Articles

Flexible Side Chain Models Improve Enrichment Rates in In Silico Screening

Daria B. Kokh and Wolfgang Wenzel*

Fachbereich C—Mathematik and Naturwissenschaften, Bergische Universität Wuppertal, Gaussstrasse 20, D-42119 Wuppertal, Germany, Institut für Nanotechnologie, Forschungszentrum Karlsruhe, Postfach 3640, 76021 Karlsruhe, Germany

Received February 29, 2008

While modern docking methods often predict accurate binding modes, affinity calculations remain challenging and enrichment rates of in silico screening methods unsatisfactory. Inadequate treatment of induced fit effects is one major shortcoming of existing in silico screening methods. Here we investigate enrichment rates of rigid-, soft- and flexible-receptor models for 12 diverse receptors using libraries containing up to 13000 molecules. For the rigid-receptor model, we observed high enrichment ($EF_1 > 20$) only for four target proteins. A soft-receptor model showed improved docking rates at the expense of reduced enrichment rates. A flexible side-chain model with flexible dihedral angles for up to 12 amino acids (3−8 flexible side chains) increased both binding propensity and enrichment rates: EF_1 values increased by ∼35% on average with respect to rigid docking. We find on average 4 known ligands in the top 10 molecules in the rank-ordered databases for the receptors investigated.

I. Introduction

Despite an enormous increase of genetic and structural data relating to a wide variety of diseases, the number of newly discovered drugs, in particular those with novel scaffolds,[1] is decreasing.[2] This opens new opportunities for virtual ligand screening, where molecules in a large database are ranked by affinity to identify some at least weakly binding molecules (leads) for further refinement. Aided by ever-increasing computational power, virtual screening is an appealing and cost-effective approach to tap into the wealth of available structural information.[3] However, despite several success stories, limitations in current in silico screening approaches restrict their accuracy and general applicability.[4,5]

The ability of structure-based docking methods to predict affinities crucially depends on the accuracy of the structural model of the complex. It is well recognized that not only ligand flexibility but also the intrinsic flexibility of the receptor plays an important role in the formation of the protein−ligand complex. Changes in the structure of the protein binding site upon ligand binding have been experimentally demonstrated for a range of therapeutically important receptors (see refs 6−8 and references therein). In many cases, the comparison of ligand-free with ligand-bound protein crystal structures is sufficient to illustrate the severity of the problem. Because crystal structures for protein−ligand complexes are difficult to obtain even if a suitable ligand is available, the development of docking methods that can dock into ligand-free crystal structures and incorporate induced-fit effects is of great importance. Rigid docking based on a single protein structure restricts the conformational search and may lead to errors in the identification of the correct binding mode and the binding affinity of novel ligands. The screening process is then biased toward compounds of high molecular similarity to the chemotype of the cocrystallized ligand.[9] In many cases, as we also illustrate below, even small side chain movements can significantly impact the results of docking simulations.[10,11]

Recently a range of models and computational tools have been developed to accommodate protein flexibility into docking methods.[8,12−18] Both the conformational variability of the binding partners prior to their association and their conformational adaptation upon binding contribute to protein and ligand conformational changes. These two closely related mechanisms are used as the starting points for modeling a variety of protein holo structures observed in experiments. Accordingly, methods for treatment of receptor flexibility in molecular docking can be categorized into two main groups: (1) methods that account for receptor reconstruction by combining multiple conformations of a target known from experiment or simulation (ensemble-docking approach) and (2) methods that sample receptor flexibility explicitly during the docking simulation (induced-fit approach).

In addition, soft-docking approaches have been proposed that modify the interactions in empirical scoring functions[19,20] (usually changing the Lennard-Jones interaction term) to adapt the size of the binding pocket for larger ligands. Because of its simplicity and speed, this approximation is very attractive and often able to identify additional ligands that cannot be docked with single-structure rigid docking.[20] However, its conformational and energetic assumptions are difficult to verify. In soft-docking methods, many known ligands reach better affinity estimates than in rigid docking but simply increasing the size of the cavity tends to generate a large number of false positives in screening applications.

Ensemble docking methods enlarge the conformational space available to ligands by postulating an ensemble of accessible, low-energy protein conformations. However, repeated docking of an entire database against a large protein structure ensemble is very time-consuming. Several algorithms have been developed that address side chain or even backbone flexibility by implementing ensembles of accessible receptor structures. Switching

* To whom correspondence should be addressed. Phone: +49-7247-82-6386. Fax: 49-7247-82-6434. E-mail: wenzel@int.fzk.de. Website: www.fzk.de/biostruct.

10.1021/jm800217k CCC: $40.75
Published on Web 09/05/2008

between the members of these ensembles is treated as a set of additional degrees of freedom that is sampled during the simulation.[21−27] Methods for the generation of the relevant ensemble of protein structures in ensemble-docking methods rely to a great extent on "knowledge-based" modeling. Given the present state of the art in protein modeling, these methods may miss relevant conformations that are not easily deduced from the experimental holo structures while including some unphysical conformations. Despite these limitations, ensemble docking methods have been shown to be quite effective in finding new lead candidates that would have been missed when docking into rigid receptors and in reproducing complex conformations of known compounds.[22−24] Unfortunately, energy differences among the receptor conformations of the simulated ensemble are generally not included in the scoring function. Therefore, the benefits of flexible receptor docking are less clear for library screening, where docking methods must be able to distinguish between true and false positives using energetic criteria. Docking into an ensemble of cavity conformations may result in lower enrichment rates than rigid docking,[22,23] which is attributed to the fact that conformational variability of the receptor may lead to unfavorable reorganization energies that can be of the same order of magnitude as the binding energy differences.[7] A term representing the complex changes of the internal energy of the receptor can be introduced into the scoring function, but its functional form must be based on additional assumptions. Approximations, based on the size of the cavity, for example, have been shown to improve the screening performance.[20,22,23,27]

Overall, the "induced-fit" approach is potentially more accurate and more promising for two reasons: (1) the conformational space of the protein is not restricted to previously known structures and may thus cover a larger set of conformations of the receptor, and (2) the method accounts naturally for changes in the internal energy of the protein, providing a consistent basis for ranking different ligands. However, incorporating induced-fit effects in a docking procedure encounters significant computational problems because protein flexibility, even restricted to the vicinity of a known active site, requires sampling of a large number of degrees of freedom in the simulation, which leads to extraordinary increases in the computational effort. To limit these computational costs and not overstrain models for receptor energy change, receptor movement is limited to small structural variations, for example, by using rotamer libraries[18,28−30] or combinatorial search,[16] by considering only conformational changes of the protein termini in the binding pocket,[31,32] or by inclusion of side-chain flexibility for a limited number of residues in the binding pocket.[33,34] While the first two approximations have been found to be problematic in some cases (it has been shown, for example, that the conformational ensemble of most side chains does not sample new rotamer conformations upon ligand binding[35]), the latter approximation needs additional information for the selection of flexible residues ("soft-points") in the protein active site. Some studies in this direction have already been undertaken: employing chemical intuition,[33] inspection of known binding modes,[36] or using molecular docking experiments.[37] In the latter study, for example, it is demonstrated for one target that conformational changes of only a few hydrophobic residues optimizing vdW interaction with inhibitor atoms are sufficient to significantly improve docking and scoring performance, while polar residues were not observed to undergo conformational changes in order to form hydrogen bonds.[37] Analysis of holo structures[35,38] also shows that in the overwhelming majority of cases only a small number of residues (mostly three or less[38]) undergo conformational changes upon ligand binding. Backbone displacement in the binding pocket is much more difficult to treat than side chain flexibility, but it has been argued that a smaller number of receptors is affected.[38] However, there are dissenting studies that suggest that flexibility should not only involve side chain rotations but also flexible portions of the backbone (in particular loops), metal atoms, or cofactors bound to the protein active site.[39] Obviously full treatment of all degrees of freedom of the complex is the ultimate development goal in docking methods. Considering general properties of the potentially flexible side chains, opinions also differ: in contrast to earlier conclusions,[37] recent studies, based on analysis of about 980 holoprotein structures, suggested[38] that large polar amino acids are more flexible than aromatic residues.

Although there is presently no universal protocol for the treatment of protein flexibility in receptor−ligand docking, most studies agree that inclusion of flexible residues in docking methods can improve docking performance. Unfortunately, investigations of the induced-fit approach have often been limited to specifically chosen receptors and mostly focused on the impact of receptor flexibility on docking accuracy, whereas ligand scoring, the real weakness of the present-day docking methodology, is treated only peripherically.[40]

In the present study, we evaluate the performance of rigid-receptor and "induced-fit" models using the FlexScreen receptor−ligand docking approach,[41] with a particular emphasis on screening efficiency. Sampling side chain conformations during the docking simulation was previously shown to improve scoring and docking of model systems.[42] To evaluate the enrichment capability of different induced-fit methods for a diverse set of receptors, we employ the recently published directory of useful decoys database (DUD[a])[43] that contains data sets for 40 different targets. DUD provides a set of annotated ligands for each receptor and a set of nonbinding decoys (from ∼900 up to 15000 molecules per target). These decoys have similar physical properties with respect to a corresponding annotated ligand in the database but have a distinct topology. The DUD database thus poses a difficult challenge for enrichment screens, because true and false positives in the database share many physical characteristics.

We also present and test a straightforward protocol for the choice of the flexible residues, which is based on the ability of the receptor structure to accommodate the set of known ligands. Tested on the DUD database, this strategy is an unbiased approach to identify the most important residues likely to be relevant for induced fit effects. The docking protocol considers all targets on the same footing, which permits the evaluation of the screening procedure for diverse targets.

II. Methods

1. Docking Method. Docking simulations have been performed with the all-atom FlexScreen receptor−ligand docking program,[41] which employs a force-field-based scoring function

$$S = \sum_{i,j}\left(\frac{R_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} - \frac{q_i q_j}{\varepsilon r_{ij}}\right) + \sum_{\text{h-bonds}} \cos(\Theta_{ij})\left(\frac{\tilde{R}_{ij}}{r_{ij}^{12}} - \frac{\tilde{A}_{ij}}{r_{ij}^{10}}\right) \quad (1)$$

similar to that of AutoDock.[44] The scoring function contains

[a] Abbreviations: EF, enrichment factor; DUD, directory of useful decoys; BEDROC, Boltzmann-enhanced discrimination of receiver operating characteristic; rmsd, root-mean-square deviation.

