sampling step of the IFD protocol dose not yield any poses with ligand RMSD below 2 Å; there are four reasonably good poses with RMSDs ranging from 2.1 to 2.5 Å. The lowest energy structure from the protein sampling step has a ligand RMSD of 2.6 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 3.0 Å (Figure 15C). The second ranked IFD structure has a ligand RMSD of 2.8 Å. The core of the ligand is very well predicted; however, the partially solvent exposed methylphenyloxazole tail of **15** is slightly shifted and flipped by 180° in the IFD structure relative to the conformation in 1fm9. There are no specific interactions between the receptor and this portion of the ligand. Excluding the methylphenyloxazole tail of the ligand gives an RMSD of 1.5 Å, which is a better indicator of the accuracy of the IFD structure. The hydrogen bonds between **15** and Ser289, His323, His449, and Try473 in 1fm9 (Figure 15B) are precisely duplicated in the IFD structure (Figure 15C).

Rigid receptor cross docking of ligand **16** (rosiglitazone) into the 1fm9 structure yields a ligand pose with an RMSD of 9.1 Å when compared to the structure of **16** in 2prg. The primary reason for this is that 1fm9 was cocrystallized with a much larger ligand (**15**), which results in an binding site that is more open

than is required to bind the smaller ligand, **16**. In particular, the side chain of Phe363 in 1fm9 is directed away from the binding site, while in 2prg, the side chain is positioned inside the binding site. The result of the more open pocket is that **16** can explore more regions of the binding site. The best scoring pose from rigid docking has the pyridine ring of **16** in the same location as the phenyl ring of the 1fm9 ligand (**15**), which is incorrect. The initial ligand sampling step of the IFD protocol yielded only 1 pose with a ligand RMSD below 2 Å (1.9 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 9.7 Å. The second ranked structure (1.0 kcal/mol higher in energy) has an RMSD of 1.9 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 1.8 Å. However, the methyl-2-pyridinylamino tail of the ligand is partially solvent exposed and in 2prg has B-factors greater than 60 $Å^2$. The ligand RMSD excluding this group is 0.4 Å. Phe282, Gly284, Leu465, and Tyr473 all move in the IFD structure to positions that are nearly identical to the 2prg crystal structure (Figure 16C). The side chain of Phe363, which starts off outside the pocket in 1fm9 (Figure 16A), partially swings into the pocket, forming interactions with **16** that are similar to that of the 2prg structure.

