

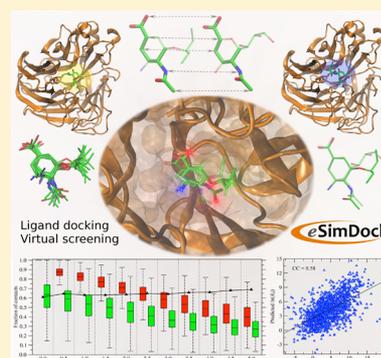
Nonlinear Scoring Functions for Similarity-Based Ligand Docking and Binding Affinity Prediction

Michal Brylinski*

Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, United States

Center for Computation & Technology, Louisiana State University, Baton Rouge, Louisiana 70803, United States

ABSTRACT: A common strategy for virtual screening considers a systematic docking of a large library of organic compounds into the target sites in protein receptors with promising leads selected based on favorable intermolecular interactions. Despite a continuous progress in the modeling of protein–ligand interactions for pharmaceutical design, important challenges still remain, thus the development of novel techniques is required. In this communication, we describe *eSimDock*, a new approach to ligand docking and binding affinity prediction. *eSimDock* employs nonlinear machine learning-based scoring functions to improve the accuracy of ligand ranking and similarity-based binding pose prediction, and to increase the tolerance to structural imperfections in the target structures. In large-scale benchmarking using the Astex/CCDC data set, we show that 53.9% (67.9%) of the predicted ligand poses have RMSD of $<2 \text{ \AA}$ ($<3 \text{ \AA}$). Moreover, using binding sites predicted by recently developed *eFindSite*, *eSimDock* models ligand binding poses with an RMSD of 4 \AA for 50.0–39.7% of the complexes at the protein homology level limited to 80–40%. Simulations against non-native receptor structures, whose mean backbone rearrangements vary from 0.5 to 5.0 \AA α -RMSD, show that the ratio of docking accuracy and the estimated upper bound is at a constant level of ~ 0.65 . Pearson correlation coefficient between experimental and predicted by *eSimDock* K_i values for a large data set of the crystal structures of protein–ligand complexes from BindingDB is 0.58, which decreases only to 0.46 when target structures distorted to 3.0 \AA α -RMSD are used. Finally, two case studies demonstrate that *eSimDock* can be customized to specific applications as well. These encouraging results show that the performance of *eSimDock* is largely unaffected by the deformations of ligand binding regions, thus it represents a practical strategy for across-proteome virtual screening using protein models. *eSimDock* is freely available to the academic community as a Web server at <http://www.brylinski.org/esimdock>.



INTRODUCTION

Due to extremely high costs of high-throughput screening, many drug discovery projects commonly employ inexpensive computations to support experimental efforts. Here, the idea is to considerably reduce the number of candidate compounds that need to be tested experimentally against a protein target of interest. As of September 2013, the ZINC database of commercially available small molecule entities for drug discovery contains 15,798,630 drug-like compounds collected from the catalogues of 209 vendors.¹ At the outset of drug development process, this vast array of drug candidates representing a broad chemical diversity must be considerably downsized to typically hundreds to thousands of molecules, which can be further subject to experimental screens. Consequently, a significant interest in virtual screening technologies continues to stimulate the development of novel computational approaches,² which currently branch into two distinct categories.³ One group of methods comprises ligand-based virtual screening tools.^{4,5} These algorithms require an initial set of known bioactive compounds, which are used to build a consensus pharmacophore model of the target receptor. Subsequently, any number of molecules can be compared to this model to identify a small subset of highly compatible, thus promising lead candidates. Individual methods within this

category employ different strategies to extract chemical information from known binders in order to construct functional pharmacophores. For example, various types of molecular fingerprints containing numerical descriptors⁶ and structural keys⁷ are widely used in 2D chemical similarity analysis.⁸ An increase in the performance of fingerprint-based methods can be achieved by using data fusion,⁹ machine learning,¹⁰ and profile scaling techniques.¹¹ Another group comprises qualitative structure–activity relationship, QSAR, models, which correlate “response” variables, e.g. bioactivity, with a set of “predictor” variables, such as the physicochemical properties or theoretical molecular descriptors of organic compounds. QSAR methods can be categorized based on dimensionality into 6 classes,¹² varying from the simplest methods using molecular properties like pK_a and $\log P$,¹³ to more sophisticated that consider ensembles of ligand configurations,¹⁴ explicit induced fit protocols,¹⁵ and various solvation models.¹⁶ Finally, shape similarity-based methods have been developed; these approaches exploit the fact that small organic molecules bind into a defined and complex site on a macromolecular target surface, thus molecular shape is an

Received: September 1, 2013

Published: October 30, 2013

important indicator of biological activity.¹⁷ A common feature of ligand-based strategies is that they require only an initial set of known binders. Moreover, it has been demonstrated that when this information is unavailable for a target of interest, evolutionarily derived compounds can be used instead.¹⁸

Contrastingly, structure-based methods do not require a priori knowledge on previously identified active compounds; however, a three-dimensional structure of the target protein must be available. At the core of these techniques is molecular docking, which predicts the optimal orientation of drug candidates with respect to the target structure; this is followed by applying a scoring function to estimate binding affinity.¹⁹ Each docking algorithm incorporates two components, which are critical to the accuracy of protein–ligand modeling: a search algorithm and a scoring function. Common search algorithms include molecular dynamics²⁰ and Monte Carlo^{21,22} simulations, genetic algorithms,^{23,24} and shape-complementarity methods²⁵ as well as systematic pose generation followed by stochastic optimization.²⁶ Docking is computationally challenging due to the size of protein–ligand conformational space, which grows exponentially with the size of docking molecules. Furthermore, a large number of commercially available drug candidates pose considerable constraints on docking times, often compromising the level of detail and the overall performance. The second component, a scoring function, estimates the likelihood of docking compounds to form a stable complex when bound to the target receptor. Most commonly used scoring terms employ physics-based force fields summing the strength of intermolecular van der Waals and electrostatic interactions,^{25,27} empirical functions counting the number of various interaction types,²⁸ and statistical potentials^{29,30} derived from large databases such as the Protein Data Bank (PDB).³¹ A comprehensive evaluation of several scoring functions for molecular docking demonstrated that despite most of them perform well identifying experimentally observed conformation among a large number of computer-generated decoys, only a few give moderate correlations between predicted binding scores and experimentally determined protein–ligand binding affinities.³² Another weakness of many scoring functions is a correlation between docking scores and compound molecular weight, which was reported in several studies.^{33,34} Certainly, both issues represent a significant problem in virtual screening because even with correctly predicted binding poses, true hits may still be missed.

Successful docking simulations typically require high-quality target receptor structures. This in turn complicates large-scale applications of molecular docking since the structures of many potential drug targets have not yet been determined by X-ray crystallography or NMR spectroscopy. Using protein models alleviates this problem; nevertheless, structural imperfections of theoretically constructed structures present a challenge to docking algorithms. For instance, it has been demonstrated that the performance of docking calculations is affected by the particular receptor conformation and that holo structures yield the highest accuracy.³⁵ Another study conducted for 41 protein–ligand complexes showed that docking reliability drops off dramatically with the decreasing quality of target structures; almost 90% of the initial docking accuracy is lost if the mean structural distortions are greater than 1.5 Å.³⁶ More recently, it has been suggested that, in general, docking accuracy depends on the structure quality within the binding site, thus modeling the conformations of the active site residues is critical for docking into homology models.³⁷ A growing body

of research indicates that using homology models in virtual screening yields compound selection better than random. It has been shown that docking to a homology model based on a template with only 30% global sequence identity to the target gives enrichment as high as that obtained from the crystal structure.³⁸ However, another study shows that the enrichment is consistently far better than random when the sequence homology within the binding site region is greater than 50%.³⁹ Furthermore, the accuracy of virtual screening calculations does not correlate well with the quality of modeled target structures, indicating that standard measures of structure quality are rather poor estimates of virtual screening confidence.³⁸ It has been suggested that alternative modeling protocols that account for conformational changes in the active site upon ligand binding should be considered.⁴⁰ A new “induced-fit” methodology was proposed to address the issue of substantial protein rearrangements upon ligand binding with encouraging results obtained from docking benchmarks against crystal structures;⁴¹ however, allowing for the flexibility of key side chains was also shown to produce a modest improvement in ligand docking against protein models.³⁸ Moreover, considering protein flexibility in ligand docking against non-native receptor structures does not improve the geometry of binding regions due to the imperfections of all-atom force fields and the rugged energy landscape of protein–ligand interaction space.⁴²

In that regard, similarity-based docking techniques have been developed to address the problem of ligand docking against particularly weakly homologous protein models containing significant structural distortions in binding regions.¹⁸ These algorithms work by superposing query ligands onto template compounds, which are either bound to the target protein or evolutionarily related structures. This strategy yields encouraging results; however, docking and scoring functions rely on purely geometrical criteria without considering molecular interactions with the target proteins. For example, a similarity to the ligand anchor was used to rank compounds in virtual screening.¹⁸ In this study, we describe a new approach, *eSimDock*, which further extends similarity-based docking. It incorporates a composite force field and various scoring functions evaluating a fit between ligands and the target binding sites, the probability of binding as well as binding affinity. The scoring functions are nonlinear, employing machine learning techniques. Using comprehensive data sets, we perform benchmarking calculations of *eSimDock* with respect to the accuracy of binding pose prediction, virtual screening, and the prediction of binding affinity. In general, *eSimDock* was developed to exploit binding information extracted from evolutionarily related protein–ligand complexes; therefore, we carry out additional benchmarks to quantify the effect of sequence homology on the accuracy of similarity-based docking. Furthermore, we also address the issue of using non-native protein structures as docking and screening targets. Specifically, we perform a series of simulations gradually decreasing the quality of target structures and measure the impact of structural deformations on the performance of ligand pose and binding affinity prediction. In comparative benchmarks, we analyze the performance of *eSimDock* and AutoDock Vina with respect to the accuracy of binding pose prediction as well as ligand ranking in virtual screening experiments. Finally, a couple of case studies are selected to demonstrate the flexibility of *eSimDock*, which not only represents a general approach to similarity-based ligand docking and virtual screening, but it can be easily tailored to

specific applications. Encouraging results indicate that eSimDock has all the characteristics required for large-scale applications at the proteome level, i.e. it is fast, robust, and fairly insensitive to structural distortions in modeled protein structures.

MATERIALS AND METHODS

Docking Data Sets. Two primary data sets are used in this study: Astex/CCDC and BindingDB. The former consists of 280 pharmacologically relevant high-quality crystal structures of protein–ligand complexes⁴³ that are widely used as a gold standard for assessing the accuracy of docking algorithms.^{44,45} The latter comprises 1,151 complexes selected from BindingDB,⁴⁶ for which both compound binding affinity (K_i) is experimentally determined and the crystal structure of the complex is available in the Protein Data Bank.³¹ Here, we also accept these crystal complexes, in which the binding ligand is not necessarily identical to the BindingDB compound but is significantly similar at the 2D Tanimoto coefficient (2D-TC) threshold of ≥ 0.5 .⁴⁷ In addition to the crystal structures of target receptors, we constructed several sets of distorted structures using an in-house software that employs local conformational Monte Carlo sampling to reach the desired RMSD from native.⁴⁸ For the Astex/CCDC data set, 10 non-native sets are generated with an average C α -RMSD close to 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, and 5.0 Å; for the BindingDB data set, we use 6 distorted receptor sets with 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0 Å C α -RMSD.