Table 1. The List of Proteins Considered in the Present Work, Number of Annotated Ligands, and Average Root-Mean-Square Deviation, rmsd, of the Docked Ligand to the Experimental Binding Pose in Simulations of the Rigid Holo Structure for Each Receptor[a]

	protein abbreviation	PDB code	protein name	no. of annotated ligands	rmsd (nm)	ΔEF_1
1	AR[b]	1xq2	androgen receptor	79	0.03	1.3
2	COX-1[c]	1p4g	cyclooxygenase 1	25	0.05	4
3	COX-2[b]	1cx2	cyclooxygenase 2	349	0.13	0.3
4	ER_agonist[b]	1l2i	estrogen receptor agonist	68	0.06	1.5
5	GPB[c]	1a8i	glycogen phosphorylase beta	52	0.06	1.9
6	GR[b]	1m2z	glutocorticoid receptor	78	0.04	1.3
7	MR[b]	2aa2	mineralcorticoid receptor	15	0.04	6.7
8	PNP[b]	1b8o	purine nucleoside phosphorylase	24	0.03	4.2
9	PR[b]	1sr7	progesterone receptor	27	0.05	3.7
10	RXRa[b]	1mvc	retinoic × receptor α	20	0.05	5.0
11	SAHH[c]	1a7a	*S*-adenosyl-homocysteine hydrolase	33	0.03	3.0
12	TK[d]	1kim	thymidine kinase	22	0.06	4.5

[a] The influence of adding or dropping a single ligand to the top percent of the database is given in the last column to judge the sensitivitiy of this value to small fluctuations in the number of selected ligands. [b] Nuclear hormone receptors. [c] Enzymes. [d] Kinases.

Lennard-Jones (first two terms), electrostatic Coulomb (third term, $\varepsilon = 4$), and angular dependent hydrogen bond (terms four and five) potentials. Protein and ligand atoms (i,j) are treated on the same footing. The Lennard-Jones and the hydrogen bond parameters have been taken from OPLSAA[45] and AutoDock,[44] respectively. The scoring function does not include solvent-related effects; it will thus differentiate between ligands where the binding energy is dominated by electrostatic/hydrogen bonding interactions and not by solvation contributions.

The docking protocol[33,41,46,47] was automated in order to consider all targets on the same footing. In addition to the protein and ligand structure (in the MOL2 format), FlexScreen requires a specification of a "docking center" (around which sampling is enhanced, but which has no effect on the energies), which we computed as the center of geometry of the native ligand extracted from the experimental holo structure of the corresponding target.

The ligands are docked using a cascadic version of the stochastic tunneling algorithm,[41,46] which samples translations of the center of mass and random rotations of the ligand as well as intramolecular conformational changes of the ligands. If selected, the dihedral angles of several receptor side chains are also sampled (manual selection in the input). In each step, FlexScreen either changes a dihedral angle of the ligand or a flexible side chain by a small, randomly chosen angle (drawn from a Gaussian distribution) or displaces the ligand by a small amount.[41] The total number of simulation steps was divided to three partitions: In the first partition 200 simulations (5000 steps each) were performed, the best five (by energy) were selected as the starting points of the second stage. After an additional 30000 steps for each conformation in the second stage, the best two are selected for a final relaxation of 75000 steps. The FlexScreen procedure thus generates two final conformations/energies for each ligand, providing an error estimate. We also tested simulations with twice the number of steps but found no notable difference in results.

FlexScreen allows continuous rotations around the single bonds of the side chains of up to 15 residues in the energy optimization procedure.[46] Additional information for the docking method as well as its applications and examination of docking performance for several systems has been reported elsewhere;[33,41,46,47] for all parameters not explicitly stated, default values were used. Depending on the number of ligand atoms and the number of flexible residues, each ligand required from 1−3 min for rigid and to 5−15 min for flexible receptor docking (IntelPC-86-64, 1.8 GHz processor).

2. Docking Database. The DUD library,[43] employed in the present analysis, contains the following data sets: (i) receptor PDB files extracted from experimental holo structures and corresponding native ligand conformations in MOL2 format; (ii) from 15 to 349 annotated ligands for the active site of each specific protein; (iii) 36 decoy molecules (excluding chirality duplicates) for every annotated ligand. The decoys have similar physical properties (such as molecular weight, cLogP, number of hydrogen donors and acceptors, number of rotatable bonds, and number of important functional groups) but different topology (comparison based on a fingerprint-based similarity analysis)[43] to ensure that, in all likelihood, the decoys bind much less to the receptor than the annotated ligands. The annotated ligands and their corresponding decoys form a library subset for each receptor, containing about 2.6% of well-binding and 97.4% of presumably nonbinding molecules. All enrichment calculations presented in this work have been carried out against these receptor-specific databases. Molecules in the DUD library are specified in MOL2 format, including partial charges, and were used without further modification for the docking FlexScreen procedure.

3. Receptor Preparation. We have analyzed holostructures of each protein with MOE[48] and selected 12 receptors listed in Table 1 with relatively small, closed binding cavities that are completely buried. These receptors present particularly challenging targets for induced-fit investigations because ligands must be accommodated in the constraints of the receptor pocket and cannot escape toward the solvent. The COX-1 receptor is the only example with a partially open pocket, which we included in this study. We experimented with several other open receptor pockets but found the induced fit problem less severe because few residues clash with known ligands.

Protonation states and receptor partial charges were prepared with the MOE program[48] using AMBER99 parameters.[49] Appropriate conformations for cofactors (phosphate, $H_2PO_4^{1-}$, and sulfate, HSO_4^{1-}, groups in PNP and TK, respectively) have been prepared with the PyMol[50] program and were incorporated into the corresponding MOL2 files. Tightly bound crystallographic water molecules, present in PDB files of the GPB target, have been maintained during the docking/scoring process and were considered as flexible, where necessary. Where required (see Results), we performed calculations with different water/cofactor configurations.

4. Criteria of Docking and Screening Efficiency. Docking and screening performance has been evaluated by computing the enrichment of annotated ligands among the top-scoring molecules of the receptor-specific database (see Section II.2). The enrichment factor (EF) of annotated ligands among the top-scoring α% of docked molecules from a database has been defined as:[8]

$$EF_\alpha = \frac{\text{(concentration of known ligands found in top-ranking subset)}}{\text{(concentration of known ligands in database)}} \quad (2)$$

i.e., the fraction of identified known ligands equals $EF_\alpha(\alpha/100)$.

In the present work, we report the values of EF_α, corresponding to enrichment factors at $\alpha = 1\%$, 2.7%, 10%, 20%, and 100%. EF_{100} equals the fraction of annotated ligands that bind to the receptor in the docking calculations (binding energy less than zero) and, therefore, shows the efficiency of ligand docking. If $EF_{100} = 100$, all known ligands "dock", i.e., attain negative binding energies. Values of $EF_{100} > 0.80$ have generally been considered as an indication of good docking efficiency. Enrichment factors at 1% and 2.7% of the top-ranking ligands in the database measure capability of the method to distinguish between true and false positives by correctly ranking their binding energies. EF_1 is often used as a criterion of screening efficiency because only a small fraction of the ligand database can be selected for experimental investigation in practice. Because in our case 1% of the database corresponds to only ~37% of the total number of annotated ligands, the maximum value of EF_1 is ~37 for all targets. Correspondingly, the maximal value for $EF_{2.7}$ is 100 by construction.

There are a number of uncertainties in any screening procedure that may lead to errors in the calculation of the enrichment factors, including, but not limited to the database composition, the accuracy of the scoring function, the quality of the sampling, the quality of the crystal structure, and uncertainties in the ligand/receptor parametrization (pH, charges, etc.). In the present study, sampling errors seem smaller than the differences of different screening procedures as verified by rescreening several receptors. However, the enrichment factor may depend very strongly on the number of identified ligands in the subset when the number of known ligands is very small. This is quantified by considering the change in EF, when a single additional ligand is either added to or dropped from the selected subset:

$$\Delta EF_\alpha = \frac{L^{\text{found}}(\alpha 0.01 N^{\text{tot}})^{-1}}{L^{\text{tot}} N^{\text{tot}-1}} - \frac{(L^{\text{found}} \pm 1)(\alpha 0.01 N)^{-1}}{L^{\text{tot}} N^{\text{tot}-1}} = \pm \frac{100}{\alpha N^{\text{lig}}} \quad (3)$$

Here L^{found} denotes the number of ligands that bind to the receptor in docking calculations, L^{tot} the total number of ligands in a database, and N^{tot} the total number of molecules in a database (ΔEF_1 values are summarized in Table 1).

In addition to the enrichment factor, we report the values of BEDROC (Boltzmann-enhanced discrimination of receiver operating characteristic) metric that has been suggested as "the best metric adapted to the early recognition problem":[51]

$$\text{BEDROC} = \frac{\int_0^1 F_a(x) w(x)\, dx}{\int_0^1 w(x)\, dx} \quad (4)$$

where $F_a(x)$equals concentration of known ligands found in a subset, and integration is over the fraction x of database; $w(x) = \exp(-\alpha x)$ is a weighting function.

BEDROC values describe the probability that annotated ligands rank higher than a decoy drawn from a hypothetical exponential probability distribution function, which approximately corresponds to the top-ranking 2% of the database for the choice of the parameter $\alpha = 0.5$.

5. Selection of Receptor Degrees of Freedom. To treat receptor flexibility for all systems at an unbiased level, some uniform scheme for choosing the flexible residues has to be implemented. We have therefore established a completely unbiased protocol that identifies potentially flexible residues by searching for clashes between known ligands and the receptor, which result in a positive binding energy (ligand does not dock). If no suitable binding conformation can be found for a known ligand in the apo conformation of the receptor, this is an indication that additional flexibility must be included to model the protein binding site correctly.

For an unbiased evaluation of the screening efficiency we have used the following procedure: (1) we compiled a list of known ligands with positive binding energy (clashing ligands) using a rigid-receptor simulation, (2) for these ligands a list of residues with large positive vdW energy (above 20 kJ/Mol) were compiled for each target (the number of residues in this list is referred as $N^{\text{f}}_{\text{total}}$), (3) these residues were ranked according to the number of large vdW terms (protein−ligand collisions) caused by this specific residue over all clashing ligands (hereafter we will refer to such lists as the "ligand-based" set of flexible residues), (4) top-ranking residues from this list, N^{f}, were treated as flexible (see data below for each screen). The choice of N^{f} will discussed below in more detail.

