A Flexible Approach to Induced Fit Docking

Sander B. Nabuurs,*,† Markus Wagener,‡ and Jacob de Vlieg†,‡

Center for Molecular and Biomolecular Informatics, Radboud University, Nijmegen, The Netherlands, and Department of Molecular Design and Informatics, N.V. Organon, Oss, The Netherlands

Received May 22, 2007

We present Fleksy, a new approach to consider both ligand and receptor flexibility in small molecule docking. Pivotal to our method is the use of a receptor ensemble to describe protein flexibility. To construct these ensembles, we use a backbone-dependent rotamer library and implement the concept of interaction sampling. The latter allows the evaluation of different orientations of ambivalent interaction partners. The docking stage consists of an ensemble-based soft-docking experiment using FlexX-Ensemble, followed by an effective flexible receptor–ligand complex optimization using Yasara. Fleksy produces a set of receptor–ligand complexes ranked using a consensus scoring function combining docking scores and force field energies. Averaged over three cross-docking datasets, containing 35 different receptor–ligand complexes in total, Fleksy reproduces the observed binding mode within 2.0 Å for 78% of the complexes. This compares favorably to the rigid receptor FlexX program, which on average reaches a success rate of 44% for these datasets.

Introduction

The major aim of computational structure-based drug design is to utilize knowledge of receptor structure to predict and optimize binding of small molecule drug candidates. Once the primary requirement, a high resolution model of the receptor structure, is available, accurately predicting the three-dimensional arrangement of the small molecules in the protein–ligand complex becomes the key objective. To tackle this so-called "docking" problem, many different programs have been developed, of which DOCK,[1] FlexX,[2] GOLD,[3] Autodock,[4] and Glide[5,6] are among the most popular. The mentioned tools are based on a range of different concepts, and each comes with its own set of strengths and weaknesses. One feature most docking programs share, however, is that they traditionally aim at positioning a flexible ligand into a rigid binding site. Computational feasibility is the main reason for utilizing a rigid protein snapshot in the docking process, as the number of degrees of freedom that have to be considered grows exponentially with the number of accessible receptor conformations.

Over the years, however, it has become increasingly clear that protein structural flexibility plays a crucial role in receptor–ligand complex formation and ideally should be considered during the drug design process.[7,8] This shift of focus from the traditional "key-and-lock" concept[9] to the "induced fit" model[10] has prompted the need for computational tools which are able to consider protein plasticity.

Theoretically, an explicit solvent molecular dynamics (MD[a]) simulation of the receptor in presence of its unbound ligand should result in the formation of the correct protein–ligand complex, while taking both receptor and ligand flexibility into account.[11] Unfortunately, imperfections in the available MD force fields and the vast computational requirements of this technique make the use of molecular dynamics simulations as a docking tool not feasible in a typical drug discovery environment. To overcome this problem, several methods have been conceived in recent years to at least partially incorporate protein flexibility into the docking process, while keeping computational demands within reasonable limits. As a result, each of the above-mentioned docking programs is nowadays able to consider some degree of receptor flexibility in one way or another.

For a detailed discussion of different approaches developed, we direct the reader to a series of recent reviews on the topic of protein flexibility in structure-based drug design.[11−14] Generally speaking, these methods can be divided in four different categories. First, soft-docking strategies consider and allow small degrees of local flexibility by attenuating the repulsion potential between ligand and receptor during the docking process. The second class of methods consists of those that explicitly evaluate side chain flexibility, for example, using rotamer libraries or optimization of side chain orientations during the docking process. Typically, these approaches assume the receptor to have a rigid backbone structure and a limited number of mobile side chains. Third, multiple receptor structures can be combined into a single receptor interaction grid, often using some kind of averaging procedure. The structures used can either be experimentally determined or computationally generated, for example, by molecular dynamics simulations. This approach allows for bigger conformational changes to occur then the aforementioned methods and is typically computationally cheaper when compared to docking into all receptor structures individually. However, the choice of the weighting scheme used for building the receptor grid is shown to have a large effect on the outcome of these approaches, and has therefore proven cumbersome. Finally, there are the ensemble-based docking approaches. As above, multiple receptor structures are used to describe protein flexibility. They are, however, not combined into a single interaction grid, but explicitly evaluated during the docking experiments, either sequentially or in parallel. The approaches that make use of multiple receptor structures do not only allow

* To whom correspondence should be addressed. Dr. Sander B. Nabuurs, CMBI 260, NCMLS, Radboud University Nijmegen, PO Box 9101, 6500 HB Nijmegen, The Netherlands. Telephone: +31-(0)412-668315. Fax: +31-(0)412-662539. E-mail: sander@nabuurs.org.

† Radboud University.

‡ N.V. Organon.

[a] Abbreviations: IFD, induced fit docking; MD, molecular dynamics; PDB, Protein Data Bank; PLP, piecewise linear potential; RMSD, root mean square deviations.

10.1021/jm070593p CCC: $37.00 © 2007 American Chemical Society
Published on Web 11/22/2007

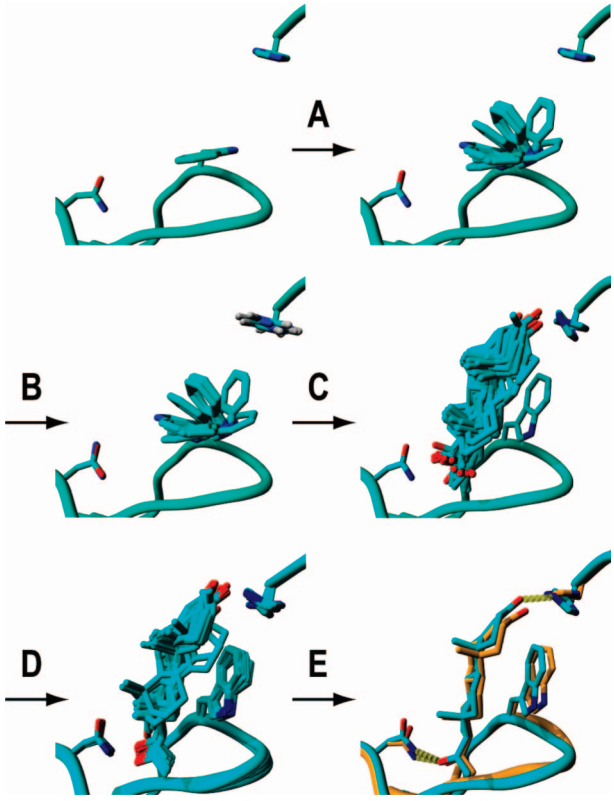


Figure 1. Visual outline of the Fleksy approach, illustrated using the induced fit docking of progesterone into the apo structure of the progesterone antibody DB3.[15] Starting with the apo form of a refined input structure, (A) potentially mobile side-chains are identified using predefined selection rules, resulting in the selection of Trp^{H100}. The conformations accessible to the selected side chains are sampled to construct a "flexible" ensemble of receptor structures. (B) Additional ambivalent interactors, like the shown asparagine and histidine residues, are also sampled. This set of structures is used as input for (C) an ensemble soft-docking experiment from which the 20 highest-ranked poses are selected and (D) submitted to a refinement procedure. Following the refinement stage, the poses are rescored using a consensus scoring function, which yields (E) the final induced fit docking receptor–ligand complex. In this example, the root-mean-square deviation (RMSD) between the modeled and cocrystallized ligand, shown in blue and orange, respectively, is 0.6 Å.

for side chain motions to occur, but can typically also take modest amounts of backbone motion into account.

The Fleksy method described here aims at combining advantages of each of the above-mentioned approaches together with several new features into a flexible and efficient pipeline for induced fit docking. In short, our approach consists of a soft-docking stage in which the ligand of interest is docked into a structural ensemble of receptor conformations, followed by a complex minimization stage for the highest ranked poses, during which both the ligand and the receptor binding site are free to move. Fleksy results in a set of minimized receptor–ligand complexes that are ranked according to a consensus scoring function based on docking scores as well as molecular dynamics force field interaction energies. The results obtained for a variety of different molecular complexes clearly demonstrate the usefulness of our method and its advantages over classical docking approaches.

Outline of the Approach. Our approach consists of several different stages, which are outlined and discussed in sequential order below. For clarity and reference, Figure 1 shows the results obtained at each of the different stages for the induced fit docking of progesterone into the apo crystal structure of the antiprogesterone antibody DB3[15] (Protein Data Bank (PDB) entry 1DBA), hereafter referred to as the DB3 receptor.

Receptor Preparation. The initial stage, during which the flexible structure ensemble is generated, starts with a receptor preparation step. Hydrogen atoms are added to the initial receptor structure, which is subsequently refined and minimized in explicit solvent (for experimental details and parameters, see the Materials and Methods). In case the receptor was cocrystallized with a ligand, the ligand is retained throughout the preparation and minimization steps. The refined receptor, with all waters and ligands removed, as shown in the first panel of Figure 1, is subsequently used as the starting point for construction of the flexible ensemble.