For each compound in the BindingDB data set, we also compiled a set of 25 ligand decoys from the ZINC7 collection of organic compounds.⁴ Here, we use a nonredundant subset of ZINC7, where the redundancy is removed at the 1D Tanimoto coefficient (1D-TC) of 0.7.⁴⁹ Decoy compounds are selected such as their average molecular weight corresponds to that of the parent binding ligand ± 50 Da. Furthermore, we make sure that these compounds are topologically dissimilar to the native ligand with a 1D-TC of < 0.3 ; the average 1D-TC across the ligand decoy sets is ~ 0.2 .

Generation of Ligand Ensembles. To speed up docking calculations, we adopt a ligand ensemble docking approach. This is a common methodology, in which first an ensemble of target compound conformations is generated, and then each conformer is docked into the binding site using rigid ligand docking. To construct a nonredundant ensemble, we generate a large set of low-energy conformations using Balloon and the MMFF94-like force field,⁵⁰ which are subsequently subject to a clustering procedure. First, we cluster conformers using a pairwise similarity threshold of 1 Å RMSD for the atomic coordinates. If the number of resulting clusters is larger than 50, we repartition them in the RMSD distance space into 50 clusters using CLUTO clustering software.⁵¹ As a result, the final ensemble comprises up to 50 low-energy conformers with a pairwise RMSD of > 1 Å.

Anchor Substructures and Chemical Matching. Similarity-based ligand docking requires two important components: an anchor substructure used as the template and a chemical matching algorithm to establish atomic equivalences between the target compound and the anchor substructure. As a chemical matching engine, we use KCOMBU, which stands for K(chemical) structure COMparison using the BUild-up algorithm.⁴⁷ This algorithm formalizes chemical matching as the maximum common subgraph problem and provides a heuristic solution; in addition to the atomic equivalences, it also

calculates a 2D-TC score that assesses the similarity between matching compounds.

In benchmarking calculations, we test three types of anchor substructures. The first one is simply a target ligand itself; this is an ideal case that delineates the upper bound of docking accuracy under perfect conditions. The second anchor type is the largest rigid fragment extracted from the target ligand. Here, the binding pose of the anchor is taken from the crystal complex structure; however, it provides only a partial coverage for the docking compound, and its flexible parts are modeled without using template coordinates as the reference positions. The last anchor type corresponds to the most practical situation, in which no ligand-bound crystal structure is available for the target receptor. Here, we detect binding pockets using eFindSite,⁵² a recently developed successor of FINDSITE,⁵³ and the target ligands as auxiliary compounds. eFindSite employs meta-threading by eThread⁵⁴ to identify evolutionary related proteins; subsequently, it finds the locations of putative binding sites and extracts template-bound compounds for use as anchor substructures in similarity-based docking. The quality of anchor substructures in terms of their chemical and structural conservation with respect to the target ligand undoubtedly depends on the level of homology between the target receptor and evolutionarily related template proteins identified by threading. Therefore, to quantify the effect of sequence homology on the accuracy of similarity-based docking, we detect binding sites and extract template ligands at different target-template sequence identity thresholds: 100%, 80%, 60%, and 40%. Furthermore, multiple anchor substructures are typically extracted from template proteins; in these cases, we use the one that has the highest chemical similarity to the target ligand measured by 2D-TC.

Similarity-Based Docking and Local Optimization. Similarity-based docking procedure implemented in eSimDock comprises two steps. First, each target ligand conformer from the ensemble is superposed onto the anchor substructure according to the atomic equivalences calculated by KCOMBU.⁴⁷ In the next step, each superposed conformer is subject to local energy minimization using AutoDock Vina.⁵⁵ This fast procedure mostly removes atom clashes keeping the docked conformation close to the anchor substructure. Also, using a nonredundant ensemble precalculated for the target ligand ensures that its internal energetically favorable conformational space is also sufficiently explored. Preprocessing input files, which includes adding partial charges and hydrogen atoms, is done using Open Babel⁵⁶ for ligands and MGL-Tools⁵⁷ for receptor proteins. All docked and energy-optimized conformers are finally subject to a scoring procedure using nonlinear machine learning models.

Nonlinear Scoring Functions. A typical scoring function for ligand docking implements a linear combination of weighted energy terms,^{25,45} however, recent studies suggest that nonlinear scoring function may yield better results.⁵⁸ Here, we developed three nonlinear scoring functions that assess fitness (how well the target ligand fits into the receptor pocket), binding (chances that the target ligand binds to the receptor), and affinity (in terms of the dissociation constant K_i). All three scoring functions use the following 9 individual structure-based energy terms: from AutoDock Vina,⁵⁵ Vina_gauss1, Vina_gauss2, Vina_repulsion, Vina_hydrophobic, Vina_hydrogen; from DrugScore,²⁹ DSX_pair, DSX_sas; from LPC,⁵⁹ LPC_complementarity; and from DFIRE for protein–ligand complexes,⁶⁰ DFIRE_pair. In addition to the structure-based

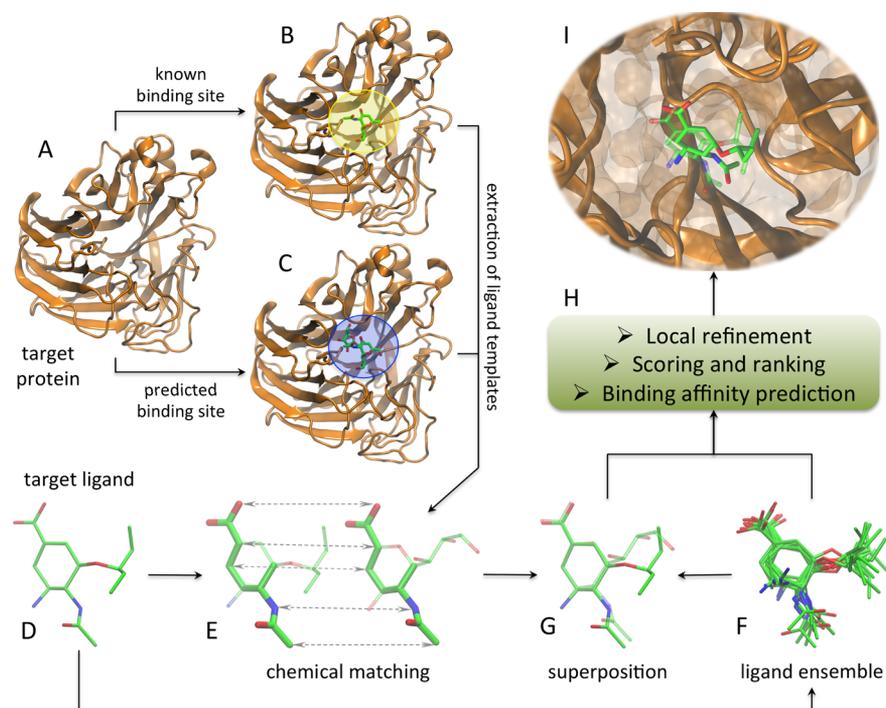


Figure 1. Flowchart of *eSimDock*. Target protein structure (A), either experimental or modeled, is used to define a ligand binding site. If the target protein is cocrystallized with small molecules, a known binding pocket is used (B), otherwise it is predicted by *eFindSite* (C). Ligand templates are extracted from the target binding pocket and chemical matching to the target ligand (D) is performed (E). An ensemble of conformations is constructed for the target ligand (F) and each conformer is superposed onto the template ligand (G). Superposed conformers are subject to local refinement followed by scoring, ranking, and binding affinity prediction by machine learning models (H). Best conformation is selected as the final docking result (I).

scoring functions, binding and affinity calculations also include a set of 5 ligand-based terms calculated for docking ligands by OpenBabel:⁵⁶ MW (molecular weight), plogp (octanol/water partitioning coefficient), PSA (polar surface area), and by MayaChemTools:⁶¹ HB_donor and HB_acceptor (number of hydrogen bond donors and acceptors, respectively).

Machine learning models are constructed by libSVM.⁶² Fitness and binding scoring functions employ Support Vector Machines (SVM) suitable for classification problems (SVC), whereas affinity prediction uses epsilon-Support Vector Regression (SVR). The training and validation of the fitness SVC model is carried out using the Astex/CCDC data set. For each complex structure, we generate a large ensemble of docking poses; native-like (positives) and misdocked (negatives) conformations are defined as these binding poses that have a heavy-atom RMSD from native of ≤ 2 Å and ≥ 3 Å, respectively. Binding SVC model is optimized on the BindingDB data set, where BindingDB compounds are positives and topologically dissimilar ZINC compounds are negatives. Affinity SVR is trained on the dissociation constants (K_i) provided by BindingDB. For each machine learning model, we perform an exhaustive grid search to find the optimal set of SVM parameters: a cost function, gamma in the kernel function (SVC and SVR), and epsilon in loss function (SVR only). Furthermore, all models are cross-validated to assess the real performance avoiding memorization effects. Specifically, leave-one-out cross-validation is applied at the receptor level, i.e. all complexes of a particular receptor with different ligands are excluded from training and then used for model validation. Finally, we constructed a couple of additional machine learning

models that are tailored for specific cases; this is described in case study sections below.

AutoDock Vina. The performance of *eSimDock* on both primary data sets, Astex/CCDC and BindingDB, is compared to AutoDock Vina.⁵⁵ Target protein structures are converted to the required PDBQT format using MGL-Tools.⁵⁷ The addition of polar hydrogens and partial charges as well as format conversion to PDBQT of ligand molecules is carried out using OpenBabel.⁵⁶ All docking simulations using Vina are performed using the default set of parameters.

Case Study: Factor Xa Data Set. The accuracy of *eSimDock* in discriminating between strong and weak binders is evaluated on a data set of 435 benzamidine inhibitors of factor Xa (fXa).⁶³ These compounds fall into two categories: weak binders with K_i higher than 1 μ M (156 molecules) and strong binders with K_i lower than 10 nM (279 molecules). First, 93 crystal structures of fXa complexed with low molecular weight inhibitors were obtained from PDB. All-against-all structure alignments identified the most “typical” fXa holo-structure with the lowest average C α -RMSD from the remaining structures (1.5 Å). This fXa structure in complex with a cis-1,2-diaminocyclohexane derivative (PDB-ID: 3iit, chain A)⁶⁴ is used for docking of benzamidine inhibitors by *eSimDock*. Next, we constructed a nonredundant set of fXa inhibitors by clustering 93 fXa-bound compounds from PDB. Here, we use the SUBSET program⁶⁵ and a 1D-TC similarity threshold of 0.7; the resulting 62 ligands are used as the set of anchor substructures for docking benzamidine inhibitors using *eSimDock*.

Here, we also demonstrate that different machine learning techniques can be used to construct a scoring function. As a

strong/weak binding classifier, we employ a Naïve Bayes model (NB) with an optimized Gaussian kernel.⁶⁶ To train and validate the discriminant capability of the NB model, we use two subsets of 290 and 145 compounds, respectively, as described in the original publication.⁶³ Similarly to SVM models, the NB model employs a set of 14 features (9 structure-based + 5 ligand-based). The performance of NB classifier is assessed by receiver operator characteristics (ROC) analysis.

Case Study: CDK2 Data Set. We also benchmark *eSimDock* in predicting the enzymatic half maximal inhibitory concentration (IC_{50}). As an example, we use the human cyclin-dependent kinase 2 (CDK2) and 903 inhibitors for which experimental IC_{50} values are available from BindingDB.⁴⁶ A high-resolution (2.0 Å) crystal structure of CDK2 complexed with ATP in the active site was obtained from PDB (PDB-ID: 1b38, chain A).⁶⁷ Next, using BLAST and CDK2 sequence, we collected from PDB 2,508 homologous proteins at the expectation value threshold of 0.01. These structures were subsequently aligned onto CDK2 structure using Fr-TM-align⁶⁸ retaining only very close structural neighbors with a TM-score⁶⁹ of ≥ 0.7 . Upon the global superposition, we extracted these kinase ligands that have geometric centers within 4 Å from the geometric center of ATP bound to CDK2. Finally, using the SUBSET program⁶⁵ and a 1D-TC similarity threshold of 0.7, we clustered these compounds to construct a nonredundant set of 488 anchor structures for docking of CDK2 inhibitors using *eSimDock*.

