JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

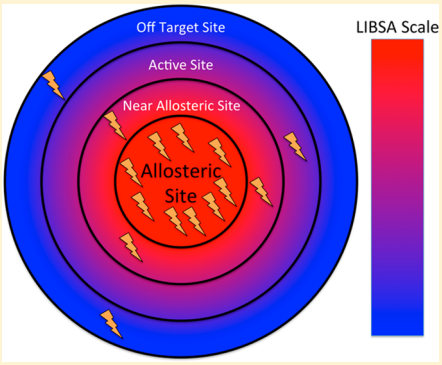
LIBSA – A Method for the Determination of Ligand-Binding Preference to Allosteric Sites on Receptor Ensembles

Harrison J. Hocker, Nandini Rambahal, and Alemayehu A. Gorfe*

Department of Integrative Biology and Pharmacology, The University of Texas Health Science Center at Houston, 6431 Fannin Street MSB 4.108, Houston, Texas 77030, United States

Supporting Information

ABSTRACT: Incorporation of receptor flexibility into computational drug discovery through the relaxed complex scheme is well suited for screening against a single binding site. In the absence of a known pocket or if there are multiple potential binding sites, it may be necessary to do docking against the entire surface of the target (global docking). However no suitable and easy-to-use tool is currently available to rank global docking results based on the preference of a ligand for a given binding site. We have developed a protocol, termed LIBSA for LIgand Binding Specificity Analysis, that analyzes multiple docked poses against a single or ensemble of receptor conformations and returns a metric for the relative binding to a specific region of interest. By using novel filtering algorithms and the signal-to-noise ratio (SNR), the relative ligand-binding frequency at different pockets can be calculated and compared quantitatively. Ligands can then be triaged by their tendency to bind to a site instead of ranking by affinity alone. The method thus facilitates screening libraries of ligand cores against a large library of receptor conformations without prior knowledge of specific pockets, which is especially useful to search for hits that selectively target a particular site. We demonstrate the utility of LIBSA by showing that it correctly identifies known ligand binding sites and predicts the relative preference of a set of related ligands for different pockets on the same receptor.



INTRODUCTION

In recent decades, the number of drugs brought to market has been declining, and new low-cost methods for drug discovery are needed.[1] While virtual screening is able to reduce the total number of ligands that need to be synthesized and screened experimentally,[2] it requires prior knowledge of the target site.[3,4] However, some of the most effective compounds on the market bind to allosteric sites,[5] and these sites are not always apparent in average structures from X-ray crystallography or nuclear magnetic resonance spectroscopy (NMR).[6] There are several computational techniques for binding site identification, including FTMap,[7] blind docking,[8,9] solvent mapping,[10] and other simulation-based methods.[11] Once the sites are determined, virtual screening (VS) can be performed to find small-molecule ligands that have the potential to become hits.[12] Typically, VS is followed by experimental assays of the highest affinity compounds. Although the application of any of these methods to one or a few conformations of a receptor is fairly straightforward, there is no simple way to find a consensus-binding site on a large number of structures or to prioritize hits by preference to a particular binding site. This is important because ligand–receptor affinity is a property of an ensemble,[13] and therefore incorporation of target flexibility is crucial for success.[14] Moreover, almost all docking methods rank ligands based on predicted binding affinities,[15] despite examples of low-affinity hits having led to potent bioactive compounds.[16] We have developed a technique called LIgand Binding Specificity Analysis (LIBSA) that quantifies (and prioritizes by) pocket-specificity following a search for allosteric ligand binding sites over an ensemble of receptor conformations.

LIBSA uses a small molecule or fragment as a probe to search for allosteric binding sites on the surface of a protein. This can be done by popular docking programs such as Autodock,[17] which in principle are capable of finding the correct binding site for the right ligand when the search area covers the entire protein surface.[9,18] However, this approach is likely to yield random docked poses. These poses typically have a low probability of occurring and thus should be filtered out. We introduce two techniques to remove such poses: a high-pass filter[19] and an algorithm that filters out ligand poses based on their frequency of occurrence. The latter favors poses with high affinity, and the high-pass filter[19] removes false-positive hits by scaling down ligand–receptor contacts that fall below a threshold value. We then use a signal-to-noise ratio (SNR),[20] defined as the relative frequency of a ligand contacting a particular region of the protein versus all other regions, to quantify binding specificity. In sum, the general procedure of LIBSA entails probing a receptor conformation with a ligand, determining the frequency of the different protein–ligand contacts, filtering out random poses, and then quantifying the consistency of binding with the SNR (Figure 1).

Received: September 18, 2013
Published: January 20, 2014

   dx.doi.org/10.1021/ci400474u | J. Chem. Inf. Model. 2014, 54, 530–538

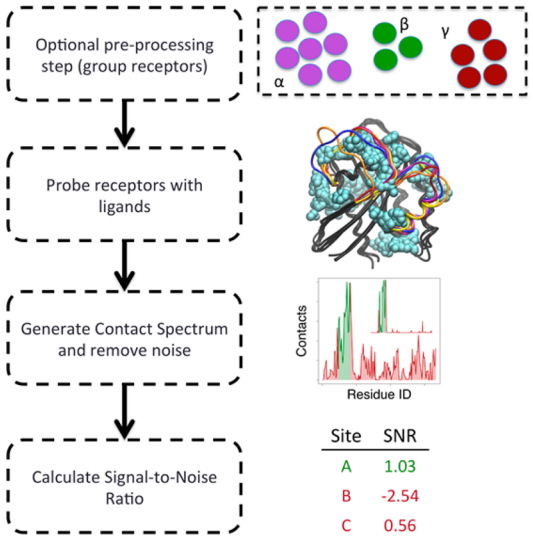


Figure 1. Schematic outline of the LIBSA protocol for the determination of ligand binding preference to allosteric sites on a single or ensemble of receptor conformations. (Left) Workflow of LIBSA. (Right top) Optional preprocessing to group protein conformers based on clustering or other methods. (Right upper middle) Depiction of several Ras conformations bound to theoretical probe ligands shown as cyan spheres. (Right lower middle) Contact spectrum generated from the probe ligands where the red peaks represent the noise and the green peaks represent the signal. A filtered spectrum is shown as an inset. (Right bottom) Table summarizing the hypothetical SNRs calculated from the contact spectrum, where pocket A is defined by the green peaks and pockets B and C lie within the red peaks.

As an initial test, we used LIBSA to correctly identify the active site on a set of receptors cocrystallized with small molecule drugs, including a group of kinase domains,[21] nuclear receptors,[22−24] the β2AR,[25] HIV-protease,[26] and K-Ras.[27,28] We then applied LIBSA to rank four related compounds by binding site preference when docked onto an ensemble of 148 experimental and simulated Ras structures.