Identifying Potentially Flexible Residues. At this point, the possibility exists to define a set of residues for which the conformational space will be sampled in the ensemble. For the selection of these "flexible" residues, several options are available. First of all, the amino acids can be selected manually based on pre-existing knowledge of the dynamic behavior of certain residues, such as the well-known "gatekeeper" residues in kinases.[16] Additionally, we incorporated an automated rule-based procedure to identify residues most likely to change conformation. Our selection rules are based on those derived and applied in previous studies[17,18] (see Materials and Methods). Application of these rules to the DB3 receptor structure results in the identification of one possibly mobile side chain in the binding pocket: tryptophan 100 (Trp^{H100}).

Rotamer Sampling. For the amino acids labeled as potentially mobile, the accessible conformational space within the apo receptor structure is sampled using a backbone-dependent rotamer library. During the generation of these new rotameric states, clashes are allowed, with the intent to maximize sampling. As a result of the high clash tolerance the newly generated structures are likely to contain a significant number of interatomic bumps. To relieve these, the altered rotamers are subjected to an energy minimization, while the positions of the remainder of the amino acids are retained. This procedure results in an ensemble of binding sites in which the selected amino-acid side chains are realistically sampled within the boundaries imposed by the remainder of the binding site, as shown for the DB3 receptor in Figure 1A.

Interaction Sampling. Besides the flexibility introduced by the rotamer sampling, an additional form of receptor flexibility is considered in our approach, which pertains to the orientation and protonation of the asparagine (Asn), glutamine (Gln), and histidine (His) side chains and the orientation of the hydroxyl hydrogen in the serine (Ser), threonine (Thr), and tyrosine (Tyr) side chains. These particular side chains are of interest as experimental difficulties associated with protein-structure determination using X-ray crystallography can lead to ambiguities in interpreting the polar interactions involving these amino acids. Several studies have shown that in macromolecular structure models, typically around 20% of the His, Asn, and Gln side chains, require a 180° flip to optimize the hydrogen bonding network.[19,20] As these side chain groups can, depending on their rotameric state, act either as a hydrogen bond donor or acceptor for incoming ligands, the choice of rotamer can to a large extent influence the docking results obtained at a later stage.

Docking procedures typically do not allow for required side chains flips to occur during the docking process. Attention of the modeler is required to rotate the appropriate torsion angle to maximize hydrogen bonding beforehand. It should, however, be noted that flipping a particular side chain might be optimal for the ligand cocrystallized in the target receptor structure, but

that same flip could well be less suited for a new ligand to be docked. To prevent such issues from arising, we simultaneously consider all flipped possibilities of the aforementioned side chains during the docking stage of our approach. This "interaction sampling" is realized through addition of the 180° flipped states of all Asn, His, and Gln amino acids to the flexible ensemble of structures considered during the docking stage. For the histidine residues, all different protonation and tautomeric states are automatically generated and added to the structure ensemble, but the evaluated possibilities can for instance also be limited to those relevant at a given pH value. The hydroxyl groups in the Ser, Thr, and Tyr amino acids are considered in the ensemble in two different states: the optimized one resulting from the receptor preparation stage and the one rotated by 180° around the relevant torsion angle.

The final structural ensemble generated for the DB3 receptor, including the results from the rotamer sampling of Trp^{H100} and the interaction sampling of Asn^{H35} and His^{L27D}, is shown in Figure 1B.

Ensemble Soft-Docking. Following construction of the ensemble of protein receptors, the next challenge is to correctly position the ligand of interest into the active site. For this we make use of the ensemble docking program FlexX-Ensemble (formerly known as FlexE).[21] Based on the different conformations of the protein receptor present in the ensemble, the program generates a so-called united protein description of the target. This is done by superposing the different members of the ensemble and constructing a rigid average structure from the most conserved structural features. For the variable regions, as introduced in the different stages described above, the structurally different conformations are explicitly evaluated and combinatorially explored during the docking process. One of the great advantages of this approach is that it can result in novel binding site geometries that consist of combined side chain orientations from different members of the input ensemble. In this way, it is for instance possible to combine a sampled side chain rotamer from one member of the ensemble together with a flipped Asn rotamer from another member into a new binding site geometry, optimally suited to accommodate the ligand under investigation.

Simultaneously, with the flexibility described in the structural ensemble, we allow for additional conformational changes in the active site by tolerating a significant overlap between ligand and protein atoms during the docking process, an approach in literature typically referred to as "soft-docking".[22] In this way, it is possible to retain certain promising ligand orientations that would otherwise be discarded too early in the processing pipeline.

As FlexX-Ensemble is an extension of the well-known and widely used FlexX program,[2] it makes use of the same underlying docking strategy and scoring functions. In our ensemble docking experiment, a large set of docking poses, together with their corresponding binding sites, is generated using the default FlexX/FlexX-Ensemble scoring function. To further improve docking accuracy, we combine the FlexX scoring function with the piecewise linear potential (PLP) scoring function[23] to rerank the generated poses. Both FlexX and PLP were recently shown to be among the best scoring functions in identifying near-native docking poses.[24] Therefore, the two are combined into a consensus scoring function[25] in which both have an equal contribution. The highest ranked poses for the DB3 receptor, together with their corresponding binding site geometries, are shown in Figure 1C.

Complex Refinement. The receptor–ligand complexes generated during the ensemble soft-docking stage are likely to contain a significant number of interatomic clashes. As such, they do not represent meaningful and physically realistic structures. To relieve these clashes and to further refine the geometry and orientation of both the docked ligand as well as the binding site residues, the highest ranked complexes from the docking experiment are subsequently refined using the Yasara program (http://www.yasara.org). The refinement consists of a short steepest descent minimization to remove the largest intermolecular and intramolecular clashes, followed a simulated annealing minimization until the force field energy converges. For the sake of speed, this minimization is by default performed in vacuo, but when preferred, it can optionally also be run in explicit solvent fully automatically.

Consensus Rescoring and Selection. Some of the receptor–ligand complexes generated by the docking experiment will probably benefit more from the refinement stage than others. Therefore, the refined complexes are rescored and reranked. To this end we again apply a consensus scoring function with contributions from the FlexX and PLP scoring functions, but this time enriched with receptor–ligand binding energies calculated from the optimized complex structures. This composite scoring function (for details the reader is referred to Materials and Methods) is used for the final ranking of the generated complex structures. The complex ranked highest in this rescoring step is typically selected as the final induced fit docking structure. Figure 1E show the highest ranked pose obtained for the DB3 receptor. The RMSD of the predicted orientation of progesterone to the orientation observed in the crystal structure (PDB entry 1DBB) is 0.6 Å.

Test Cases. To further assess the effectiveness of our Fleksy approach, we evaluated its performance under different circumstances likely to occur in a drug design environment. We demonstrate the ability of our algorithm to accurately dock similar compounds, for example, from within a chemical series or derivatives from a shared structural scaffold, into several receptor structures that were cocrystallized with one such compound. This is a situation that typically occurs within a lead optimization project in which many similar compounds are synthesized, but crystal structures are only determined for few. First, we compare the performance of our approach to the default FlexX algorithm for a set of 10 receptor–ligand complexes of the mineralocorticoid receptor (MR). Each of the cocrystallized ligands is docked into the receptor in which it was crystallized (self-docking) and into the nine other MR structures in the dataset (cross-docking). Second, we perform a similar set of docking experiments using five protein kinase B-selective (PKB) inhibitors that were crystallized in complex with protein kinase A (PKA). The studied compounds all originate from the same chemical series and differ only in their substitutions at a single position. These substitutions are, however, sufficient to change the conformation of the ATP pocket and induce movements of the flexible glycine-rich loop,[26] such that they are likely to pose a challenge to docking algorithms not allowing receptor flexibility. Third, we show the ability of our methodology to dock ligands that require significant induced fit when docked into a receptor cocrystallized with a structurally different ligand, a situation likely to occur during the lead finding stages of a drug design project. To this end, our approach is tested using a large set of cross-docking experiments on 10 different pharmaceutically

relevant receptor classes, all of which were used as induced fit docking test cases in a recent study.[18]

Materials and Methods

Structure Preparation. Three-dimensional coordinates of the receptors were obtained from the PDB[27] and, when required, structurally aligned using the MOTIF program,[28] as embedded in YASARA (http://www.yasara.org). Prior to the optimization, the internal hydrogen bonding network of the receptor was optimized using the hydrogen bonding network optimization algorithm implemented in WHAT IF.[19] To optimize hydrogen positions, remove interatomic bumps, and correct the covalent geometry, all structures were energy-minimized in explicit solvent with the Yamber2 force field.[29] During this minimization, all atoms were allowed to move. A 7.86 Å force cutoff and the particle mesh Ewald algorithm[30] were used to treat long-range electrostatic interactions. After removal of conformational stress by a short steepest descent minimization, the procedure continued by simulated annealing (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence was reached, that is, no energy improvement was found for 200 steps.

Ligand Preparation. Three-dimensional ligand input structures for all docking experiments were generated using the CORINA program.[31] The generated structures were saved as MOL2 files.