The SVR model employing a set of 14 features used to estimate the binding affinity is retrained here to predict IC_{50} values for CDK2 inhibitors instead. To minimize the risk of memorization, we partitioned the set of 903 benchmarking inhibitors into 252 clusters using a 1D-TC threshold of 0.7. Leave-one-out cross-validation is done at the cluster level rather than for individual inhibitors, i.e. a whole cluster is removed from the training set, the SVR model is constructed, and IC_{50} values are predicted for inhibitors from the excluded cluster. This procedure is repeated for all clusters. The accuracy is assessed by calculating a cross-validated correlation between experimental and predicted IC_{50} values across the entire set of CDK2 inhibitors.

RESULTS AND DISCUSSION

Overview of *eSimDock*. Similarity-based ligand docking by *eSimDock* represents a hierarchical approach that comprises several consecutive steps as depicted in Figure 1. Docking procedure starts from the selection of the target protein structure (Figure 1A), which can be either solved experimentally or predicted from its amino acid sequence. If the experimental structure is selected and the target protein is cocrystallized with a small molecule, the target binding site is known (Figure 1B). For experimental structures in the apo conformational state as well as for protein models, binding pockets are predicted using *eFindSite* (Figure 1C). In the next step, ligand templates are extracted from the target binding site for use as anchor substructures in similarity-based docking. For experimental holo structures, cocrystallized compounds are used as the anchor substructures, otherwise *eSimDock* employs compounds extracted from evolutionarily related ligand-bound templates identified by *eFindSite*. Subsequently, atomic equivalences are established between the target (Figure 1D) and the template ligands in order to select the best anchor substructure (Figure 1E, anchor is shown as thick solid sticks).

In this study, we benchmark the performance of *eSimDock* using different anchor substructures; however, in real applications typically multiple ligand templates are identified. In that case, the anchor substructure is selected based on the highest chemical similarity to the target ligand. To explore the internal conformational space of the target ligand, an ensemble of low-energy conformations is generated (Figure 1F). Note that the ensemble is constructed only once for a given ligand and can be reused in docking simulations against different target receptors. Each conformer is superposed onto the anchor substructure according to the atomic equivalences reported by chemical matching (Figure 1G, target ligand is solid) and subject to a refinement procedure to remove atom clashes and locally optimize the interactions with the target protein. Next, nonlinear models are applied to each minimized conformation to evaluate fitness and binding scores and to predict the binding affinity (Figure 1H). Conformers are ranked based on the fitness score, and the top-ranked conformation is selected as the final docking result (Figure 1I, solid; experimental binding pose is transparent).

We can clearly identify several components of *eSimDock* that are critical for the docking accuracy and, consequently, affect the reliability of binding affinity prediction. For example, the quality of the target structure (experimental or modeled) will likely have an effect on the docking outcome. Moreover, the chemical similarity of ligand templates to the target compound is crucial for positioning the latter correctly within the binding pocket. When ligand templates are extracted from evolutionarily related proteins, we may expect the docking accuracy to depend on the level of protein homology. Also, since *eSimDock* employs only a local energy minimization, the ensemble of conformers precalculated for the target ligand should contain biologically relevant conformations. Finally, the accuracy of nonlinear scoring functions is critical for the correct evaluation of fitness and binding scores, leading to successful virtual screening experiments. These factors are assessed in the subsequent sections.

Pseudoflexible Ligand Docking. Mimicking ligand internal flexibility by ensemble docking is a commonly used technique for the modeling of ligand-protein interactions.^{22,70} In this approach, first an ensemble of low-energy conformations is precalculated, and then each conformer is docked using a rigid ligand docking protocol. The main advantage is the computational speed, since computing different ligand conformations explicitly is replaced by a look-up table; for example, a speed-up of 100 compared to docking a single flexible conformation has been reported in large-scale virtual screening of 117,000 compounds.⁷⁰ To date, numerous algorithms have been developed to generate sets of low-energy conformations for small molecules, e.g. Balloon,⁵⁰ TrixX Conformer Generator,⁷¹ and Frog2.⁷² Critical for the docking success is the presence of at least one native-like conformation in the generated ensemble. Table 1 shows RMSD values for the closest and farthest rotamer in the ensemble from the native ligand conformation using Astex/CCDC ligands. For small ligands with up to 5 rotatable bonds, the closest conformation is typically within 0.77 Å heavy-atom RMSD from native. Even for larger compounds having up to 15 rotatable bonds, there is at least one native-like rotamer in the ensemble with an RMSD of < 2.0 Å on average. This analysis suggests that, in most cases, pseudoflexible ligand docking calculations have a chance to explore biologically relevant ligand rotamers.

Table 1. Characteristics of Ligand Ensembles Generated for Docking Compounds in the Astex/CCDC Data Set^b

# rotatable bonds	molecular weight	min RMSD ^a [Å]	max RMSD ^a [Å]
1–5	251.9 ± 102.9	0.77 ± 0.36	1.74 ± 0.72
6–10	413.4 ± 105.6	1.24 ± 0.45	3.14 ± 0.86
11–15	521.7 ± 154.6	1.84 ± 0.55	4.17 ± 0.71
16–20	570.1 ± 161.0	2.21 ± 0.72	4.74 ± 0.82
>20	698.2 ± 96.0	2.73 ± 0.70	5.22 ± 0.42

^aCalculated over heavy atoms. ^bCompounds are grouped by the number of rotatable bonds. Min and max RMSD values report the average similarity ± standard deviation of the closest and farthest rotamer in the ensemble to the native ligand conformation, respectively.

Recognition of Native-like Docking Conformations. A successful docking approach should have a strong capability to discriminate between native-like and misdocked ligand conformations. In binding pose prediction, eSimDock employs a set of 9 structure-based features, which are combined using a nonlinear machine learning model. Model training and validation is carried out for a set of native-like and misdocked decoys generated for protein–ligand complexes in the Astex/CCDC data set. First, we assess how well the average values of individual scores separate both sets of conformations. Figure 2 shows that the most effective scoring functions are DSX_pair and Vina_gauss1, with a median score for native-like (misdocked) conformations of −19.7 (619.9) and 66.0 (114.3), respectively. In addition, weakly discriminatory

functions include Vina_repulsion, DFIRE_pair, and DSX_sas, with the respective values of 13.2 (199.2), 0.9 (0.6), and −14.3 (−5.6). This analysis shows that individual scoring terms should have good discriminant capabilities against misdocked conformations; below, we demonstrate that their nonlinear combination using machine learning is highly accurate in picking up native-like docking conformations.

Figure 3 shows the ROC analysis of the separation of native-like conformations from misdocked decoys across trajectories

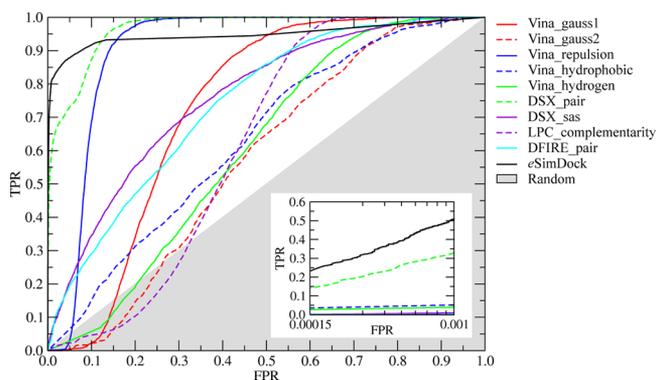


Figure 3. ROC plot for the recognition of native-like conformations in docking ensembles by eSimDock compared to its individual scoring functions. TPR – true positive rate, FPR – false positive rate. Inset: zoom in at the very beginning of the ROC space at FPR of <0.001.

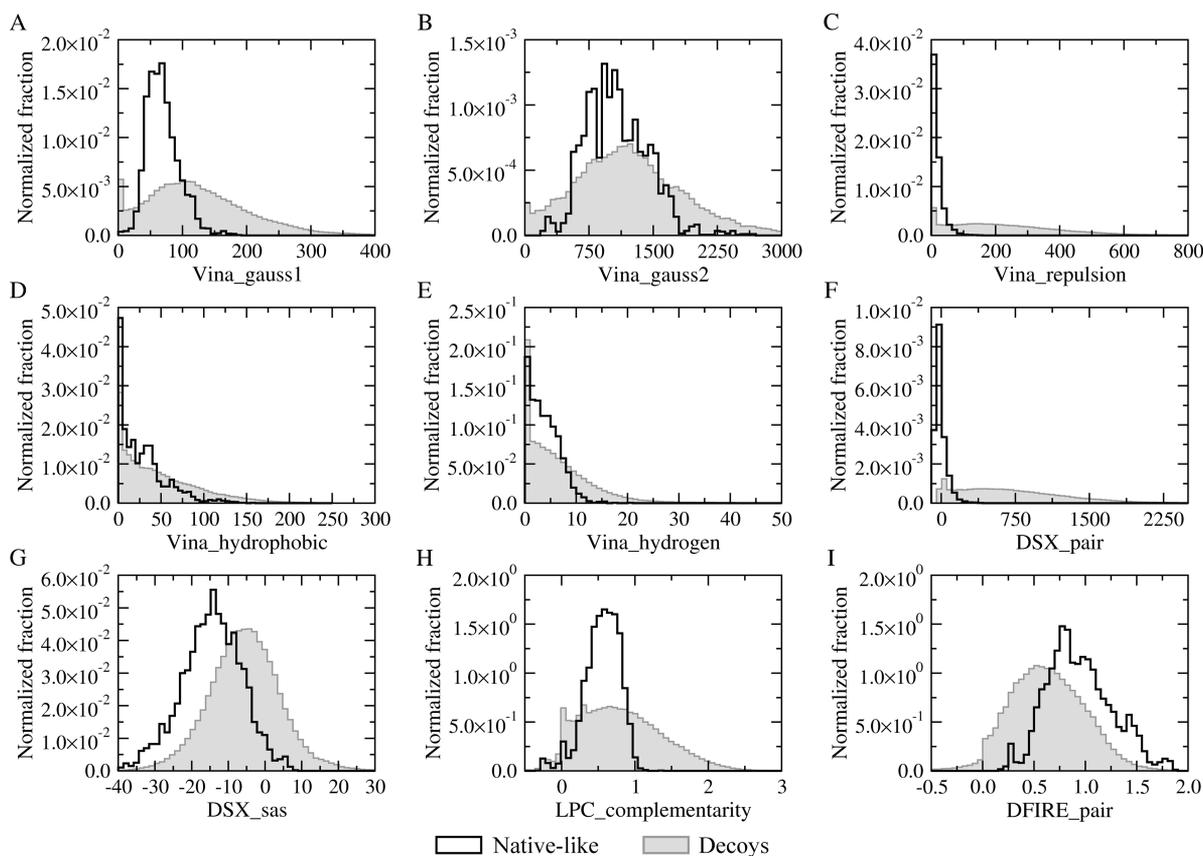


Figure 2. Distribution of the energy component values over a set of native-like conformations as well as misdocked decoys. (A) Vina_gauss1, (B) Vina_gauss2, (C) Vina_repulsion, (D) Vina_hydrophobic, (E) Vina_hydrogen, (F) DSX_pair, (G) DSX_sas, (H) LPC_complementarity, and (I) DFIRE_pair.