MATERIALS AND METHODS

Generating Library of Receptor Conformers for Docking. We applied LIBSA on two sets of structural ensembles. The first involved 10 protein−ligand complexes from the Protein Data Bank (PDB) representing five different protein families (G-proteins, G-protein coupled receptors, kinases, nuclear receptors, and proteases). In each case, the cocrystallized ligand was removed and used as a probe for blind docking to search for the binding site on the cocrystal conformation (Table 1) and, when available, the corresponding *apo* structure (Table 2). The second ensemble containing 148 wild type and mutant H- and K-Ras conformers was derived from molecular dynamics (MD) simulations (80), crystallography (66), and NMR (2 averaged and energy-minimized structures). The MD simulations were conducted as described in the SI and representative structures were identified as follows. 10 ps-separated snapshots from trajectories were clustered using Cα atom positions with the leader-RMSD-based algorithm implemented in Wordom.[29] The RMSD cutoff for clustering ranged from 1.3 to 1.8 Å so that the top 10 clusters contained 92−99% of the conformations in each trajectory (see the SI). The resulting 80 cluster centroids were taken as representative conformations of the large phase space sampled by the simulations. For an additional and more global characterization of the ensemble, we regrouped the structures into five clusters in principal component (PC) space with the K-means algorithm.[30]

Molecular Docking. The proteins and ligands were prepared for docking by first removing ions and water molecules and building missing protein atoms using the CHARMM27 force field[31] and VMD;[32] missing nucleotide hydrogen atoms were built using AutoDock Tools[17] or OpenBabel.[33] The structure of ligands andrographolide (AGP), 3,19-(2-bromobenzylidene) andrographolide (SRJ09), 3,19-(3-bromobenzylidene) andrographolide (SRJ10), and 3,19-(3-chloro-4-fluorobenzylidene) andrographolide (SRJ23)[8] were prepared using the CHARMM generalized force field (CGenFF36),[34] as described before.[8] Then, nonpolar hydrogen atoms were condensed into their respective heavy atoms, and atomic charges were assigned using the Gasteiger−Marsili method.[35] "Blind docking"[18] was carried out with AutoDock 4.2[17] with the aid of the AutoDock tools (ADT) package,[17] keeping the ligand flexible and the receptor rigid. The search space was a cubic box covering the entire surface and centered on the geometric center of the protein, extending 10 Å beyond the protein edge in each direction. The termination criterion

Table 1. Performance of LIBSA on Known Protein−Ligand Complexes[a]

ligand	receptor	PDB ID	box vol. (Å^3)	Pop (S&W)	# Lig Tors	RMSD (Å)	SNR	
							no filter	affinity filter
Nilotinib	p38 MAPK	3GP0	2.1e5	225 (0.25)	7	0.72	1.12	1.15
Gleevec	c-Abl Kin.	1IEP	1.9e5	225 (0.25)	7	1.34	1.42	1.43
Tamoxifen	ERα	3ERT	2.2e5	150 (0.10)	10	1.46	2.06	3.75
Raloxifene	ERα	2QXS	2.1e5	150 (0.10)	9	0.95	1.50	1.76
estradiol	ERα	1QKT	2.1e5	150 (0.10)	2	1.01	3.64	3.64
BI-167107	β2AR	3SN6	2.5e5	225 (0.25)	8	1.48	1.15	1.07
Indinavir	HIV Pr.	1HSG	1.3e5	225 (0.25)	14	1.41	0.96	0.94
benzimidazole	K-Ras	4DSU	1.3e5	150 (0.10)	0	0.94	1.10	3.48
0QV	K-Ras	4EPW	1.1e5	150 (0.10)	3	0.74	1.17	3.54
0QW	K-Ras	4EPT	1.0e5	150 (0.10)	3	0.71	1.24	1.74

[a]Selected protein−ligand complexes from the protein databank were used to illustrate the ability of LIBSA to identify the active site and ligand pose. SNR values were calculated with the affinity filter and without any filter. The AutoDock ligand population size (Pop), probability of performing a Solis and Wets (S&W) local energy minimization, rotatable bonds in the ligand (Lig. Tors), volume of the search box, and the lowest docked RMSD relative to the crystal pose are listed. Kin. = kinase, Pr. = protease.

dx.doi.org/10.1021/ci400474u | J. Chem. Inf. Model. 2014, 54, 530−538

Table 2. LIBSA Analysis of Test Ligands Docked onto Crystal and MD-Derived *apo* Structures[a]

ligand	receptor	structure	box vol. (Å^3)	Pop (S&W)	#Lig. Tors	SNR
Docking onto an *apo* Structure with Open or Semiopen Binding Site						
estradiol	ERα	2B23	1.5e5	150 (0.10)	2	3.61
0QW	H-Ras	2CL0, 2RGB, MD	1.1−1.2e5	150 (0.10)	3	1.11, 1.03, 0.55
BZI	H-Ras	2CL0, 2RGB, MD	1.1−1.2e5	150 (0.10)	0	1.05, 1.00, 1.58
0QV	H-Ras	2CL0, 2RGB, MD	1.1−1.2e5	150 (0.10)	3	0.87, 0.72, 0.86
Gleevec	p38 MAPK	1WFC	2.5e5	225 (0.25)	7	0.75
Nilotinib	p38 MAPK	1WFC	2.5e5	225 (0.25)	7	0.71
Docking onto an *apo* Structure with Occluded or Small Binding Pocket						
BI-1667107	β2AR	2R4S	1.9e6	225 (0.25)	8	−0.04
Raloxifene	ERα	2B23	1.5e5	150 (0.10)	9	−0.16
Tamoxifen	ERα	2B23	1.5e5	150 (0.10)	10	−0.17
Global Cross Docking						
Nilotinib	c-Abl Kin.	1IEP	1.9e5	225 (0.25)	7	1.05
Gleevec	p38 MAPK	3GP0	2.1e5	225 (0.25)	7	1.05

[a]Because the available Ras ligands are reportedly nonspecific for the highly homologous Ras isoforms,[27,28] for a stringent test of LIBSA we docked ligands solved with the K-Ras isoform on two PDB and an MD *apo* structure of H-Ras. These structures were chosen because they possess a visually discernible pocket that is similar to that in the ligand-bound K-Ras PDB structures 4DSU, 4EPT, or 4EPW.

for each docking run was set to be either 10,000 LGA generations or 10^9 energy evaluations, whichever comes first. The genetic algorithm (GA) population size and the Solis and Wets local search probability were optimized for each system depending on the volume of the search area, the accessibility of the pocket, and the number of rotatable bonds in the ligand. Additional details of the docking parameters are listed in Tables 1 and 2.