Selection of Flexible Residues. In this work, the selection of flexible residues for the induced fit dataset is based on a set of rules that were previously defined.[17,18] Residues whose rotameric states are sampled were selected using the following criteria: (1) All binding site residues whose side chain atoms deviate more than 2.5 Å from the nearest atom in the corresponding amino acid in another crystal structure of the same receptor. (2) All binding site residues with multiple occupancies or missing density. (3) If the unit cell contains multiple independently refined copies of the receptor, all binding site residues whose side chain atoms deviate more than 1.5 Å from the nearest atom in the corresponding amino acid in another copy of the receptor in the unit are selected. (4) All binding site residues containing atoms with a B-factor greater than 40 Å^2. The residues are ranked by the highest B-factor occurring within each side chain.

The residues that were selected as flexible for the induced fit docking dataset are shown in Table 1 of the Supporting Information.

Rotamer Sampling. The conformational space of the selected amino acid side chains is simultaneously sampled using internal coordinates by the YASARA program (version 6.12.1). A nonredundant subset of the PDB (90% sequence identity cutoff, resolution better than 2.5 Å) is searched for stretches of five scaffold residues that have a similar sequence and backbone conformation compared to those in the starting structure. Conformational similarity is judged by calculating the root-mean-square deviation (RMSD) over all those dihedral angles that have not been selected for sampling and remain frozen. Sequence similarity is measured using the BLOSUM62 matrix.[32] The two similarity scores are each mapped to the interval [0...1], where 0 corresponds to the minimal score and 1 to the maximal score, which the local template sequence can theoretically achieve in any gapless alignment. The final score is then the product of structural and sequence similarity, that is, they are weighted equally. This procedure results in top-scoring hits that all share high structural and sequence similarity.

The maximum allowed sum of interatomic bumps is set to 8.0 Å; this is to allow for partial overlap between the simultaneously generated rotamers and the input structure to maximize sampling. By default, a total of 15 side chain rotamers are generated for each of the amino acids selected as flexible. This procedure results in 15 different binding sites in which all flexible amino acids are sampled. Finally, to alleviate inter-residual clashes, the generated binding sites are subjected to an energy minimization, during which the nonflexible regions of the protein are kept fixed.

Interaction Sampling. For all nonflexible side chains, the orientation in the refined input structure is considered as the initial state. Subsequently, for the Asn, Gln, and His side chains, the relevant φ torsion angle is flipped by 180° and the resulting structure is added to the ensemble. Additionally, for the His side chains, all four different ionic and tautomeric states are generated for each of the two flipped states and described in the structure ensemble. For the Ser, Thr, and Tyr side chains, the orientation of the hydroxyl group is rotated 180° around the relevant optimized torsion angle, as obtained after the receptor preparation stage and also added to the structure ensemble.

FlexX Docking. The program FlexX[2] version 1.13.5, as implemented in the 7.1 release of the SYBYL package, is used as the reference docking program to which we compare our approach. Standard parameters of FlexX are used for iterative growing and subsequent scoring of docking poses. For all docking experiments, the active-site atoms of a receptor are defined as those atoms within a radius of 8.0 Å from the ligand cocrystallized with that particular receptor. Receptor description files used by FlexX were automatically generated from the receptor PDB coordinates.

Ensemble Soft-Docking. The program FlexX-Ensemble,[21] as embedded in version 1.13.5 of the program FlexX,[2] is used for the ensemble soft-docking stage. The binding sites generated during the rotamer and interaction sampling stages are used as input for FlexX-Ensemble.

To enable soft-docking, two of the default FlexX-Ensemble parameters are adjusted. First, the maximum allowed overlap between receptor and ligand is raised from 2.5 Å^3, which is the default value, to 6.0 Å^3. This allows the ligand to partially protrude into the receptor to maximize its interactions. Furthermore, in FlexX-Ensemble, hydrogen atoms attached to carbon atoms are not explicitly evaluated in the overlap tests conducted during a docking experiment. Instead, a united atom radii model is used in which the van der Waals radius of a carbon atom is incremented by a given number. By default, this number is set to 0.1 Å, which is lowered to 0.05 Å in our approach to allow for additional overlap to occur.

A drawback of the FlexX-Ensemble program is that it only generates coordinates for the docked poses and the amino acids surrounding it and does not result in a final protein structure for each generated docking pose. As a complete protein structure is required for our complex refinement stage, the output generated by FlexX-Ensemble is automatically read in by our program and used to reconstruct the complete pose-specific receptor structure from the input ensemble. These regenerated pose specific receptor coordinates are subsequently used together with the generated docking poses in the complex refinement stage.

In the docking experiment, 100 docking solutions are generated using the FlexX scoring function. The solutions are subsequently also scored using the PLP scoring function, as implemented in FlexX. A consensus scoring function combining these two score values with equal weight is subsequently used to select the 20 most highly ranked docking poses and their corresponding binding site geometries for the complex minimization stage.

Complex Refinement. The complex optimization consists of an in vacuo energy minimization with the YASARA program using the Yamber2 force field.[29] After removal of conformational stress by a short steepest descent minimization, the procedure continues by simulated annealing (time step 2 fs, atom velocities scaled down by 0.9 every 10th step) until convergence is reached, that is, no energy improvement is found for 200 steps. During the simulated annealing stage, all amino acids that have no atoms within 5 Å of the initial orientation of the ligand are kept fixed.

Consensus Scoring. To rescore the refined poses, we propose a consensus scoring function combining the FlexX, PLP scoring functions with a molecular dynamics force field interaction energy calculated from the minimized structural complex. The receptor–ligand interaction energy is defined as

$$E_{\text{interact}} = (E_{\text{int complex}} - (E_{\text{int receptor}} + E_{\text{int ligand}})) + (E_{\text{sol complex}} - (E_{\text{sol receptor}} + E_{\text{sol ligand}})) \quad (1)$$

where E_{int} is the internal force field energy and E_{sol} is a first-order approximation of the solvation energy as calculated by the YASARA program using the Yamber2 force field.

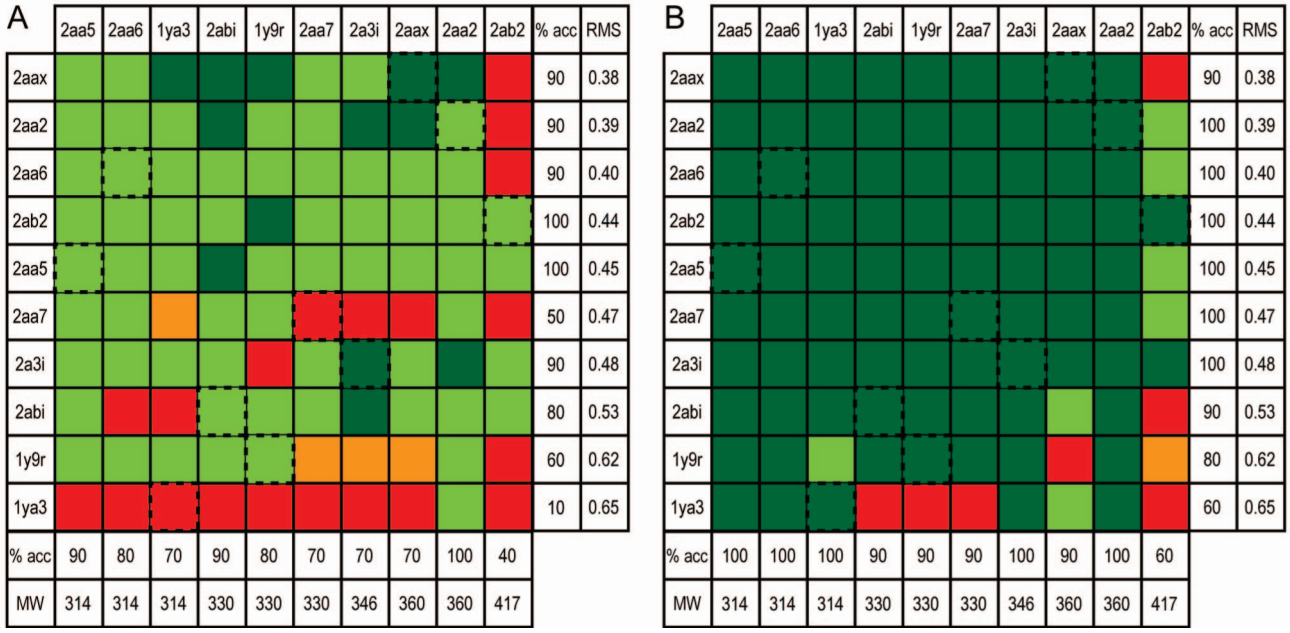


Figure 2. Results of 200 self-docking and cross-docking experiments performed using 10 mineralocorticoid receptor structures and their respective cocrystallized ligands for (A) FlexX and (B) Fleksy. The receptor structures are shown vertically and are ranked according to their average pairwise all-heavy-atom binding site RMSD to all other structures in the dataset, as shown in the RMS columns. The binding site was defined as all residues within 4.5 Å of at least one of the crystallized ligands in the dataset. The ligands are shown horizontally and are ranked according to their molecular weight, as shown in the MW rows. Both receptors and ligands are labeled with the PDB code of the complex in which they were crystallized. The accuracy of the highest ranked pose for each experiment is indicated using color. An excellent pose (RMSD ≤ 1.0 Å), dark green; a good pose (RMSD ≤ 2.0 Å), light green; a poor pose (RMSD ≤ 3.0 Å), orange; and a bad pose (RMSD ≥ 3.0 Å), red. For all receptors and ligands, the overall performance (defined as the percentage of highest ranked poses, with a RMSD ≤ 2.0 Å) is indicated in the "% acc" columns and rows, respectively. Self-docking experiments are indicated using dashed lines.