Thermolysin. Rigid receptor cross docking of ligand **17** (benzyloxycarbonyl-D-glutamic, Z-D-glutamic acid) into the 1kjo structure yields a ligand pose with an RMSD of 3.5 Å when compared to the structure of **17** in 1kr6. The primary reason for this is that 1kjo was cocrystallized with a smaller ligand (Z-L-threonine, **18**), which has the side-chain of Asn112 in a position that blocks the proper binding orientation of the backbone carboxylate group of **17**. As a result, the side chain
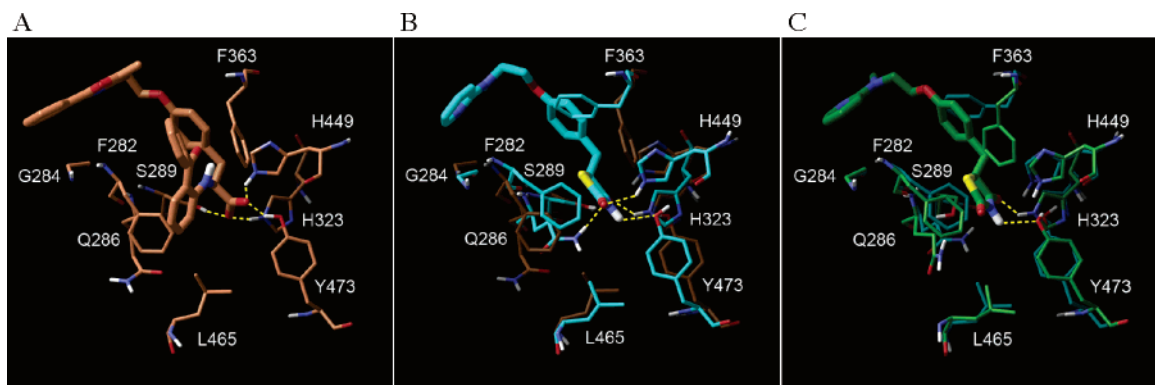
Figure 16. (A) Crystal structure of the binding site of 1fm9 with **15**; (B) crystal structure of the binding site of 2prg with **16** (rosiglitazone) superimposed on the 1fm9 structure; (C) top ranked IFD structure in which **16** was docked into 1fm9. The binding site of 2prg is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
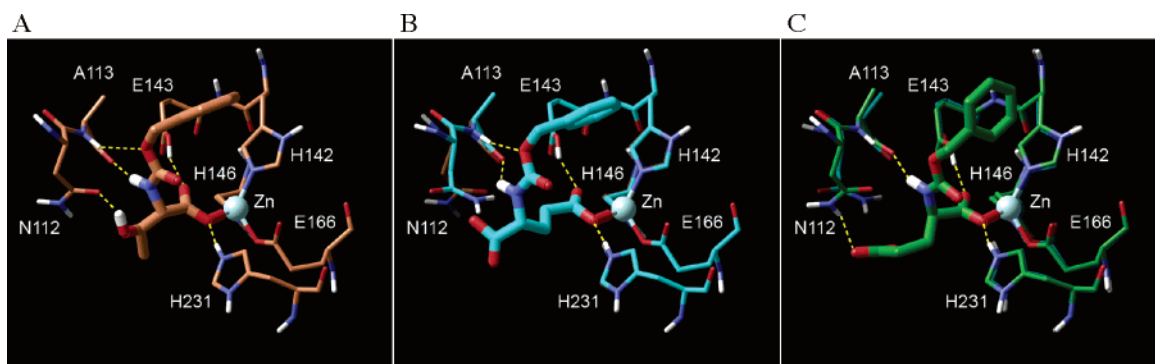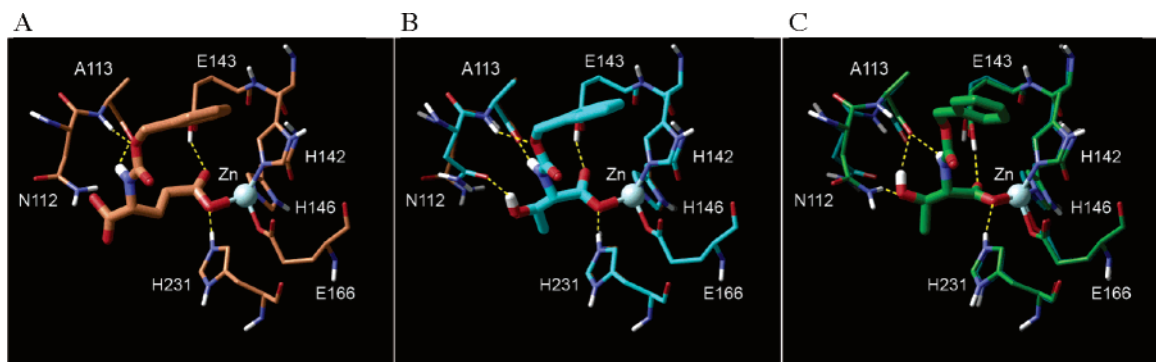

Figure 17. (A) Crystal structure of the binding site of 1kjo with **18** (Z-L-threonine); (B) crystal structure of the binding site of 1kr6 with **17** (Z-D-glutamic acid) superimposed on the 1kjo structure; (C) top ranked IFD structure in which **17** was docked into 1kjo. The binding site of 1kr6is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.