Affinity Filtering: Noise Reduction Using Docking Scores. It is well-known that ranking by predicted affinity is a major source of false positives (error of ∼2 kcal/mol),[17] and currently there is no simple way to identify potential hits with low predicted affinity. Therefore, we have developed an approach that favors binding consistency over affinity by putting more weight on the frequency of occurrence of a particular docking score rather than the magnitude of the score. A key concept behind affinity filtering is that docking scores can be used as a metric for pose uniqueness and not just ranking. The steps for performing the filtering are as follows: (i) Generate a histogram of the binding scores, which we refer to as an *affinity spectrum*, from the AutoDock DLG file (in kcal/mol). (Since AutoDock gives the scores with a precision out to the hundredth decimal we used a relatively fine binning width of 0.05 kcal/mol.) (ii) Identify the peaks that represent frequently sampled affinity space based on a simple cutoff relative to the maximum value within each spectrum. We found that a threshold of 40% of the maximum peak value yields the best compromise between elimination of false positives and retention of alternative poses at a given site. We therefore collected all peaks whose height is ≥40% of the maximum peak height and referred to them as *explicit peaks*. (iii) To account for the fact that scoring functions are error prone, we added all peaks whose AutoDock score is within a certain percentage of that of the explicit peaks, which we call *auxiliary peaks* (i.e., peaks in neighboring bins each side of the explicit peaks). For example, using a sampling window of 1% (as used throughout this paper) and a hypothetical explicit peak with an average energy score of −10.0 kcal/mol in a spectrum of bin width 0.05 kcal/mol, we include the peaks in the first 2 bins each side of the explicit peak ((|(−10.0 × 0.01)|)/0.05) = 2). (iv) Write out the structures that gave rise to the explicit and auxiliary peaks; this requires keeping track of each docked conformation in step (i).

This procedure effectively eliminates low frequency high affinity poses while assigning more weight on the high frequency high affinity poses. Moreover, poses that are similar in affinity to the high frequency ones but appear less frequently can be captured through the incorporation of auxiliary peaks around the explicit peaks.

Binding Site Identification and Scoring with SNR. The affinity filtering described above identifies docked poses that are well sampled in affinity space irrespective of their binding site on the receptor. To identify a ligand-binding site, we use histograms of the frequency with which a ligand contacts residues on the receptor, followed by an analysis of the signal-to-noise ratio (SNR, see below). This entails two simple steps. First, generate a histogram of residue contact frequencies by counting the number of times any heavy atom of a residue lies within 4.0 Å of any ligand heavy atom during repeated docking runs. In this histogram, which we refer to as a *contact spectrum*, each bin corresponds to a single residue, and for each ligand there can be as many contact spectra as there are receptor conformations. Second, define a surface patch of interest (e.g., from the dominant peaks in the contact spectrum or from prior knowledge about the biochemical/structural features of the protein). Then, the binding preference of a ligand to the patch is quantified by SNR (eq 1), where signal refers to peaks that lie within the binding patch and noise denotes the peaks that lie outside the binding patch:

$$\mathrm{SNR} = \log_{10}\left[\frac{N^{\mathrm{noise}}\sum u_i}{N^{\mathrm{signal}}\sum u_j}\right] \quad (1)$$

Here u_i refers to the peak height at residue i and is indexed over all the residues that lie within the binding patch (total number = N^{signal}), and u_j is the peak height at residue j indexed over all the residues that lie outside the binding patch (total number = N^{noise}). Since the function is divergent when the noise term is zero, we arbitrarily set the noise floor to 0.0001; SNR is set to 0.0 when the signal is zero.

This approach allows for scanning the protein surface for allosteric ligand binding sites, ranking ligands by their preference for a given site, or establishing threshold SNR