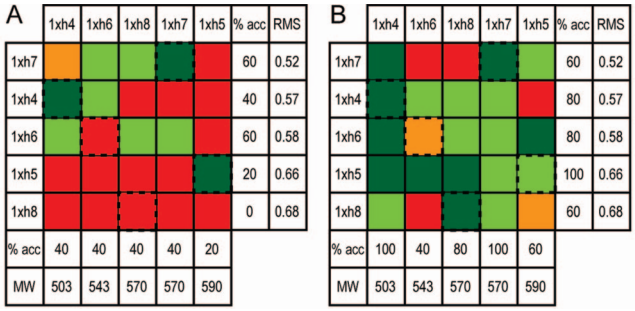


Figure 3. The results of 50 self-docking and cross-docking experiments performed using five protein kinase A receptor structures and their respective cocrystallized ligands for (A) FlexX and (B) Fleksy. Labeling, ranking, and color schemes as in Figure 2.

The different scoring terms are combined in the consensus function using the so-called scaling method,[33] where the score of each model was scaled to a number between 0.0 and 1.0 for each of the different scoring functions applied. Subsequently, the three scores are combined in a consensus score; this time, however, not with equal weights. The consensus score is defined as

$$S_{\text{consensus}} = (w_{\text{PLP}} \times S_{\text{PLP}} + w_{\text{FlexX}} \times S_{\text{FlexX}} + w_{\text{interact}} \times S_{\text{interact}})/(w_{\text{PLP}} + w_{\text{FlexX}} + w_{\text{interact}}) \quad (2)$$

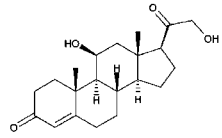
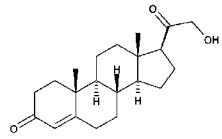
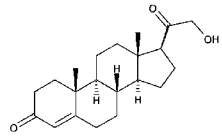
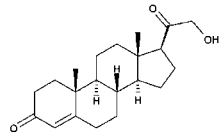
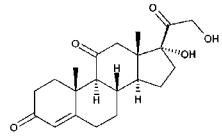
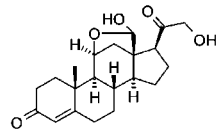
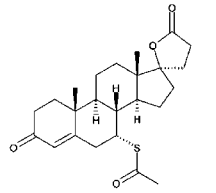
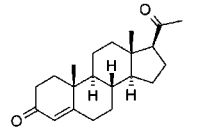
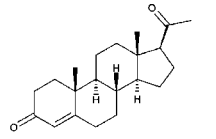
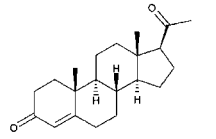
where $S_{\text{consensus}}$ is the consensus score, S_{PLP} is the normalized PLP score, S_{FlexX} is the normalized FlexX score, and S_{interact} is the normalized interaction energy. Through an empirical Monte Carlo optimization on a benchmark set of self-docking experiments,[34] we arrived at the following weights: $w_{\text{PLP}} = 5.0$, $w_{\text{FlexX}} = 1.0$, and $w_{\text{interact}} = 0.5$.

Computing Time. The average computation time for a complete induced fit docking calculation using our approach, resulting in 20 ranked induced fit complexes, is typically around 1 h on a single 3.4 Ghz Intel Pentium IV processor. The computationally most expensive step in our protocol is the complex optimization stage, which can take up to a few minutes per docking pose. However, these calculations are independent and, as such, they are readily parallelized using multiple processors.

Results

Mineralocorticoid Receptor Dataset. The 10 receptor–ligand complexes in the mineralocorticoid dataset consist of similar MR receptor structures crystallized with moderately different ligands, all of which are based on a steroidal scaffold.[35–37] The chemical structure of the ligands, the mutations present in the different receptor structures, and the PDB codes of the original complexes are shown in Table 1. All structures were first prepared and minimized, as described above, and subsequently used in the docking experiments. The results obtained for the self-docking and the cross-docking experiments using both FlexX and Fleksy are shown in Figure 2. While FlexX is already very effective at docking the different ligands into different MR structures, it is clearly outperformed by Fleksy for this dataset. Using FlexX, 13% of the highest ranked docking poses of the in total 100 docking experiments have a RMSD to their corresponding crystal structure below 1.0 Å; an additional 63% have a RMSD below 2.0 Å. Thus, FlexX results in 76% of the highest ranked poses having a RMSD below 2.0 Å. Fleksy, however, performs markedly better: 84% of the highest ranked poses have a RMSD below 1.0 Å, and an additional 8% have a RMSD below 2.0 Å. This results in an overall percentage of 92% of the highest ranked poses having a RMSD below 2.0 Å to their corresponding crystal structure. In only 2% of the cases does Fleksy result in a poorer binding mode than the one obtained using FlexX: in both the R_{2ABI}−L_{2AB2} and

Table 1. Chemical Structures and Names of the Mineralocorticoid Receptor–Ligands in the MR Dataset

receptor PDB entry (resolution)	ligand structure	ligand name	receptor mutations
2a3i (2.0 Å)		corticosterone	C808S
2aa7 (2.2 Å)		deoxycorti-costerone	C808S
2abi (2.3 Å)		deoxycorti-costerone	C910A
1y9r (2.0 Å)		deoxycorti-costerone	S810L C910A
2aax (1.8 Å)		cortisone	C808S S810L
2aa2 (2.0 Å)		aldosterone	C808S
2ab2 (1.9 Å)		spironolactone	C808S S810L
1ya3 (2.3 Å)		progesterone	S810L C910A
2aa5 (2.2 Å)		progesterone	C808S
2aa6 (2.0 Å)		progesterone	C808S S810L

$R_{1Y9R}-L_{2AAX}$ cross-docking experiments the ligand is incorrectly positioned in the binding site.

The different rows and columns in Figure 2 clearly show that some ligands and receptors are less permissible to our docking

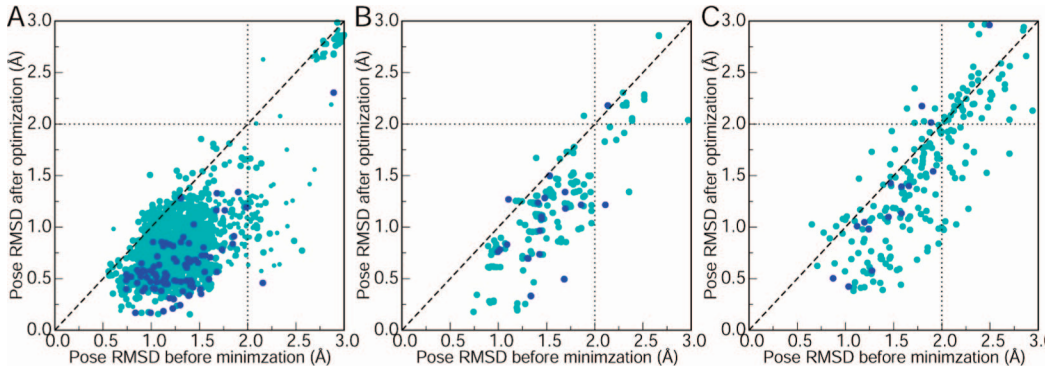
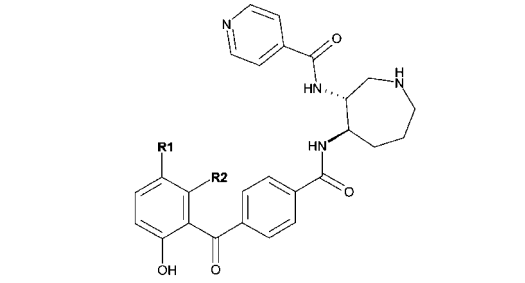


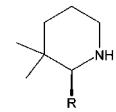
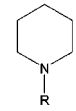
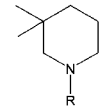
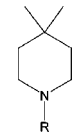
Figure 4. Improvement in pose accuracy during the complex minimization stage, shown are all obtained poses with an RMSD ≤ 3.0Å after the FlexX-Ensemble soft-docking stage for (A) 100 docking experiments of the mineralocorticoid dataset (1689 poses shown in total), (B) the 25 docking experiments of the protein kinase A dataset (388 poses), and (C) the 20 docking experiments of the diverse IF dataset consisting of 10 different receptors (230 poses). The highest ranked poses for each docking experiment after the final consensus scoring are indicated in dark blue, all others in light blue.

experiments than others. The overall trend that can be observed is that obtaining correct docking poses in a cross-docking experiment becomes more difficult as the target binding site becomes more structurally dissimilar compared to the other sites in the dataset (see Figure 2). Similarly, cross-docking larger ligands is typically more difficult than docking smaller ones. The largest and chemically most different ligand in the dataset is spironolactone, crystallized in PDB entry 2AB2, and as such it poses a challenge to both approaches. FlexX is able to accurately position this ligand into 4 out of 10 structures; Fleksy can do so for 6 out of 10 MR receptor structures. Positioning the 10 MR ligands in the 1YA3 and 2AA7 receptor structures poses a challenge for FlexX (the docking accuracies of these receptors are 10% and 50%, respectively). For the 1YA3 structure Fleksy manages to dock 6 out of 10 of the most highly ranked docking poses with a RMSD below 2.0 Å, and for the 2AA7 receptor, it performs even better: all of the highest ranked poses have a RMSD below 2.0 Å.