Obviously, the proposed procedure depends on the availability of ligands that are known to bind to the receptor. Here we used it to have an unbiased selection of flexible residues for the DUD data set. However, in most practical applications, we presume that this method would be refined by additional inspection of potentially flexible residues. To investigate the possibility to choose potentially flexible residues without knowledge of ligands, we have experimented with a "decoy" based approach, where the whole decoy set (instead of just the known ligands), is used to identify potentially clashing residues.

The number of potentially flexible residues in the "ligand-based list" ($N^{\text{f}}_{\text{total}}$) for each protein, shown in Table 3, can run as high as 30, depending on the number of annotated ligands for the protein in the database. We note that $N^{\text{f}}_{\text{total}}$ most likely overestimates the number of side chains that must be considered as flexible to accommodate a specific ligand, simply because (i) a large number of residues are involved in interaction with the ligand simultaneously, and (ii) residues are selected for all annotated ligands simultaneously. On the other hand, some additional residues that do not have close contact with any known ligand may need to be treated as flexible in order to optimize polar contacts. Despite these limitations the "ligand-based" list should improve the success rate of the docking protocol and is an unbiased receptor-specific selection of a set of potentially flexible residues.

To examine the dependence of the selected set of potentially flexible residues on the number of available ligands (training set effect), we applied the selection procedure to different randomly chosen subsets of annotated ligands. We compared the ranked-ordered set of potentially flexible residues (top 5 and top 10 side chains) determined on the basis of subsets with the "optimal" set of potentially flexible residues obtained using all available ligands. The selection protocol is shown to be quite robust: for the AR receptor, for example, 19 molecules of 79 ligands do not dock into the apo structure, resulting in an "optimal" set of five most important flexible residues: MET745, PHE764, LEU880, THR877, and MET749. When we reduced the size of the training set to 40 randomly chosen ligands, only nine dock, but four of the five "optimal" residues are again selected: (MET749 is replaced by ARG752). By reducing the

Table 2. Comparison of the Docking and Scoring Efficiency for Rigid-, Soft- (0.25 nm shift), and Flexible-Receptor Docking: Fraction of Identified Ligands (EF_{100}), Enrichment of Known Ligands (EF_x, $x = 20$, 10, 2.7, 1) at x% of Rank-Ordered Database, and BEDROC[51] Metric Values against the Receptor-Specific Databases (See Text)[a]

protein	receptor model/N^f	EF_{100}	EF_{20}	EF_{10}	$EF_{2.7}$	EF_1	EF_1[43]	BEDROC
AR	rigid	*0.77*	*3.3*	*5.7*	*12.5*	*9.4*	*15*	*0.13*
	soft/22							0.11
	flexible/8	0.90	3.8	6.1	10.6	10.9		0.21
		0.85	3.6	6.5	16.8	25.3		
COX-1	rigid	*0.64*	*2.8*	*3.0*	*5.9*	*9.6*	*3*	*0.08*
	soft/16	0.84	2.9	4.4	9.4	12.5		0.13
	flexible/9	0.88	3.7	6.0	15.7	30.3		0.24
COX-2	rigid	*0.64*	*3.2*	*5.2*	*13.5*	*25.5*	*20*	*0.20*
	soft/36	0.90	3.2	5.1	14.2	25.5		0.21
	flexible/3	0.66	3.3	5.6	14.6	27.1		0.21
ER_agonist	rigid	*0.97*	*3.5*	*6.0*	*17.5*	*27.6*	*5*	*0.25*
	soft/4	0.99	3.7	7.0	17.9	28.6		0.24
	flexible/4	0.96	3.6	6.2	17.7	30.5		0.26
GPB	rigid	*0.88*	*3.8*	*7.0*	*18.3*	*18.8*	*5*	*0.19*
	soft/16	0.95	3.9	7.3	10.9	15.0		0.14
	flexible/6	0.93	3.9	7.0	18.4	26.5		0.21
GR	rigid	*0.40*	*1.9*	*3.3*	*8.8*	*14.9*	*20*	*0.14*
	soft/25	0.58	1.8	2.9	5.2	9.1		0.09
	flexible/6	0.46	2.1	3.2	8.9	17.5		0.15
MR	rigid	*0.80*	*3.5*	*6.9*	*12.7*	*26.6*	*25*	*0.30*
	soft/11	0.80	2.8	5.7	12.4	18.6		0.22
	flexible/6	0.80	3.6	6.4	15.6	30.0		0.32
PNP	rigid	*0.52*	*2.3*	*4.1*	*10.4*	*20.2*	*18*	*0.18*
	soft/21	0.92	3.2	5.4	15.3	26.0		0.21
	flexible/3	0.75	2.1	4.1	10.8	21.1		0.19
PR	rigid	*0.76*	*1.9*	*2.2*	*4.5*	*8.6*	*0*	*0.08*
	soft/10	0.93	1.6	3.0	1.3	3.4		0.07
	flexible/7	0.80	2.3	3.5	5.3	10.4		0.09
RXRa	rigid	*0.55*	*1.9*	*2.0*	*3.7*	*9.6*	*15*	*0.07*
	soft/12	0.95	2.7	1.9	4.4	9.2		0.08
	flexible/11	0.80	3.3	6.0	11.3	21.7		0.19
SAHH	rigid	*0.79*	*3.7*	*6.4*	*13.7*	*18.6*	*10*	*0.15*
	soft/10	1.0	3.9	6.0	9.2	11.9		0.12
	flexible/2	0.91	4.1	7.4	16.7	18.7		0.17
TK	rigid	*0.82*	*3.5*	*6.5*	*15.6*	*18.4*	*0*	*0.20*
	soft/8	0.86	3.3	6.0	14.4	20.8		0.21
	flexible/3	0.84	3.4	5.0	13.0	19.7		0.19

[a] The values of EF_1 for the receptor specific decoy sets were extracted from Figure 3 of ref 43 for comparison. N^f designates the number of shifted or flexible side chains, the list of residues for each receptor is given explicitly in Table 3.

training set again by 50% to 20 known ligands only three of the five most important residues survive (only 3 of the 20 ligands do not dock): MET745, PHE764, and THR877. However, even in this case, the choice of the top 5 or 10 of flexible residues is far from random. As can be expected, the uncertainty of potentially flexible residues increases as the subset size shrinks. Similar results were obtained for other receptors with more than 25 annotated ligands and more then 10 residues in the list of potentially flexible residues in the "optimal" list: for all these receptors (see Table 4), about 3−4 residues among the five top-ranking potentially flexible residues can be found even for rather small training sets of ligands.

III. Results

1. Prediction of Native Ligand Binding Modes. We first docked the native ligands of each receptor to investigate the accuracy of the protocol in prediction of the ligand binding mode observed in the crystallographic structures. These results are summarized in Table 1. They show that rigid-receptor docking simulations against holo structures yield root-mean-square deviations (rmsd) of less than 0.06 nm for all but one receptor (COX-2, rmsd = 0.13 nm).

2. Rigid-Receptor Docking. Table 2 summarizes the key results of rigid-receptor docking (in italic): For EF_{100}, we find that between 40% (worst case: GR) and 97% (best case: estrogen receptor agonist) of the known ligands dock to their respective receptor. Obviously, if the receptor conformation does not permit docking of all known ligands, many prospective drug candidates are eliminated from the screen, completely independent of the quality of the docking method and of the scoring function. A representative example of such a difficult case is shown in Figure 1, where several clashes with atoms in the holo structure preclude docking of several known ligands. On average, 71.5% of ligands dock successfully into the rigid receptor structures.

Qualitatively, the enrichment performance of the rigid-receptor screen can be classified using the thresholds of ref 43: screens with $EF_1 > 20$ were considered as "good", $10 < EF_1 < 20$ as "medium", and $EF_1 < 10$ as "poor". $EF_1 = 1$ corresponds to a random selection of ligands. The enrichment factors of ligands reported in ref 43 for the receptors considered here are "good" for 3 cases ($EF_1 > 20$), while 4 cases performed "medium" and 5 "poor" ($EF_1 < 10$) by these standards. Among the latter cases, for two targets (PR and TK), the value of EF_1 is almost zero. Enrichment data for one challenging (PR, $EF_1 = 8.6$) and one successful example (ER-agonist $EF_1 = 27.6$) are illustrated in Figure 2A. The distribution of molecules from the database (including ligands) as a function of the binding energy is shown in Figure 2B,C. The figure clearly demonstrates that the compounds with the best binding energies are mostly known ligands: specifically about 75% of the top 1% of molecules for the ER-agonist are annotated ligands, whereas for the PR target, this value is only about 23%. While the latter case is not fully satisfactory, both results are much higher than random selection (EF = 1).

"Good" docking, as well as screening performance ($EF_{100} > 0.8$ and $EF_1 > 20$), was achieved for two targets: the ER-agonist and MR. There are three further targets (GPB, SAHH, TK) where docking performance can be described as "medium" ($EF_1 \sim 18$ and $EF_{100} \sim 0.8$). Interestingly, enrichment (EF_1) is "medium" or even "good" for three proteins (COX-2, GR, PNP), where more than 20% of the annotated ligands do not dock at all: for example, COX-2 and PNP, where only 68% and 56% known ligands bind, still have enrichment factors of $EF_1 = 25.8$ and 20.2, respectively. This suggests that there are different classes of ligands, some dock well and the relative ranking is correct, while others do not dock at all. The remaining four targets (AR, COX-1, PR, RXRα) show quite "poor" docking and scoring performance. It is important that there is no example where docking is "good" ($EF_{100} > 0.8$), but scoring is "poor" ($EF_1 < 10$). This is commensurate with the assumption that the method is able to differentiate well between different classes of structurally dissimilar ligands: if a known ligand does not dock well, neither do all associated decoys. For such cases ligand-induced receptor reconstruction must be considered to increase the number of ligands that successfully dock into an apoprotein structure.