dx.doi.org/10.1021/ci400474u | J. Chem. Inf. Model. 2014, 54, 530−538

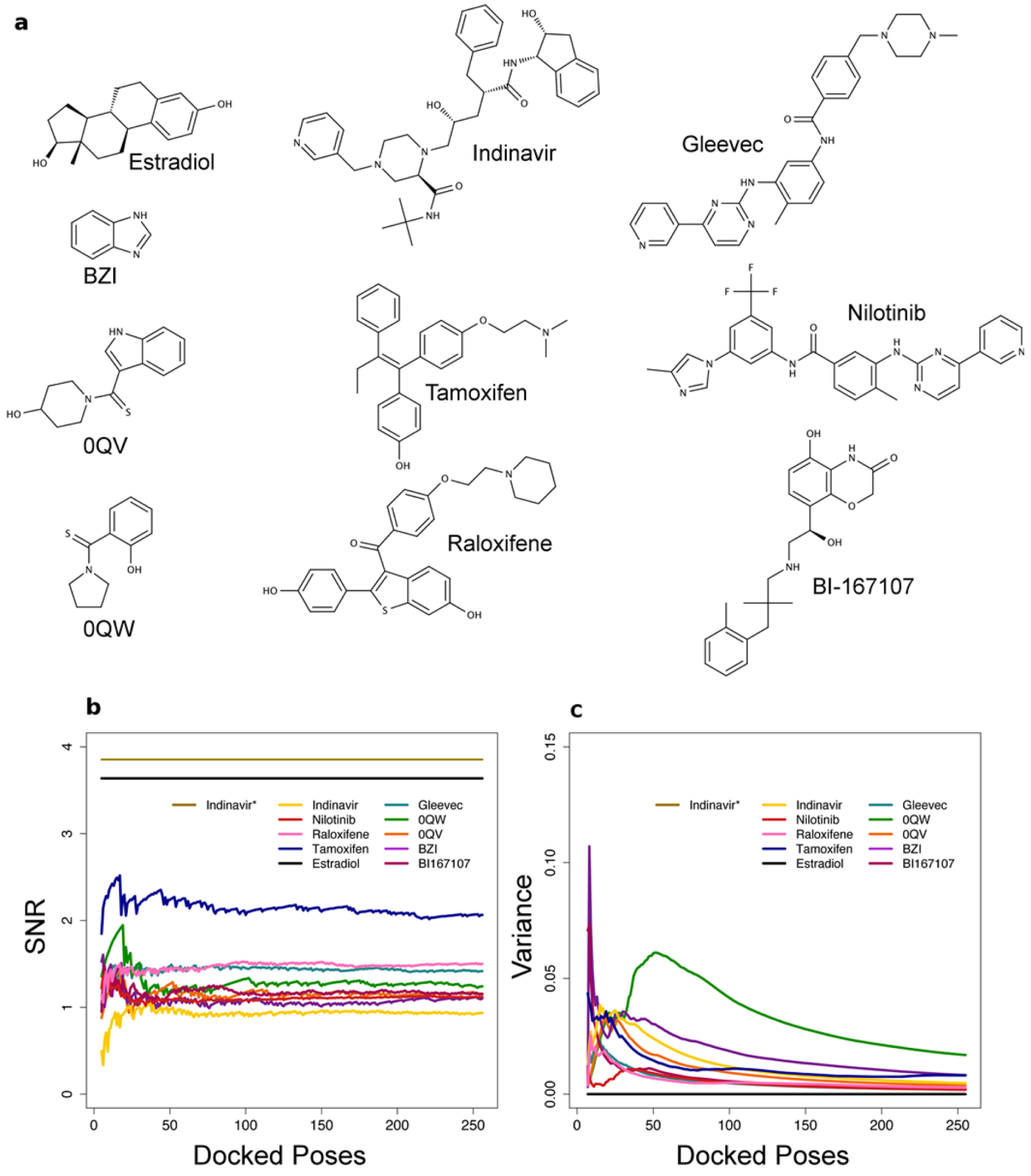


Figure 2. Global docking followed by Signal-to-Noise Ratio (SNR) analysis for our test cases. (a) Chemical structure of ligands used to evaluate the performance of LIBSA on known protein−ligand complexes. (b) Convergence of SNR values during docking (5−256 docked poses without filters) to the value reported in Table 1 (see No filter column). (c) The cumulative variance of SNR decreases as the number of poses increases. Estradiol, which has only 2 rotatable bonds, and Indinavir*, whose torsions were fixed, targeted a single site and therefore yielded a constant SNR regardless of the number of docked poses used. BZI = benzimidazole, 0QV = (4-hydroxypiperidin-1-yl)(1H-indol-3-yl)methanethione, 0QW = (2-hydroxyphenyl)(pyrrolidin-1-yl)methanethione.

values below which binding could be deemed nonspecific (see Results and Discussion). One caveat is that this procedure cannot be easily automated for receptors whose ligand binding site(s) is not well-defined. One solution could be to build a series of patches over the entire surface and compare the final SNRs or use prior knowledge about the target including interaction sites or sites of post-translational modification (e.g., acetylation, phosphorylation, ubiquitination).

High-Pass Filter: Noise Reduction Based on Contact Frequency. Concepts from digital signal processing such as high-pass filters can be used to directly remove docking noise from a contact spectrum by treating it as a discrete signal. Here we use the Butterworth filter,[19] though other types of filters could also be used. A fifth order Butterworth high-pass filter with a critical value equal to the Golden ratio conjugate[36,37] (or golden section taken as 0.618) efficiently removes less frequent

dx.doi.org/10.1021/ci400474u | J. Chem. Inf. Model. 2014, 54, 530−538

contact signals and allows for predominant peaks to be more easily identified. The formulation we used is as follows

$$H(u) = \frac{1}{\sqrt{1 + \left(\frac{u_c^2}{u^2}\right)^n}} \quad (2)$$

where u_c is the golden section, u is the input spectrum, and n is 5. Note that in this analysis, the contact spectra should all be internally normalized to the maximum value in order to use a consistent cutoff across the data. Moreover, using an appropriate cutoff is crucial because, if the cutoff value is set too low, the noise will be amplified and if it is set too high the signal is not properly enhanced. A contact spectrum can be constructed from a single ligand−receptor pair or by combining the results from multiple ligand−receptor pairs into a single histogram. In either case, the general procedure involves generating a contact spectrum as described in the previous section, applying the high-pass filter and computing the SNR.

Although high-pass and affinity filters serve somewhat different purposes and operate on different spectrum types, their end result is the same: remove docking noise. Therefore, either can be used in many situations. One advantage of applying a high-pass filter on contact spectra is that it makes no assumption about the relationship between affinity and ligand-binding residues (i.e., affinity is completely removed from the postdocking analysis).

Consensus Ligand Binding Site Identification with Ensemble SNR. In the context of docking, the most notable difference between NMR/X-ray crystallographic and MD data is the number of structures involved. The former is typically one or several to dozens, while the latter can be in the millions, which complicates the search for emergent binding sites. Ensemble SNR, which is a simple extension of the procedures discussed above, can help tackle this problem. The key here is to cluster the receptor conformations into N groups (denoted α, β, γ in Figure 1) based on either global or local structural features, such as solvent accessible surface area, principal components, pocket volume, backbone RMSD, etc. Then a single contact spectrum can be built by combining the data from all members of a cluster (e.g., α), which can then be used to compute SNR for use in consensus-binding site identification. Thus, this approach utilizes the results from multiple, independent docking experiments and produces a single number that describes how frequently a ligand binds to a given region. Such an ensemble-averaged SNR can improve prediction as the results are not contingent upon a single structure but on multiple structures.[14,38] Its major drawback lies in the assumption that the pocket is present on most if not all of the structures within a cluster. It is thus prudent to first examine individual SNRs to ensure that the desired pocket is present in the majority of the cluster members before proceeding to generating an ensemble-averaged SNR.

RESULTS AND DISCUSSION

1). Validation of LIBSA. *Blind Redocking of Cocrystallized Ligands.* A set of 10 protein−ligand complexes (Figure 2a) was used to assess the ability of the method to reliably identify known ligand binding sites. Table 1 shows that calculation of SNR from the raw docking data yields values equal to or greater than 1.0 for each system. While SNR > 0.0 (see eq 2) can be regarded as a successful identification of the binding site, SNR ≥ 1.0 is even better because this means that there are at least 10-times more hits in the active site than elsewhere on the protein. Importantly, preceding the SNR calculations by affinity filtering led to an even larger (i.e., better) SNR for half of the test cases.

Many docking algorithms including AutoDock are sensitive to the total number of rotatable bonds on the ligand.[39] It is therefore expected that the number of flexible torsions will likewise affect postprocessing by SNR. Indeed, the smallest SNR was found for Indinavir, which has the largest number of active torsions (14). This is likely because it overwhelmed the search algorithm in AutoDock, as suggested by the dramatically larger SNR (3.85) obtained when the ligand is docked with all its torsions fixed to their crystal structure values (see Indinavir* in Table 1 and Figure 2b). Therefore, the complexity of the probe to be used in LIBSA will depend on the capability of the program used for docking. Similarly, convergence of SNR should depend on the number of docked poses generated, as shown by the profile of the SNR calculated from contact spectra with 5−256 docked poses (Figure 2b). Excluding estradiol and Indinavir* whose SNR was invariant (Table S2), convergence was achieved (Figure 2b) and the variance plateaued (Figure 2c) after about 100 runs, with the coefficient of variation being 0.01−0.03 for the additional 100−256 runs.

Overall, LIBSA identified the right binding site (Table 1 and Figure 2) and recovered the correct ligand pose with reasonably small root-mean-square deviations (RMSD) from the crystal poses (Table 1). Moreover, affinity filtering increased the SNR for half of the ligand−protein pairs without affecting the rest, supporting our expectation that low probability poses are characteristic of docking noise.

Blind Docking of Known Ligands on apo Structures. To examine if LIBSA could identify the correct binding site on structures solved without the probe ligand, we blind-docked estradiol on ERα, Gleevec, and Nilotinib on p38 MAPK, BI-167107 on β2AR, and Tamoxifen and Raloxifene on ERα. Indinavir was excluded due to its large number of rotatable bonds, as discussed in the previous section. Of these, visual analysis suggested that the ligand-binding pocket on the *apo* structure of p38 is open and appears suitable for binding, whereas that of ERα is small and β2AR's is closed. As shown in Table 2, LIBSA predicted a positive binding preference of ∼0.7 for both p38 ligands and 3.61 for the small ligand estradiol on ERα. In contrast, the SNR is negative when the pocket is closed (β2AR) or the ligand is too large to fit in the pocket (Tamoxifen/Raloxifene on ERα). This result provides additional evidence that LIBSA can discriminate between favorable and unfavorable binding.