Protein Kinase A Dataset. The protein kinase A (PKA) dataset consists of five different azepane derivatives,[38] which were synthesized to address selectivity issues between protein kinase A and protein kinase B.[26] As such, these ligands provide a relevant test case originating from a real life lead optimization project. The chemical structures of the five inhibitors are shown in Table 2, together with the PDB identifier of their coordinates in complex with protein kinase A. The results of the 25 self-docking and cross-docking experiments, as performed using both FlexX and Fleksy are shown in Figure 3. The five inhibitors are ATP competitive and all bind to the ATP pocket, the cleft between the small N-lobe and the larger C-lobe of protein kinases. The overall orientation of the inhibitors in the adenine and ribose pocket is similar, but the benzophenone moieties with their substituents occupy markedly different positions, inducing also movements of the glycine-rich loop.[26] As a result of this, this dataset clearly poses a challenge for the FlexX program. In no more than 36% of the docking experiments does FlexX result in a pose with a RMSD below 2.0 Å at rank one. Our Fleksy approach manages to produce a RMSD for the highest ranked pose below 2.0 Å in 76% percent of the cases, which can be considered a significant improvement. We see a similar pronounced difference between the two approaches if we look at the occurrence of poses with a RMSD below 1.0 Å. When looking at the highest ranked poses, FlexX generates 12% with a RMSD below 1.0 Å; Fleksy reaches a level of 36% at this threshold. If all 20 poses resulting from each docking run are considered, we find for FlexX a modest increase to 16% of the experiments, with at least one pose generated with a RMSD below 1.0 Å; for Fleksy, however, 60% of the performed docking experiments yield at least one pose below 1.0 Å (data not shown). Also, for this dataset, it should be noted that there are two occurrences where FlexX performs better than Fleksy.

Table 2. Chemical Structures of the Protein Kinase A Inhibitors in the PKA Dataset



PDB entry (resolution)	R1	R2
1xh4 (2.5 Å)		H
1xh5 (2.1 Å)		F
1xh6 (1.9 Å)		H
1xh7 (2.5 Å)		H
1xh8 (1.6 Å)		H

Induced Fit (IF) Dataset. Finally, we test Fleksy using a set of 20 cross-docking experiments for 10 different pharmaceutically relevant receptors that were used in a recent study as induced fit docking test cases.[18] In these docking experiments, structurally very diverse ligands were cross-docked. Therefore, the option to consider a selected set of binding residues as flexible is applied here. The selected residues are kept identical

Table 3. Induced Fit Docking Results Obtained for 40 FlexX and Fleksy Cross-Docking Experiments on a Diverse Test Set of Different Pharmaceutically Relevant Drug Targets[a]

	FlexX		Fleksy	
cross-docking experiment	RMSD pose at rank 1 (Å)	rank 1st pose with RMSD ≤ 2.0 Å	RMSD pose at rank 1 (Å)	rank 1st with RMSD ≤ 2.0 Å
		CDK-2		
1aq1–1dm2	5.1	4 (1.9 Å)	**5.3**	**2 (1.2Å)**
1dm2–1aq1	2.3	-	**1.4**	**1**
		COX-2		
3pgh−1cx2	2.4	-	**1.0**	**1**
1cx2–3pgh	**1.2**	**1**	2.0	1
		ER		
1err−3ert	1.9	1	**1.1**	**1**
3ert−1err	1.5	1	**1.4**	**1**
		Factor Xa		
1ksn−1xka	8.2	-	**1.4**	**1**
1xka−1ksn	2.3	-	**2.2**	**2 (1.9Å)**
		HIV-RT		
1rth−1c1c	6.3	-	6.1	-
1c1c−1rth	7.0	-	5.4	-
		Neuramidase		
1nsc−1a4q	4.7	9 (1.8 Å)	**1.5**	**1**
1a4q−1nsc	7.2	10 (1.5 Å)	**0.5**	**1**
		PPARγ		
1fm9–2prg	8.0	-	**3.0**	**7 (1.8Å)**
2prg−1fm9	12	-	10	-
		Thermolysin		
1kr6–1kjo	1.6	1	**1.1**	**1**
1kjo−1kr6	7.7	-	7.7	-
		TK		
1kim−1ki4	6.9	-	**0.4**	**1**
1ki4–1kim	6.8	2 (1.2 Å)	**1.1**	**1**
		CDK-2		
1buh−1dm2	4.9	-	**1.0**	**1**
		Antibody DB3		
1dba−1dbb	9.3	-	**0.6**	**1**

[a] The cross-docking experiments are labeled using two PDB codes; the first refers to the entry from which the receptor was taken and the second refers to the entry from which the ligand is obtained. For each cross-docking experiment, the RMSD of the highest ranked pose and, if obtained, the rank of the first pose with an RMSD ≤ 2.0 Å are shown. The results of the best performing method are indicated in bold (if at least one pose with an RMSD below 2.0 Å is obtained).

to those applied in the previous study and are show in Supporting Information, Table 1. The docking results obtained for these receptors are shown in Table 3, again for both FlexX and Fleksy. For this set of cross-docking experiments, FlexX produces a pose with a RMSD below 2.0 Å at rank 1 in 4 out of 20 experiments. Fleksy results in 13 poses at the highest rank, with a RMSD below 2.0 Å. In two cases, however, we find a pose with a RMSD below 2.0 Å ranked second best, and in one additional case the first pose below 2.0 Å is found at the seventh position in the ranked list of poses. When all generated docking poses are taken into account, Fleksy manages to produce a docking pose with a RMSD below 2.0 Å in 80% of the cross-docking experiments, for FlexX this number is 40%.

In four cross-docking experiments belonging to three different receptors, HIV-RT, PPARγ, and thermolysin, Fleksy fails to produce a docking pose with a RMSD below 2.0 Å. Apparently, these are truly challenging docking problems to solve because, also in the previously mentioned study,[18] three cross-docking experiments related to the same three receptors failed to produce docking poses with a RMSD below 2.0 Å. For each of these complexes, it can be understood why both the FlexX and FlexX-Ensemble programs fail to arrive at an acceptable solution. In the case of the HIV-RT receptor, the ligand in its correct orientation does not make any hydrogen bonds to the receptor.

