JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

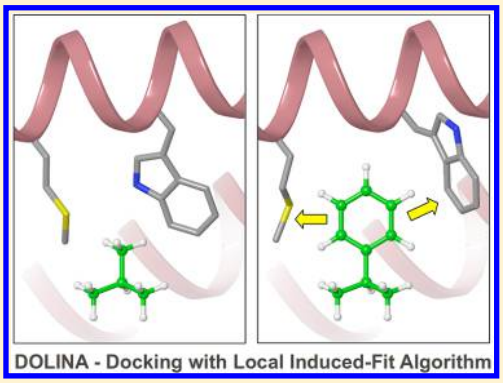
DOLINA − Docking Based on a Local Induced-Fit Algorithm: Application toward Small-Molecule Binding to Nuclear Receptors

Martin Smieško*

Department of Pharmaceutical Sciences, University of Basel, Klingelbergstrasse 50, 4056 Basel, Switzerland

Supporting Information

ABSTRACT: Docking algorithms allowing for ligand and − to various extent − also protein flexibility are nowadays replacing techniques based on rigid protocols. The algorithm implemented in the Dolina software relies on pharmacophore matching for generating potential ligand poses and treats associated local induced-fit changes by combinatorial rearrangement of side-chains lining the binding site. In Dolina, ligand flexibility is not treated internally, instead a pool of low-energy conformers identified in a conformational search is screened for extended binding-pose candidates. Grouping rearranged residues in sterically independent families and side-chain conformer clustering are employed to achieve efficient use of the computational resources along with a good accuracy of the generated poses. Dolina was applied toward docking of small-molecule ligands to three different nuclear receptor ligand binding domains for which in total 18 high-resolution crystal structures were used as reference. The selected nuclear receptors feature a deeply buried ligand-binding site where local induced-fit is to be expected, particularly for receptor antagonists. For each receptor, a crystal structure with a cocrystallized small steroid ligand (template) was chosen as a target system, to which several synthetic ligands of different sizes were docked. Poses within an *RMSD* of 2.0 Å from the crystal reference pose were generated in 91% of the cases. In 28%, the pose with the lowest *RMSD* to the reference pose was ranked as the top one, and in 76% it was ranked among the top five poses. Detailed descriptions of the docking algorithm and observed results are included. Dolina is available free of charge for academic institutions.



INTRODUCTION

Computational chemistry and molecular modeling techniques have become an integral part of modern drug design. Based on the three-dimensional structure of the macromolecular target, molecular docking can help in designing both selective and affine ligands. Such small molecules may, upon binding to the protein, trigger changes in the 3D structure of the macromolecule associated with some kind of biologically relevant response or effect (e.g., nuclear receptor: dissociation of repressor molecule, dimerization). Structural changes induced upon ligand binding − induced-fit − can be of different magnitude. In this account, the term "local induced fit" refers to a rearrangement of side-chains at and near the binding site, without affecting the protein backbone conformation or the overall shape of the macromolecular target.

Molecular docking methods treating the protein target as a rigid body can be used with some limitations for optimizing congeneric series of ligands with a conservative substitution pattern or for identification of new similarly sized scaffolds. However, if a medicinal chemist − aiming at increasing selectivity or affinity of a lead compound − would like to exploit new binding site pockets composed of flexible side-chains, a docking algorithm supporting induced-fit should be employed.[1] Otherwise there would be a risk that quantitative structure−activity relationships (QSAR) analysis could be misleading because of (50−70%) incorrect poses.[2] Simple geometry optimization might not correctly rearrange side-chains as the optimized structure typically reaches the closest local minimum. Molecular dynamics simulations have the potential to address small as well as large-scale conformational changes[3,4] but usually require a considerable amount of time for setup and run, which render this technique rather impractical for screening larger counts of ligand candidates.

The importance of induced-fit for molecular docking can be underlined by the fact that all major software[5,6] nowadays offer some kind of treatment of protein flexibility,[7−9] thus opening the possibility for protein rearrangement upon ligand binding. Approaches addressing the protein flexibility can be classified based on the underlying philosophy to the following: soft potentials,[10,11] rotamer exploration,[12−14] multiple protein structures and ensemble docking,[2,15] and hybrid methods. Each of the approaches has it strengths and weaknesses regarding accuracy, computational costs, and reliability; that is why several research groups have focused their attention to development and application of specialized docking algorithms and protocols.[16−23]

Received: February 8, 2013
Published: June 3, 2013

dx.doi.org/10.1021/ci400098y | *J. Chem. Inf. Model.* 2013, 53, 1415−1423

The main objective of this study was to investigate whether poses of structurally diverse ligands observed in crystal structures, which were shown to cause the rearrangement of side-chains at protein binding sites, can be identified based on pharmacophore matching and combinatorial solution of induced-fit. The idea behind employing the pharmacophore matching for pose generation is that the binding site usually does not completely rearrange upon binding of even chemically diverse ligands, and some protein pharmacophore features remain at a constant relative orientation, e.g. features assigned to side-chains stabilized by several interactions or to rigid backbone elements. Such pharmacophore features can serve as a "navigation" for positioning of different low-energy ligand conformers in the binding site. After a valid pose (devoid of close contacts with the protein backbone) is obtained a combinatorial scan of energetically favorable side-chain rotamers in the ligand's vicinity is performed to ensure optimal interaction with the pose and other side-chains. The combinatorial explosion associated with exploring a large number of degrees of freedom is tackled by grouping residues to sterically independent families and by clustering side-chain rotamers.

MATERIALS AND METHODS

Three nuclear receptors − the androgen (AR), mineralocorticoid (MR), and progesterone (PR) receptor − were chosen as targets for benchmarking the performance of Dolina. Nuclear receptors are suitable targets for simulating induced fit, because their binding site is located in the interior of the protein, and consequently the binding of voluminous ligands is likely to be associated with changes of the binding site topology. For each of the receptors, high resolution crystal structures with ligands of various size are available from the PDB.[24] As it is desirable that the docking algorithm provides satisfactory results for both small and large ligands, with the latter likely triggering induced-fit changes, molecules of different size and constitution were included in the data set. The natural ligands, testosterone for the AR and progesterone for the PR, and a drug molecule prednisone for the MR were chosen as template molecules. Details on the ligands and proteins are given in Table 1.