For a further and more stringent test we docked three ligands that were solved with K-Ras (0QW, 0QV, and BZI) onto two X-ray and one MD *apo* structures of the homologous protein H-Ras. The H-Ras structures were chosen because they display a pocket similar to that seen in the ligand-bound K-Ras. LIBSA yielded SNR values of ∼0.6−1.6 for these pairs (Table 2), showing that our tools are robust and applicable to diverse problems. Finally, cross docking of the chemically somewhat similar Gleevec and Nilotinib (Tanimoto coefficient of 0.6)[40] on c-Abl and p38 kinases resulted in SNR = 1.05 for both. Thus LIBSA was able to recognize the potential of ligands with similar chemical signatures to have similar binding profiles.

In sum, LIBSA was able to correctly identify binding sites on *apo* structures derived from crystallography or MD (Table 2), though the success rate is somewhat smaller than in "re-docking" (Table 1). Taken together, these results highlight the

dx.doi.org/10.1021/ci400474u | J. Chem. Inf. Model. 2014, 54, 530−538

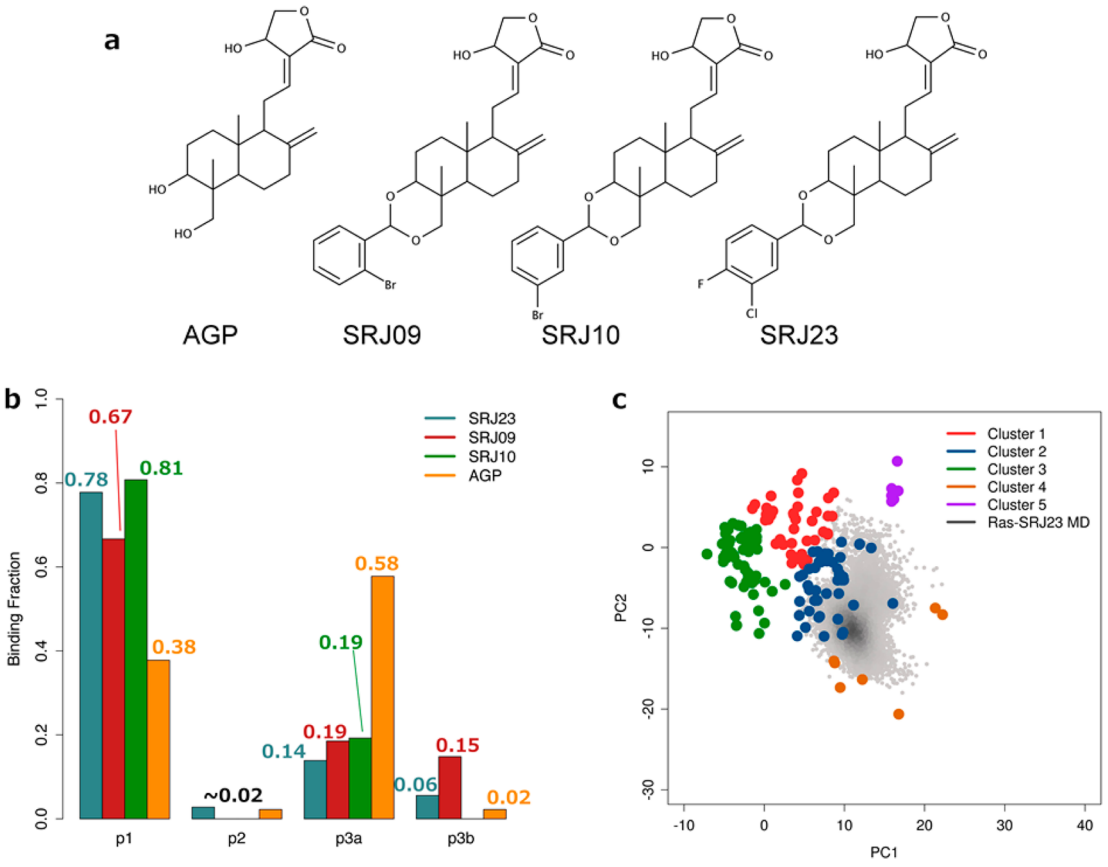


Figure 3. LIBSA of four highly related ligands docked onto an ensemble of Ras conformers. (a) Chemical structure of Andrographolide, SRJ09, SRJ10, and SRJ23. (b) Distribution of the hits with SNR >1.0 at sites p1, p2, p3a, and p3b, showing preference of the SRJ ligands for p1 (~70−80%) and Andrographolide for p3a (58%). (c) PC projection of the 148 Ras conformations used for docking (clustered into 5 groups based on their first two PCs) as well as conformers from a trajectory of p1-bound SRJ23-K-RasQ61H complex (gray shade). Clusters 2 and 4, which overlay well with conformers of SRJ23-bound K-RasQ61H, bind ligands at p1 according to the ensemble SNRs shown in Figure 4.

potential of global docking combined with our analytic tools to numerically describe ligand-binding preference, identify putative binding sites, and filter out off target poses.

2). Ranking Ligands and Classifying Receptor Conformations by Binding Specificity. Here we used an ensemble of 148 Ras conformers (see Methods) and four ligands (Figure 3a) from our previous work (AGP, SRJ09, SRJ10, and SRJ23)[8] to illustrate how chemically similar yet pharmacologically different[8,41] compounds may exhibit distinct pocket and conformation preferences when analyzed by LIBSA. Focusing only on four previously described sites,[9] we calculated SNR after removing nonspecific hits by affinity filtering. The results (Table 3) show that the four compounds predominantly target p1 or p3a (SNR > 0.0) and rarely visit pockets 2 and p3b. In fact, SNR ≥1.0 was obtained only for p1 and p3a. The distribution of the scores in Figure 3b further shows that the SRJ compounds target p1 on 70−80% of the conformers, whereas AGP targets p1 and p3a on 40% and 60% of the conformers. Binding at p2 and p3b is negligible. Thus, the four ligands can be ranked by their preference for p1: SRJ10 > SRJ23 > SRJ09 > AGP or p3a: AGP ≫ SRJ10 ≈ SRJ09 > SRJ23. This shows that LIBSA can effectively identify binding sites preferred by a specific ligand and prioritize ligands by their preference for a given pocket.

To check if AGP and its derivatives preferentially (SNR ≥1.0) hit a given site on the same set of conformers, we used the Jaccard similarity coefficient, J, defined here as the ratio between the number of conformers targeted by both ligands a and b ($n_{a,b}$) and the total number of conformers targeted by either ligand ($n_a + n_b$):

$$J = \frac{n_{a,b}}{n_a + n_b} \tag{3}$$

Table 4 shows that the rather similar SRJ compounds target the same pocket on the same set of conformers about 50% of the time on average, whereas the chemically more divergent AGP has less in common with the SRJs in terms of both its pocket and conformation preference. This is consistent with our observation from cross docking on Abl and p38 and highlights yet another utility of LIBSA.