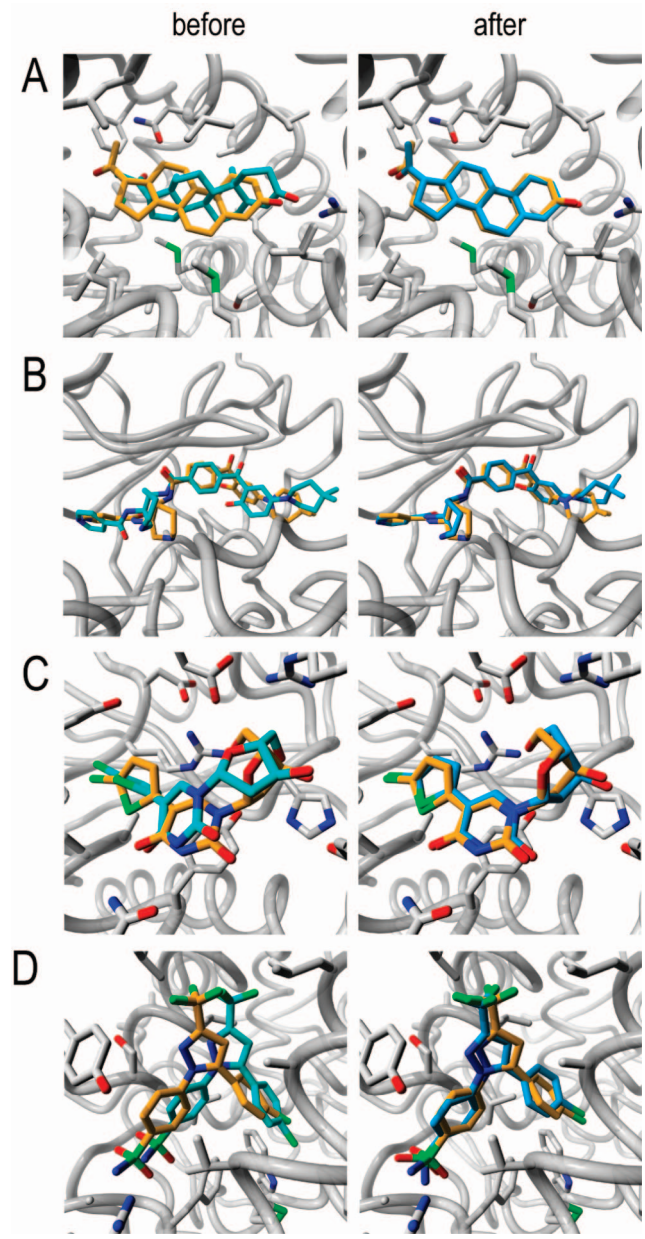


Figure 5. Examples of the effect of complex minimization stage on pose accuracy for poses taken from the (A) $R_{1ya3}-L_{2aa5}$ cross-docking experiment from the MR dataset (RMS improvement = 1.7 Å), (B) the $R_{1xh6}-L_{1xh8}$ cross-docking experiment from the PKA dataset (RMS improvement = 0.9 Å) and the cross-dockings (C) $R_{1kim}-L_{1ki4}$ (RMS improvement = 1.1 Å), and (D) $R_{3pgh}-L_{1cx2}$ (RMS improvement = 1.1 Å) from the IF dataset. The poses prior to the minimization are shown in the left panels (light blue) and the orientations after minimization are shown in those on the right (dark blue). The orientation of the ligand as observed in the corresponding crystal structure is shown in orange. The receptor crystal structure is shown in gray. For clarity, side chains surrounding the ligand have been omitted from (B).

As such, the receptor does not provide any anchoring points for the FlexX/FlexX-Ensemble base placement algorithm, which therefore never arrives at the correct solution. In the PPARγ cross-docking experiment ($R_{2prg}-L_{1fm9}$), the ligand to be docked (GI262570) is a relatively large and extended compound with 12 rotatable bonds. It has previously been suggested that compounds with so many rotatable bonds are indeed difficult to handle using the FlexX/FlexX-Ensemble approach.[39] The cross-docking experiment in thermolysin ($R_{1kjo}-L_{1kr6}$) fails due to the presence of an arginine side chain in the binding pocket

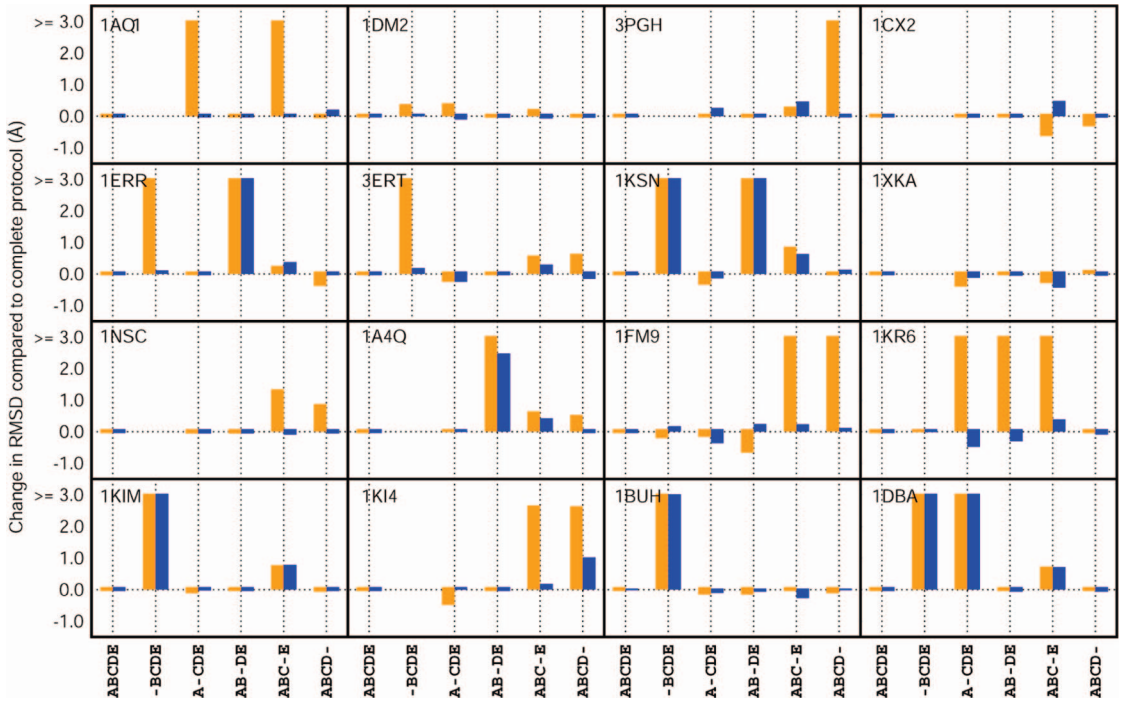


Figure 6. Results of the one-off sensitivity analysis in which systematically each of the five steps in the Fleksy procedure was omitted from the docking pipeline presented in Figure 1. Shown on the vertical axes are the changes in RMSD compared to those obtained in the cross-docking experiments using the complete Fleksy procedure as reported in Table 3. Along the horizontal axes the different steps included in the protocol are shown, where A represents the rotamer selection and sampling stage, B represents the interaction sampling stage, C represents the use of a softened van der Waals potential during docking, D represents the complex optimization stage, and E represents the use of consensus scoring with multiple scoring functions throughout the protocol. Each panel is labeled with the PDB entry code of the structure that provides the receptor coordinates for the cross-docking experiments, as shown in Table 3. The orange and blue bars indicate the results obtained for the highest ranked pose and the best pose generated, respectively.

where the ligands need to bind. The ligand, benzyloxycarbonyl-D-glutamic-*Z*-D-glutamic acid, contains two carboxylic acid moieties. Neither interact with the nearby arginine side chain in the crystallized complex with thermolysin, however, all obtained docking poses have a carboxylic acid moiety positioned close to this arginine side chain. Despite the occurrence of three problematic datasets, Fleksy clearly manages to realize an improvement in docking performance on the IF dataset when compared to a regular docking approach.

Contribution of the Complex Optimization Stage. The complex optimization stage following the soft-docking experiment is a crucial step to reach an accurate prediction of the orientation of the ligand under investigation. Figure 4 shows the RMSDs preceding and following the refinement stage for the three complete datasets described above. Shown are all obtained docking poses with a RMSD value after the soft docking stage smaller than or equal to 3.0 Å. From Figure 4, it is evident that for the majority of the generated docking poses the minimization stage results in a clear improvement in pose accuracy. For the MR dataset, the PKA dataset, and the IF dataset, the percentage of poses for which the accuracy improves is 95, 91, and 77%, respectively. Interestingly, the amount of pose improvement appears to be, within the evaluated window of 0.0 to 3.0 Å, relatively independent of its initial accuracy. Even poses that, before minimization, already have an acceptable orientation (e.g., RMSD $\leq$ 1.5 Å), typically show significant improvement as a result of the complex optimization. The average improvement for poses with an RMSD $\leq$ 1.5 Å before refinement is 0.37 $\pm$ 0.29 Å, 0.41 $\pm$ 0.25 Å, and 0.31 $\pm$ 0.36 Å for the MR, PKA, and IF datasets, respectively. Figure 5 shows several examples of the effect of the complex minimization on both initially reasonably accurate and initially less accurate poses from the three different datasets. From these results, it is clear that the effect of the pose minimization stage is not limited to slight conformational changes and minor structural rearrangements, but can also result in some quite dramatic translations and rotations of the complete ligand. It should be kept in mind that the initial orientations result from the soft-docking stage during which quite severe van der Waals clashes are allowed, and as such, the generated poses are often required to move to relieve these clashes.

Assessing the Contribution of the Different Stages. The Fleksy approach is a multistep protocol aimed at accurately docking small molecules into a protein receptor. To assess the contribution of each of the different stages of the protocol, we performed a one-off sensitivity analysis in which each of the five individual stages, as indicated in Figure 1A−E, was in turn removed from the protocol. The analysis was performed on the 16 cross-docking experiments from the induced fit dataset for which at least one pose with an RMSD of 2.0 Å or better was generated (see Table 3). For all datasets, the Fleksy workflow was executed an additional five times, each time omitting one of the five different steps in the procedure. The change in RMSD values of both the highest ranked pose and of the most accurate pose generated are shown in Figure 6. The obtained results show that for the 16 different evaluated cross-docking experiments here, different stages of our protocol are found to be crucial, both for generating accurate poses as well as correctly ranking the generated poses.

For example, in nine cross-docking experiments, residues have been selected as being potentially flexible and, as such, they are sampled in step A of the full protocol. Omission of this step in these nine docking experiments leads to several different results. In four cases (R_{1ksn}, R_{1kim}, R_{1buh}, and R_{1dba}),

no pose with a RMSD below 2.0 Å is generated anymore, in two cases (R_{1err} and R_{3ert}), good poses are generated but they are no longer the highest ranked poses, and in the final three docking experiments (R_{1dm2}, R_{1fm9}, and R_{1kr6}), only very modest changes compared to the complete protocol are observed. This illustrates that considering receptor flexibility is not only useful in generating accurate poses, but can also contribute to achieving an accurate ranking of the generated poses.