It is known that the presence of cofactors and crystal water can significantly affect the scoring results (see, for example, ref 52). Our observations clearly support this conclusion. We have considered the PNP and GPB receptors with various poses of a cofactor/crystal-water bound to the active site. The $PO_4H_2^{1-}$

Table 3. Receptor Specific Sets of Flexible Residues: Multiple Sets of Flexible Residues are Shown in Brackets for Each Receptor, Residues Involved in Polar Contacts with Annotated Ligands, Number of Residues in a "Ligand-Based" List of Flexible Residues (N^{f}_{total}), and Approximate Number of Polar Contacts for Each Annotated Ligand (N^{p})[a]

protein	flexible residues	polar contacts	N^{f}_{total}	N^{p}
AR	[[[MET745, PHE764, LEU880, MET749, **THR877**], **ARG752**, LEU704, LEU873], MET780, VAL746]	ARG752, GLN711, ASN705, LEU794, THR877	22	1–3
COX-1	[[[LEU531, **ALA527**, ILE523, VAL349], LEU352, PHE518, VAL116], SER353, LEU384]	TYR385, ARG120, SER530, TYR355, ALA527	20	1–3
COX-2	[[VAL523, **SER353**, VAL349], **TYR355**, **ALA527**, **LEU352**, LEU531]	ARG120, GLN192, HIS90, ARG513, LEU352, SER353, SER530, TYR385, TYR355, MET522, GLU524, GLY526, PHE518, ALA527	29	2–4
ERagonist	[MET343, THR347, **LEU387**, LEU391]	GLY521, ARG394, GLU352, HIE524, LEU346, LEU387	7	2–3
GPB	[[**LEU136**, LYS574], **HIS377**, **GLU672**, **ASN284**]	LEU136, LYS135, GLY135, ASN284, GLU672, ASP283, HIS377, LLP1, GLY675, THR378, TYR573	18	7–10
GR	[[[**LEU732**, **LEU563**, MET646, PHE623], **MET560**, MET601, **TYR739**], **GLN570**, TYR735, GLY567, MET604, PHE745]	MET560, LEU732, ARG611, THR739, GLN570, ASN564, GLN642, LEU563	24	1–2
MR	[LEU769, MET845, ALA773, TRP806, PHE829, LEU960], **ASN770**, **GLN776**]	GLN776, ARG817, ASN770, THR945, LEU938	16	2–4
PNP	[[**MET219**, PHE200, VAL260], **ALA116**, VAL217]	ASN243, GLU201, ALA116, MET219, HIS257, TYR88, SER33, TYR192, SER220, PO_4H_2	22	4–6
PR	[[[CYS891, **THR894**, MET759, **ARG766**], **LEU763**], LEU714, **SER792**, TYR890, ASN719, PHE905, GLY722]	ARG766, LEU763, GLN725, ASN719, THR894, SER792	16	0–1
RXRa	[[PHE313, **PHE439**, ILE268], TRP305, LEU436, LEU309, ILE310, **GLN275**, VAL342, ILE345, VAL265]	ALA328, ARG316, ASN306, PHE346, LEU325, GLN275	21	2–3
SAHH	[[**HIS353**, THR157], [[LEU344, LEU347, MET358]	GLU59, MET351, HIS55, THR57, ASP131, HIS353, GLU156, ASP190, HIS353, THR157, ASP181, NAH-433	17	5–7
TK	[[**GLN125**, MET128, ALA167], ILE100, TYR172]	GLN125, TYR101, GLU225, ARG222, GLU83, ARG176, HIS58	11	4–5

[a] The flexible residues involved in polar contacts with protein active-side residues are shown in bold.

Table 4. Comparison of Different Sets of Potentially Flexible Residues Chosen on the Basis of Different Subsets of Annotated Ligands[a]

			$N_{AL} = 40$			$N_{AL} = 20$			all decoys		
	N_{AL}	N^{nd}_{AL}	N^{nd}_{AL}[b]	M_{10}[b]	M_5[b]	N^{nd}_{AL}[b]	M_{10}[b]	M_5[b]	N^{nd}_{D}	M_{10}	M_5
AR	79	19	9	6	4	3	6	3	885	8	2
COX-2	349	145	15	8	4	8	7	3	4053	6	3
GR	78	45	22	8	4	12	8	12	1175	7	3
COX1	25	9				6	8	4	208	8	3
SAHH	33	7				4	8	4	886	6	4

[a] N_{AL}: number of annotated ligands in a training set; N^{nd}_{AL} (N^{nd}_{D}): number of annotated ligands (decoys) in a training set that do not dock; M_{10} (M_5): number of flexible residues found in top 10 (5) flexible residues in training sets that are the same as found in top 10 (5) flexible residues in the total set of annotated ligands. The last column compares the "optimal" ligand-based list with the "optimal" decoy-based list. [b] Averaged numbers are obtained from 4 subsets of randomly chosen annotated ligands.

cofactor in the PNP receptor structure, for example, can be orientated either with oxygen (orientation A, see Figure 3) or with hydrogen atoms (orientation B) pointing toward the ligand. This orientation significantly affects screening performance, enrichment changes from $EF_1 = 20.2$ for orientation A to $EF_1 = 16.1$ for orientation B, while overall binding is largely unaffected [$EF_{100} = 52$ (A) and 54 (B)]. Concerning the GPB target, only screening calculations against the holo structure with crystal water molecules show "good" enrichment of annotated ligands ($E_1 = 18.8$, $EF_{100} = 0.88$, see Table 2). As soon as the water is removed from the target structure, potential ligands have more space for binding, which results in a larger number of known ligands that dock ($EF_{100} = 0.94$). However, the selectivity of the screen becomes notably worse ($EF_1 = 6.7$).

3. Soft-Receptor Docking. One of the technically simplest ways to model receptor flexibility is to allow protein residues to be shifted away from the ligand in order to adapt the pocket size. Many methods also reduce the vdW radii or similar "overlap parameters" in the scoring function to achieve the same effect. Here we investigate a "soft-receptor" approximation, shifting every residue from the "ligand-based" list of the flexible residues that is involved in at least two collisions with known ligands (one collision in the case of PR). The residue is shifted by a displacement Δ along a straight line connecting the center of geometry of the ligand atoms and the protein residue atoms away from the location of collision. We experimented with several values of the shift parameter Δ (see Figure 4) and found no strong dependence of the docking performance for a large range of physically reasonable values of Δ. Ultimately a uniform shift value of $\Delta = 0.25$ nm was chosen for all receptors for the data reported in Table 2. We found that the reconstruction of binding pockets at $\Delta = 0.25$ nm does not notably influence the binding poses of the native ligands (average error in rmsd $\sim$0.025 nm with respect to the rigid-docking values listed in Table 1).

The enrichment results at $\Delta = 0.25$ nm for the reconstructed receptors are shown in Figure 5. As can be expected, the soft-receptor model is very effective in finding binding poses for ligands that did not dock in the rigid-receptor calculations. The magnitude of the shift correlates with the number of ligands (EF_{100}) that bind to the receptor as illustrated in Figure 4. Although the improvement differs among targets, more than 80% of known ligands dock at $\Delta = 0.25$ nm for all receptors,

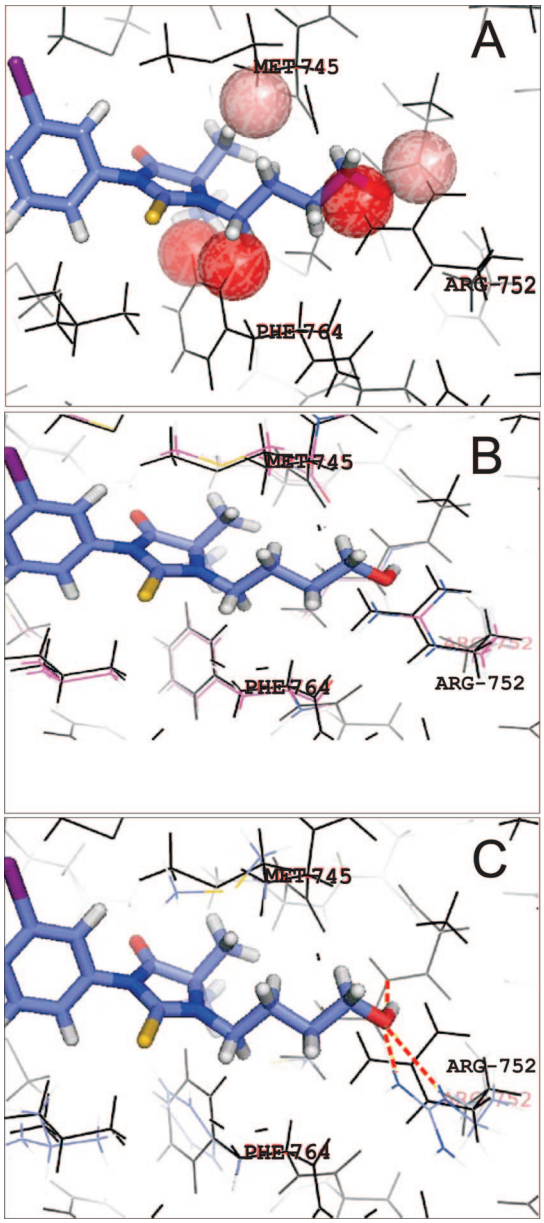

Figure 1. View of a native ligand (shown as blue sticks) bound to AR receptor (side chains are shown in black). (A) protein−ligand contacts that cause large vdW repulsion energies are illustrated by red spheres (bright red indicates protein−ligand distance less than 1Å; light red, distances of 1−2 Å) for three residues: MET-745, PHE-764, and ARG-752. (B) Soft-receptor model: residues shown in red are shifted by 2.5 Å to reduce vdW energy clashes. (C) Three problematic residues are treated as flexible (shown in blue) and their poses are optimized in flexible docking. The flexible docking approach not only allows minimization of vdW repulsion but also the formation of hydrogen bonds between the ligand and ARG-752 residue (shown by dashed line) that where not possible in the original conformation.

except for GR. If the shift parameter is increased to 0.50 nm, EF_{100} increases to 0.83 even for GR.