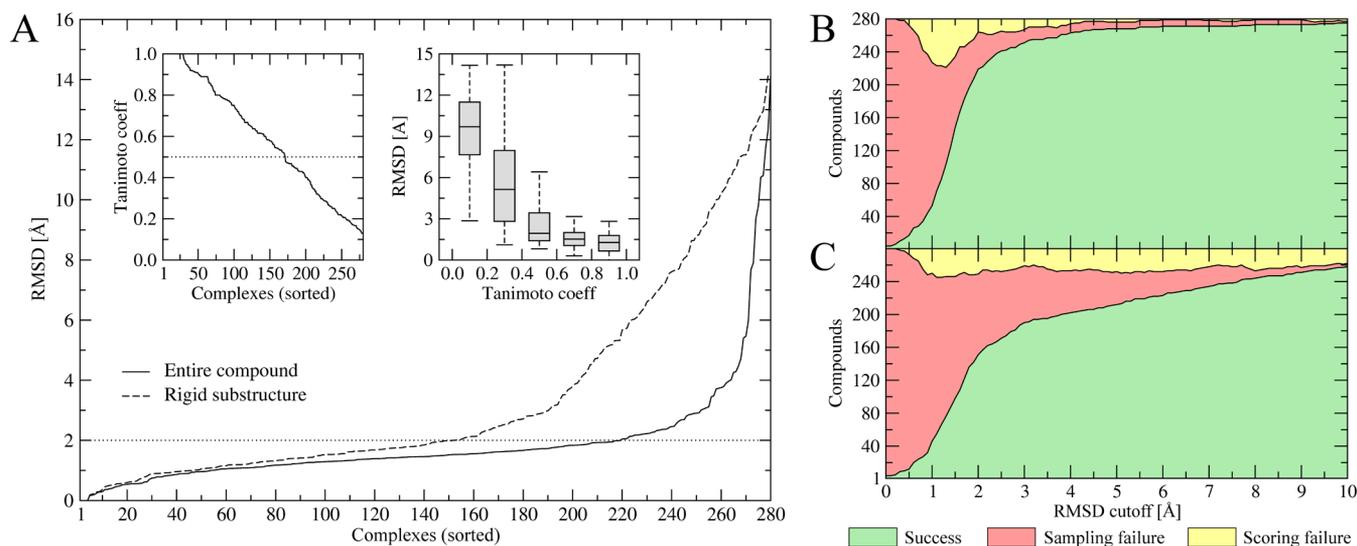


Figure 4. Docking accuracy of *eSimDock* for the Astex/CCDC data set. (A) Heavy atom RMSD from the crystal binding pose for the top-ranked conformation; the simulations employ as an anchor either the ligand itself or its largest rigid substructure. Left inset in A: 2D-TC between the entire ligand and its largest rigid substructure across the Astex/CCDC data set. Right inset in A: the dependency of the accuracy of the top-ranked pose on the 2D-TC. (B) RMSD spectrum plot for similarity-based docking using the entire ligand; (C) RMSD spectrum plot for docking using the largest rigid substructure. In B and C, the RMSD cutoff displayed on the *x*-axis is used to calculate the number of successfully docked compounds, sampling failures and scoring failures, which are shown in green, red, and yellow, respectively.

generated during docking simulations. In general, all individual scoring functions identify native-like poses better than a purely random selection; however, *DSX_pair* is the most effective, followed by *Vina_repulsion*, *DSX_sas*, and *DFIRE_pair*, which is consistent with previous results. More importantly, a nonlinear machine learning model designed to assess ligand fitness performs better than any individual scoring function. This is particularly noticeable in the inset of Figure 3; at the expense of only 0.1% false positive rate, the sensitivity of *eSimDock* is 50%, which is ~20% higher than the best single scoring function, *DSX_pair*.

Accuracy of Binding Pose Prediction. A typical docking procedure often generates multiple acceptable solutions; however, from a practical point of view, the most important is the accuracy of the top-ranked conformation selected as the final docking result. In Figure 4A, we show the distribution of RMSD from crystal calculated over ligand heavy atoms for complexes constructed by *eSimDock* for the Astex/CCDC data set. Using the entire ligand as the anchor substructure, the heavy atom RMSD vs crystal for the top-ranked docking conformations is <2 Å (<3 Å) for 78.2% (89.6%) of the cases. When only the largest rigid substructure extracted from the native ligand is used, 53.9% (67.9%) of the modeled ligands have RMSD of <2 Å (<3 Å). This can be explained by a chemical similarity of rigid substructures to native ligands. Left inset in Figure 4A shows that only in 61.1% of the cases, 1D-TC between the two molecules is >0.5. As a rule, a high 1D-TC is needed to achieve a high accuracy of the final docked pose; this relation is shown in the right inset of Figure 4A, where the median RMSD is below 2 Å (1.5 Å) only when these rigid substructures, whose 1D-TC is at least 0.5 (0.7), are used.

In general, the success of docking simulations depends on two factors: a sampling efficiency and the accuracy of a scoring function. RMSD spectrum plots provide a convenient assessment tool to decouple sampling and scoring failures in binding pose prediction.⁷³ Both characteristics are analyzed in Figures 4B and 4C using as an anchor the entire ligand and the largest

rigid substructure, respectively. As might be expected, for a small RMSD cutoff of 1 Å, the docking outcome in the majority of cases is classified as sampling failures; this is because no ligand conformations are generated within this RMSD range. Nevertheless, for a more realistic RMSD cutoff of 2 Å, sampling and scoring failures in docking using the entire ligand as an anchor account for only 16.1% and 5.7% of the cases, respectively. This demonstrates that the nonlinear scoring function implemented in *eSimDock* is highly effective in recognizing native binding poses. When the largest rigid substructure is used as an anchor, the number of sampling and scoring failures is roughly doubled (35.0% and 11.1%, respectively); however, in most of the cases (53.9%), the ligand binding pose is predicted within 2 Å RMSD. Moreover, relaxing the RMSD cutoff to 3 Å increases the success rate to 67.9%. Approximately three times more sampling failures than scoring errors suggest that further improvements of *eSimDock* should focus on enhancing the sampling efficiency rather than the scoring function.

Evolutionarily Derived Anchor Substructures. Simulations using the largest rigid substructure as an anchor are useful for e.g. docking of series of derivatives of a compound whose experimental complex structure with a target protein is available. There are many examples of such applications in lead optimization, where sets of derivatives are constructed by substituting the R-group(s) of a prototype compound with different chemical moieties.⁷⁴ However, for a target protein not cocrystallized with ligands that are sufficiently similar to docking compounds, we need to consider a different source of anchor substructures. One possibility is to derive them from evolutionarily related holo-proteins. Here, we use *eFindSite*, a recently developed template-based binding site predictor,⁵² which detects conserved binding sites across sets of evolutionarily related proteins. It also extracts template-bound ligands for use as anchor substructures in similarity-based docking. First, we analyze the accuracy of *eFindSite* for the Astex/CCDC data set in terms of the prediction of pocket

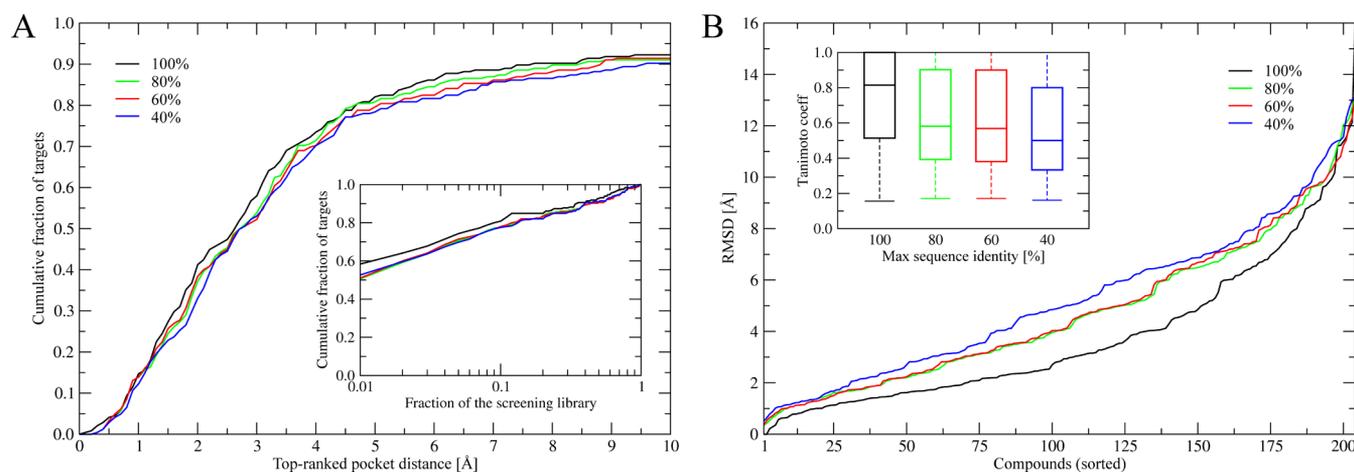


Figure 5. Docking results for the Astex/CCDC data set using template ligands identified by *eFindSite*. (A) Cumulative fraction of targets for which the top-ranked binding site is predicted within a distance displayed on the *x*-axis. Inset in A: cumulative fraction of targets for which the native ligand was ranked within the top fraction of the screening library shown on the *x*-axis. (B) Heavy atom RMSD from the crystal binding pose for the top-ranked conformation constructed using template ligands. Inset in B: the distribution of 1D-TC between native and template ligands extracted from predicted pockets. *eFindSite* calculations were performed excluding homologous templates at several sequence similarity thresholds shown in different colors.

location and the similarity of extracted template-bound ligands to these compounds complexed with Astex/CCDC proteins. We note that *eFindSite* predictions are obtained at different target-template sequence identity thresholds. Figure 5A shows that using a maximum pairwise sequence identity of 100%, 80%, 60%, and 40%, the top-ranked pocket center is predicted within 4 Å from the geometric center of bound ligand in the experimental complex in 73.5%, 71.4%, 70.2%, and 70.2% of the cases, respectively. This clearly demonstrates that the performance of *eFindSite* using only weakly homologous templates (<40% sequence identity) is almost the same as that obtained by including closely related templates (>40%). The chemical similarity of template-bound ligands to the native ligand can be assessed by fingerprint-based virtual screening; here we use a representative and nonredundant at a 1D-TC of 0.8 ZINC12¹ collection of 244,659 small organic compounds. We add each Astex/CCDC ligand to the library and then rank all compounds using *eFindSite* and consensus fingerprints constructed from template-bound ligands. Inset plot in Figure 5A shows that this procedure ranks the native ligand in the top 1% (10%) of the screening library in 58.4% (80.8%), 50.6% (77.6%), 51.0% (78.0%), and 52.7% (77.6%) of the cases, including only these templates whose sequence identity to the target protein is no more than 100%, 80%, 60%, and 40%, respectively. As expected, these results are in good agreement with the accuracy of binding site prediction; it has been previously observed that ligand ranking tends to be more accurate for correctly predicted pockets.⁵³

From now on, we will be focusing on these Astex/CCDC pockets that are predicted within 4 Å from the experimental pocket center. Figure 5B presents the accuracy of similarity-based ligand docking using *eSimDock* and evolutionarily derived anchor substructures at different levels of homology between the target and template proteins. The fraction of targets for which the ligand binding pose is predicted within 2 Å (4 Å) using these templates whose sequence identity is no more than 100%, 80%, 60%, and 40% is 34.8% (65.2%), 21.1% (50.0%), 20.6% (48.5%), and 15.2% (39.7%), respectively. We note that two factors need to be considered assessing these results. First, the predicted pocket center is up to 4 Å away

from the experimental pocket, which may cause the predicted ligand binding pose to be shifted with respect to the experimental conformation. Second, we use anchor substructures extracted from evolutionarily related complexes, whose chemical similarity to docking compounds may vary. This is shown in the inset to Figure 5B. As one might expect, decreasing homology between the target and template proteins results in lower chemical similarity between docking ligands and the anchor substructures used in *eSimDock* with a median 1D-TC of 0.81, 0.58, 0.57, and 0.50 for sequence identity threshold of 100%, 80%, 60%, and 40%, respectively. This is consistent with previous benchmarking results reported in the right inset to Figure 4A, which shows that docking accuracy falls off with decreasing 1D-TC between the target ligand and the anchor substructure used in similarity-based docking. Nevertheless, even in the most difficult situation when only weakly homologous templates are available, *eSimDock* still generates fairly accurate ligand binding poses in many cases. These could be a good starting point for further postdocking binding pose refinement calculations.⁷⁵