Preparation of the Protein Structures. For each target protein a crystal structure cocrystallized with a small natural (AR and PR) or synthetic (MR) ligand was selected (template complex; Table 1, highlighted in bold). Each of the three template complexes was processed using the Protein Preparation Wizard script within the Schrodinger Maestro modeling environment.[26] The remaining complexes (reference complexes; Table 1) were rigidly aligned based on the backbone atoms to the template complex, and the $RMSD_{backbone}$ of the best fit was calculated. The structure of the reference complexes was not modified in any manner.

For the ligand molecules, the quantum chemical charges were calculated using AMSOL[27] and charge model 1 (CM1).[28] Before minimizing each of the ligand−protein complexes using the Yeti force-field[29] to remove the internal strain associated with addition of the hydrogen atoms, the optimal orientation of hydrogen bonded residues was automatically found using Yeti.[30]

Ligand Preparation. Ligands cocrystallized with the AR (12), the MR (2), and the PR (4) were prepared for input to Dolina (Figures S1−S3, Supporting Information). In order to address ligand flexibility, a pool of low-energy conformers for each docked molecule was obtained in conformational analysis using two different force-fields − OPLS-2005 and AMBER*[31] − as implemented in MacroModel.[32] An energy window of 10 kcal/mol was used, and at maximum 100 conformers were saved. The individual conformers were considered different if the *RMSD* calculated on their heavy atoms was higher than 0.5 Å. For each ligand, 5000 iterations of the mixed torsional/low-mode sampling were performed, and the generated conformers were minimized into stable structures. Implicit solvent conditions (water) were used. In order to better explore the impact of conformer geometry on the docking, an additional set of conformers was obtained by optimizing the OPLS-2005 conformers in AMSOL using the SM5.4A solvation model,[33] the AM1 level of theory,[34] and the water as solvent.

Table 1. Protein−Ligand Crystal Structures Used for Benchmarking

PDB ID	ligand ID	ligand volume [Å^3][a]	$RMSD_{backbone}$ [Å]
Androgen Receptor			
2AM9	**TES**	**993**	**n/a**
2AMB	17H	1050	0.27
2AX9	BHM	863	0.33
2AXA	FHM	1093	0.38
2HVC	LGD	945	0.38
2IHQ	LG7	923	0.49
2NW4	8NH	871	0.55
2PNU	ENM	1367	0.38
3B66	B66	1142	0.39
3B67	B67	1117	0.35
3B68	B68	1228	0.47
3G0W	LGB	990	0.40
3RLJ	RLJ	1126	0.42
Mineralocorticoid Receptor			
2AAX	**PDN**	**1019**	**n/a**
3VHU	SNL	1212	0.56
3VHV	LD1	1015	0.50
Progesterone Receptor			
1A28	**STR**	**1034**	**n/a**
1SR7	MOF	1351	0.57
3HQ5	GKK	1105	0.56
3KBA	WOW	1168	0.51
4APU	A2K	1552	0.59

[a]Ligand volumes were calculated by QuikProp.[25]

Rigid Docking of Crystal Ligand Conformers to Flexible Protein. The crystal conformers of the ligands were extracted from their aligned protein−ligand complexes and pseudorandomly oriented in the space by converting their coordinates to the Z-matrix format and back to the Cartesian coordinates. The ligands were then docked rigidly into the target protein, and the *RMSD* toward reference pose was calculated. No changes to ligand conformers were allowed; however, protein side-chains could be rearranged to accommodate the ligand in the binding site.

Flexible Docking of Ligands to Flexible Protein. Up to 100 low-energy conformers identified in the conformational search were used as input for Dolina. The conformational search procedure automatically modifies the orientation and the geometry of the ligand and thus removes all potentially biasing information inherited from its original crystal structure. Therefore, no additional spatial operation was necessary. In Dolina, only the positions of the rotatable polar hydrogen atoms (−SH and −OH groups) are optimized.

THE DOLINA DOCKING ALGORITHM

The flowchart of the docking algorithm is depicted in Figure 1.

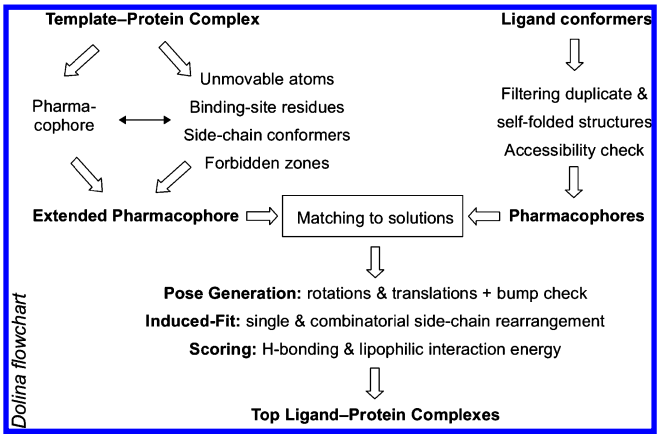


Figure 1. Overview of the task flow of the Dolina docking algorithm.

Conformers of the ligand to be docked are analyzed for chemical correctness, and duplicate as well as self-folded conformers within the low-energy conformer pool are filtered out. Our analysis of 4451 druglike ligands (with a molecular weight of 150−600) retrieved from high resolution crystals (resolution ≤1.6 Å) in the PDB showed that cocrystallized poses are almost exclusively represented by extended conformers devoid of intramolecular π-stacking or hydrophobic interactions. To prioritize such conformers, the average number of closely neighboring nonbonded heavy atoms (separated by at least 5 bonds, distance shorter than the sum of their van der Waals radii +1.0 Å cutoff) and its standard deviation are calculated on the whole conformer pool. Self-folded conformers, indicated by a higher number of neighboring nonbonded atoms compared to the average corrected with a cutoff value (0.25 * standard deviation; minimum 0.1), are filtered out. Besides ensuring that extended conformers are used, this procedure also decreases the number of possible solutions and generated poses, which have to be scored.