As mentioned earlier, affinity filtering can be replaced or complemented by a high-pass digital filter if groups of structurally related receptor conformations exhibit similar tendencies to bind small molecules. We thus divided the Ras conformers into 5 groups with K-means clustering using Euclidean distances based on principal components, analyzed each group separately with the high-pass filter and SNR, and compared the results. This led to the following observations (Figures 3, 4, and S2 and Table 3): (i) the SRJ compounds have a preference (SNR ≥ 1.0) for p1 in clusters 2 and 4 and p3a in clusters 1 and 3. (ii) In contrast, AGP favors p1 only in cluster 4 and p3a in clusters 1, 3, and 5. (iii) Cluster 5 shows no tendency to bind any of the SRJ compounds. Thus, a combination of structure clustering, high-pass filtering, and

dx.doi.org/10.1021/ci400474u | J. Chem. Inf. Model. 2014, 54, 530−538

Table 3. LIBSA-Predicted Ensemble-Average Binding Preferences of AGP and Its Derivatives for Specific Pockets on Ras[a]

		SNR			
ligand	Ras cluster	p1	p2	p3a	p3b
	1	−2.56	−2.24	**3.80**	***0.53***
	2	***0.83***	***0.13***	***0.63***	−0.13
AGP	3	−1.27	−0.39	**1.66**	***0.36***
	4	**1.59**	***0.12***	−3.54	−3.89
	5	−1.39	−0.14	**1.31**	−0.22
	1	−1.70	−2.25	**1.67**	***0.68***
	2	**1.07**	0.03	−0.02	***0.40***
SRJ09	3	−0.30	−0.26	**1.10**	***0.75***
	4	**1.52**	***0.15***	−4.17	−3.91
	5	***0.36***	−0.43	***0.73***	***0.44***
	1	−1.60	−1.95	**2.65**	***0.91***
	2	**1.15**	***0.13***	−0.22	***0.13***
SRJ10	3	−0.40	−0.23	**1.20**	***0.58***
	4	**1.54**	***0.10***	−3.81	−3.65
	5	***0.33***	−0.29	***0.70***	***0.13***
	1	−1.81	−2.04	**1.75**	***0.82***
	2	**1.47**	***0.29***	−0.39	−0.29
SRJ23	3	−0.19	−0.06	***0.98***	−0.01
	4	**1.60**	***0.16***	−5.00	−5.51
	5	***0.28***	−0.26	***0.77***	−1.14

[a]Receptor clusters were defined using K-means clustering based on principal components (see text). SNR scores were calculated after applying high-pass filter to the raw data. Based on the definition of SNR (eq 2) and the benchmark data in Tables 1 and 2, SNR ≥ 1.0 (highlighted in bold) implies high preference for a given pocket, 0 < SNR < 1.0 (bold-italic) moderate preference and SNR ≤ 0.0 not favored.

Table 4. Ligand Binding Similarity among MD-Derived Ras Conformers[a]

	AGP	SRJ09	SRJ10	SRJ23
AGP		**0.40** (10/25)	**0.41** (11/27)	**0.41** (13/32)
SRJ09	**0.11** (3/27)		**0.56** (14/25)	**0.48** (15/31)
SRJ10	**0.11** (3/27)	**1.0** (5/5)		**0.49** (16/33)
SRJ23	**0.11** (3/27)	**0.43** (3/7)	**0.43** (3/7)	

[a]The data represent quantification of the ability of closely related ligand pairs to target a given pocket on the same set of Ras conformers, calculated as the ratio between the number of Ras conformers targeted by both ligands and the total number of conformers targeted by either ligand. Upper triangle: pocket p1, lower triangle: pocket p3a.

SNR calculation can isolate reasonably well a group of receptor conformers that preferentially bind a similar set of ligands.

It should be noted that the ensemble contact spectrum used here does not represent a single conformation but multiple receptor conformations in tandem (see Methods). Additionally, application of the high-pass filter does not directly correlate with the simple inclusion or exclusion of docked poses, as does affinity filtering. As a result, the scaled spectrum cannot be represented by a single structure. That said, the receptor conformations identified by a high-pass filter plus SNR as favoring p1 binding (clusters 2 and 4) resemble those from molecular dynamics of SRJ23-K-RasQ61H (Figure 3c and ref 8). They are also consistent with those found by looking at the high-specificity ligand−receptor pairs in a piecewise fashion (discussed above). Thus, irrespective of the specific filter employed, LIBSA provides information on the receptor conformations that are best suited to binding small molecules at a given pocket.

3). Optimizing Binding Affinity. It is important to note that whereas LIBSA can potentially identify a drug core and/or receptor conformations that are suitable for ligand binding, it is unlikely to yield high affinity hits with desired biochemical properties. It is therefore important that LIBSA is followed up with site-directed VS to optimize potential hits or identify new ones. For example, a library of compounds could be generated using programs such as DAIM,[42] and a pocket can be characterized as preferentially binding a molecule with LIBSA. Then other methods such as BOMB,[43] GANDI,[44] and BROOD[45] can be used to generate and optimize additional compounds which can then be screened against only the most relevant receptor conformations. Thus the initial library can be built into a targeted ligand library that is tailored toward a particular pocket on a particular receptor conformation.

CONCLUDING REMARKS

We have presented a computational framework for identifying a consensus-binding site on a single or ensemble of receptor conformations. The method relies on three simple yet novel techniques to remove nonrelevant docked poses and compute binding preference. The first tool, affinity filtering, removes low frequency docked poses based on the distribution of affinity scores. The second is a high-pass filter that scales down protein−ligand contacts based on their probability of occurrence. The third technique combines probing the surface of a target protein by blind docking with the concept of signal-to-noise-ratio (SNR) to identify allosteric sites that have the potential to bind small molecule ligands. Computation of SNR can be preceded with any of the two filtering methods and can be applied on data generated from blind docking with any docking algorithm or other methods that can scan the surface of a target receptor with druglike molecular probes. The resulting protocol, termed LIBSA, provides a metric for identifying binding sites and ranking ligands by their consistency of binding to these sites. We have demonstrated the usefulness of this approach by applying it on a diverse set of known ligands and their receptors as well as a small set of related ligands docked onto a large ensemble of Ras conformers with multiple binding sites. LIBSA was able to correctly identify the known active sites as the preferred binding site for the respective ligands and predicted the preference of each of our four test ligands for a particular pocket on the Ras structures.