Tolerance to Distortions in Receptor Structures. A desired feature of a ligand docking approach is its tolerance to structural deformations of target proteins. This considerably expands the application range to proteome-wide docking studies, where only computationally generated protein models are available for most of the target receptors. To evaluate the sensitivity of *eSimDock* to protein structure distortions, we perform docking simulations against a series of non-native conformations constructed for Astex/CCDC proteins. Table 2 provides the characteristics of these receptor conformations in terms of the global structure quality measured by $C\alpha$ -RMSD⁷⁶ and TM-score⁶⁹ as well as a local heavy-atom RMSD calculated over binding residues. Comparative protein structure modeling typically results in models whose $C\alpha$ -RMSD from native is 2–2.5 Å.^{77,78} At this global structure quality, the local geometry of ligand-binding regions is fairly well preserved with all-atom RMSD of <2 Å on average, see Table 2. However, weakly homologous models generated by e.g. threading and fold recognition are frequently less accurate with a $C\alpha$ -RMSD of 4–6 Å.⁷⁹ Binding pockets in such structures can be severely

Table 2. Quality of the Distorted Receptor Structures Constructed for the Astex/CCDC Data Set^c

receptor data set	structure quality			fitness (docking)
	$C\alpha$ -RMSD ^a	TM-score ^a	pocket ^b	
set 0.0				0.80 ± 0.36
set 0.5	0.75 ± 0.36	0.98 ± 0.02	0.44 ± 0.13	0.61 ± 0.43
set 1.0	1.13 ± 0.28	0.96 ± 0.02	0.78 ± 0.19	0.45 ± 0.42
set 1.5	1.58 ± 0.29	0.93 ± 0.03	1.15 ± 0.27	0.28 ± 0.39
set 2.0	2.05 ± 0.28	0.89 ± 0.04	1.50 ± 0.34	0.20 ± 0.33
set 2.5	2.50 ± 0.31	0.82 ± 0.05	1.85 ± 0.41	0.13 ± 0.29
set 3.0	3.03 ± 0.48	0.82 ± 0.06	2.28 ± 0.56	0.11 ± 0.25
set 3.5	3.47 ± 0.35	0.78 ± 0.07	2.64 ± 0.54	0.08 ± 0.22
set 4.0	3.96 ± 0.37	0.75 ± 0.08	2.99 ± 0.66	0.09 ± 0.23
set 4.5	4.47 ± 0.41	0.72 ± 0.08	3.44 ± 0.77	0.06 ± 0.18
set 5.0	4.95 ± 0.42	0.69 ± 0.08	3.79 ± 0.77	0.05 ± 0.17

^aGlobal, calculated over backbone $C\alpha$ atoms. ^bLocal RMSD calculated over heavy atoms of binding residues. ^cLast column reports the average fitness score estimated by *eSimDock* in docking simulations against each distorted receptor data set.

deformed to 3–4 Å RMSD presenting a significant challenge to ligand docking algorithms.

Figure 6A reports the accuracy of ligand docking using *eSimDock* against theoretical protein models whose structure quality progressively deteriorates. As opposed to self-docking simulations, where we hope for docked conformations with 0 Å

RMSD, docking against non-native receptor structures have an upper bound for the accuracy of ligand pose prediction. This is a consequence of the distorted geometry of ligand binding regions; here, the upper bound can be estimated by directly transferring ligand conformation from the crystal to the modeled structure upon the optimal superposition of binding residues. Moreover, it is more meaningful to measure the docking accuracy using the fraction of correctly predicted protein–ligand atomic contacts instead of heavy atom RMSD. In that regard, Figure 6 shows that the upper bound expressed as the fraction of native contacts measured using transferred ligands correlates well with the overall structure quality of target proteins and decreases linearly from 0.88 to 0.40 for $C\alpha$ -RMSD increasing from 0.5 Å to 5.0 Å. The accuracy of *eSimDock* assessed using the same evaluation metric drops from 0.59 to 0.27 for receptor models distorted to 0.5 Å and 5.0 Å $C\alpha$ -RMSD, respectively. However, the ratio of docking accuracy and the estimated upper bound is at a constant level of ~0.65 across the entire data set of non-native receptor structures demonstrating that the performance of *eSimDock* is unaffected by the deformations of ligand binding regions. This remarkable and important feature of similarity-based docking arises from the global structure alignments of ligand-bound templates onto target proteins, which position anchor substructures in approximately correct orientations. Interestingly, Table 2 also shows that despite the implemented scoring function assessing ligand “fitness” correctly recognizes structural distortions in

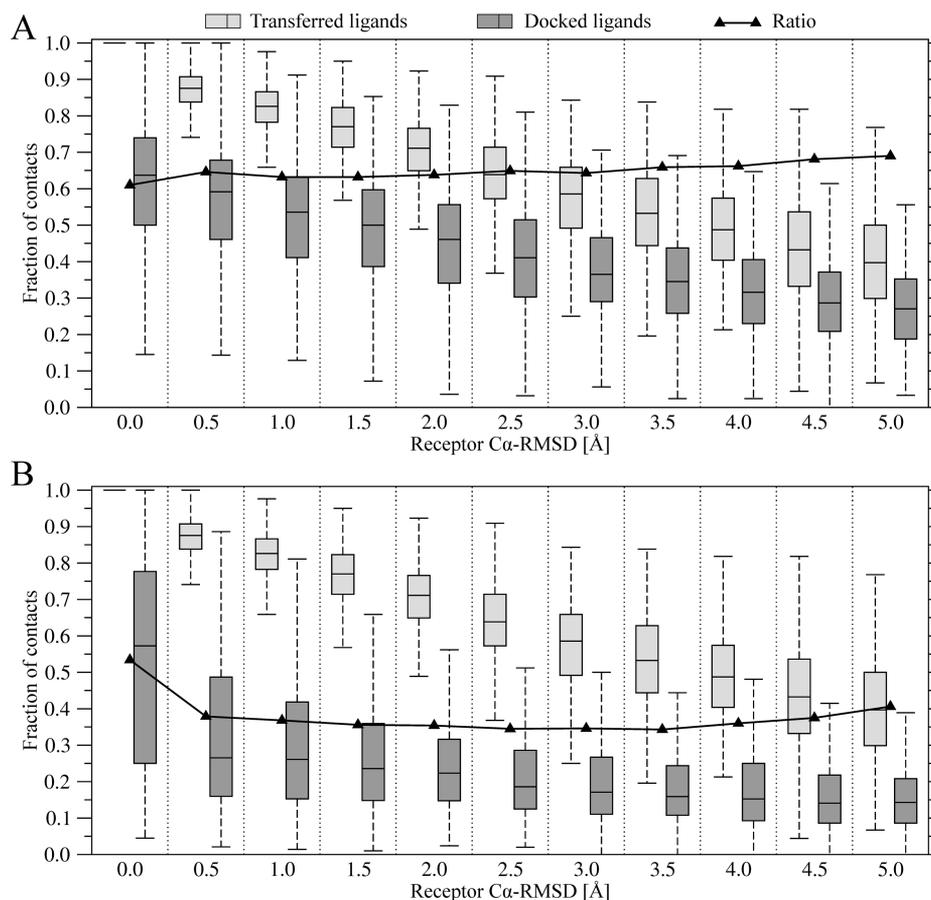


Figure 6. Docking accuracy with respect to structural distortions in target receptors for (A) *eSimDock* and (B) AutoDock Vina. The accuracy is evaluated by the fraction of correctly predicted interatomic native contacts. Transferred ligands estimate the upper bound for the docking accuracy. The ratio of the docking accuracy and the corresponding upper bound is shown as solid black triangles/lines.

Table 3. Accuracy of Virtual Screening by *eSimDock* Compared to Its Individual Component Scoring Functions As Well As to *AutoDock Vina*^a

scoring function	enrichment factor				BEDROC				AUC	ACT-50
	1%	2%	5%	10%	5%	10%	20%	30%		
Vina_gauss1	0.44	0.36	0.48	0.64	0.19	0.09	0.04	0.03	0.53	0.48
Vina_gauss2	2.40	1.82	1.27	1.11	0.25	0.15	0.10	0.08	0.55	0.45
Vina_repulsion	0.51	1.93	3.29	3.60	0.53	0.36	0.23	0.17	0.80	0.15
Vina_hydrophobic	1.02	1.02	0.87	0.93	0.19	0.11	0.07	0.05	0.50	0.52
Vina_hydrogen	0.36	0.51	0.67	0.88	0.24	0.13	0.06	0.04	0.57	0.41
DSX_pair	19.37	15.77	10.12	6.67	0.74	0.66	0.60	0.59	0.88	0.05
DSX_sas	5.39	4.77	3.71	2.91	0.43	0.31	0.24	0.21	0.70	0.23
LPC_complementarity	0.87	0.55	0.64	0.73	0.22	0.11	0.05	0.04	0.59	0.42
DFIRE_pair	1.31	1.49	1.54	1.55	0.30	0.18	0.11	0.09	0.61	0.36
<i>eSimDock</i>	24.91	24.47	17.52	9.19	0.94	0.92	0.91	0.91	0.95	0.02
<i>AutoDock Vina</i>	6.92	4.95	3.26	2.48	0.37	0.27	0.21	0.20	0.63	0.32

^aResults are assessed in terms of the enrichment factors and BEDROC values calculated for the top 1, 2, 5, and 10% of the screening library as well as the area under the enrichment curve (AUC) and the top fraction of the screening library that contains 50% of binding compounds (ACT-50).

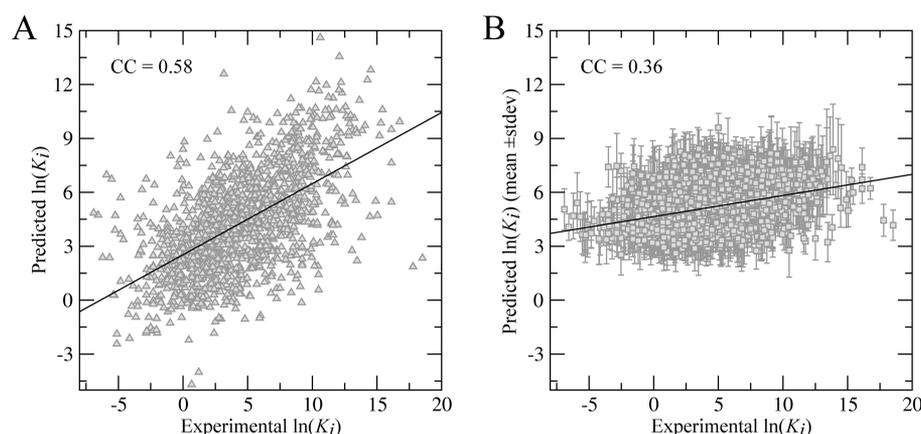


Figure 7. Correlation between experimental and predicted binding affinity for BindingDB compounds. (A) Binding affinity is predicted for the binding compounds. (B) Binding affinity is predicted for nonbinding decoys. In B, the mean binding affinity and the corresponding standard deviation is calculated from docking a set of 25 decoys per one binding compound. CC is the Pearson correlation coefficient.

target binding sites, the generated ligand conformations are fairly accurate considering the maximum theoretical docking accuracy.