Pharmacophores. For each valid ligand conformer to be docked, a pharmacophore is constructed. In Dolina, hydrogen-bond donors and acceptors, rings, lipophilic substituents (based on their volume − small: e.g. $-CH_3$, medium: e.g. $-CHCl_2$, big: e.g. *tert*-butyl), and π-systems as well as carbon linker atoms are the features included in the pharmacophore. For hydrogen-bond donors and acceptors, the accessibility check is performed by placing a probe with a radius of a counterpart atom in the optimal position. For donors, an oxygen atom-sized probe is placed at an optimal distance along the hydrogen bond extension vector; for acceptors hydrogen atom-sized probes are placed at an optimal distance along the lone-pair vectors. If a steric bump with any atom of the ligand molecule and the probe is identified there, the corresponding pharmacophore feature is deactivated. Additionally, for ring- and lipophilic group-based pharmacophore features a vector property is generated (a ring plane normal with origin at the ring centroid or a mass-averaged extension vector with origin at the carrier atom for lipophilic groups, respectively).

Template. To identify the binding site Dolina requires a small ligand molecule − referred to as the template − provided along with the target protein structure (typically, the cocrystallized ligand). For the template molecule, a pharmacophore is generated based on the single supplied pose. All identified hydrogen bonding-related pharmacophore features are inspected for the existence of a counterpart at the protein. If no counterpart is found within a given radius, the pharmacophore feature is deactivated because it is not directly involved in a stabilizing intermolecular interaction and might thus jeopardize the pharmacophore matching process.

Target Protein Structure. The protein structure is thoroughly analyzed and found cofactor molecules (e.g., heme), backbone atoms including C_β atoms, and entire proline and bridged cysteine residues are marked as unmovable. Water molecules forming at least two hydrogen bonds with unmovable partners (e.g., bridging two parts of the backbone) or at least three hydrogen bonds to any protein partners and having the total hydrogen bonding energy more favorable than −6.0 kcal/mol are marked as structural, i.e. unmovable.

Residues having a side-chain atom within a given distance (by default 4.0 Å) from any atom of the template ligand are marked as binding-site residues. In the next step, the side-chains of all binding site residues are subjected to a systematic conformational analysis. A search at torsion angles in ±30° range (10° steps) around the most preferred χ-values (gauche+, trans, gauche−) is performed, as for side-chains present in the binding site a slight deviation to standard torsions values are allowed because of the ligand presence. Bump-free side-chain conformers (no steric clashes to unmovable atoms or residues outside of the binding site) are stored. For each binding site residue, the stored conformers are grouped into clusters based on their *RMSD*, so that each cluster is represented by a conformer having an *RMSD* lower than 0.5 Å to the maximum number of the similar conformers.

Sterically forbidden regions − a useful addition to the 3D pharmacophore − are searched around each binding site residue. They indicate portions of space within the binding site, which are permanently occupied by some side-chain atom, despite residue flexibility. Such regions are most frequently found near the position of side-chain γ-atoms with restricted mobility, e.g. due to vicinity of a backbone loop, and the presence of any docked ligand atom should be avoided there. In Dolina, forbidden regions are represented by dummy atoms added to the list of unmovable atoms.

Pharmacophore Matching. Next, the pharmacophores generated for the docked ligand conformers are matched to the pharmacophore of the template molecule. Identical as well as similar pharmacophore features (e.g., small lipophilic feature and medium lipophilic feature, π−system and aromatic ring) are matched to form pharmacophore solutions. Pharmacophore features with a vector property are matched only if the angle of their vectors is smaller than a threshold value (by default 30°). Solutions with the maximum number of matching pharmacophore features are sought for, but those having up to two pharmacophore features less than the maximum are also accepted.

Pose Generation. The accepted solutions are used to obtain an initial pose of the docked molecule in the binding site. First, the centroids of both template and docked ligand pharmacophores are calculated and superimposed. Systematic rotations (initial round: 10° step, refinement round: 1° step in a ± 10° range) along three principal axes are used to obtain the best fit of all matched pharmacophore features, characterized by the lowest separation between the matched pairs and lowest angle deviation of their vector properties. The rotations

dx.doi.org/10.1021/ci400098y | J. Chem. Inf. Model. 2013, 53, 1415−1423

producing the best fit are then applied to the docked conformer to obtain its pose in the binding site.

If the pose bumps into any unmovable atom of the protein structure (including the dummy atoms) fine rotations within a ± 40° range and translations within ±0.75 Å range along principal axes are applied to reposition it so that it becomes free of bumps. If this procedure fails, the corresponding pose is discarded.

Induced Fit. Side-chains of the binding site residues containing hydrogen-bond donors or acceptors as well as side-chains whose original conformer bumps to the ligand pose are rearranged to account for possible induced-fit in several steps.

First, the side-chains of residues, which can form hydrogen bonds or salt bridges among themselves, are grouped into families. In each family, all side-chains are then rearranged simultaneously in a combinatorial manner, so that the interaction energy is the most favorable and there are the least possible (optimally: no) bumps with the pose and among side-chains. In the combinatorial search, up to six residues can be rearranged simultaneously. This should be sufficient for most cases, as the analysis of the structures deposited within the PDB showed that only a small number of residues in the binding site undergo changes upon binding (in 85% of the proteins three or less residues are moved).[35] In Dolina, the maximum number of side-chain conformer combinations is limited: if the number of combinations exceeds 10^6 (default), for the residues having the most side-chain conformers only cluster representative conformers are used in the combinatorial search, which helps to keep combinatorial explosion under control. Should there be still too many possible combinations even if for all residues only cluster representatives are used, such a pose is discarded and excluded from further processing. This typically indicates that an unreasonable pose, which bumps to too many residues, was generated. In the case that for some residue only cluster representatives had to be used, an additional refinement of the induced-fit is performed, when for that particular residue only conformers belonging to the cluster of the representative conformer are used.

Next, for every not yet rearranged bumping or hydrogen-bonding capable side-chain the most favorable conformer having the least bumps to ligand is found.

In the last step, the rearranged side-chains are checked for inter-residual bumps and, if bumps are found, grouped into sterically independent families of residues, i.e. residues of one family cannot sterically affect residues of another family. In each family, all residue conformers are searched again in combinatorial manner for the best possible arrangement with the lowest number of bumps and the most favorable interaction energy. A typical family contains three to four residues, which effectively decreases the degrees of freedom and the computational costs. If, after induced-fit rearrangement, a docked pose or any residue has more than five bumps to any other residue, such pose is discarded. Finally, the positions of ligand's rotatable polar hydrogen atoms are explored by systematic rotations to yield the most favorable interaction with the protein.

