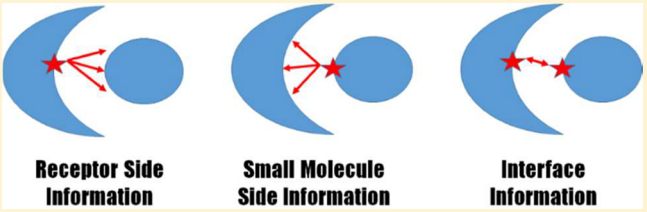
JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING
Cite This: J. Chem. Inf. Model. 2018, 58, 225−233
Review
pubs.acs.org/jcim

Predictive Power of Different Types of Experimental Restraints in Small Molecule Docking: A Review

Darwin Y. Fu and Jens Meiler*

Department of Chemistry Vanderbilt University Nashville, Tennessee 37235, United States

ABSTRACT: Incorporating experimental restraints is a powerful method of increasing accuracy in computational protein small molecule docking simulations. Different algorithms integrate distinct forms of biochemical data during the docking and/or scoring stages. These so-called hybrid methods make use of receptor-based information such as nuclear magnetic resonance (NMR) restraints or small molecule-based information such as structure−activity relationships (SARs). A third class of methods directly interrogates contacts between the protein receptor and the small molecule. This work reviews the current state of using such restraints in docking simulations, evaluates their feasibility across broad systems, and identifies potential areas of algorithm development.



KEYWORDS: *Experimental restraints, Small molecule docking, SAR, Structure-based drug discovery, Receptor side information, Small molecule information, Interface information, Molecular similarity, Hybrid methods, Computer-aided drug design*

PROTEIN−SMALL MOLECULE DOCKING AND STRUCTURE-BASED DRUG DISCOVERY

Computational protein−small molecule docking is an important step in the pipeline for structure-based design of small molecule drugs. Successful prediction of binding position is necessary to delineate critical interactions for improving selectivity and/or efficacy. Popular docking algorithms such as AutoDOCK,[1] DOCK,[2] GLIDE,[3] GOLD,[4] and RosettaLigand[5] have diverse methods for representing, sampling, and scoring the molecular interface. These techniques, along with structure-based virtual screening, have aided advances in drug hit discovery and lead optimization.[6]

Community assessments of docking software have generally displayed success in recovering near-native binding poses. Davis et al. found that accurate binding poses were found for all targets in a GlaxoSmithKline compound collection, but the overall success rate varied dramatically among systems. Furthermore, no algorithm consistently outperformed the others across all systems.[7] The CSAR 2012 benchmark demonstrated features such as protein structure minimization, histidine tautomeric states correction, pregenerated small molecule conformations, native small molecule training, and substructure based restraints correlated positively with docking success. However, binding affinity prediction and relative ranking of active small molecules remains the most challenging aspect in the field and during this experiment in particular.[8]

A common theme across docking assessments was the benefit afforded by restraints derived from experimental data.[7−10] The potential for such hybrid/integrative methods has already been reported for protein structure prediction, which benefits from leveraging nonatomic resolution structural biology methods such as cryo-electron microscopy[11] or electron paramagnetic resonance.[12] Hybrid methods for small molecule docking can be categorized by data type and by integration point. Experimental information can be classified as receptor structure-based, small molecule-based, or interface-based. Although the same experimental methods may be used in all three, only interface-based measurements directly identifies a protein-small molecule interaction. Data derived restraints can be incorporated as part of the sampling and/or scoring process.

THREE TYPES OF RESTRAINTS AND IMPACTS ON DOCKING ACCURACY

In order to demonstrate the different power among the restraint types, a simulated restrained docking benchmark was conducted using RosettaLigand[5] on the PDBBind Core Set.[13] The PDBBind Core Set is a collection of 65 high affinity protein−small molecule complexes. This subset of the PDBBind database has been previously used for assessing docking accuracy.[9] Receptor-based restraints were represented by restraining three randomly chosen binding pocket residues contact any small molecule heavy atom. Similarly, small molecule restraints promoted contacts for three small molecule heavy atoms. Interface data were simulated by restraining three randomly chosen pairs of interacting atoms between the receptor and the small molecule. A contact was defined as an interatomic distance less than 4 Å, which includes most commonly observed molecular interactions.[14] For each test, the small molecule was initially subject to a random reorientation and translation within a 5 Å sphere. An additional test was conducted with minimal initial perturbation as a representation of using the binding mode of a similar small molecule as a guide

Received: July 13, 2017
Published: December 29, 2017

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233

to initial placement. A set of 2500 docking trajectories were completed for each protein−small molecule test case under each restraint condition. The models were analyzed for percentage of native-like small molecule binding modes using a 2.0 Å RMSD cutoff.