The interaction sampling stage (step B) is applied in all cross-docking experiments. Omitting this stage results in two cases (R_{1aq1} and R_{1kr6}) in an incorrect ranking of the generated poses and in one case (R_{1dba}) in a complete absence of poses with a RMSD below 2.0 Å. The latter is readily understood from Figure 1. Progesterone in its final orientation makes two hydrogen bonds to both a glutamine and a histidine. For the histidine to be able to make this hydrogen bond with progesterone, it needs to be flipped 180° with respect to its orientation in the refined receptor structure used as input. With the interaction sampling disabled, the histidine will remain in an orientation unsuitable for ligand binding and, as such, no accurate poses are obtained.

Disabling the softened van der Waals potential results in three cases (R_{1err}, R_{1ksn}, and R_{1a4q}) where no accurate poses are generated anymore. In one experiment (R_{1kr6}), at least one correct pose is generated, but it is no longer identified as the highest scoring one. This clearly shows that the soft-docking stage of the protocol is indeed important, but not solely responsible for the performance of the approach presented here. Removing the complex optimization stage and consensus scoring stages from the protocol does not in any of the test cases result in an absence of poses with a RMSD below 2.0 Å. However, in a number of cases, it does result in a less accurate ranking of the generated poses, resulting in the highest ranking pose often having a much poorer RMSD when compared to the full protocol with all steps included.

Overall, the presented sensitivity analysis clearly illustrates there is not one particular single stage of our protocol that governs the performance of the Fleksy pipeline, but that all stages combined are indeed required to reach an optimum overall performance over the wide variety of different cross-docking experiments benchmarked here.

Discussion and Future Work. A well-known drawback of docking programs that do not allow for receptor flexibility is their sensitivity to relatively small changes in the target structure. Figure 2A clearly shows that with an increasingly different rigid receptor structure, docking a ligand correctly also becomes increasingly different. For example, for the MR dataset we obtain for FlexX a Pearson's correlation coefficient of −0.78 between overall receptor docking accuracy and the average pairwise RMSD of the receptor to all others in the cross-docking dataset, which clearly illustrates this effect (data not shown). Using Fleksy, we observe that the receptor that is most different from all others is also the most difficult one to dock, but overall, the amount of structural dissimilarity that can be tolerated in a cross-docking experiment is much higher when compared to the default FlexX protocol.

The aforementioned structural differences between different models of the same receptor, either uncomplexed or complexed with a ligand, can originate from different sources. One obvious source is bound ligands that can induce modestly to markedly different binding site conformations depending on their interactions with the receptor binding site. However, another source of binding site dissimilarity between different models can be the structural uncertainty or heterogeneity of the crystal structures deposited in the PDB.[40] The latter probably also plays a role in the MR dataset for the R_{2aa7}, R_{2abi}, and the R_{1y9r} structures, all bound to deoxycorticosterone, and the R_{1ya3}, R_{2aa5}, and the R_{2aa6} structures, all bound to progesterone. From the results we obtained, it is clear that allowing for a limited amount of protein plasticity can be very beneficiary in such cases, as we see a much improved overall accuracy when compared to the reference rigid receptor docking algorithm.

Over the last few years, several approaches aimed at taking protein flexibility into account in small molecule docking have been presented. The Fleksy method introduced here aims at combining the advantages of existing techniques together with several new features into an efficient pipeline aimed at induced fit docking. One of the earliest attempts at considering protein flexibility in small molecule docking has been the use of decreased van der Waals radii.[22] Despite its simplicity, over time and during this work, this has proven to be both a powerful and affective approach. In the Fleksy approach, the ensemble soft-docking stage is always complemented by a powerful complex optimization stage, which in many cases significantly improves the accuracy of the generated docking poses. As a result, Fleksy creates realistic clash-free models of the receptor–ligand complex, which in follow-up studies can easily be used in other modeling or analysis programs.

Fleksy uses the FlexX-Ensemble ensemble docking program[21] as its docking engine. A powerful feature of the FlexX-Ensemble program is that it can construct novel binding sites from members of the input ensemble; this is in contrast to many other approaches that either do not generate a ligand complementary binding site model[41] or select one model from a set of binding sites given as input.[42] A binding site alone is, however, not sufficient for the subsequent optimization stage. Thus, we extended this FlexX-Ensemble feature by generating a complete receptor–ligand complex to be used in the next processing step.

Using a structural ensemble as a model for protein flexibility enabled us to implement the concept of interaction sampling. This allows for the automatic consideration of Asn, Gln, and His side chains and the hydroxyl groups in Ser, Thr, and Tyr in multiple orientations and protonation states during the ensemble docking stage. As a result of this, it is no longer required to preset the orientation and protonation of these side chains prior to docking, as currently required in nearly all docking approaches. In a similar fashion, Fleksy has the option to evaluate markedly different side chain rotamers, sampled using a backbone-dependent rotamer library, during the docking stage. To the best of our knowledge, only the SLIDE[43] and GOLD[3] docking programs possess similar features. In the SLIDE program, however, side chain motions are designed to remove receptor–ligand van der Waals overlap rather than thoroughly search conformational space.[44] GOLD by default evaluates different side chain hydroxyl orientations, and the program allows for the definition of multiple rotatable side chain torsion angle ranges. It does, however, not have the possibility to assess different protonation states, for example, for histidines, and includes a less advanced backbone independent rotamer library.[45]

In some respects, our approach is similar to a recently introduced induced fit docking (IFD) protocol.[18] Both are multistep protocols that make use of soft-docking strategies and consider a limited set of residues as flexible. However, there are some pronounced differences in the strategies chosen to tackle the protein flexibility problem in small molecule docking. A first difference is that the IFD protocol uses a single structure, with predefined side chain orientations, both as input and during the docking stage, whereas Fleksy is able to use multiple

structures as input and docks into an ensemble of receptors, enabling the interaction sampling described above. Second, the IFD protocol first mutates flexible residues to Ala, and following the docking, the deleted residues are modeled back (which limits the number of residues that can be kept flexible to three[18]). In contrast to this, Fleksy is able to consider multiple sampled side chain orientations during the actual docking experiments. Finally, the IFD protocol requires two cycles of docking, scoring, and refinement, whereas for the Fleksy approach, a single cycle of docking, scoring, and refinement suffices, which results in lower computational requirements and a faster protocol.

As discussed, our method has some clear strengths, but one should also be aware of its current limitations. One hurdle that needs to be taken for full consideration of receptor plasticity is the incorporation of larger degrees of protein backbone movement. Limited amounts of backbone movement can be allowed by adding multiple receptor conformations into the structure ensemble used for the molecular docking stage, obtained either from crystallography or molecular simulations. In the original FlexX-Ensemble publication, loop movements of up to 1.5 Å were successfully included in the structure ensembles,[21] in a more a recent docking study, docking success using FlexX-Ensemble with loop deviations up to 2.5 Å was reported.[46] Fortunately, in many cases, only modest amounts of backbone motion, often smaller than those mentioned here, are sufficient to accurately model ligand-induced changes, as shown throughout this paper.

In this work we used an ensemble of structural models to describe receptor flexibility. Previously, several studies have already suggested that structure ensembles provide an efficient means of describing protein flexibility. Together with the observation that small structural differences can have pronounced effects on docking performance, this poses a strong argument for considering multiple structures in docking studies. Recently, it was demonstrated that even a modestly sized ensemble of experimental structures can capture a representative subset of true native-state dynamics.[40,47] These results highlight both the importance as the feasibility of accounting for native-state protein dynamics via protein structure ensembles.

As mentioned above, in addition to ours, several other methods exist for which a subset of residues is selected as potentially flexible. Examples of these include IFREDA[41] and a recently presented induced fit docking protocol.[18] In this work, we have applied the same set of empirical rules as applied in the latter work to select flexible residues for the induced fit docking dataset. A useful extension of our approach would be the inclusion of more sophisticated techniques to select residues to be considered as flexible. These could for instance include analyses of structural deformability,[48,49] sequence conservation,[50,51] or structural features of binding site residues.[52]

Finally, another foreseeable extension of the Fleksy approach is the inclusion of crystal waters in the generated structure ensembles. Many cases are known where water molecules play a crucial role in receptor–ligand interactions. In a recent docking study, it was also shown that inclusion of crystal waters[53] had a pronounced effect on docking performance in a large set of self-docking experiments. However, the presence of a given crystal water might be optimal for one ligand and it could well be suboptimal for another. Inclusion of observed (or possibly even predicted) crystal waters in the structure ensemble used in the soft-docking stage would allow the FlexX-Ensemble program to include these water when required or preferred by the ligand under investigation. As such, this approach could provide an alternative to other techniques currently available to treat water molecules in small molecule docking.[54,55]

Summary and Conclusions

In this paper, we present Fleksy, a new approach capable of considering both ligand as well as receptor flexibility in a small molecule docking experiment. The method results in a ranked list of minimized docking poses, with each pose complemented by a corresponding, custom receptor structure. We validated our method with more than 300 self-docking and cross-docking experiments using 35 different receptor–ligand complexes taken from the PDB. The obtained results clearly show the ability of our docking pipeline to consider receptor plasticity during small molecule docking. As such, the method is suitable to be readily implemented within a computational drug discovery environment.