Pose Scoring. Each valid pose is assigned a score by summing up ligand−protein and protein−protein interaction energy (total energy). Arbitrary weight factors were chosen for ligand−protein interaction energy (0.875) and protein−protein energy (0.125). Interaction energies are composed of two principal components − hydrogen bonding energy and hydrophobic interaction energy.

The hydrogen-bonding energy is calculated based on the Yeti force-field parameters (well depths and optimal distances for a 10−12 potential) and contains a term correcting the interaction energy for linearity ($\cos^2 \alpha_{X\text{-}H\cdots Y}$).

The hydrophobic interaction energy is calculated by a function scoring the pairwise interaction between ligand and protein atoms. Below or at the distance corresponding to the sum of atoms' van der Waals radii the function returns the product of atoms' lipophilicity coefficients. At distances larger than the sum of van der Waals radii the function returns product of atoms' lipophilicity coefficient divided by the steep falloff factor (eq 1):

$$f = e^{(distance - r_{i_{vdW}} - r_{j_{vdW}})^3} \quad (1)$$

Lipophilicity coefficients were obtained for 42 non-hydrogen atom types recognized in Dolina by fitting them to reproduce experimentally measured octanol−water partition coefficients of 449 small organic molecules (method based on simple summing of atom contributions).[36]

Filtering and Sorting. The following procedures are applied to filter out poses representing an unreasonable binding mode. First of all, the solution should be free of bumps. While bumps to unmovable parts are unacceptable, a small number of ligand−side-chain bumps may be tolerated, as they typically resolve within a few cycles of geometry optimization. Up to five bumps to the ligand from a single side-chain are tolerated, otherwise the pose is discarded. For all valid poses, bumps to all side-chains are summed, and an average is calculated. Poses having more than an average number of bumps are discarded. Average values are calculated also for binding, complex, and total energy, and if a pose is worse by 10.0, 20.0, or 25.0 kcal/mol (arbitrary threshold values) than the average, respectively, it is discarded. Finally, all valid poses are sorted by their total energy, and the top ranked ligand−protein complexes are saved.

RESULTS AND DISCUSSION

Rigid Docking of Crystal Ligand Conformers to Flexible Protein. For 15 out of 18 ligands (83%) Dolina generated a pose with *RMSD* to the reference pose lower than 1.0 Å, and for three ligands the *RMSD* was still lower than the generally accepted threshold value of 2.0 Å (Table 2). In 14 cases (78%), the pose with the best *RMSD* was ranked as top pose. In all 18 cases (100%), the pose with the best *RMSD* to the reference pose could be found within the top five generated poses.

The results confirm the validity of both docking and induced-fit algorithm and the good sensitivity of the scoring function implemented in Dolina. However, one should keep in mind that even if the pseudorandomly oriented crystal conformer of the ligand could be properly docked into the binding site and top-ranked, the result might be still favorably biased by the (correct) crystal conformer itself.

Flexible Docking of Ligands to Flexible Protein. The results of flexible docking runs (Table 3) indicate algorithm performance in a real-world application, when no prior knowledge about the correct ligand pose is available. Dolina was able to generate a pose with the *RMSD* to the reference lower than 2.0 Å and rank it among the top ten poses for the large majority of ligands. Satisfactory results were obtained in

Table 2. Results of the Rigid Docking[a]

reference PDB ID	*RMSD* [Å]	pose rank
2AMB	0.48	1
2AX9	0.85	1
2AXA	0.76	1
2HVC	1.53	1
2IHQ	0.57 (0.98)	4 (1)
2NW4	0.76	2
2PNU	0.94	1
3B66	0.53	2
3B67	0.46	1
3B68	0.56	1
3G0W	0.67	1
3RLJ	0.95	1
3VHU	0.89	1
3VHV	0.87	1
1SR7	1.45	1
3HQ5	1.62	1
3KBA	0.85	3
4APU	0.72	1

[a]Numbers in brackets indicate another pose having an *RMSD* below 2.0 Å ranked above the best pose.

49 out of 54 runs (91%), and only in five out of 54 runs (9%) the *RMSD* of the best pose was worse than 2.0 Å. In 15 out of 54 Dolina flexible docking runs (28%), the pose with the lowest *RMSD* to the reference pose was ranked as the top one, and in 41 cases out of 54 (76%), it was ranked among the top 5 poses. No failures were encountered.

The number of generated pharmacophore features for the template molecules was 10 for **TES**, 13 for **PDN**, and 9 for **STR** (Table S1, Supporting Information). Correctly, no protein-donor features were generated for the 11-oxo group of the template ligand **PDN**, because no hydrogen-bonding partner could be found at its binding pocket. For the docked ligand molecules the number of pharmacophore features ranged from 8 (**LGD**) to 25 (**B67**). Various conformers of the same ligand can have a different maximum number of pharmacophore features, because some of the features are sterically inaccessible at certain geometries. The number of matched pharmacophore features effectively used for obtaining (fitting) the best pose ranged from three to eight (4.3 on average). In 32 out of 54 cases (59%) the best pose was generated by fitting as little as three pharmacophore features indicating that despite low similarity between the template and the docked ligand, the native or near-native docking pose can be found.

In this study, the force-field used to generate conformers had only a moderate impact on docking results, and when the conformer input to Dolina was limited to 25 extended conformers, there was a high probability of obtaining at least one crystal-like pose among the top 10 poses. In 51 out of 54 cases (94%), the pool of low energy conformers contained a conformer highly similar to the one found in the crystal structure of the protein−ligand complex (*RMSD* < 1.0 Å calculated on the heavy atoms; Table S2, Supporting Information). This observation validates using of the low energy conformers as binding pose candidates in docking of similarly sized and flexible ligands, without the need to address their flexibility explicitly at runtime. The worst conformers-to-crystal pose *RMSD* observed was 1.60 Å for the ligand **B68** in combination with the AMBER* force field, which represents the only case where a rather poor conformer input could have compromised the docking (the best pose *RMSD* was 3.64 Å, rank 6). In the case of the ligand **B67**, the best matching conformer from the pool generated with the OPLS-2005 force field also had a rather high *RMSD* of 1.41 Å, nevertheless a