carboxylate in the rigid receptor cross docking incorrectly interacts with Arg203. The initial ligand sampling step of the IFD protocol yields no poses with ligand RMSDs below 2.0 Å (all are between 3.6 and 7.5 Å) and are mostly in similar binding modes to the rigid receptor cross docking pose described above, i.e., having one carboxylate interacting with the bound Zn ion and the other in an incorrect binding mode. The lowest energy structure from the protein sampling step has a ligand RMSD of 5.8 Å and the lowest RMSD pose of 3.4 Å is 11.5 kcal/mol higher in energy. All structures from this stage have one carboxylate interacting with the bound metal ion while in all cases but one, the other carboxylate is interacting with Arg203, which is not observed in the 1kr6 crystal structure. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 3.2 Å. The second and third ranked IFD structures each have ligand RMSDs of 3.3 Å. In all three of these structures the positions of the two carboxylates are switched in the binding site with respect to the 1kr6 structure (Figure 17C). Due to the diacid nature of **17**, it is possible to make almost identical interactions with the protein when the ligand carboxylates swap positions. In fact, the top ranked pose has a ligand RMSD of 1.6 Å when excluding the six atoms from the two carboxylates. Three of the hydrogen bonds between **17** and the 1kr6 receptor as well as the metal interaction are mimicked in the top ranked IFD structure (Figure 17C). Additionally, the backbone carboxylate picks up a hydrogen bond with the side chain of Asn112 that is not seen in the 1kr6 structure. One hydrogen bond to the **17** amide is lost because the shorter carboxylate group is ligated to the Zn, which brings the ligand closer to the Zn and away from the backbone NH of Ala113.

Rigid receptor cross docking of ligand **18** (N-benzyloxycarbonyl-L-threonine, Z-L-threonine) into the 1kr6 structure yields a ligand pose with an RMSD of 1.1 Å when compared to the structure of **18** in 1kjo. Therefore, it appears that there are no major induced fit effects involved in binding **18** to 1kjo. The primary reason for this is that 1kr6 was cocrystallized with a similar but slightly larger ligand (**17**), which allows the **18** ligand to successfully dock without encountering any receptor clashes. The initial ligand sampling step of the IFD protocol yielded seven poses with ligand RMSDs below 2 Å (ranging from 1.1 to 1.4 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 2.2 Å while the second ranked structure (6.5 kcal/mol higher in energy) has an RMSD of 2.3 Å and the third ranked structure (7.3 kcal/mol higher in energy than the lowest energy structure) has an RMSD of 1.6 Å. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 1.3 Å, which came from the lowest energy structure in the protein sampling stage. The second and third ranked IFD structures have ligand RMSDs of 1.1 and 1.6 Å, respectively. The overall movement needed to accommodate the **18** ligand in the 1kr6 structure is minimal, as evidenced from the very low RMSD obtained from rigid receptor cross docking. In the top ranked IFD structure the primary electrostatic and nonpolar interactions with the receptor are accurately reproduced (Figure 18C).

Thymidine Kinase. Rigid receptor cross docking of ligand **19** (5-bromothienyldeoxyuridine) into the 1kim structure yields a ligand pose with an RMSD of 4.7 Å when compared to the structure of **19** in 1ki4. The primary reason for this is that 1kim was cocrystallized with a smaller ligand (deoxythymidine, **20**), which allows the side chains of Tyr132 and Ala168 to protrude

Figure 18. (A) Crystal structure of the binding site of 1kr6 with **17** (Z-D-glutamic acid); (B) crystal structure of the binding site of 1kjo with **18** (Z-L-threonine) superimposed on the 1kr6 structure; (C) top ranked IFD structure in which **18** was docked into 1kr6. The binding site of 1kjo is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.
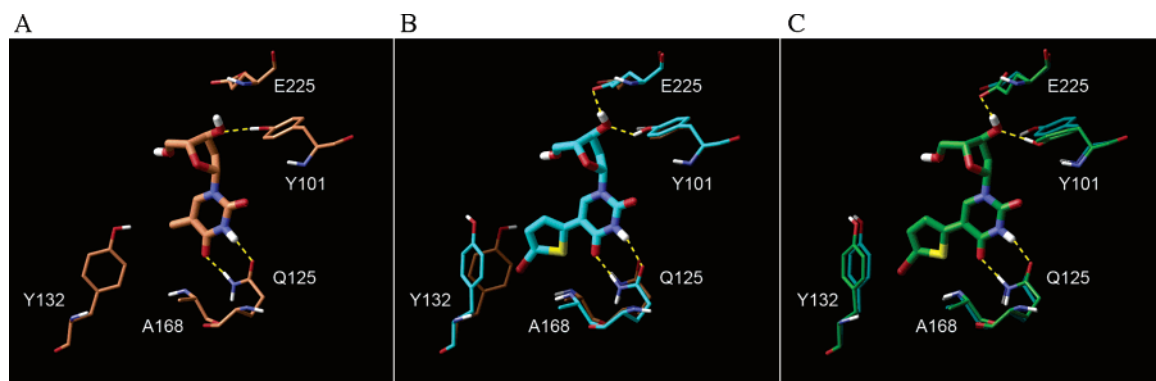
Figure 19. (A) Crystal structure of the binding site of 1kim with **20** (deoxythymidine); (B) crystal structure of the binding site of 1ki4 with **19** (5-bromothienyldeoxyuridine) superimposed on the 1kim structure; (C) top ranked IFD structure in which **19** was docked into 1kim. The binding site of 1ki4 is also shown for comparison. Only residues that undergo significant movement or are hydrogen bonded to the ligand are shown.