Virtual Screening Benchmarks. Ligand ranking capabilities of *eSimDock* are assessed in virtual screening benchmarks using the BindingDB data set. This set comprises 1,151 protein–ligand complexes with experimentally measured binding affinities. For each active compound, we also include 25 topologically and chemically different ligand decoys whose molecular weight is comparable to that of binding molecules. Note that this is a standard practice used in cheminformatics to create high-quality benchmarking sets for virtual screening such as DUD⁸⁰ or its recently enhanced version, E-DUD.⁸¹ The results are assessed by several metrics that account for different aspects of the ranking performance. One of the most commonly used, enrichment factor (EF) measures how many active compounds are recovered in the top-ranked fraction (1%, 2%, 5%, and 10%) of the screening library. Boltzmann-enhanced discrimination of receiver operating characteristic, or BEDROC, was developed to address the “early recognition” problem of ranking actives early in an ordered list.⁸² Compared to EF, it is also sensitive to ranking performance before and after the selected cutoff and retains the statistical significance of ROC. The area under the enrichment curve (AUC) is used to assess the overall performance of virtual screening over the

entire screening library. Finally, ACT-50 corresponds to the top fraction of the library that needs to be scanned through to recover half of the active compounds.

Table 3 shows ligand ranking accuracy comparing *eSimDock* to its individual component scoring functions. For the latter, the most effective is *DSX_pair* with an EF^{10%}, BEDROC₂₀, AUC, and ACT-50 of 6.67, 0.60, 0.88, and 0.05, respectively. Other sensitive scoring functions include *DSX_sas*, *Vina_repulsion*, and *DFIRE_pair*. These results are consistent with our previous analysis of discriminant capabilities against misdocked conformations; see Figures 2 and 3. Interestingly, individual scoring functions have different characteristics; for instance, *Vina_repulsion* gives lower EF 1–5% than *DSX_pair*, but it is more accurate when assessed by BEDROC 5–10, AUAC, and ACT-50. Table 3 also demonstrates that this variety can be beneficially exploited using nonlinear machine learning models. As a result, *eSimDock* systematically outperforms its individual scoring functions offering high compound ranking capabilities, which should be sufficient for successful virtual screening applications.

Binding Affinity Prediction. Virtual screening benchmarks evaluate the relative ranking of bioactive compounds with respect to nonbinding decoys. Here, we also investigate how reliably the actual binding affinity can be predicted for active molecules. *eSimDock* uses SVR, machine learning for

regression problems, to estimate binding affinity (K_i) from individual docking scores. As shown in Figure 7A, Pearson correlation coefficient (CC) between experimental and predicted by *eSimDock* K_i values is 0.58. We note that this correlation is measured for a representative data set of 1,151 complexes using leave-one-out cross-validation at the receptor level. According to the scale provided by Salkin,⁸³ this value can be interpreted as a moderate/strong positive correlation. A common drawback of many scoring functions is that predicted binding affinities are highly correlated with the molecular weight of testing compounds.^{33,34} We look into this issue in Figure 7B by plotting the correlation between experimental K_i values for BindingDB compounds and these predicted for their isomolecular weight decoys. Encouragingly, CC drops to 0.36, which can be interpreted as a weak relationship clearly demonstrating that the previous high CC of 0.58 develops from specific protein–ligand interactions rather than just the number of contacts. Furthermore, we also show in Table 4 that

Table 4. Pearson Correlation Coefficients between Experimental and Predicted Binding Affinities for Binding Compounds^a

scoring function	binding compds	decoy compds
Vina_gauss1	0.27	0.29
Vina_gauss2	0.36	0.32
Vina_repulsion	0.16	0.30
Vina_hydrophobic	0.28	0.30
Vina_hydrogen	0.09	0.02
DSX_pair	0.11	0.30
DSX_sas	0.18	0.01
LPC_complementarity	0.01	0.23
DFIRE_pair	0.16	0.15
<i>eSimDock</i>	0.58	0.36

^aThe accuracy of binding affinity prediction by *eSimDock* is compared to its individual component scoring functions. The last column reports correlation coefficient between experimental affinities of binding compounds and these predicted for nonbinding decoys.

eSimDock more accurately predicts binding affinity than its component scoring functions. Best performing single scores are from Vina: Vina_gauss2, Vina_hydrophobic, and Vina_gauss1, which give weak CC for binding (decoy) compounds of 0.36 (0.32), 0.28 (0.30), and 0.27 (0.29), respectively.

Similar to the assessment of binding pose prediction, we also quantify the impact of structural distortions in the target pockets on binding affinity prediction. As shown in Table 5, we

use non-native receptor structures whose global $C\alpha$ -RMSD (TM-score) to crystal ranges from 0.52 Å (0.99) to 2.66 Å (0.86); the corresponding deformations of ligand binding regions are in the range of 0.32–1.57 Å. As we previously demonstrated for the Astex/CCDC data set, fitness and binding probability estimates from machine learning classifiers are correlated with the degree of structural distortions in protein targets; see docking results in Table 5. CC for the experimental vs predicted binding affinity also drops as the target structures become more distorted; however, these relationships still have a moderate strength for up to a $C\alpha$ -RMSD of 3.0 Å. This analysis indicates that *eSimDock* predicts binding affinity with a fair accuracy even against modeled receptor structures, which is a nontrivial result.

Examples of Successful Docking Calculations. Figure 8 illustrates three representative examples from the Astex/CCDC data set: dihydrofolate reductase (1aoeB), carbonic anhydrase I (1azmA), and neuraminidase (1a4gA). Experimental binding sites of 1aoeB, 1azmA, and 1a4gA with bound ligands are shown in Figure 8A along with pockets (global receptor structures) distorted to 1.4 Å (1.6 Å), 2.0 Å (3.0 Å), and 3.4 Å (4.4 Å) all-atom (backbone $C\alpha$) RMSD, respectively. Using crystal receptor structures and the ligand itself (the largest rigid substructure) as the docking anchor, the predicted ligand binding poses have an RMSD of 1.69 Å (1.75 Å), 0.82 Å (1.19 Å), and 1.53 Å (1.81 Å), respectively; see Figure 8B (Figure 8C). Figure 8D shows the most similar ligand templates identified by *eFindSite* at a sequence identity threshold of 80%, 60%, and 40%, respectively. These are subsequently used to dock query ligands using *eSimDock*; the generated binding poses presented in Figure 8E have an RMSD from crystal of 1.87 Å, 1.31 Å, and 1.58 Å, respectively. The last panel, Figure 8F, shows ligand poses predicted using the distorted receptor structures. Here, the docking accuracy measured by the fraction of recovered native protein–ligand contacts (the corresponding upper bound) is 0.70 (0.87), 0.44 (0.63), and 0.45 (0.52), respectively. These examples depict the performance of *eSimDock* using different quality receptor structures, various anchor substructures, and template ligands extracted at different homology levels.

Comparison to AutoDock Vina. Using both primary data sets, Astex/CCDC and BindingDB, we compare the performance of *eSimDock* to AutoDock Vina,⁵⁵ which is one of the most widely used programs for structure-based virtual screening. Figure 6B shows the accuracy of Vina against crystal structures (0 Å $C\alpha$ -RMSD) as well as a series of non-native conformations constructed for Astex/CCDC proteins (0.5–5.0

Table 5. Docking of BindingDB Compounds to Non-Native Structures^c

receptor set	structure quality			docking results		
	$C\alpha$ -RMSD ^a [Å]	TM-score ^a	pocket RMSD ^b [Å]	fitness	binding	CC
set 0.0				0.80 ± 0.35	0.71 ± 0.36	0.58
set 0.5	0.52 ± 0.06	0.99 ± 0.00	0.32 ± 0.16	0.69 ± 0.42	0.59 ± 0.41	0.55
set 1.0	1.02 ± 0.09	0.97 ± 0.01	0.67 ± 0.24	0.64 ± 0.40	0.51 ± 0.40	0.54
set 1.5	1.44 ± 0.13	0.94 ± 0.02	0.91 ± 0.33	0.60 ± 0.43	0.44 ± 0.41	0.53
set 2.0	1.85 ± 0.16	0.92 ± 0.03	1.15 ± 0.41	0.49 ± 0.44	0.34 ± 0.39	0.50
set 2.5	2.25 ± 0.19	0.89 ± 0.04	1.37 ± 0.49	0.38 ± 0.43	0.27 ± 0.37	0.49
set 3.0	2.66 ± 0.21	0.86 ± 0.04	1.57 ± 0.58	0.33 ± 0.42	0.24 ± 0.35	0.46

^aGlobal, calculated over backbone $C\alpha$ atoms. ^bLocal RMSD calculated over heavy atoms of binding residues. ^cData set characteristics include both global and local structure quality of the receptor proteins. Docking results report the estimated fitness and binding scores as well as Pearson correlation between experimental and predicted binding affinity (CC).

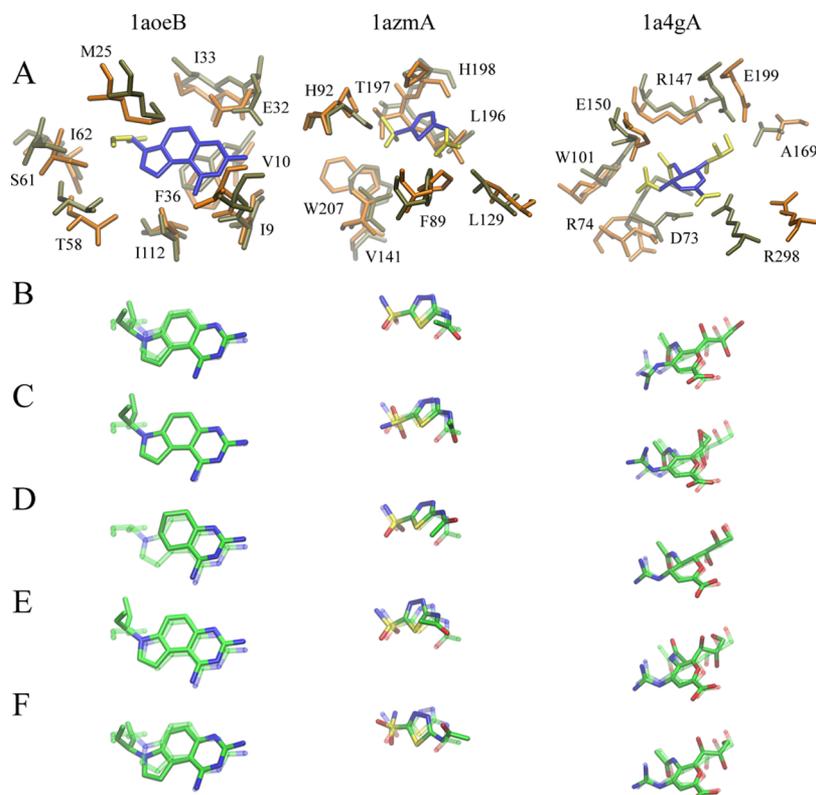


Figure 8. Examples of successful docking simulations using *eSimDock*. Three cases are shown: dihydrofolate reductase (1aocB), carbonic anhydrase I (1azmA), and neuraminidase (1a4gA). (A) Close look at the crystal binding pose: selected binding residues are shown as tan sticks and labeled, the largest rigid substructure of a ligand is colored in blue with the remaining parts shown in yellow. Equivalent binding residues in the distorted receptor structures are shown as orange sticks. Top-ranked docked conformation using as an anchor (B) the entire ligand and (C) the largest rigid substructure. (D) The most similar ligand template identified by *eFindSite*. (E) Target ligand docked using the template ligand shown in D. (F) Target ligand is docked into the distorted binding site (orange sticks in A) rather than the receptor crystal structure. In B–F, the reference crystal binding pose is transparent and the docked conformation (template ligand in D) is solid; both molecules are colored by atom type.