Table 3. Results of the Flexible Docking

	OPLS 2005 conformers		AMBER* conformers		AMSOL conformers	
reference PDB ID	*RMSD* [Å]	pose rank	*RMSD* [Å]	pose rank	*RMSD* [Å]	pose rank
2AMB	0.57	1	0.58	1	0.65	1
2AX9	0.91	1	1.09	2	1.07	2
2AXA	1.62	4	0.98	1	0.63	1
2HVC	1.64	2	1.72	1	0.97	2
2IHQ	1.01	5	0.92	8	1.15	3
2NW4	1.07	3	1.04	1	0.95	4
2PNU	1.50	1	1.29 (1.63)	5 (2)	1.38	2
3B66	1.41	10	1.20	1	0.82	4
3B67	1.68	3	1.87	3	0.67	2
3B68	2.27	8	3.64	6	1.15	1
3G0W	1.21 (1.42)	9 (1)	1.78	5	1.17 (1.23)	9 (3)
3RLJ	1.84	9	1.27	1	1.18	1
3VHU	0.77	1	1.00	6	0.88 (1.05)	3 (1)
3VHV	0.78	2	1.09	5	0.87	8
1SR7	1.64 (1.76)	2 (1)	1.80	3	1.66	4
3HQ5	2.60	6	1.19	4	2.97	6
3KBA	1.56	9	1.37	3	2.65	1
4APU	1.08	1	1.17	3	1.15	2
RMSD <2.0 Å	89%		94%		89%	
best = top pose[a]	28%		33%		22%	
best pose in top 5[a]	67%		83%		78%	

[a]Only poses with an *RMSD* to the reference pose lower than 2.0 Å are included. Numbers in brackets indicate another pose having an *RMSD* below 2.0 Å ranked above the best pose.

dx.doi.org/10.1021/ci400098y | *J. Chem. Inf. Model.* 2013, 53, 1415−1423

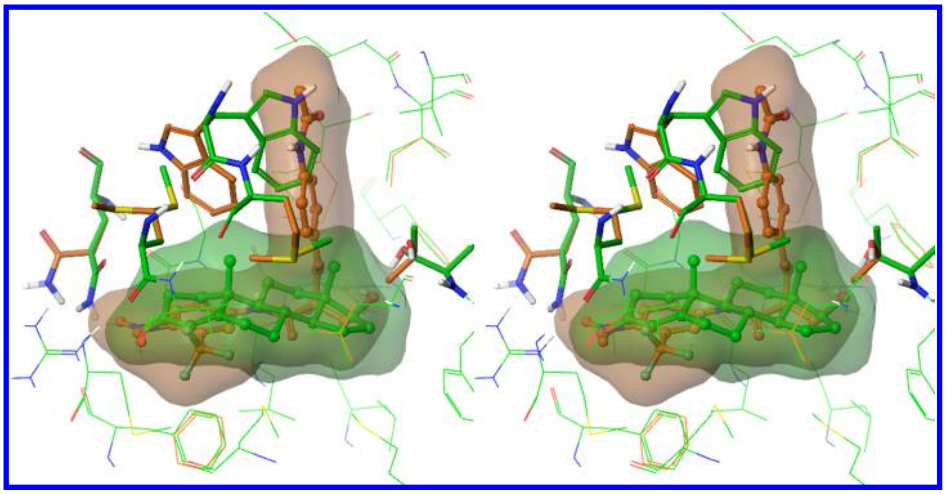


Figure 2. Stereoview of the induced-fit changes observed for ligand **B68**. Green carbon atoms denote the initial structure of AR (lines) with bound testosterone (ball and stick, green surface). Orange carbon atoms denote Dolina-generated structure (lines) with bound **B68** (ball and stick, orange surface). Substantially rearranged residues are shown as sticks.

near-native pose could be generated by Dolina (*RMSD* 1.68 Å, rank 3).

At the AR, the most voluminous ligand **ENM**, which was originally designed to impede repositioning of the mobile helix 12 blocking the ligand dependent transactivation function,[37] could still be docked by the local induced-fit (i.e., no backbone repositioning) algorithm of Dolina with a satisfactory result (*RMSD* of ∼1.4 Å). Comparing its protein crystal structure (PDB entry 2PNU) with the one obtained for a small agonist reveals an average displacement of helix C_α atoms of 1.5 Å. However, Dolina had some difficulty finding a reasonable pose for a group of the selective androgen receptor modulators (SARMs: **B66**, **B67**, **B68**, **RLJ**) using conformers generated by both OPLS-2005 and AMBER* force-fields. This can be best demonstrated with **B68**, the most voluminous ligand in the group. Proper accommodating of the **B68** into the binding site, associated with a slight backbone repositioning observed in the crystal structure (PDB entry 3B68), requires a conformer with a high similarity to the crystal pose, as the protein structure cocrystallized with a small agonist (template) does not provide too much space for its bulky 4-acetaminophenyl group (Figure 2). Unfortunately, both force-fields produced conformers in which the 3-(trifluoro)-4-nitrophenylamide scaffold was not planar (in contrast to crystal pose geometry). Liquid-phase geometry optimization in AMSOL rectified such distorted geometry, which resulted in a good and highly ranked pose (*RMSD* of 1.15 Å, rank 1) compared to the OPLS-2005 and the AMBER* poses (2.27 Å, rank 8 and 3.64 Å, rank 6, respectively). A detailed analysis of the **B68** docking runs showed, that some OPLS-2005 conformers could be in fact fitted to the template producing two near-native poses (*RMSD*s 1.57 and 1.77 Å), but both were discarded because of too many bumps with the binding site side-chains. The absence of a conformer similar to the crystal pose in the AMBER* conformer pool (as explained earlier) caused that no bump-free solution with an *RMSD* lower than 2.0 Å could be fitted to the binding site. **B68** (optimized with AMSOL) represents a borderline example, when a bump-free pose similar to the reference was generated using local induced-fit, even if adjusting part of the protein backbone would be desirable. A successful docking (or "squeezing") result in this case can be explained by a positive synergism of small deviations in both conformer geometry and pose orientation along with scaled van der Waals radii. A beneficial, but less pronounced, effect of the AMSOL-rectified scaffold geometry on the *RMSD* and pose rank could be observed also for other ligands with *N*-phenypropionamide scaffold (**FHM**, **B66**, **B67**, and **RLJ**).

Binding of the ligands **SNL** and **LD1** at the MR does not involve any interaction with the protein backbone. In all docking runs with Dolina a pose with a comparably low *RMSD* to the reference pose could be found at the MR.