Because the energy correction accounting for receptor reconstruction is omitted in the soft-receptor model, it is not surprising that this method is not so successful with regard to the enrichment performance (Figure 5, Table 2). At shift values ranging from $\Delta = 0.15$ nm to $\Delta = 0.25$ nm, the soft-receptor model slightly improves enrichment at the EF_1 level in comparison to rigid docking for five receptors (AR, COX-1, TK, ER-agonist, and PNP) (Figure 4). However, even for these targets, EF_1 is lowered at $\Delta > 0.25$ nm and at $\Delta = 0.5$ nm

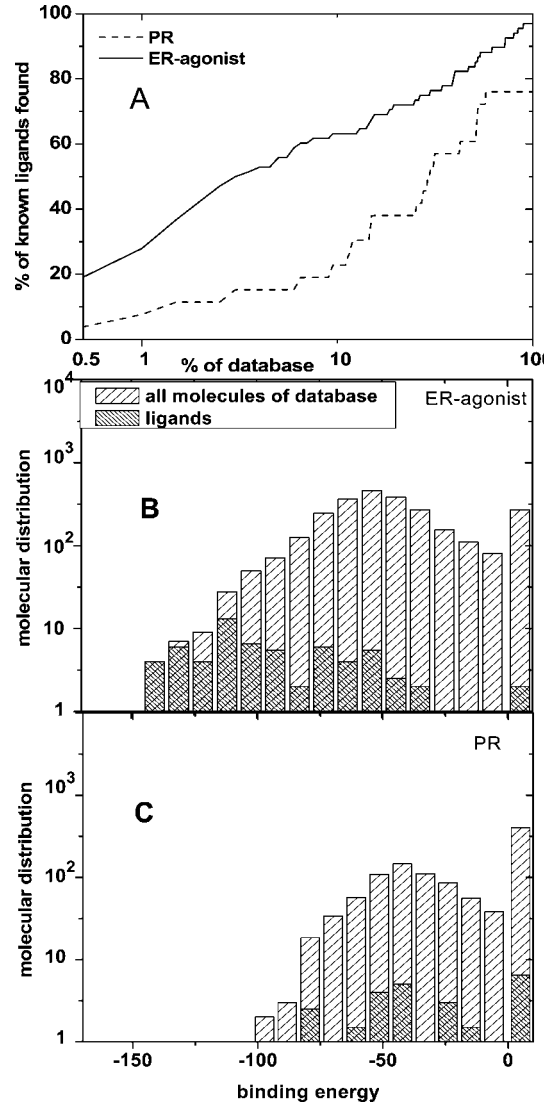

Figure 2. Example of docking enrichment plots for successful (ER-agonist, $EF_1 = 27.6$) and relatively poor (PR, $EF_1 = 8.6$) scoring: (A) enrichment plot, (B,C) distributions of ligands/all molecules (including ligands) vs their binding energy obtained in rigid-docking calculations.

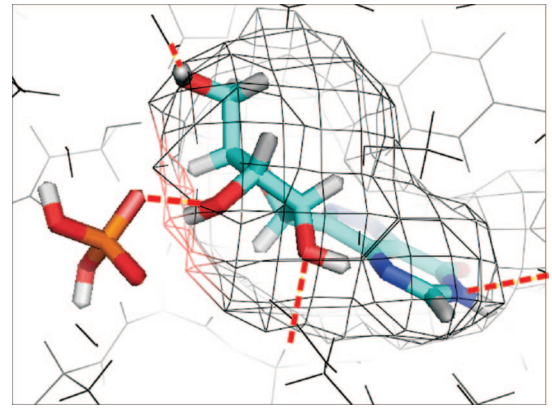

Figure 3. An annotated ligand docked to the PNP receptor with $PO_4H_2^{1-}$ cofactor (shown by sticks) oriented with the oxygen atoms toward the binding site (illustrated by the grid); polar contacts are shown by dashed lines.

becomes worse in comparison with rigid-docking results. For all other receptors, the soft-receptor model leads to a decrease

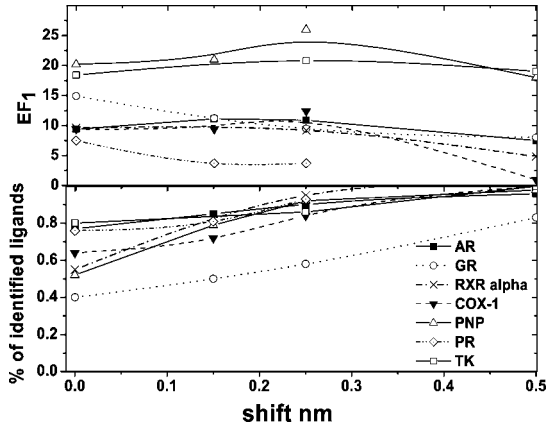

Figure 4. Variation of enrichment factors for several representative targets with the shift value in the soft-docking model (see Soft-Docking section in Results).

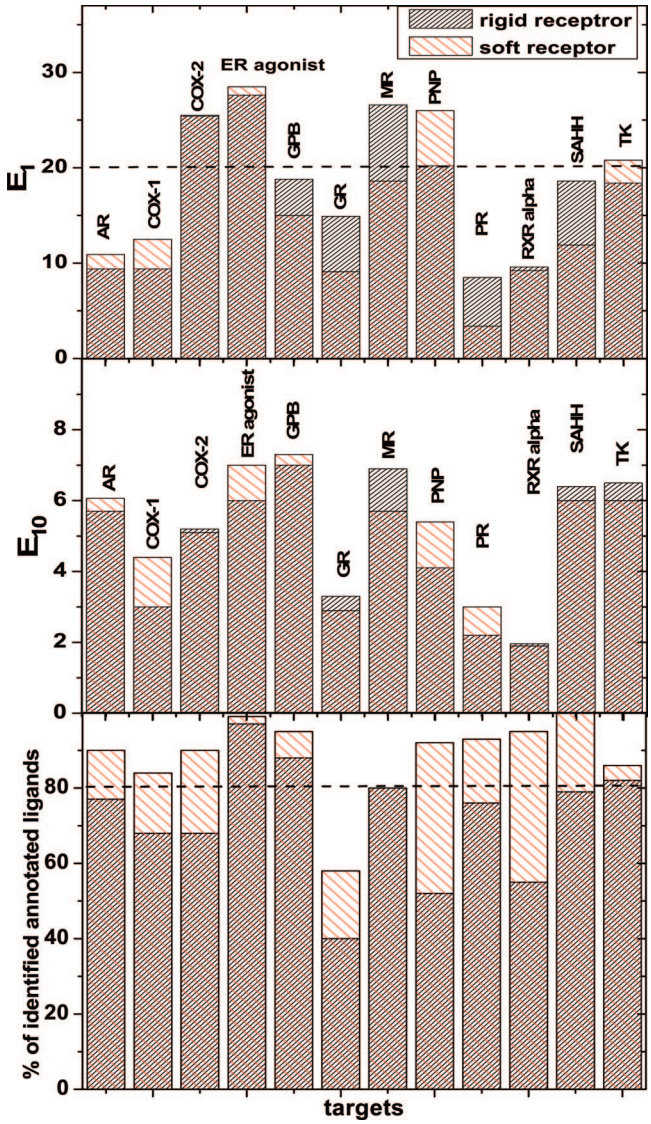

Figure 5. Comparison of enrichment factors of rigid- and soft-receptor models for 12 receptors. The corresponding number of the shifted residues for each receptor is given in Table 2.

of ligand enrichment. Overall we find $EF_1 > 20$ in four cases, as with the rigid-receptor screen and $10 < EF_1 < 20$ for five (previously four) targets. Enrichment at 10% of the top-ranking database shows better results than in rigid-docking for seven targets (Figure 5).

Variation of number of shifted residues may affect both docking and scoring performance of soft docking. This, however, does not change the general tendency (e.g., if only part of residues from "ligand-based" list is used, the enrichment factors attain intermediate values between those for rigid docking and those for soft docking with all residues taken into account—they never improve beyond the date shown in the table).

These results indicate that the soft-receptor reconstruction often leads to a loss of specificity of the binding site and therefore reduces the ability of the method to distinguish between true and false positives. One of the physical reasons for the reduction of selectivity in the soft receptor model is the neglect of an energy penalty for target reconstruction because ligands are permitted to optimize their binding energy in the cavity, without having to carry the energetic cost to create that space. The selection of shifted residues on the basis of docking calculations for a whole set of annotated ligands, can overestimate the degree of receptor reconstruction for each individual ligand and, therefore, additionally decrease method selectivity.

4. Flexible-Receptor Docking. Next we investigated a flexible receptor model in which selected side chains can rotate continuously about single bonds during the docking simulation. As before, the same set of flexible residues is chosen from residues of the "ligand-based" set (see Table 3). We have experimented with various choices of the number of flexible side chains N^f. This number should be chosen as small as possible for reasons of computational efficiency but large enough to describe the most relevant receptor degrees of freedom. The number of flexible residues that lead to the largest EF_1 values (noted hereafter as an optimal set of flexible residues, N^f_{opt}) varies from 3 to 9 for all receptors (except for RXR-α, where $N^f = 11$ leads to optimal scores and SAHH, which shows largest EF_1 at $N^f = 2$).

The fraction of annotated ligands (EF_{100}) that bind to a receptor increases monotonically with the number of flexible residues (see Figure 6). This suggests that, as in the soft-receptor model, side chain flexibility increases the conformational space sufficiently to accommodate larger ligands. Using just 2−6 flexible residues increases the percentage of identified ligands by 20−50%, bringing this value close to 80% for most of the targets (see also Figure 7). For PNP and RXR-α, for example, using just three flexible residues improves EF_{100} by about 40%.

For all but one receptor, the flexible-receptor model increases enrichment rates in comparison with the rigid receptor model (see Figure 7). The scoring performance is now "good" ($EF_1 > 20$) for 8 of 12 targets (in comparison to four in the case of rigid docking) and "medium" ($10 < EF_1 < 20$) in the remaining four cases. In contrast to rigid-receptor docking, where four cases resulted in "poor" screening results, we now find $EF_1 > 10$ for all targets. The most impressive improvement is observed for the RXR-α, AR, and COX-1 targets, where EF_1 more than doubles, going from the rigid to the flexible receptor simulations. For other receptors, EF_1 increases by 10−50% relative to the rigid docking. It is worth noting that the flexible-receptor model also helps to improve docking and screening performance of the hydrophobic ligands of the PR receptor. In this case, however, even increasing of EF_1 by ∼50%, does not allow us to reach a "good" scoring performance because the PR receptor has only few pronounced collision points (residues that often cause sterical clashes). The only exception is the SAHH target, whose EF_1 value does not increase going from rigid- to flexible-

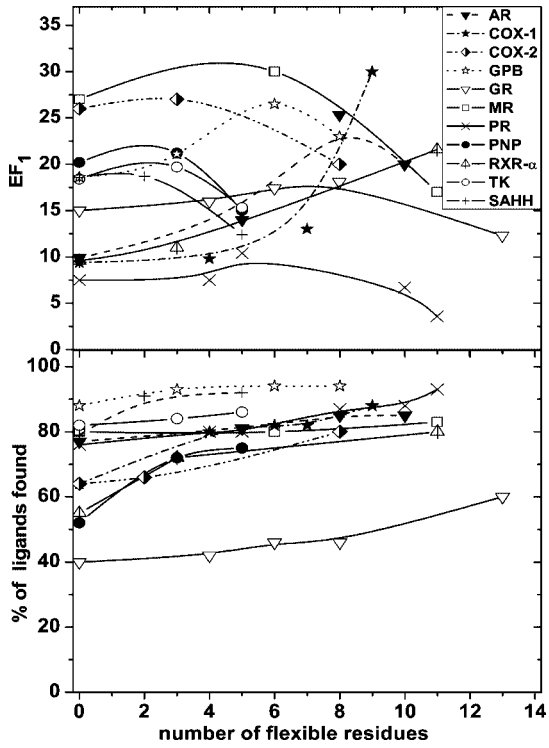

Figure 6. Comparison of enrichment factors for the flexible-receptor model for different sets of selected flexible residues (N^f) (see Table 3 and text).

receptor docking. For the TK receptor, EF_1 is slightly improved but $EF_{2.7}$ is not. In the latter case, the choice of the flexible residues is complicated by the small number of vdW clashes. Moreover, according to ref 54, crystal water molecules may participate in protein−ligand binding for the TK receptor, which introduces an additional difficulties when docking against an apo structure without water.