Figure 1 shows the distribution of sampling success rates across test systems for each of the restraint conditions. The

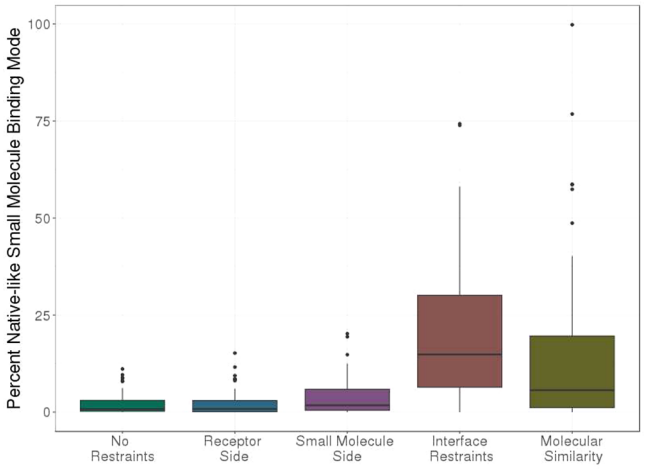


Figure 1. Boxplots of simulated restrained docking sampling efficiency. Boxplots show distribution of percent native-like binding modes observed across 65 PDBBind Core Set test cases.

largest improvements are seen in docking with interface based restraints and in using molecular similarity to restrict the starting position. This makes intuitive sense as both restrict the small molecule rotational orientation in additional to its translational location in the binding pocket. Reinforcing specific interatomic distances also restricts the small molecule conformational flexibility. The applicability of molecular similarity as a docking restraint depends on similar binding exhibited by similar small molecules, a question explored later in this review.

Table 1 shows some common types of experimental information in addition to examples of programs that makes use of each type of restraint. Referenced methods are discussed in further detail in the following sections.

■ PROTEIN RECEPTOR-BASED DATA FOR DOCKING

Receptor-based, also referred to as structure-based, data are derived from observed changes or effects on the protein alone. Protein structures in the absence of (apo) or in complex with the small molecule (holo) determined via X-ray crystallography are the most straightforward form of structure-based data. Small molecules can be directly docked into receptor crystal structures or, if such structures are unavailable, into homology models. Docking into holo crystal structures is generally more accurate than docking into apo crystal structures or comparative models.[15] In testing a nitroreductase protein−-small molecule target from CASP 11, Huwe et al. found few dockings to comparative model structures that were superior to docking to the experimental crystal structure. However, the comparative model docking managed to capture specific contacts 72.7% of the time.[16] Bordogna et al. found a Spearman's correlation coefficient of 0.66 between RMSD accuracy of the comparative model and the accuracy of the docking simulation for a diverse test set.[17] In high-throughput

Table 1. Receptor-Based (Blue), Small Molecule-Based (Red), and Interface-Based (Green) Experimental Data[a]

Technique	Information
X-Ray Crystallography	Static model of apo protein or protein-small molecule complex
Nuclear Magnetic Resonance	Ensemble of protein models
Chemical Shift Perturbations	Protein residue signal changes with varying small molecule concentrations
Nuclear Overhauser Effect (Intra-protein)	Protein structural changes upon small molecule binding
HD-Exchange Mass Spectrometry	Protein amide hydrogens buried by addition of small molecule
Isothermal Titration Calorimetry	Binding affinity changes due to protein mutations or small molecule alterations
Structure – Activity Relationships (DoMCoSAR[40])	Correlation between small molecule modifications and binding affinity
Consensus Molecular Shapes (SABRE[39])	3D structural overlay of multiple active small molecules
Protein-Mutation/Ligand-Modification Coupling	Specific contacts based on correlation of protein and ligand modifications ("Double Mutant Cycle")
Inter-ligand NOEs (INPHARMA[58])	Relative orientations of two small molecule binders
Protein-Ligand NOEs	Distances between protein-small molecule contact points
Ligand Molecular Similarity (HybridDock[60], LigBEnD[61])	Binding mode of a related small molecule to the same protein target

[a]Example programs for particular methods are given in parentheses.

docking, or virtual screening, applications, comparative models are capable of similar enrichment rates as their crystal structure template counterparts.[18,19] A common theme across the assessments is that traditional measures of comparative modeling ease, such as sequence similarity between template and target, does not correlate with subsequent docking success. The success rate for docking into homology models can be improved by up to 70% by using holo experimental templates crystallized with small molecules of similar chemotypes.[20] Careful validation of the input crystal structure, particularly in regards to proper orientation and placement of the small molecule, should be performed prior to using the structure in computational drug discovery efforts. Any modifications to the target protein in the crystallization process, including biologically irrelevant mutations or inserted constructs, should also be considered.[21]

Receptor structures may also be derived from nuclear magnetic resonance (NMR) spectroscopy. An ensemble of conformations is generally provided to capture the flexibility observed in structures obtained by NMR spectroscopy. Alternatively, NMR spectroscopy may be utilized to obtain information on protein−small molecule interface contacts. Chemical shift perturbations[22] are observed for specific residues upon small molecule binding, while intraprotein Nuclear Overhauser Effects (NOEs)[23−25] reflect structural changes within the protein. Distance restraints derived from these two sources are based on the assumption that changes are due to interactions with the small molecule. Protein-focused methods can help define the receptor binding pocket but do not necessarily give information on the small molecule binding mode. This type of information generally translate to a restraint favoring small molecule positons that are within a certain distance of the contact residue.[26] Orts et al. demonstrated the use of protein-mediated NOE data for two competitively binding small molecules as a postdocking scoring filter that can improve accuracy by 2 orders of magnitude.[23] Onila et al.