On the other hand at the PR, satisfactory results for both ligands with the 4-(benzyl(pyrrolidin-3-yl)amino)benzonitril scaffold (**GKK** and **WOW**) could be obtained only with conformers using one particular force-field. The benzyl group of these two ligands induces relocation of the backbone near F-794 (C_α separation >0.7 Å), which is not present in the crystal structure of the protein bound with the small template agonist. Conformers that readily fit into the limited space in the binding site were only found in the pools generated by the AMBER* force-field. In the case of the ligand **GKK**, one of the conformers closely resembles the crystal pose, but those generated using the other two force-fields have slightly different geometries (the best conformer-to-pose *RMSD* are the following: 0.28 Å with AMBER*, 0.88 Å with OPLS-2005, 0.86 Å with AMSOL; Table S2, Supporting Information). However, in the case of the ligand **WOW**, the similarity of conformers toward the crystal pose is approximately the same with all three force-fields (*RMSD*s 0.67−0.85 Å), but a good pose could be found only with the OPLS-2005 and the AMBER* force-fields. This suggests that in extreme cases the docking might be very sensitive, and even delicate differences among the ligand conformers can significantly affect the outcome. The backbone relocation for the ligand **MOF** is even more pronounced due to the bulky furoate group, (C_α separation ∼2.0 Å). Dolina had to compensate such a substantial deficit of space and associated steric bumps by shifting the pose away from the backbone, which resulted in rather poor *RMSD*s (∼1.7 Å) obtained for this ligand, regardless of the force-field used. Interestingly, for the most voluminous ligand in the data set, **A2K**, fairly accurate and high-ranked poses could be generated with all three force-fields, because its protruding pyridin-3-ylphenyl substituent induces

Table 4. Rotamer Analysis of the Binding Site Residues

	number of residues[a] in the reference crystal structure			number of different residues[a] in complex with the best pose having a different rotamer					
				OPLS-2005		AMBER*		AMSOL	
PDB ID	total	h-phil[b]	h-phob[b]	h-phil[b]	h-phob[b]	h-phil[b]	h-phob[b]	h-phil[b]	h-phob[b]
2AMB	21	5	16	2	2	1	2	1	2
2AX9	17	5	12	0	1	1	1	1	1
2AXA	22	6	16	1	4	2	2	2	2
2HVC	20	5	15	1	5	1	6	2	5
2IHQ	19	5	14	3	4	3	4	3	4
2NW4	19	5	14	3	4	1	4	3	4
2PNU	26	7	19	2	3	1	4	2	2
3B66	22	6	16	2	4	2	3	3	2
3B67	22	6	16	2	2	3	4	0	3
3B68	24	7	17	2	6	2	6	3	2
3G0W	21	5	16	2	3	3	3	3	6
3RLJ	22	6	16	2	4	2	3	3	2
3VHU	23	8	15	1	2	0	6	2	4
3VHV	21	7	14	3	5	3	3	4	3
1SR7	24	8	16	1	4	3	5	3	5
3HQ5	21	7	14	0	8	2	6	0	8
3KBA	23	8	15	3	4	2	4	3	5
4APU	26	8	18	1	5	2	5	2	5
average	21.8	6.3	15.5	1.7	3.9	1.9	3.9	2.2	3.6

[a]With a movable side-chain atom within 5.0 Å from the ligand, i.e. binding site residues. [b]h-phil = hydrophilic residues: S, T, C, D, E, Q, N, H, Y, W, R, K; h-phob = hydrophobic residues: V, L, I, F, M.

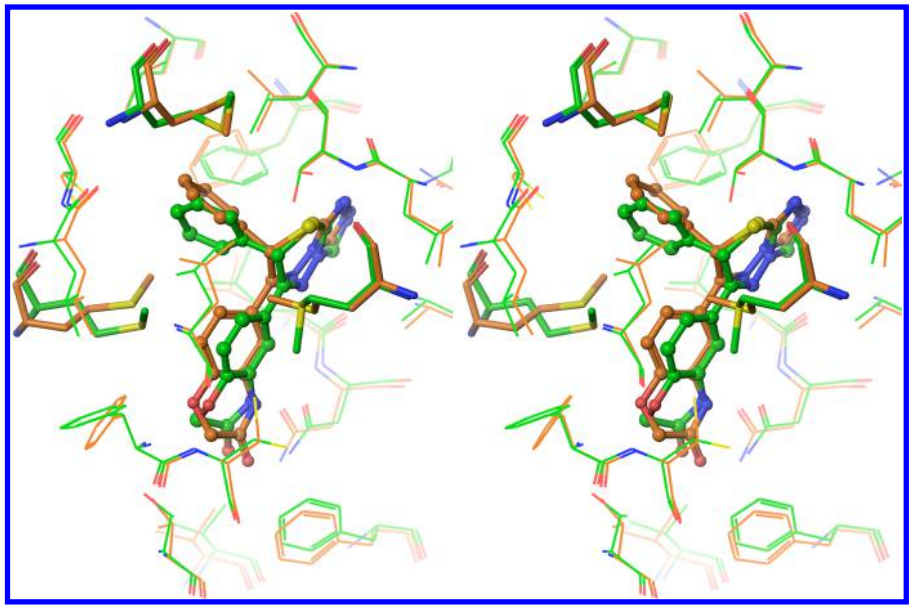


Figure 3. Stereoview of the reference ligand−protein complex (PDB entry 3VHV, carbon atoms in green) compared to Dolina generated complex (carbon atoms in orange). Binding site rotamers of methionines (stick representation) are different but occupy essentially the same space in the close vicinity of the ligand (ball and stick representation).

only a slight backbone relocation in a different binding pocket (PDB entry 4APU).