into the binding site. This blocks the larger ligand **19** from binding in the correct pose to 1kim. The initial ligand sampling step of the IFD protocol yielded 14 poses with ligand RMSDs below 2 Å (ranging from 0.5 to 1.4 Å). The lowest energy structure from the protein sampling step has a ligand RMSD of 0.3 Å. The top four structures from this step all have RMSDs below 1.0 Å and are within 15.1 kcal/mol of the lowest energy structure. The top ranked structure from the final ligand resampling and scoring step has a ligand RMSD of 0.4 Å. The second and third ranked IFD structures both have ligand RMSDs of 1.0 Å. In the top ranked IFD structure, Tyr132 adopts a side-chain conformation that is nearly identical to that of the 1ki4 structure (Figure 19C). The χ_1 and χ_2 angle of Tyr132 start at −55.1 and 153.0° in 1kim, respectively. In 1ki4, these dihedral angles are −69.2 and 89.8 and in the top ranked IFD structure, the angles are −69.4 and 86.3°. A more subtle induced fit effect is apparent with Ala168 in which the methyl side chain bends slightly away from the pocket in the IFD structure to accommodate the large **19** ligand. The four hydrogen bonds between the ligand and receptor in 1ki4 are reproduced in the top ranked IFD structure.

Rigid receptor cross docking of **20** (deoxythymidine) into the 1ki4 structure yields a ligand pose with an RMSD of 0.5 Å when compared to the structure of **20** in 1kim. Induced fit docking produces a similarly good pose with an RMSD of 1.2 Å.

Computation Times. The average time required for a complete IFD calculation is 5 h on a single 1.6 GHz AMD64 Opteron processor. On 20 such processors the calculation takes only about 30 min. The initial Glide docking step on a single processor takes about 10 min. The Prime protein sampling step takes about 10 min per complex; the job can be distributed, so

for 20 complexes on 20 processors, the calculation requires a total of 10 min. The final docking step takes about 10 min per complex; this job can also be distributed, and since there are no more than 20 complexes, the calculation takes 10 min on 20 processors.

Discussion

The binding sites that we have examined are diverse, with regard to both size and polarity, and similarly bind many different types of ligands. Furthermore, many of the test cases, which were selected specifically because the receptor pairs exhibited large structural differences, required significant simultaneous motions of multiple side chains in order to modify (open or close) a pocket that was otherwise blocked. The results presented above demonstrate that the induced fit algorithm we have developed satisfies the objectives proposed at the beginning of this paper; it is robust across a wide range of targets, can be applied in an automated fashion, and completes using an acceptable amount of computation time. The structures generated by the protocol, arguably, are of sufficient quality to be used for subsequent structure-based drug design efforts (although this will have to be shown explicitly, by performing virtual screening and lead optimization studies using induced fit structures as a starting point). At the very least, these modeled receptor/ligand complexes can be visually inspected by modelers and medicinal chemists to obtain qualitative ideas about how to modify lead compounds, just as would be the case if the cocrystallized structure was obtained experimentally.