For both cases of small (for example, by ∼5% for AR) and relatively large (for example, by ∼50% for RXR-α) improvement of EF_{100}, EF_1 almost doubles with respect to rigid-docking. For very large numbers of flexible side chains ($N^f > 6$), enrichment performance is degraded. This may result from sampling problems (the number of steps in the simulation remains fixed) and approximations in the scoring function, which must balance ligand-induced protein energy changes with the binding energy (note: the internal energy change of the protein is included in the binding energy).

IV. Discussion

1. Flexible- vs Soft-Receptor Screening. Figures 4 and 5 summarize the most important results of the soft-docking simulations. The data show that the soft-receptor approximation tends to decrease the enrichment rates (EF_1) despite an increase of the number of known ligands that bind to the receptor. Nevertheless, soft-receptor calculations at small shift ($\Delta \leq 0.25$ nm) may be useful for a first-stage screening: soft docking enables one to find binding conformation of some molecules that do not dock in rigid docking while it eliminates most molecules that will not dock at all. This reduces the size of the database that needs to be considered to ∼20% of its initial size while keeping most true positives in the reduced database. Indeed, as can be seen in Table 2, for 8 of 12 considered receptors, the fraction of identified annotated ligands in the top 20% of the database (that is defined as $EF_{20} \times 20/100$) in the soft-docking model is close to that obtained with the rigid-receptor model against the total database (the value of EF_{100}).

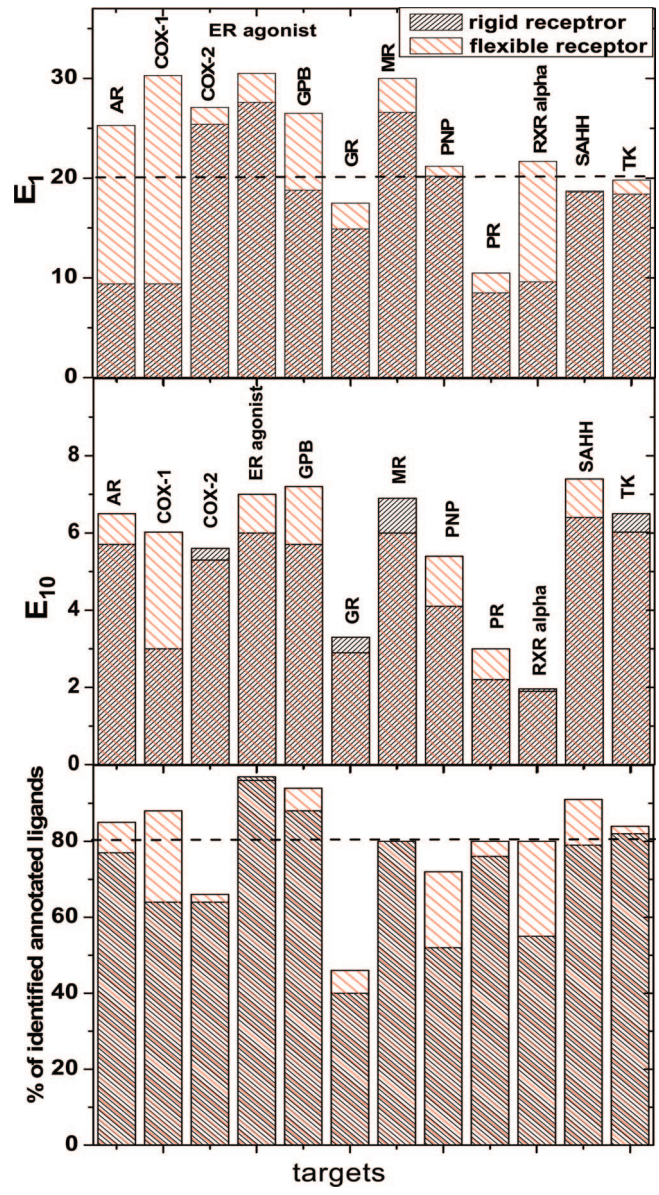

Figure 7. Comparison of enrichment factors of rigid- and flexible-receptor models for all 12 receptors. The number of flexible residues for each receptor is given in Table 2.

Flexible receptor screens not only increase the number of known ligands that bind (EF_{100}) but also lead to a significant improvement of ligand scoring for most targets. Thus, conformational sampling of protein residues that often cause clashes with annotated ligands allows difficult ligands to bind while still discriminating between true and false positives. Figure 8 illustrates for two examples that the distribution of "positives" in the database clearly changes its weight toward smaller energies in flexible-receptor screens. In terms of enrichment, the EF_1 value increases by ∼35% on average from rigid- to flexible-receptor model. Figure 9 illustrates scoring results of rigid- and flexible-receptor docking obtained in the present work in comparison with results for the same database reported in ref 43 (docking was performed with DOCK 3.5.54). Good enrichment ($EF_1 > 20$), obtained for 8 of 12 receptors (first two columns in Figure 9), means that more than 54% of molecules selected in the top-ranking 1% of the database are known ligands for these targets. In fact, in the four most

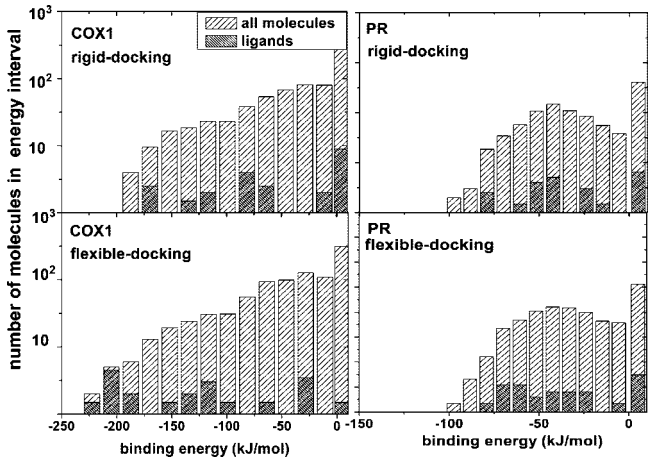

Figure 8. Number of known ligands (dense pattern) and total database size (sparse pattern) vs binding energy found in rigid-docking (upper plots) and flexible-docking (lower plots) simulations for the COX-1 and PR protein targets. Both targets show poor enrichment in rigid-docking; high and medium (COX-1 and PR, respectively) enrichment is achieved in flexible-receptor docking.

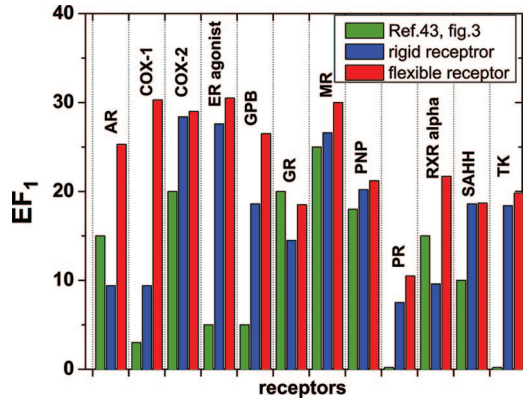

Figure 9. Comparison of the enrichment factor EF_1 for rigid- and flexible-docking in this investigation with the data reported in ref 43.

successful cases (COX-1, COX-2, ER-agonist, and MR) this value is even above 81%. For the remaining four targets with medium enrichment ($10 < EF_1 < 20$), the fraction of annotated ligands in the top 1% of the database is above 27%. There are a variety of reasons that may account for the difference in the performance of the two methods: different docking algorithms were used as well as different scoring functions (no explicit H-bond term in DOCK, but a solvation term instead) and slightly different receptor models (treatment of cofactors and crystal water).

As an alternative to enrichment factors, comparative screening performance of rigid-, soft-, and flexible-docking can be assessed using the BEDROC metric values listed in Table 2. BERDOC values interpolate, for the chosen parameter $\alpha = 0.5$ of the weighting function in eqs 1 and 2, early enrichment data of EF_1 and $EF_{2.7}$. For all receptors except TK, where BERDOC stays stable, this analysis agrees completely with the enrichment values in support of our conclusions: flexible-receptor screens outperform rigid- and soft-receptor models.

2. Importance of Polar Interactions and Solvent Effect. To rationalize these results, we have investigated the influence of polar interactions on docking performance. To this purpose we have summarized the average number of polar contacts per annotated ligand (N_p) in Table 3 for all considered targets. The docking performance, E_{100}, shows little correlation with the number of polar contacts because the capability of the method to find some binding conformation mainly depends on the binding pocket size. The worst enrichment rate (EF_1) is observed for the PR target that, in comparison with other receptors, has the most hydrophobic binding site, with only one or even no polar contacts between known ligands and receptor. This indicates that the explicit hydrogen-bonding term in the scoring function may give an important differential contribution to the binding energies. We find that top-scoring ligands have more polar contacts than the average ligand from the database. On the other hand, enrichment rates for highly polar receptors (as for example, GPB) do not tend to be higher than for receptors with only few polar contacts (for example, MR). Therefore, the specific hydrogen bonding potential seems to contribute an essential component to the binding energy, but it does not seem to overemphasize hydrogen bonding in general.

We have investigated the solvation effects by considering simple SASA (solvent accessible surface area) based corrections to the screening function considered in this investigation for several receptors but found only insignificant changes in the enrichment rates. This may result from inaccuracies in the models used but also from the particular construction of the DUD decoy sets; in the databases considered here, decoys are constructed from known ligands to share the physical characteristics of the original molecule, but in a different topology. Many solvent models, including the SASA based models we have considered, compute the solvation energy as the sum of fragments. If ligands and decoys arise by permuting the same fragments, the change in solvation energy will be small. As a result, differential effects arising from solvation effects may be particularly small for the DUD databases.