The analysis of the side-chain rotamers at the binding site in complexes with the best poses showed that even if the ligand pose closely matches to the reference one, rotamers of the residues that rearranged due to the ligand's presence in the binding site can differ from those in the reference protein crystal structure (Table 4). On average, slightly less than two (OPLS-2005: 1.7, AMBER*: 1.9, AMSOL: 2.2) out of six (6.3) hydrophilic binding site residues had a different rotamer than in the reference crystal structure (a difference of two corresponding torsion angles χ greater than 60°). When only directly interacting hydrophilic residues are considered (i.e., those having a side-chain atom within the radius of 3.5 Å from any ligand atom; default 5.0 Å), on average around one (OPLS-2005: 1.1, AMBER*: 1.3, AMSOL: 1.5) out of four (4.39) residues is represented by a different rotamer. In the case of the hydrophobic residues, on average less than four (OPLS-2005: 3.9, AMBER*: 3.9, AMSOL: 3.6) out of 15 (15.5) binding site residues were different. There seem to be several causative

dx.doi.org/10.1021/ci400098y | J. Chem. Inf. Model. 2013, 53, 1415−1423

factors. One of them is a high number of the variable side-chain torsions in the binding site, which cannot always be explored in an exhaustive manner. In complex cases (conformer-rich residues, a high number of them), only conformers representing certain clusters are used, which might decrease the accuracy of side-chain positioning in the available space between the pose and the unmovable protein atoms (cf. extreme case of the ligand **ENM** with 26 binding site residues; Table S3, Supporting Information). On average, a difference greater than 60° was found for less than three torsion angles in the hydrophilic (OPLS-2005: 2.2, AMBER*: 2.4, AMSOL: 2.8; cf. Table S4, Supporting Information) and around six torsion angles in the hydrophobic binding site residues (OPLS-2005: 5.8, AMBER*: 6.2, AMSOL: 5.7).

Some residues with increased conformational freedom (particularly methionine, Figure 3) can occupy the available space in the binding site by means of several different rotamers with similar energy. This phenomenon is frequently observed in crystal structures as disordered residues. Methionine represented 23.7% of the binding site residues of the three nuclear receptors in this study but accounted for more than 40% of the rotamer differences (Table S5, Supporting Information). For comparison, leucine represented 24.2% of the binding site residues but accounted for less than 15% of the observed differences, and the remaining 12 residue types with a flexible side-chain (*n.b.* three residue types - aspartate, glutamate, and lysine - are not present in the binding sites of AR, MR, and PR) shared the remaining 45% of differences without any significant outlier. There are several reasons explaining the low consistency of the methionine rotamers generated by Dolina and those found in the reference structures. In the crystal structure of AR PDB entries 2AMB and 2PNU, the binding site residues M-895 and M-780 are disordered, suggesting a lack of conformational preference. Moreover, three different rotamers of the residue M-895 can be found (g-/g-/g-, g-/g+/g-, and g-/t/t) in three different crystal structures (2AM9, 2HVC, and 2IHQ), in which no apparent ligand induced-fit toward this residue can be assumed. Even more variability is observed in the crystal structures in which the ligands trigger induced fit changes (four additional rotamers). In the case of the binding site residue M-745, the rotamer selection was strongly affected by the ligand-induced rearrangement of the nearby W-741 residue, which in some cases was not positioned properly due to its size and limited space in the binding site.

In order to introduce some tolerance for close contacts, van der Waals radii of atoms in the Dolina are scaled by a factor of 0.95 below the value at which the Lennard-Jones potential becomes repulsive ($0.95 * 0.891 * (r_{i_{vdW}} + r_{j_{vdW}})$). However, bulky aromatic side-chains (especially of tryptophan) are still quite sensitive to fine changes in the position of the ligand in the binding site. As Dolina tries to generate bump-free poses, in certain situations a rotamer devoid of close contacts to ligand or other protein atoms is preferred to the one having better interaction energy. In the case of the hydrophilic residues forming direct H-bond interactions with the ligand, selecting a bump-free nevertheless incorrect rotamer can negatively affect pose scoring due to a significant loss of the ligand–protein interaction energy, which is the most important component of the total score. This explains why some of the near-native poses (AR: **B66**, **RLJ**; PR: **WOW**, the OPLS-2005 force-field) are ranked low in the pose list.

Due to the reasons discussed above, a visual inspection aimed at congested binding pockets, structure minimization, or molecular dynamic simulations are recommended methods for postprocessing of the Dolina protein–ligand output structures.

CONCLUSION

The results obtained with Dolina demonstrate that docking based on matching pharmacophores of ligand conformers with extended pharmacophore derived from the ligand–protein complex followed by rearranging of binding site side-chains can be utilized for generating poses similar to those found in crystal structures. The approach was shown to work surprisingly well also for diverse ligands requiring substantial local induced-fit (or even slight backbone relocation, which is not accounted for) with several side-chains involved and newly created cavity volume of hundreds of Å^3. A simple, but sensitive scoring function is able to identify the most reasonable poses and rank them at the top of the pose list. However, this study does not give any answer to whether it is conformational selection or induced-fit which drives ligand binding, as Dolina, in principle, combines features of both approaches into one algorithm.

A minor disadvantage of Dolina is that it depends on the input of ligand conformers, which to a certain extent can have impact on the quality and ranking of generated poses. However, this can be viewed also as an advantage, if the user would like to perform docking using some particular conformers, e.g. those based on small molecule crystallography or NMR experiments. In terms of performance, Dolina is comparable to other established software treating docking and induced-fit at comparable level of complexity, but it certainly cannot be classified as high throughput docking program. A typical docking run (ligand of similar size as template, 25 input conformers) consumes approximately 30 min of CPU time on a 2.5 GHz processor.

Dolina along with documentation and an application example is available for academic institutions upon request for Linux and MacOS X platforms.

Note: In the Slovak language, the word "dolina" means "valley". One-letter amino acid codes are used in the manuscript.

ASSOCIATED CONTENT

Supporting Information

Structural formulas of templates and docked ligands, additional tables, comparison to commercial software. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +41 61 267 15 54. Fax: +41 61 267 15 52. E-mail: martin.smiesko@unibas.ch.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The author thanks Angelo Vedani for valuable suggestions and proofreading the manuscript.

ABBREVIATIONS

PDB, Protein Data Bank; *RMSD*, root mean square deviation

REFERENCES

(1) Kokh, D. B.; Wenzel, W. Flexible side chain models improve enrichment rates in in silico screening. *J. Med. Chem.* **2008**, *51*, 5919−5931.

(2) Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol.* **2008**, *18*, 178−184.

(3) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How does a drug molecule find its target binding site? *J. Am. Chem. Soc.* **2011**, *133*, 9181−9183.

(4) Flores, S. C.; Gerstein, M. B. Predicting protein ligand binding motions with the conformation explorer. *BMC Bioinf.* **2011**, *12*, 417−429.