Acknowledgment. The authors thank Prof. Dr. Matthias Rarey for stimulating discussions, Dr. Ramy Farid and Dr. Gerd Räther for discussions on the IF dataset, Scott Lusher for critical reading of the manuscript, and Marijn Sanders for his help in data analyses. S.B.N. acknowledges financial support of the BioRange program of The Netherlands Bioinformatics Centre (NBIC), which is supported by a BSIK grant through The Netherlands Genomics Initiative (NGI).

Supporting Information Available: Supplementary Table 1 describing the induced fit docking dataset. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.
(2) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
(3) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–748.
(4) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
(5) 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.
(6) 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.
(7) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213–218.
(8) Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery* **2003**, *2*, 527–541.
(9) Fischer, E. Über die Optischen Isomeren des Traubenzuckers, der Gluconsäure und der Zuckersäure. *Ber. Dtsch. Chem. Ges.* **1890**, *23*, 2611.
(10) Koshland, D. E. Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98–104.
(11) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531–568.
(12) Carlson, H. A. Protein flexibility and drug design: How to hit a moving target. *Curr. Opin. Chem. Biol.* **2002**, *6*, 447–452.
(13) Teodoro, M. L.; Kavraki, L. E. Conformational flexibility models for the receptor in structure based drug design. *Curr. Pharm. Des* **2003**, *9*, 1635–1648.
(14) Barril, X.; Fradera, X. Incorporating protein flexibility into docking and structure-based drug design. *Expert Opin. Drug Discovery* **2006**, *1*, 335–349.
(15) Arevalo, J. H.; Stura, E. A.; Taussig, M. J.; Wilson, I. A. Three-dimensional structure of an anti-steroid Fab′ and progesterone-Fab′ complex. *J. Mol. Biol.* **1993**, *231*, 103–118.

(16) Noble, M. E.; Endicott, J. A.; Johnson, L. N. Protein kinase inhibitors: Insights into drug design from structure. *Science* **2004**, *303*, 1800–1805.
(17) Anderson, A. C.; O'Neil, R. H.; Surti, T. S.; Stroud, R. M. Approaches to solving the rigid receptor problem by identifying a minimal set of flexible residues during ligand docking. *Chem. Biol.* **2001**, *8*, 445–457.
(18) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–553.
(19) Hooft, R. W.; Sander, C.; Vriend, G. Positioning hydrogen atoms by optimizing hydrogen-bond networks in protein structures. *Proteins* **1996**, *26*, 363–376.
(20) Word, J. M.; Lovell, S. C.; Richardson, J. S.; Richardson, D. C. Asparagine and glutamine: Using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* **1999**, *285*, 1735–1747.
(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) Jiang, F.; Kim, S. H. "Soft docking": Matching of molecular surface cubes. *J. Mol. Biol.* **1991**, *219*, 79–102.
(23) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: Conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317–324.
(24) Velec, H. F.; Gohlke, H.; Klebe, G. DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* **2005**, *48*, 6296–6303.
(25) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.
(26) Breitenlechner, C. B.; Friebe, W. G.; Brunet, E.; Werner, G.; Graul, K.; Thomas, U.; Kunkele, K. P.; Schafer, W.; Gassel, M.; Bossemeyer, D.; Huber, R.; Engh, R. A.; Masjost, B. Design and crystal structures of protein kinase B-selective inhibitors in complex with protein kinase A and mutants. *J. Med. Chem.* **2005**, *48*, 163–170.
(27) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
(28) Vriend, G.; Sander, C. Detection of common three-dimensional substructures in proteins. *Proteins* **1991**, *11*, 52–58.
(29) Krieger, E.; Darden, T.; Nabuurs, S. B.; Finkelstein, A.; Vriend, G. Making optimal use of empirical energy functions: Force-field parameterization in crystal space. *Proteins* **2004**, *57*, 678–683.
(30) Essman, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Phys. Chem. B* **1995**, *103*, 8577–8593.
(31) Gasteiger, J.; Rudolph, C.; Sadowski, J. Automatic generation of 3D atomic coordinates for organic molecules. *Tetrahedron Comput. Methodol.* **1990**, *3*, 537–547.
(32) Eddy, S. R. Where did the BLOSUM62 alignment score matrix come from. *Nat. Biotechnol.* **2004**, *22*, 1035–1036.
(33) Oda, A.; Tsuchida, K.; Takakura, T.; Yamaotsu, N.; Hirono, S. Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *J. Chem. Inf. Model.* **2006**, *46*, 380–391.
(34) Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* **1999**, *37*, 228–241.
(35) Bledsoe, R. K.; Madauss, K. P.; Holt, J. A.; Apolito, C. J.; Lambert, M. H.; Pearce, K. H.; Stanley, T. B.; Stewart, E. L.; Trump, R. P.; Willson, T. M.; Williams, S. P. A ligand-mediated hydrogen bond network required for the activation of the mineralocorticoid receptor. *J. Biol. Chem.* **2005**, *280*, 31283–31293.
(36) Fagart, J.; Huyet, J.; Pinon, G. M.; Rochel, M.; Mayer, C.; Rafestin-Oblin, M. E. Crystal structure of a mutant mineralocorticoid receptor responsible for hypertension. *Nat. Struct. Mol. Biol.* **2005**, *12*, 554–555.
(37) Li, Y.; Suino, K.; Daugherty, J.; Xu, H. E. Structural and biochemical mechanisms for the specificity of hormone binding and coactivator assembly by mineralocorticoid receptor. *Mol. Cell* **2005**, *19*, 367–380.
(38) Breitenlechner, C. B.; Wegge, T.; Berillon, L.; Graul, K.; Marzenell, K.; Friebe, W. G.; Thomas, U.; Schumacher, R.; Huber, R.; Engh, R. A.; Masjost, B. Structure-based optimization of novel azepane derivatives as PKB inhibitors. *J. Med. Chem.* **2004**, *47*, 1375–1390.
(39) Kramer, B.; Rarey, M.; Lengauer, T. CASP2 experiences with docking flexible ligands using FlexX. *Proteins* **1997**, *Suppl 1*, 221–225.
(40) Furnham, N.; Blundell, T. L.; DePristo, M. A.; Terwilliger, T. C. Is one solution good enough. *Nat. Struct. Mol. Biol.* **2006**, *13*, 184–185.
(41) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
(42) Huang, S. Y.; Zou, X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins* **2007**, *66*, 399–421.
(43) Schnecke, V.; Swanson, C. A.; Getzoff, E. D.; Tainer, J. A.; Kuhn, L. A. Screening a peptidyl database for potential ligands to proteins with side-chain flexibility. *Proteins* **1998**, *33*, 74–87.
(44) Zavodszky, M. I.; Kuhn, L. A. Side-chain flexibility in protein−ligand binding: The minimal rotation hypothesis. *Protein Sci.* **2005**, *14*, 1104–1114.
(45) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. The penultimate rotamer library. *Proteins* **2000**, *40*, 389–408.
(46) Polgar, T.; Keseru, G. M. Ensemble docking into flexible active sites. Critical evaluation of FlexE against JNK-3 and beta-secretase. *J. Chem. Inf. Model.* **2006**, *46*, 1795–1805.
(47) Best, R. B.; Lindorff-Larsen, K.; DePristo, M. A.; Vendruscolo, M. Relation between native ensembles and experimental structures of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10901–10906.
(48) Kovacs, J. A.; Chacon, P.; Abagyan, R. Predictions of protein flexibility: First-order measures. *Proteins* **2004**, *56*, 661–668.
(49) Garzon, J. I.; Kovacs, J.; Abagyan, R.; Chacon, P. DFprot: A webtool for predicting local chain deformability. *Bioinformatics* **2007**, *23*, 901–902.
(50) Landau, M.; Mayrose, I.; Rosenberg, Y.; Glaser, F.; Martz, E.; Pupko, T.; Ben-Tal, N. ConSurf 2005: The projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **2005**, *33*, W299–W302.
(51) Glaser, F.; Rosenberg, Y.; Kessel, A.; Pupko, T.; Ben-Tal, N. The ConSurf-HSSP database: the mapping of evolutionary conservation among homologs onto PDB structures. *Proteins* **2005**, *58*, 610–617.
(52) Gunasekaran, K.; Nussinov, R. How different are structurally flexible and rigid binding sites? Sequence and structural features discriminating proteins that do and do not undergo conformational change upon ligand binding. *J. Mol. Biol.* **2007**, *365*, 257–273.
(53) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J. Med. Chem.* **2007**, *50*, 726–741.
(54) Rarey, M.; Kramer, B.; Lengauer, T. The particle concept: Placing discrete water molecules during protein−ligand docking predictions. *Proteins* **1999**, *34*, 17–28.
(55) Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Nissink, J. W.; Taylor, R. D.; Taylor, R. Modeling water molecules in protein−ligand docking using GOLD. *J. Med. Chem.* **2005**, *48*, 6504–6515.

JM070593P