The success of our composite scoring function is highly encouraging, and our expectation is that it will work in the great majority of induced fit docking problems. However, one would expect in a small fraction of cases that the correct structure will

not be the most highly ranked. The most common situation would be a single, specific low lying alternative solution that the scoring function ranks as lower in energy. In some such cases, visualization and experience with the binding modes of the receptor in question can help to choose between two such structures. An alternative is to employ both structures in a virtual screening or lead optimization effort, until more experimental data is available.

The present results only consider induced fit problems in which the primary changes in the receptor structure are restricted to side chains, and in which backbone motions are relatively small. However, there are of course many receptors which can exhibit large changes in loop conformations upon ligand binding, for example kinases, where the activation loop or hinge region can adopt different conformations at a relatively low energetic cost. The induced fit methodology that we describe here can be applied to such problems by introducing loop prediction into the protocol.

We have chosen in this paper to evaluate the quality of our induced fit results via superposition with the crystal structure of the target complex, and evaluation of ligand RMSDs between the calculated and target structures; the results so obtained demonstrate that, at the very least, the induced fit structures are of sufficient quality to enable visual inspection to benefit a lead optimization effort. However, the quantitative question of performance of induced fit structures in a virtual screening experiment, with regard to enrichment factors and ability to yield correctly docked poses of active compounds other than that used to generate the induced fit (e.g., of compounds that correctly docked into the crystal structure of the target complex), has not yet been established. Tests of this type are straightforward to carry out and in our view represent the most critical evaluation of any induced fit protocol.

In summary, we have developed the first practical, robust, and accurate induced fit methodology that is suitable for immediate deployment in a modern drug discovery environment. The most important measure of the effectiveness of the methodology will come from actual use in structure-based drug design projects; our most recent version of the software, in which many steps have been automated, will facilitate use by a wide range of groups in the biotechnology and pharmaceutical industry. Feedback from these applications will enable a more precise evaluation of the capabilities and limitations of our approach.

Supporting Information Available: Table of protein RMSDs. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Brooijmans, N.; Kuntz, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335−373.

(2) 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*, 1739−1749.

(3) 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*, 1750−1759.

(4) Shoichet, B. K.; Kuntz, I. D. Matching Chemistry And Shape In Molecular Docking. *Protein Eng.* **1993**, *6*, 723−732.

(5) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1−5.

(6) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein−ligand docking. *Proteins-Struct. Funct. Genet.* **1999**, *37*, 228−241.

(7) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein−ligand docking using GOLD. *Proteins-Struct. Funct. Genet.* **2003**, *52*, 609−623.

(8) Gschwend, D. A.; Good, A. C.; Kuntz, I. D. Molecular docking towards drug discovery. *J. Mol. Recognit.* **1996**, *9*, 175−186.

(9) Teague, S. J. Implications of protein flexibility for drug discovery. *Nature Rev. Drug Discovery* **2003**, *2*, 527−541.

(10) 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.

(11) Carlson, H. A.; McCammon, J. A. Accommodating Protein Flexibility in Computational Drug Design. *Mol. Pharmacol.* **2000**, *57*, 213−218.

(12) Jiang, F.; Kim, S. H. Soft Docking − Matching Of Molecular-Surface Cubes. *J. Mol. Biol.* **1991**, *219*, 79−102.

(13) Nakajima, N.; Higo, J.; Kidera, A.; Nakamura, H. Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation. *Chem. Phys. Lett.* **1997**, *278*, 297.

(14) Wasserman, Z. R.; Hodge, C. N. Fitting an inhibitor into the active site of thermolysin: A molecular dynamics case study. *Proteins-Struct. Funct. Genet.* **1996**, *24*, 227.

(15) Dinola, A.; Roccatano, D.; Berendsen, H. J. C. Molecular-Dynamics Simulation Of The Docking Of Substrates To Proteins. *Proteins-Struct. Funct. Genet.* **1994**, *19*, 174.

(16) Mangoni, R.; Roccatano, D.; Di Nola, A. Docking of flexible ligands to flexible receptors in solution by molecular dynamics simulation. *Proteins-Struct. Funct. Genet.* **1999**, *35*, 153−162.

(17) Carlson, H. A. Protein flexibility and drug design: how to hit a moving target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447−452.