(5) Biesiada, J.; Porollo, A.; Velayutham, P.; Kouril, M.; Meller, J. Survey of public domain software for docking simulations and virtual screening. *Hum. Genomics* **2011**, *5*, 497−505.

(6) B-Rao, C.; Subramanian, J.; Sharma, S. D. Managing protein flexibility in docking and its applications. *Drug Discovery Today* **2009**, *14*, 394−400.

(7) Lill, M. A. Efficient incorporation of protein flexibility and dynamics into molecular docking simulations. *Biochemistry* **2011**, *50*, 6157−6169.

(8) Sinko, W.; Lindert, S.; McCammon, J. A. Accounting for receptor flexibility and enhanced sampling methods in computer-aided drug design. *Chem. Biol. Drug Des.* **2013**, *81*, 41−49.

(9) Durrant, J. D.; McCammon, J. A. Computer-aided drug-discovery techniques that account for receptor flexibility. *Curr. Opin. Pharmacol.* **2010**, *10*, 770−774.

(10) Ferrari, A. M.; Wei, B. Q.; Costantino, L.; Shoichet, B. K. Soft docking and multiple receptor conformations in virtual screening. *J. Med. Chem.* **2004**, *47*, 5076−5084.

(11) Jiang, F.; Kim, S. H. Soft docking": Matching of molecular surface cubes. *J. Mol. Biol.* **1991**, *219*, 79−102.

(12) Koska, J.; Spassov, V. Z.; Maynard, A. J.; Yan, L.; Austin, N.; Flook, P. K.; Venkatachalam, C. M. Fully automated molecular mechanics based induced fit protein-ligand docking method. *J. Chem. Inf. Model.* **2008**, *48*, 1965−1973.

(13) Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE). *J. Comput.-Aided Mol. Des.* **2008**, *22*, 311−325.

(14) Meiler, J.; Baker, D. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins* **2006**, *65*, 538−548.

(15) Korb, O.; Olsson, T. S. G.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C. Potential and limitations of ensemble docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262−1274.

(16) Takaya, D.; Yamashita, A.; Kamijo, K.; Gomi, J.; Ito, M.; Maekawa, S.; Enomoto, N.; Sakamoto, N.; Watanabe, Y.; Arai, R.; Umeyama, H.; Honma, T.; Matsumoto, T.; Yokoyama, S. A new method for induced fit docking (GENIUS) and its application to virtual screening of novel HCV NS3−4A protease inhibitors. *Bioorg. Med. Chem.* **2011**, *19*, 6892−6905.

(17) Wagener, M.; de Vlieg, J.; Nabuurs, S. B. Flexible protein-ligand docking using the Fleksy protocol. *J. Comput. Chem.* **2012**, *33*, 1215−1217.

(18) Kalid, O.; Warshaviak, D. T.; Shechter, S.; Sherman, W.; Shacham, S. Consensus induced fit docking (cIFD): methodology, validation, and application to the discovery of novel Crm1 inhibitors. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 1217−1228.

(19) Gabrielsen, M.; Kurczab, R.; Ravna, A. W.; Kufareva, I.; Abagyan, R.; Chilmonczyk, Z.; Bojarski, A. J.; Sylte, I. Molecular mechanism of serotonin transporter inhibition elucidated by a new flexible docking protocol. *Eur. J. Med. Chem.* **2012**, *47*, 24−37.

(20) Bottegoni, G.; Kufareva, I.; Totrov, M.; Abagyan, R. Four-dimensional docking: A fast and accurate account of discrete receptor flexibility in ligand docking. *J. Med. Chem.* **2009**, *52*, 397−406.

(21) Davis, I. W.; Baker, D. ROSETTALIGAND docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009**, *385*, 381−392.

(22) Park, I.-H.; Li, C. Dynamic ligand-induced-fit simulation via enhanced conformational samplings and ensemble dockings: A survivin example. *J. Phys. Chem. B* **2010**, *114*, 5144−5153.

(23) Flick, J.; Tristram, F.; Wenzel, W. Modeling loop backbone flexibility in receptor-ligand docking simulations. *J. Comput. Chem.* **2012**, *33*, 2504−2515.

(24) ProteinDataBank. http://www.rcsb.org/ (accessed September 2012).

(25) *QikProp*, version 3.5; Schrödinger, LLC: New York, NY, 2012.

(26) *Maestro*, version 9.3; Schrödinger, LLC: New York, NY, 2012.

(27) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Thompson, J. D.; Winget, P.; Lynch, B. J.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL*, version 7.0; University of Minnesota: Minneapolis, 2003; based in part on AMPAC-version 2.1 by Liotard, D. A.; Healy, E. F.; Ruiz, J. M.; Dewar, M. J. S..

(28) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. Class IV charge models: a new semiempirical approach in quantum chemistry. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 87−110.

(29) Vedani, A.; Huhta, D. W. A new force field for modeling metalloproteins. *J. Am. Chem. Soc.* **1990**, *112*, 4759−4767.

(30) *Yeti*. http://www.biograf.ch/downloads/YetiX_Documentation.zip (accessed September 2012).

(31) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.

(32) *MacroModel*, version 9.9; Schrödinger, LLC: New York, NY, 2012.

(33) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Model for aqueous solvation based on class IV atomic charges and first solvation shell effects. *J. Phys. Chem.* **1996**, *100*, 16385−16398.

(34) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1993**, *115*, 5348−5348.

(35) Najmanovich, R.; Kuttner, J.; Sobolev, V.; Edelman, M. Side-chain flexibility in proteins upon ligand binding. *Proteins* **2000**, *39*, 261−268.

(36) Section 16: Health and safety information. In *CRC handbook of chemistry and physics*, 87th ed.; Lide, D. R., Ed.-in-Chief; CRC Taylor and Francis: Boca Raton, FL, 2006; pp 16-41−16-45.

(37) Cantin, L.; Faucher, F.; Couture, J.-F.; de Jesus-Tran, K. P.; Legrand, P.; Ciobanu, L. C.; Frechette, Y.; Labrecque, R.; Singh, S. M.; Labrie, F.; Breton, R. Structural characterization of the human androgen receptor ligand-binding domain complexed with EM5744, a rationally designed steroidal ligand bearing a bulky chain directed toward helix 12. *J. Biol. Chem.* **2007**, *282*, 30910−30919.