Our approach is similar in spirit with an earlier study that has shown that binding consistency is a necessary condition for successful docking using multiple runs of a genetic algorithm.[46] However, the previous study has focused on identifying a consistent binding mode at a particular site, whereas the goal of LIBSA was to simultaneously identify a binding site and ligands that bind consistently to that site. Furthermore, in principle, more rigorous methods such as MD[47] can be used for binding site identification and scoring. For example, Huang et al used microsecond scale explicit solvent MD simulations to show that Darunavir binds to HIV protease in a completely different mode than that found in the starting crystal structure.[48] However, while potentially more accurate, this approach is too expensive to be used for screening tens of thousands of probes against a library of receptor conformers. The computationally

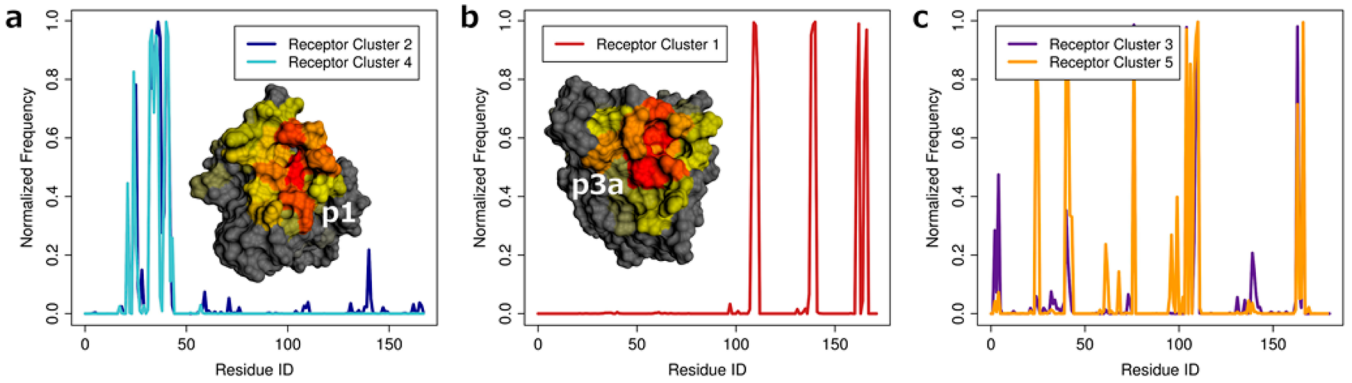


Figure 4. Ensemble contact spectra for Ras using SRJ23 as probe (normalized by the largest score). (a) The dominant peak in the ensemble-averaged contact spectra of clusters 2 and 4 yielded a p1 SNR > 1.0 after noise removal by a high-pass filter. (b) The four large peaks in the ensemble-averaged contact spectrum of cluster 1 yielded a p3a SNR > 1.0 after noise removal by a high-pass filter. (c) The cluttered spectra for clusters 3 and 5 yielded insignificant SNRs even after noise removal due to nonspecific binding. Insets in (a) and (b) illustrate binding to pockets p1 and p3a using the surface of representative Ras structures colored by contact spectrum (red-to-black: high to low binding probability).

much more efficient LIBSA can be used alone or in conjunction with MD, depending on need and the necessary tradeoff between accuracy and computational efficiency. Moreover, each of the techniques described in this work can be used either for analyzing the binding mode of known drugs retrospectively or for a prospective design of new inhibitors.

■ ASSOCIATED CONTENT

Supporting Information

Additional information regarding the MD protocol, summary of trajectories, SNR convergence plots with and without the affinity filter, unfiltered ensemble contact spectra, and a summary of the SNR values. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: 713-500-7538. Fax: 713-500-7444. E-mail: Alemayehu.G.Abebe@uth.tmc.edu.

Author Contributions

H.J.H., N.R., and A.A.G. performed research. H.J.H., N.R., and A.A.G. contributed reagents. H.J.H. and A.A.G. designed research and wrote the paper.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Dr. Hualin Li for suggesting the Golden Ratio conjugate for noise reduction and other helpful discussions and the Texas Advanced Computing Center for computational resources. H.J.H. acknowledges support from the Keck Gulf Coast Consortia Training in Pharmacological Sciences and the National Institute of General Medical Sciences (grant #T32GM089657). We thank the University of Texas Medical School at Houston for financial support.

■ ABBREVIATIONS

MD, molecular dynamics; SNR, signal-to-noise ratio; LIBSA, ligand binding specificity analysis; NMR, nuclear magnetic resonance spectroscopy

■ REFERENCES

(1) Scannell, J. W.; Blanckley, A.; Boldon, H.; Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discovery* **2012**, *11*, 191−200.

(2) Shima, F.; Yoshikawa, Y.; Ye, M.; Araki, M.; Matsumoto, S.; Liao, J.; Hu, L.; Sugimoto, T.; Ijiri, Y.; Takeda, A.; Nishiyama, Y.; Sato, C.; Muraoka, S.; Tamura, A.; Osoda, T.; Tsuda, K.-i.; Miyakawa, T.; Fukunishi, H.; Shimada, J.; Kumasaka, T.; Yamamoto, M.; Kataoka, T. In silico discovery of small-molecule Ras inhibitors that display antitumor activity by blocking the Ras−effector interaction. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 8182−8187.

(3) Schneider, G. Virtual screening: an endless staircase? *Nat. Rev. Drug Discovery* **2010**, *9*, 273−276.

(4) Yuriev, E.; Agostino, M.; Ramsland, P. A. Challenges and advances in computational docking: 2009 in review. *J. Mol. Recognit.* **2011**, *24*, 149−164.

(5) Christopoulos, A. Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. *Nat. Rev. Drug Discovery* **2002**, *1*, 198−210.

(6) Schames, J.; Henchman, R.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, A. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem.* **2004**, *47*, 1879−1881.

(7) Brenke, R.; Kozakov, D.; Chuang, G.-Y.; Beglov, D.; Hall, D.; Landon, M. R.; Mattos, C.; Vajda, S. Fragment-based identification of druggable 'hot spots' of proteins using Fourier domain correlation techniques. *Bioinformatics* **2009**, *25*, 621−627.

(8) Hocker, H. J.; Cho, K.-J.; Chen, C.-Y. K.; Rambahal, N.; Sagineedu, S. R.; Shaari, K.; Stanslas, J.; Hancock, J. F.; Gorfe, A. A. Andrographolide derivatives inhibit guanine nucleotide exchange and abrogate oncogenic Ras function. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 10201−10206.

(9) Grant, B.; Lukman, S.; Hocker, H. J.; Sayyah, J.; Brown, J. H.; McCammon, J. A.; Gorfe, A. A. Novel allosteric sites on Ras for lead generation. *PLoS One* **2011**, *6*, e25711.

(10) Verkhivker, G. M.; Bouzida, D.; Gehlhaar, D. K.; Rejto, P. A.; Freer, S. T.; Rose, P. W. Algorithms for computational solvent mapping of proteins. *Proteins: Struct., Funct, Bioinf.* **2003**, *53*, 201−219.

(11) Bakan, A.; Nevins, N.; Lakdawala, A. S.; Bahar, I. Druggability assessment of allosteric proteins by dynamics simulations in the presence of probe molecules. *J. Chem. Theory Comput.* **2012**, *8*, 2435−2447.