3. Flexibility of Polar vs Hydrophobic Groups. We have also investigated the relative influence of the flexibility of polar/apolar groups on the performance of the receptor model. Some studies suggest that polar groups tend not to assume novel conformations upon ligand binding, while hydrophobic sidechains may move to accommodate the ligand.[39,53] Our results show that flexibility of polar groups is as important as that of the hydrophobic residues. Although the selection criteria of flexible residues in the present method are based on vdW energy only, both hydrophobic and polar residues are included in the flexible residue set (see Table 3). As a result, movement of both types of residues is important for binding mode optimization. One example is shown in Figure 1C, where polar side chains clearly move to form favorable interactions with a ligand. It is important to note that the formation of these favorable interactions depends on the continuous side chain flexibility in the model. The side chain dihedral angles must stay within a few degrees of the optimal position to recover the full hydrogen bonding energy. Accurate description of hydrogen bonding is an important prerequisite for selectivity. Therefore, exclusion of polar amino acids from the set of flexible residues, as proposed in ref 37, may lead to serious degradation of screening performance. However, we see no evidence that large polar amino acids are more flexible than aromatic ones, as suggested in ref 39.

4. Advantages and Limitations of Flexible-Receptor Screening. The present work is one of the first investigations where ligand-induced receptor reconstruction is simulated on the basis of only one apoprotein structure. The proposed approach, therefore, improves upon static ensemble-docking methods because its applicability does not depend on the number and quality of available holo structures, which are unavailable for many systems of biomedical interest. Instead, the receptor

reorganization energy is included on the same footing as the ligand binding energy. The present method thus incorporates one important contribution to the overall binding energy not present in other methods (see for example refs 21−24), which may contribute to the improved enrichment rates.

The present flexible-receptor model samples receptor conformations on the basis of chemical/physical properties of the training set of experimentally known ligands. Regarding the input basis, this method is similar the Raptor receptor-modeling approach,[36] which makes use of a ligand training set with known affinity to a specific protein to build a physicochemical field and receptor topology for each individual ligand. It is, however, difficult to compare the two methods because quite different examples and criteria of method validation are used. Published applications of this method[36,54] succeeded in the accurate prediction of affinities for sets of selected ligands but did not focus on the discrimination of decoys with similar physical/chemical properties. Our results, covering a diverse set of proteins, demonstrate that the FlexScreen flexible-docking approach is highly transferable among protein−ligand complexes.

The docking protocol will benefit from further improvements in the algorithm for the selection of the flexible side chains, in particular for cases where no or little experimental information is available. As one possible way to define flexible residues in the absence of the experimental data, we have experimented with a protocol that uses decoys in a precursor rigid-receptor screen to create a "decoy-based" list of flexible residues to the "ligand-based" one. One can see from Table 4 that the top-scoring residues of the "decoy-based" list share many common elements with the "ligand-based" one. Specifically, the number of "correctly" found flexible residues among top-ranking 5 and 10 ones is similar to that obtained from the small set of annotated ligands ($N_{AL} = 20$). This result suggests that DUD decoys (i.e., molecules with physical properties similar to known ligands), or the library as a whole, may be used as a "training set" for prediction of possible flexible residues. Screens performed with a five-residue decoy-based set of the AR receptor achieved similar enrichment ($EF_1 = 17.7$ with the "decoy-based" list as opposed to $EF_1 = 14$ using the "ligand-based" list; see Methods). We also generated a list of potentially flexible residues using only a random subset of 50% of the available ligands to test the dependence on the "training set", we again find 4 of the 5 top-ranking residues as before (MET749 is replaced by ARG752). Using this five-residue ligand-based list for the selection of flexible residues, we obtain an $EF_1 = 16.4$, indicating that the selection procedure is quite robust under changes of the protocol. We will perform further investigations to establish optimal selection protocols for the flexible side chains.

The most important finding of this investigation is summarized in Figure 10, which shows the number of known ligands (hits) among the 50 top-scoring ligands for each receptor. The data demonstrate a high population of true hits in the absolute top positions of the receptor-specific databases, comprising up to 13000 molecules each. This summary reflects the hit rate in a thought experiment where potential ligands would be synthesized without further analysis solely based on the predictions of the docking algorithm. It is encouraging that for 11 of the 12 receptors the top-ranking molecule is a "hit". Synthesizing 5/50 molecules for each receptor would generate 4/17 "hits" on average. These data demonstrate that accurate enrichment simulations incorporating receptor flexibility have matured to generate molecules with at least medium affinity with high certainty for a wide array of receptors of present-day interest.

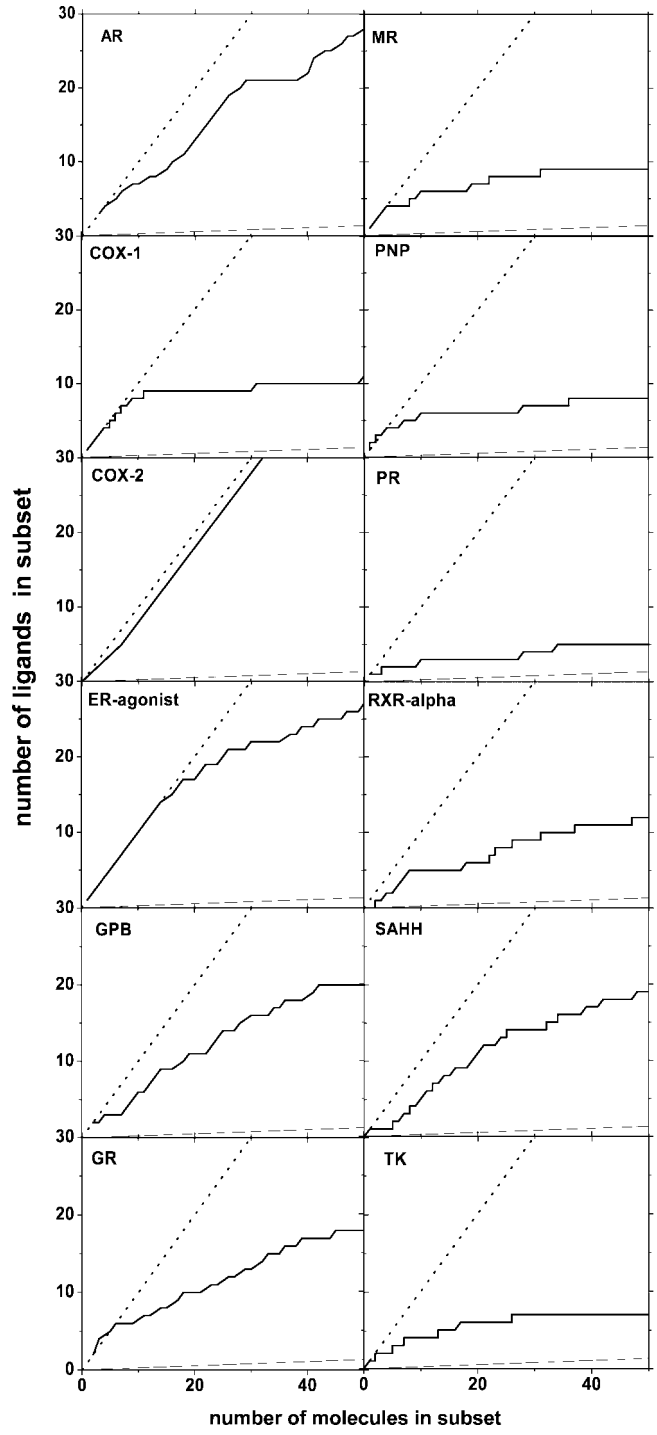

Figure 10. Absolute number of hits (known ligands) vs the total number of molecules in the rank-ordered subset of the top 50 molecules of each database in flexible-receptor docking calculations. Dotted lines show the result for perfect enrichment, the full lines the cumulative number of known ligands found in the screen, the dashed lines show random curves.

We also note that a different protocol for the selection of flexible residues should be employed for very open receptor pockets. Preliminary investigations for such receptors reveal that both the ligand-based or decoy-based approaches have difficulty to rationally select potentially flexible residues as essentially all ligands can somehow be accommodated in the vicinity of binding site. In practice, we believe this is not a problem because visual inspection of the binding site will guide a receptor-specific selection of the residues. As a word of caution, we do not believe

that there is a fully automated protocol for flexible side chain selection that will work for all receptors: visual inspection of the residues in the list of potentially flexible residues is always recommended. Choosing all side chains in the vicinity of the binding site is computationally very demanding, thus judicious inspection of the proposed list of potentially flexible side chains is probably best in the foreseeable future.

If a sufficient number of ligands are known, Figure 6 suggests an interesting possibility to select the optimal subset in a quasi-automated fashion: we note that the enrichment rate EF_1 has near optimal values as soon as the docking rate EF_{100} saturates with the number of flexible residues. Based on this observation, which holds even when EF_{100} never reaches 100, one should perform simulations for the known ligands only (which is inexpensive) with increasing sets of flexible residues chosen from the rank-ordered ligand-based list of potentially flexible residues until EF_{100} saturates and use this subset for a full screen of the entire database.

V. Conclusions

We have carried out a comparative analysis of docking and scoring performance of rigid-, soft- and flexible-receptor models using a biophysical forcefield-based docking approach. The latter two models accommodate ligand-induced protein reconstruction by moving residues or rotating of receptor side chains that are most often involved in steric clashes between the protein and known ligands.

We find significant limitations of rigid receptor models, which for some receptors fail to bind even 50% of the known ligands to the apostructure of the protein. Obviously such ligands can never be successfully identified in subsequent enrichment studies. The enrichment rate did not correlate with docking performance and was good only for four receptors. For the other receptors enrichment rates remain poor, in high correlation with a previous study of the same database.[43]

Both soft- and flexible-receptor approximations increase the conformational space available to potential ligands and thus increase for 11/9 (for soft and flexible receptor, respectively) the number of known ligands, which bind to their respective receptors, to more than 80% (compared to the targets in rigid docking). However, the soft-docking model fails to significantly increase the enrichment rates because not only known ligands but also "false positives" dock with equal efficiency.