Å α -RMSD). Focusing on the crystal structures, the accuracy of Vina is only slightly lower than that of *eSimDock*; the median fraction of correctly predicted protein–ligand atomic contacts is 0.57 and 0.64, respectively. However, the performance of Vina considerably deteriorates with the increasing magnitude of structural deformations in target structures. The ratio of docking accuracy and the estimated upper bound is at a level of ~ 0.35 across the set of distorted receptor structures. This analysis clearly demonstrates that the performance of traditional ligand docking drops off dramatically with the decreasing quality of target structures, which is in line with previous reports.^{35,36} Contrastingly, as discussed above, the ratio of docking accuracy and upper bound for *eSimDock* is at a constant level of ~ 0.65 across the entire data set (Figure 6A).

Next, using the BindingDB data set, we evaluate the performance of AutoDock Vina in ligand ranking and compare the results to those obtained for *eSimDock*. We note that this data set comprises 1,151 protein–ligand complexes and includes 25 topologically different isomolecular weight ligand decoys per active compound. Table 3 shows that in general, Vina performs better than most individual scoring functions used in *eSimDock*, with the exception of these from DrugScore,²⁹ DSX_pair and DSX_sas. For example, the enrichment factor calculated for the top 1% of the screening library is 7 times better than random, AUC is 0.63, and ACT-50 is 0.32, which shows that Vina ranks at least half of the binding molecules within the top one-third of the library. Nevertheless, the accuracy of *eSimDock* is significantly higher than this

obtained from Vina; compare the last two rows in Table 3. We can identify two factors responsible for the better performance of *eSimDock*. First, it employs DrugScore potentials, which have remarkably high ranking capabilities; see DSX_pair in Table 3. Second, nonlinear scoring functions implemented in *eSimDock* effectively combine individual scoring terms enhancing ranking accuracy in virtual screening. We conjecture that replacing the linear combination of weighted energy terms in Vina with nonlinear machine learning models should further improve its ranking capabilities, in analogy to recent successful modifications of the scoring function of the docking program eHiTS.⁵⁸

Case Study: fXa Inhibitors. In addition to calculations using two representative data sets, Astex/CCDC and BindingDB, we discuss two case studies to demonstrate that *eSimDock* represents a general methodology, which can be conveniently adapted for more specific purposes. The first case study evaluates the performance of *eSimDock* in discriminating between 279 strong and 156 weak inhibitors of fXa whose experimentally determined binding affinities are <10 nM and >1 μ M, respectively. 62 molecules bound to fXa are used as a set of anchor substructures for similarity-based docking of 435 benzamidine inhibitors by *eSimDock*. The three nonlinear scoring functions implemented in *eSimDock* that assess fitness, binding, and affinity employ SVM models. In this case study, we also demonstrate that other machine learning techniques, e.g. Naïve Bayes classifiers, can be used. NB models the distribution of numerical scores using a mean and the

corresponding standard deviation for each feature. These statistical dispersion characteristics across the training set of 290 compounds are reported in Table 6 for all individual

Table 6. Average (\pm Standard Deviation) Values of Individual Scoring Terms Calculated Across the Sets of Strongly and Weakly Binding Inhibitors of Factor Xa

scoring function	strong binders	weak binders
Vina_gauss1	130.90 \pm 38.52	84.33 \pm 36.38
Vina_gauss2	1663.54 \pm 279.06	1330.71 \pm 311.94
Vina_repulsion	130.31 \pm 110.12	81.22 \pm 94.22
Vina_hydrophobic	57.35 \pm 25.99	38.76 \pm 25.71
Vina_hydrogen	4.59 \pm 2.99	3.65 \pm 1.66
DSX_pair	361.78 \pm 360.73	195.49 \pm 286.55
DSX_sas	-21.80 \pm 9.55	-16.45 \pm 9.02
LPC_complementarity	0.85 \pm 0.30	0.78 \pm 0.23
DFIRE_pair	1.07 \pm 0.21	1.08 \pm 0.15

scoring functions. We note that these values are calculated from binding poses predicted by *eSimDock*; there is a clear difference between the individual distributions obtained for strong and weak binders. Figure 9 shows that this is efficiently exploited by

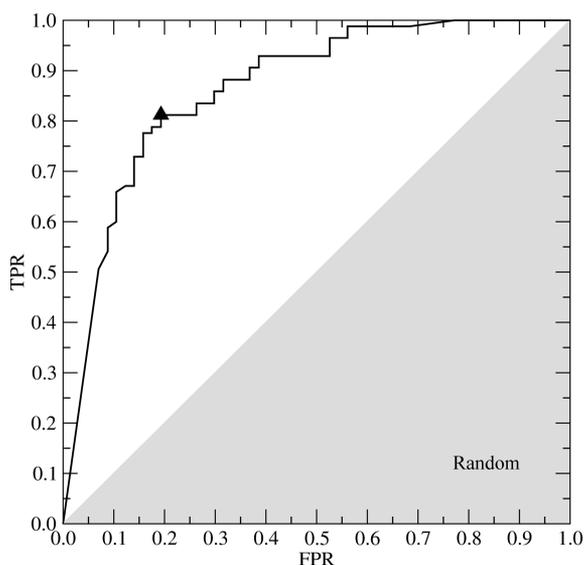


Figure 9. ROC plot for the discrimination between strong and weak binders of fXa using a Naïve Bayes classifier implemented in *eSimDock*. A solid black triangle corresponds to the best performance that maximizes the accuracy. The gray area shows the performance of a random classifier.

machine learning and the performance of the NB classifier is quite high. Specifically, the optimal posterior probability cutoff depicted by a black triangle in Figure 9 yields a true positive rate of 0.81 at the expense of a false positive rate of 0.19. This corresponds to the accuracy, specificity, and precision of 0.81, 0.81, and 0.86, respectively, demonstrating a very good performance of *eSimDock* in discriminating between strong and weak inhibitors of fXa.

Case Study: CDK2 Inhibitors. The second example selected in this study considers the prediction of IC_{50} values for a diverse set of 903 CDK2 inhibitors. Here, anchor structures are extracted from homologous proteins identified by BLAST. As a nonlinear scoring function, we use the SVR model developed for binding affinity prediction, which is now

retrained on IC_{50} against CDK2 rather than K_i values. Figure 10 shows that the cross-validated Person correlation coefficient

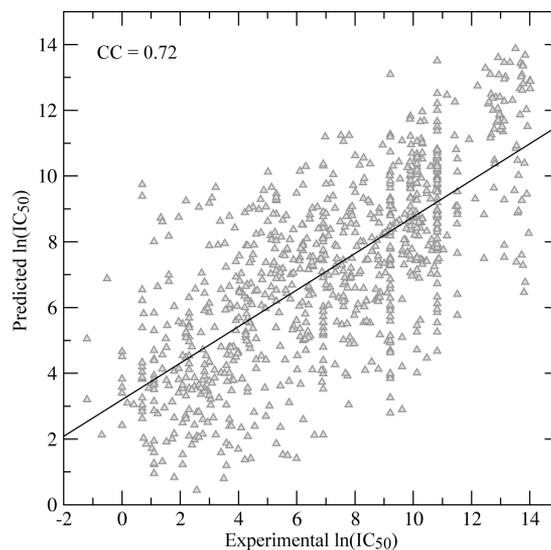


Figure 10. Correlation between experimental and predicted IC_{50} for 903 inhibitors of the human cyclin-dependent kinase 2, CDK2. CC is the Pearson correlation coefficient.

is 0.72, thus the SVR model implemented in *eSimDock* maintains a high accuracy when it is modified for specific applications.

CONCLUSIONS

In this study, we describe the development and benchmarking of *eSimDock*, a new approach to similarity-based ligand docking and binding affinity prediction. A unique feature of *eSimDock* is a set of nonlinear machine learning-based scoring functions constructed to evaluate the confidence of binding pose prediction, calculate the likelihood that a query ligand forms a stable complex with the receptor protein, and predict binding affinity. In large-scale benchmarks using representative data sets of protein–ligand complexes, we demonstrate that machine learning models are specific and accurate, systematically outperforming individual scoring functions. As a similarity-based approach, *eSimDock* requires template ligands to preposition docking compounds; we show that this information can be extracted from related proteins at various levels of homology. Another important feature of *eSimDock* is its remarkably high tolerance to structural distortions in the target structures, thus protein models with varying structure quality can be effectively utilized in virtual screening calculations. It is possible that the insensitivity to structural deformations in binding regions will affect the selectivity among homologous members within the same family. Nevertheless, even close homologues typically have different amino acid composition of binding pockets, which may be sufficient to maintain high selectivity; this will be investigated in future studies. Finally, we assess the performance of *eSimDock* in selected case studies demonstrating that it can be easily adapted for specific applications and different nonlinear scoring functions are effective as well.

Based on the overall algorithm design and benchmarking results, we identify two types of primary applications *eSimDock* is particularly suitable for. First application is a rapid evaluation of a series of compounds having similar scaffold but different

side groups. At the outset of the docking experiment, at least one experimental target structure cocrystallized with a scaffold compound is required. When binding affinities are measured for a subset of testing compounds, nonlinear scoring functions implemented in eSimDock can be easily retrained on new experimental data substituting the default scoring scheme with target-specific machine learning models. This notable feature allows eSimDock to be tailored to specific projects, as demonstrated in this study for fXa and CDK2. The second application is ligand docking to theoretical receptor structures. Computational efficiency of similarity-based docking together with encouraging results obtained from docking simulations against distorted targets and evolutionarily derived anchor substructures indicate that eSimDock can be integrated with evolution/structure-based approaches to ligand binding prediction providing a virtual screening component to across-proteome function annotation projects.

eSimDock is freely available to the academic community as a Web server at <http://www.brylinski.org/esimdock>. The Web site also provides all data sets and benchmarking results reported in this paper as well as another case study on inhibitors of heat shock protein 90 in the form of step-by-step instructions for using online services.

AUTHOR INFORMATION

Corresponding Author

*E-mail: michal@brylinski.org.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This study was supported by the Louisiana Board of Regents through the Board of Regents Support Fund [contract LEQSF(2012-15)-RD-A-05]. The author is grateful to Dr. Yaoqi Zhou for providing DFIRE software. Portions of this research were conducted with high performance computational resources provided by Louisiana State University (HPC@LSU, <http://www.hpc.lsu.edu>) and the Louisiana Optical Network Institute (LONI, <http://www.loni.org>).