(12) 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.* **2005**, *49*, 5912−5931.

(13) Benedix, A.; Becker, C. M.; de Groot, B. L.; Caflisch, A.; Bockmann, R. A. Predicting free energy changes using structural ensembles. *Nat. Methods* **2009**, *6*, 3−4.
(14) Lin, J.-H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.* **2002**, *124*, 5632−5633.
(15) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discovery* **2004**, *3*, 935−49.
(16) Barreiro, G.; Kim, J. T.; Guimarães, C. R. W.; Bailey, C. M.; Domaoal, R. A.; Wang, L.; Anderson, K. S.; Jorgensen, W. L. From docking false-positive to active anti-HIV agent. *J. Med. Chem.* **2007**, *50*, 5324−5329.
(17) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785−2791.
(18) Hetenyi, C.; Vanderspoel, D. Blind docking of drug-sized compounds to proteins with up to a thousand residues. *FEBS Lett.* **2006**, *580*, 1447−1450.
(19) Butterworth, S. On the theory of amplifiers. *Wireless Engineer* **1930**, *7*, 536−541.
(20) Vary, P.; Eurasip, M. Noise suppression by spectral magnitude estimation —mechanism and theoretical limits. *Signal Processing* **1985**, *8*, 387−400.
(21) Nagar, B.; Bornmann, W. G.; Pellicena, P.; Schindler, T.; Veach, D. R.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* **2002**, *62*, 4236−4243.
(22) Shiau, A. K.; Barstad, D.; Loria, P. M.; Cheng, L.; Kushner, P. J.; Agard, D. A.; Greene, G. L. The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **1998**, *95*, 927−937.
(23) Bruning, J. B.; Parent, A. A.; Gil, G.; Zhao, M.; Nowak, J.; Pace, M. C.; Smith, C. L.; Afonine, P. V.; Adams, P. D.; Katzenellenbogen, J. A. Coupling of receptor conformation and ligand orientation determine graded activity. *Nat. Chem. Biol.* **2010**, *6*, 837−843.
(24) Gangloff, M.; Ruff, M.; Eiler, S.; Duclaud, S.; Wurtz, J. M.; Moras, D. Crystal structure of a mutant hERα ligand-binding domain reveals key structural features for the mechanism of partial agonism. *J. Biol. Chem.* **2001**, *276*, 15059−15065.
(25) Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **2011**, *477*, 549−557.
(26) Chen, Z.; Li, Y.; Chen, E.; Hall, D. L.; Darke, P. L.; Culberson, C.; Shafer, J. A.; Kuo, L. C. Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *J. Biol. Chem.* **1994**, *269*, 26344−26348.
(27) Maurer, T.; Garrenton, L. S.; Oh, A.; Pitts, K.; Anderson, D. J.; Skelton, N. J.; Fauber, B. P.; Pan, B.; Malek, S.; Stokoe, D.; Ludlam, M. J. C.; Bowman, K. K.; Wu, J.; Giannetti, A. M.; Starovasnik, M. A.; Mellman, I.; Jackson, P. K.; Rudolph, J.; Wang, W.; Fang, G. Small-molecule ligands bind to a distinct pocket in Ras and inhibit SOS-mediated nucleotide exchange activity. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 5299−5304.
(28) Sun, Q.; Burke, J. P.; Phan, J.; Burns, M. C.; Olejniczak, E. T.; Waterson, A. G.; Lee, T.; Rossanese, O. W.; Fesik, S. W. Discovery of small molecules that bind to K-Ras and inhibit Sos-mediated activation. *Angew. Chem., Int. Ed. Engl.* **2012**, *124*, 6244−6247.
(29) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* **2007**, *23*, 2625−2627.
(30) Hartigan, J. A.; Wong, M. A. Algorithm AS 136: A k-means clustering algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1979**, *28*, 100−108.
(31) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, A.; Gao, J.; Gao, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E. I.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(32) Humphrey, W.; Dalke, A.; Schulten, K. VMD - visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33−38.
(33) O'Boyle, N.; Banck, M.; James, C.; Morley, C.; Vandermeersch, T.; Hutchison, G. Open Babel: an open chemical toolbox. *J. Cheminform.* **2011**, *3*, 1−14.
(34) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(35) Gasteiger, J.; Marsili, M. Iterative partial equilization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron* **1980**, *36*, 3219−3228.
(36) Dunlap, R. A. *The golden ratio and Fibonacci numbers*; World Scientific: Singapore, 1997.
(37) Boroden, C. *Fibonacci Trading: How to Master the Time and Price Advantage*; McGraw Hill Professional: New York, 2008.
(38) Amaro, R.; Baron, R.; McCammon, J. A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 693−705.
(39) Chang, M. W.; Ayeni, C.; Breuer, S.; Torbett, B. E. Virtual screening for HIV protease inhibitors: a comparison of AutoDock 4 and Vina. *PLoS One* **2010**, *5*, e11955.
(40) Manley, P. W.; Stiefl, N.; Cowan-Jacob, S. W.; Kaufman, S.; Mestan, J.; Wartmann, M.; Wiesmann, M.; Woodman, R.; Gallagher, N. Structural resemblances and comparisons of the relative pharmacological properties of imatinib and nilotinib. *Bioorg. Med. Chem.* **2010**, *18*, 6977−6986.
(41) Jada, S. R.; Matthews, C.; Saad, M. S.; Hamzah, A. S.; Lajis, N. H.; Stevens, M. F. G.; Stanslas, J. Benzylidene derivatives of andrographolide inhibit growth of breast and colon cancer cells in vitro by inducing G1 arrest and apoptosis. *Br. J. Pharmacol.* **2008**, *155*, 641−654.
(42) Kolb, P.; Caflisch, A. Automatic and efficient decomposition of two-dimensional structures of small molecules for fragment-based high-throughput docking. *J. Med. Chem.* **2006**, *49*, 7384−7392.
(43) Jorgensen, W. L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.* **2009**, *42*, 724−733.
(44) Dey, F.; Caflisch, A. Fragment-based de novo ligand design by multiobjective evolutionary optimization. *J. Chem. Inf. Model.* **2008**, *48*, 679−690.
(45) Wang, L.-h.; Evers, A.; Monecke, P.; Naumann, T. Ligand based lead generation - considering chemical accessibility in rescaffolding approaches via BROOD. *J. Cheminf.* **2012**, *4*, 1−1.
(46) Cecchini, M.; Kolb, P.; Majeux, N.; Caflisch, A. Automated docking of highly flexible ligands by genetic algorithms: a critical assessment. *J. Comput. Chem.* **2004**, *25*, 412−422.
(47) Durrant, J.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71.
(48) Huang, D.; Caflisch, A. How does darunavir prevent HIV-1 protease dimerization? *J. Chem. Theory Comput.* **2012**, *8*, 1786−1794.