(18) Knegtel, R. M. A.; Kuntz, I. D.; Oshiro, C. M. Molecular docking to ensembles of protein structures. *J. Mol. Biol.* **1997**, *266*, 424−440.

(19) Broughton, H. B. A method for including protein flexibility in protein−ligand docking: Improving tools for database mining and virtual screening. *J. Mol. Graphics Modelling* **2000**, *18*, 247−+.

(20) 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-Struct. Funct. Genet.* **2002**, *46*, 34−40.

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

(22) Kua, J.; Zhang, Y.; McCammon, J. A. Studying enzyme binding specificity in acetylcholinesterase using a combined molecular dynamics and multiple docking approach. *J. Am. Chem. Soc.* **2002**, *124*, 8260−8267.

(23) Kua, J.; Zhang, Y. K.; Eslami, A. C.; Butler, J. R.; McCammon, J. A. Studying the roles of W86, E202, and Y337 in binding of acetylcholine to acetylcholinesterase using a combined molecular dynamics and multiple docking approach. *Protein Sci.* **2003**, *12*, 2675−2684.

(24) Frimurer, T. M.; Peters, G. H.; Iversen, L. F.; Andersen, H. S.; Moller, N. P. H.; Olsen, O. H. Ligand-induced conformational changes: Improved predictions of ligand binding conformations and affinities. *Biophys. J.* **2003**, *84*, 2273−2281.

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

(26) Ota, N.; Agard, D. A. Binding mode prediction for a flexible ligand in a flexible pocket using multi-conformation simulated annealing pseudo crystallographic refinement. *J. Mol. Biol.* **2001**, *314*, 607−617.

(27) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft Docking and Multiple Receptor Conformations in Virtual Screening. *J. Med. Chem.* **2004**, *47*, 5076−5084.

(28) Meagher, K. L.; Carlson, H. A. Incorporating protein flexibility in structure-based drug discovery: Using HIV-1 protease as a test case. *J. Am. Chem. Soc.* **2004**, *126*, 13276−13281.

(29) Schrödinger, LLC: Portland, OR.

(30) Jacobson, M. P.; Kaminski, G. A.; Friesner, R. A.; Rapp, C. S. Force field validation using protein side chain prediction. *J. Phys. Chem. B* **2002**, *106*, 11673−11680.

(31) Jacobson, M. P.; Friesner, R. A.; Xiang, Z. X.; Honig, B. On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **2002**, *320*, 597−608.

(32) Cheney, D. L.; Langley, D. R.; Mueller, L. Protein ensemble − Based lead docking: A comparison of FLO, GLIDE, GOLD and ICM in cross docking scenarios. *Abstracts Of Papers*; The 227th National Meeting of the American Chemical Society, American Chemical Society: Washington, DC, 2004; U1029−U1029.

(33) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins-Struct. Funct. Bioinformatics* **2004**, *57*, 225−242.

(34) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351−367.

(35) Andrec, M.; Harano, Y.; Jacobson, M. P.; Friesner, R. A.; Levy, R. M. Complete protein structure determination using backbone residual dipolar couplings and side chain rotamer prediction. *J. Struct. Funct. Genomics* **2002**, *2*, 103−111.

(36) Yang, A. S.; Honig, B. An integrated approach to the analysis and modeling of protein sequences and structures. I. Protein structural alignment and a quantitative measure for protein structural distance. *J. Mol. Biol.* **2000**, *301*, 665−678.

(37) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.

(38) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.

(39) Ghosh, A.; Rapp, C. S.; Friesner, R. A. Generalized born model based on a surface integral formulation. *J. Phys. Chem. B* **1998**, *102*, 10983−10990.

(40) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. The SGB/NP Hydration Free Energy Model Based on the Surface Generalized Born Solvent Reaction Field and Novel Non-Polar Hydration Free Energy Estimators. *J. Comput. Chem.* **2002**, *23*, 517−529.

(41) Arevalo, J. H.; Hassig, C. A.; Stura, E. A.; Sims, M. J.; Taussig, M. J.; Wilson, I. A. Structural analysis of antibody specificity. Detailed comparison of five Fab'-steroid complexes. *J. Mol. Biol.* **1994**, *241*, 663−690.