REFERENCES

- (1) Irwin, J. J.; Shoichet, B. K. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–82.
- (2) Reddy, A. S.; Pati, S. P.; Kumar, P. P.; Pradeep, H. N.; Sastry, G. N. Virtual screening in drug discovery – a computational perspective. *Curr. Protein Pept. Sci.* **2007**, *8*, 329–51.
- (3) Jain, A. N. Virtual screening in lead discovery and optimization. *Curr. Opin. Drug Discovery Dev.* **2004**, *7*, 396–403.
- (4) Geppert, H.; Vogt, M.; Bajorath, J. Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205–16.
- (5) Ripphausen, P.; Nisius, B.; Bajorath, J. State-of-the-art in ligand-based virtual screening. *Drug Discovery Today* **2011**, *16*, 372–6.
- (6) Xue, L.; Bajorath, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Comb. Chem. High Throughput Screening* **2000**, *3*, 363–72.
- (7) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273–80.
- (8) Willett, P. Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today* **2006**, *11*, 1046–53.
- (9) Seal, A.; Yogeewari, P.; Sriram, D.; Consortium, O.; Wild, D. J. Enhanced ranking of PknB Inhibitors using data fusion methods. *J. Cheminf.* **2013**, *5*, 2.
- (10) Plewczynski, D.; Spieser, S. A.; Koch, U. Performance of machine learning methods for ligand-based virtual screening. *Comb. Chem. High Throughput Screening* **2009**, *12*, 358–68.
- (11) Xue, L.; Godden, J. W.; Stahura, F. L.; Bajorath, J. Profile scaling increases the similarity search performance of molecular fingerprints containing numerical descriptors and structural keys. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1218–25.
- (12) Verma, J.; Khedkar, V. M.; Coutinho, E. C. 3D-QSAR in drug design—a review. *Curr. Top. Med. Chem.* **2010**, *10*, 95–115.
- (13) Free, S. M., Jr.; Wilson, J. W. A mathematical contribution to structure-activity studies. *J. Med. Chem.* **1964**, *7*, 395–9.
- (14) Martins, J. P.; Barbosa, E. G.; Pasqualoto, K. F.; Ferreira, M. M. LQTA-QSAR: a new 4D-QSAR methodology. *J. Chem. Inf. Model.* **2009**, *49*, 1428–36.
- (15) Vedani, A.; Dobler, M. 5D-QSAR: the key for simulating induced fit? *J. Med. Chem.* **2002**, *45*, 2139–49.
- (16) Vedani, A.; Dobler, M.; Lill, M. A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. *J. Med. Chem.* **2005**, *48*, 3700–3.
- (17) Ebalunode, J. O.; Zheng, W. Molecular shape technologies in drug discovery: methods and applications. *Curr. Top. Med. Chem.* **2010**, *10*, 669–79.
- (18) Brylinski, M.; Skolnick, J. FINDSITE(LHM): a threading-based approach to ligand homology modeling. *PLoS Comput. Biol.* **2009**, *5*, e1000405.
- (19) Gschwend, D. A.; Good, A. C.; Kuntz, I. D. Molecular docking towards drug discovery. *J. Mol. Recognit.* **1996**, *9*, 175–86.
- (20) Dror, R. O.; Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 13118–23.
- (21) Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM - A new method for protein modeling and design: applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.* **1994**, *15*, 488–506.
- (22) Brylinski, M.; Skolnick, J. Q-Dock: low-resolution flexible ligand docking with pocket-specific threading restraints. *J. Comput. Chem.* **2008**, *29*, 1574–1588.
- (23) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1–5.
- (24) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **1997**, *267*, 727–48.
- (25) Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 411–428.
- (26) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–49.
- (27) Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- (28) Bohm, H. J. Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from de novo design or 3D database search programs. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 309–23.
- (29) Neudert, G.; Klebe, G. DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. *J. Chem. Inf. Model.* **2011**, *51*, 2731–45.
- (30) Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* **1999**, *42*, 791–804.

- (31) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–42.
- (32) Wang, R.; Lu, Y.; Wang, S. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem.* **2003**, *46*, 2287–303.
- (33) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793–806.
- (34) Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L., 3rd Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.* **2004**, *47*, 3032–47.
- (35) McGovern, S. L.; Shoichet, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895–907.
- (36) Erickson, J. A.; Jalaie, M.; Robertson, D. H.; Lewis, R. A.; Vieth, M. Lessons in molecular recognition: the effects of ligand and protein flexibility on molecular docking accuracy. *J. Med. Chem.* **2004**, *47*, 45–55.
- (37) Bordogna, A.; Pandini, A.; Bonati, L. Predicting the accuracy of protein-ligand docking on homology models. *J. Comput. Chem.* **2011**, *32*, 81–98.
- (38) Kairys, V.; Fernandes, M. X.; Gilson, M. K. Screening drug-like compounds by docking to homology models: a systematic study. *J. Chem. Inf. Model.* **2006**, *46*, 365–79.
- (39) Oshiro, C.; Bradley, E. K.; Eksterowicz, J.; Evensen, E.; Lamb, M. L.; Lanctot, J. K.; Putta, S.; Stanton, R.; Grootenhuis, P. D. Performance of 3D-database molecular docking studies into homology models. *J. Med. Chem.* **2004**, *47*, 764–7.
- (40) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-Protein Coupled Receptors suitable targets? *Proteins* **2003**, *50*, 5–25.
- (41) Sherman, W.; Day, T.; Jacobson, M. P.; Friesner, R. A.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **2006**, *49*, 534–53.
- (42) Davis, I. W.; Baker, D. RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **2009**, *385*, 381–92.
- (43) Nissink, J. W.; Murray, C.; Hartshorn, M.; Verdonk, M. L.; Cole, J. C.; Taylor, R. A new test set for validating predictions of protein-ligand interaction. *Proteins* **2002**, *49*, 457–71.
- (44) Mancera, R. L.; Kallblad, P.; Todorov, N. P. Ligand-protein docking using a quantum stochastic tunneling optimization method. *J. Comput. Chem.* **2004**, *25*, 858–64.
- (45) Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins* **2003**, *52*, 609–23.
- (46) Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res.* **2007**, *35*, D198–201.
- (47) Kawabata, T. Build-up algorithm for atomic correspondence between chemical structures. *J. Chem. Inf. Model.* **2011**, *51*, 1775–87.
- (48) Bindewald, E.; Skolnick, J. A scoring function for docking ligands to low-resolution protein structures. *J. Comput. Chem.* **2005**, *26*, 374–83.
- (49) Daylight Chemical Information Systems, I. Daylight Theory Manual. <http://www.daylight.com/dayhtml/doc/theory/index.html> (accessed November 5, 2013).
- (50) Vainio, M. J.; Johnson, M. S. Generating conformer ensembles using a multiobjective genetic algorithm. *J. Chem. Inf. Model.* **2007**, *47*, 2462–74.
- (51) Rasmussen, M. D.; Deshpande, M. S.; Karypis, G.; Johnson, J.; Crow, J. A.; Retzel, E. F. wCLUTO: a Web-enabled clustering toolkit. *Plant Physiol.* **2003**, *133*, 510–6.
- (52) Brylinski, M.; Feinstein, W. P. eFindSite: improved prediction of ligand binding sites in protein models using meta-threading, machine learning and auxiliary ligands. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 551–67.
- (53) Brylinski, M.; Skolnick, J. A threading-based method (FINDSITE) for ligand-binding site prediction and functional annotation. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 129–34.
- (54) Brylinski, M.; Lingam, D. eThread: a highly optimized machine learning-based approach to meta-threading and the modeling of protein tertiary structures. *PLoS One* **2012**, *7*, e50200.
- (55) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–61.
- (56) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: an open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
- (57) 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*, 2785–91.
- (58) Kinnings, S. L.; Liu, N.; Tonge, P. J.; Jackson, R. M.; Xie, L.; Bourne, P. E. A machine learning-based method to improve docking scoring functions and its application to drug repurposing. *J. Chem. Inf. Model.* **2011**, *51*, 408–19.
- (59) Sobolev, V.; Sorokine, A.; Prilusky, J.; Abola, E. E.; Edelman, M. Automated analysis of interatomic contacts in proteins. *Bioinformatics* **1999**, *15*, 327–32.
- (60) Zhang, C.; Liu, S.; Zhu, Q.; Zhou, Y. A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes. *J. Med. Chem.* **2005**, *48*, 2325–35.
- (61) Sud, M. MayaChemTools. <http://www.MayaChemTools.org> (accessed November 5, 2013).
- (62) Chang, C. C.; Lin, C. J. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 27.
- (63) Fontaine, F.; Pastor, M.; Zamora, I.; Sanz, F. Anchor-GRIND: filling the gap between standard 3D QSAR and the GRIND-INdependent descriptors. *J. Med. Chem.* **2005**, *48*, 2687–94.
- (64) Yoshikawa, K.; Kobayashi, S.; Nakamoto, Y.; Haginoya, N.; Komoriya, S.; Yoshino, T.; Nagata, T.; Mochizuki, A.; Watanabe, K.; Suzuki, M.; Kanno, H.; Ohta, T. Design, synthesis, and SAR of cis-1,2-diaminocyclohexane derivatives as potent factor Xa inhibitors. Part II: exploration of 6–6 fused rings as alternative S1 moieties. *Bioorg. Med. Chem.* **2009**, *17*, 8221–33.
- (65) Voigt, J. H.; Bienfait, B.; Wang, S.; Nicklaus, M. C. Comparison of the NCI open database with seven large chemical structural databases. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 702–12.
- (66) Keselj, V. Naive Bayes Classification. <http://search.cpan.org/~vlado/> (accessed November 5, 2013).
- (67) Brown, N. R.; Noble, M. E.; Lawrie, A. M.; Morris, M. C.; Tunnah, P.; Divita, G.; Johnson, L. N.; Endicott, J. A. Effects of phosphorylation of threonine 160 on cyclin-dependent kinase 2 structure and activity. *J. Biol. Chem.* **1999**, *274*, 8746–56.
- (68) Pandit, S. B.; Skolnick, J. Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score. *BMC Bioinf.* **2008**, *9*, 531.
- (69) Zhang, Y.; Skolnick, J. Scoring function for automated assessment of protein structure template quality. *Proteins* **2004**, *57*, 702–10.
- (70) Lorber, D. M.; Shoichet, B. K. Flexible ligand docking using conformational ensembles. *Protein Sci.* **1998**, *7*, 938–50.
- (71) Griewel, A.; Kayser, O.; Schlosser, J.; Rarey, M. Conformational sampling for large-scale virtual screening: accuracy versus ensemble size. *J. Chem. Inf. Model.* **2009**, *49*, 2303–11.
- (72) Miteva, M. A.; Guyon, F.; Tuffery, P. Frog2: efficient 3D conformation ensemble generator for small compounds. *Nucleic Acids Res.* **2010**, *38*, W622–7.
- (73) Brozell, S. R.; Mukherjee, S.; Balias, T. E.; Roe, D. R.; Case, D. A.; Rizzo, R. C. Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 749–73.
- (74) Gobbi, S.; Cavalli, A.; Rampa, A.; Belluti, F.; Piazzi, L.; Paluszczak, A.; Hartmann, R. W.; Recanatini, M.; Bisi, A. Lead optimization providing a series of flavone derivatives as potent

nonsteroidal inhibitors of the cytochrome P450 aromatase enzyme. *J. Med. Chem.* **2006**, *49*, 4777–80.

(75) Pencheva, T.; Lagorce, D.; Pajeva, I.; Villoutreix, B. O.; Miteva, M. A. AMMOS: automated molecular mechanics optimization tool for in silico screening. *BMC Bioinf.* **2008**, *9*, 438.

(76) Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr., Sect. A: Cryst. Phys., Diffr., Theor. Gen. Crystallogr.* **1976**, *32*, 922–3.

(77) Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. Evaluation of comparative protein modeling by MODELLER. *Proteins* **1995**, *23*, 318–26.

(78) Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Res.* **2003**, *31*, 3381–5.

(79) Jaroszewski, L.; Rychlewski, L.; Zhang, B.; Godzik, A. Fold prediction by a hierarchy of sequence, threading, and modeling methods. *Protein Sci.* **1998**, *7*, 1431–40.

(80) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking sets for molecular docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.

(81) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *J. Med. Chem.* **2012**, *55*, 6582–6594.

(82) Truchon, J. F.; Bayly, C. I. Evaluating virtual screening methods: good and bad metrics for the "early recognition" problem. *J. Chem. Inf. Model.* **2007**, *47*, 488–508.

(83) Chung, M. K. Correlation Coefficient. In *Encyclopedia of Measurement and Statistics*; Salkin, N. J., Ed.; Sage Publications: London, 2007; pp 189–201.