The flexible receptor model shows a significant improvement in enrichment rates with respect to rigid docking for 11 of 12 targets: on average 62% of the molecules in the top percent of the database are known ligands. Considering the top 10/50 molecules in the database the FlexScreen induced-fit approach selects a high number (4/17 on average) of known ligands in the top-scoring ligands of the screen. This high rate of true positives indicates that accurate present-day in silico screening methods have matured to a high degree of selectivity for databases of moderate size (~10000 molecules). The performance of an in silico screen depends on a number of parameters, including the similarity of the ligands to the decyos (or inactives) in the database. As a result, for sets where actives and ligands are more similar than in DUD, screening performance could be worse, but in sets where actives and ligands are less similar, performance could be even better. For this reason, the absolute performance of the screening method will vary with the system under consideration.

These results suggest that scoring performance can be notably improved by sampling the conformations of several side chains of an active site with continuously rotatable single bonds. This algorithm does not require empirical assumptions for finding possible soft points. If only a limited number of known ligands is available, compounds with similar physical properties, such as those constructed by the DUD procedure, may be used to explore the active site and create a list of soft spots. Obviously, if several protein structures are available, these may be used for refinement of the list of flexible residues.

The soft docking model may be used for prescreening because it can reduce database size by up to 80% without loss of important compounds. It should be noted that this procedure may be even more effective in prescreening standard databases because the decoys in the DUD database are designed to match the physical characteristics of the known ligands rather well. If the database contains many molecules without any correlation to ligands with high affinity, the success of prescreening may even be better.

Acknowledgment. This work was supported by the Deutsche Forschungsgemeinschaft (WE 1863/14-1). We also acknowledge the use of the computational facilities at the Barcelona Supercomputing Center. The program is available from the authors upon request (wenzel@int.fzk.de or http://www.fzk.de/biostruct).

References

(1) Hopkins, A. L.; Groom, C. R. The druggable genome. *Nat. Rev. Drug Discovery* **2002**, *1*, 727–730.
(2) Bolten, B. M.; DeGregorio, T. From the analyst's couch. Trends in development cycles. *Nat. Rev. Drug Discovery* **2002**, *1*, 335–336.
(3) Ghosh, S.; Nie, A.; An, J.; Huang, Z. Structure-based virtual screening of chemical libraries for drug discovery. *Curr. Opin. Chem. Biol.* **2006**, *10*, 194–202.
(4) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.
(5) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
(6) Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev.* **2003**, *2*, 527–541.
(7) Lazaridis, T.; Masunov, A.; Gandolfo, F. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins* **2002**, *47*, 194–208.
(8) Mohan, V.; Gibbs, A. C.; Cummings, M. D.; Jaeger, E. P.; DesJarlias, R. L. Docking: successes and challenges. *Curr. Pharm. Des.* **2005**, *11*, 323–333.
(9) Rabal, O.; Schneider, G.; Borrell, J.; Teixido, J. Structure-Based Virtual Screening of FGFR Inhibitors: Cross-Decoys and Induced-Fit Effect. *Biodrugs* **2007**, *21*, 31–45.
(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 neuraminidas. *J. Comput.-Aided Mol. Des.* **1999**, *13*, 547–562.
(11) Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.* **2007**, *371*, 1118–1134.
(12) Carlson, H. A.; McCammon, J. A. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, *57*, 213–218.
(13) Ahmed, A.; Kazemi, S.; Gohlke, H. Protein flexibility and mobility in structure-based drug design. *Front. Drug Des. Discovery* **2007**, *3*, 455–475.
(14) Teodoro, M. L.; Kavraki, L. E. Conformational flexibility models for. the receptor in structure based drug design. *Curr. Pharm. Des.* **2003**, *9*, 1635–1648.
(15) Teague, S. J. Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discovery* **2003**, *2*, 527–541.
(16) Lakomek, N. A.; Carlomagno, T.; Becker, S.; Griesinger, C.; Meiler, J. A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *Proteins: Struct., Funct., Bioinform.* **2006**, *65*, 538–548.
(17) Zhao, Y.; Sanner, M. F. FLIPDock: docking flexible ligands into flexible receptors. *Proteins* **2007**, *68*, 726–737.
(18) Meiler, J.; Baker, D. ROSETTALIGAND: protein−small molecule docking with full side-chain flexibility. *Proteins.* **2006**, *65*, 538–548.
(19) (a) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. Flexible ligand and receptor docking with a continuum solvent model and soft-core energy

function. *J. Comput. Chem.* **2003**, *24*, 1637–1656. (b) Jiang, F.; Kim, S. H. "Soft docking": matching of molecular surface cubes. *J. Mol. Biol* **1991**, *219*, 79–102.
(20) Ferrari, A. M.; Wei, B. O.; Costantion, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076–5084.
(21) Cavasotto, C. N.; Kovacs, J. A.; Abagyan, R. A. Representing receptor flexibility in ligand docking through relevant normal modes. *J. Am. Chem. Soc.* **2005**, *127*, 9632–9640.
(22) Barril, X.; Morley, S. D. Unveiling the full potential of flexible receptor docking using multiple crystallographic structures. *J. Med. Chem.* **2005**, *48*, 4432–4443.
(23) Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. Testing a flexible-receptor docking algorithm in a model binding site. *J. Mol. Biol.* **2004**, *337*, 1161–1182.
(24) Huang, Sh.-Y.; Zou, X. Ensemble docking of multiple protein structures: considering protein structural variations in molecular docking. *Proteins: Struct., Funct., Bioinform.* **2007**, *66*, 399–421.
(25) Bowman, A. L.; Lerner, M. G.; Carison, H. A. Protein Flexibility and Species Specificity in Structure-based drug discovery: dihydrofolate reductase as a test system. *J. Am. Chem. Soc.* **2007**, *129*, 3634–3640.
(26) 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.
(27) Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **1992**, *255*, 178–183.
(28) Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol.* **1994**, *235*, 345–356.
(29) Desmet, J.; Wilson, I. A.; Joniau, M.; De Maeyer, M.; Lasters, I. Computation of the binding of fully flexible peptides to proteins with flexible side chains. *FASEB J.* **1997**, *11*, 164–172.
(30) Frimurer, T. M.; Peters, G. H.; Iversen, L. F.; Andersen, H. S.; Moller, N. P.; Olsen, O. H. Ligand-induced conformational changes: improved predictions of ligand binding conformations and affinities. *Biophys. J.* **2003**, *84*, 2273–2281.
(31) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. Protein-ligand docking: current status and future challenges. *Proteins* **2006**, *65*, 15–26.
(32) Kellenberger, E.; Rodrigo, J.; Muller, D.; Rognan, D. Comparative evaluation of eight docking tools for docking and Virtual Screening accuracy. *Proteins* **2004**, *57*, 225–242.
(33) Fischer, B.; Merlitz, H.; Wenzel, W. Increasing Diversity in in Silico Screening with Target Flexibility. *Lect. Notes Comput. Sci.* **2005**, *695*, 186–197.
(34) (a) Schnecke, V.; Swanson, G.; Getzoff, E.; Tainer, J.; Kahn, L. A. Screening a peptidyl database for potential ligands to proteins including side-chain flexibility. *Proteins* **1998**, *33*, 74–87. (b) Schnecke, V.; Kuhn, L. A. Virtual screening with solvation and ligand-induced complementarity. *Persp. Drug Discovery Des.* **2000**, *20*, 171–190.
(35) Zavodszky, M. I.; Kuhn, L. A. Side-chain flexibility in protein−ligand binding: the minimal rotation hypothesis. *Protein Sci.* **2005**, *14*, 1104–1114.
(36) Lill, M. A.; Vedani, A.; Dobler, M. Raptor: combination duel-shell representation, induced-fit simulation, and hydrophobicity scoring in receptor modeling. Application toward the simulation of structurally diverse ligand sets. *J. Med. Chem.* **2004**, *47*, 6174–6186.
(37) 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.
(38) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-chain flexibility in proteins upon ligand binding. *Proteins* **2000**, *39*, 261–268.
(39) 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.
(40) Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
(41) (a) Merlitz, H.; Wenzel, W. Comparison of stochastic optimization methods for receptor−ligand docking. *J. Chem. Phys. Lett.* **2002**, *362*, 271–277. (b) Merlitz, H.; Burghardt, B.; Wenzel, W. Application of the stochastic tunneling method to high throughput database screening. *J. Chem. Phys. Lett.* **2003**, *370*, 68–73.
(42) Merlitz, H.; Wenzel, W. Impact of receptor conformation on in silico screening performance. *J. Chem. Phys. Lett.* **2004**, *390*, 500–505.
(43) Huang, N.; Schoichet, B. K.; Irwing, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789−6801, http://blaster.docking.org/dud/.
(44) Morris, G. M.; Goodsell, D. S.; Halliday, R.; 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. *Comput. Chem.* **1998**, *19*, 1639–1662.
(45) Jorgensen, W. L.; McDonald, N. A. Development of an all-atom force field for heterocycles. Properties of liquid pyrrole, furan, diazoles, and oxazoles. *J. Mol. Struct.* **1997**, *424*, 145–155.
(46) (a) Wenzel, W.; Hamacher, K. Stochastic Tunneling Approach for Global Minimization of Complex Potential Energy Landscapes. *Phys. Rev. Lett.* **1999**, *82*, 3003–3007. (b) Merlitz, H.; Wenzel, W. High Throughput in Silico Screening against Flexible Protein Receptors. *Lect. Notes Comput. Sci.* **2004**, *3045*, 465–472.
(47) Fischer, B.; Basili, S.; Merlitz, H.; Wenzel, W. Accuracy of binding mode prediction with a cascadic stochastic tunneling method. *Proteins: Struct., Funct., Bioinform.* **2007**, *68*, 196–204.
(48) *Molecular Operating Environment (MOE) version 20003.02*; Chemical Computing Group Inc.: Montreal, 2003.
(49) Wang, J. M.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
(50) DeLano, W. L. *The PyMOL Molecular Graphics System 2002*; DeLano Scientific: Palo Alto, CA, 2002; http://www.pymol.org.
(51) Truchon, J.-F.; Bayly, Ch. I. Evaluating virtual screening methods: good and bed metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.
(52) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895–2907.
(53) Davis, A. M.; Teague, S. J. Hydrogen bonding, hydrophobic interactions, and failure of the rigid receptor hypothesis. *Angew. Chem., Int. Ed.* **1999**, *38*, 736–749.
(54) (a) Lill, M. A.; Winger, F.; Vedani, A.; Ernst, B. Impact of induced fit on ligand binding to the androgen receptor: a multidimensional QSAR study to predict endocrine-disrupting effects of environmental chemicals. *J. Med. Chem.* **2005**, *48*, 5666–5674. (b) Vedani, A.; Dobler, M.; Lill, M. A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.* **2005**, *48*, 3700–3703.

JM800217K