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233

extended this method to directly use NMR data during docking by simultaneously optimizing poses for both small molecules, which improved docking in a test set of weakly bound cAMP-dependent protein kinase complexes. However, the results were highly dependent on obtaining proper orientation of the protein side chains.[24] Cala et al. reviews further experimental details for NMR characterization of protein-small molecule contacts.[26]

Other methods for localizing small molecule binding site interactions include hydrogen/deuterium exchange mass spectrometry (HDX-MS) and isothermal titration calorimetry (ITC) used in conjunction with mutagenesis. HDX-MS relies on the different exchange rates for exposed versus buried amide hydrogen atoms to identify protein residues covered up by small molecule binders.[27] Mouchlis et al. used HDX-MS on protein backbone amides in conjunction with docking to determine binding modes of phospholipase A_2 inhibitors.[28] ITC is used to measure the thermodynamic components of binding affinity: enthalpy and entropy. Structural information is inferred from binding affinity changes following protein mutagenesis or small molecule modification. It is generally assumed that such changes is due to the impact of the alteration on small molecule binding.[29] To quantify the reliability of using binding affinity changes from mutagenesis experiments for binding site localization, the Platinum Database[30] of single mutants is examined to see if affinity change alone is feasible as a binary predictor of binding distance. The Platinum Database is a manually curated collection of 1000 mutation−structure pairs for assessing the impact of mutations on small molecule binding. Figure 2 shows affinity change measured by relative

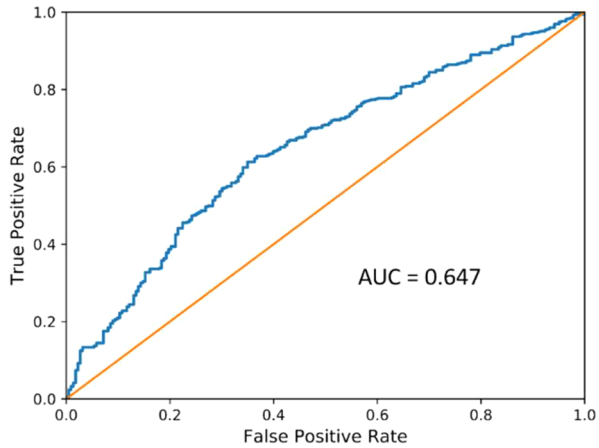


Figure 2. ROC curve using single mutant small molecule affinity fold change as binary classifier for whether mutated residue interacts with small molecule (distance <5 Å). The calculated area under the curve is shown with 1.0 being a perfect classifier.

folds as a weak classifier (area under curve = 0.647) for distance to small molecule. This is not surprising as there are often allosteric effects that explain differential binding to mutant receptors. Interface-based experimental techniques such as a double mutant cycle or cross-linking may be better ways to confirm a direct interaction. This is particularly true when mutational changes are measured in terms of biological response instead of binding affinities, for which change may be due to a direct impact on binding or due to downstream structural changes.[31,32]

One important component of receptor-based information is the use of ensemble methods to account for receptor flexibility. Rueda et al. demonstrated improvement of cross-docking results using binding site ensembles to represent protein flexibility.[33] In particular, ensembles of two or more proteins, enhanced for proteins cocrystallized with chemically similar small molecules, performed better on average than single docking or randomly enumerated ensembles.[34] These conformational ensembles can also be derived computationally using the relaxed complex scheme (RCS), a series of molecular dynamics simulations to pregenerate low energy conformations.[35] Experimental data can then filter the conformations to avoid docking efficiency decrease stemming from having a large number of ensembles structures.[36] Sinko et al. and Feixas et al. further discuss RCS and other experimental methods to account for protein flexibility in drug design applications.[37,38]

One missing area is ensemble methods designed to work with protein mutants rather than just protein conformers. Developing such an algorithm would allow the use of receptor-based SAR data.

SMALL MOLECULE-BASED DATA FOR DOCKING

Small molecule-based information takes advantage of binding data through quantitative structure−activity relationships (QSARs). QSAR is traditionally part of virtual screening applications when no receptor information is available. These models generally make use of 2D small molecule property descriptors or 3D small molecule shape fitting without constructing a receptor model. Certain docking algorithms can incorporate comparisons of known small molecule binders in generating putative binding modes. SABRE is a method that generates a consensus molecular shape density function from multiple bioactive small molecules. Candidate molecules are then shape fitted using chemical substructures as opposed to the entire molecule at once.[39] Adherence to experimental SAR can also be used as a filtering step, though this is tricky as scoring functions generally do not rank order compounds well. DoMCoSAR is a docking algorithm that selects the most commonly observed binding modes and utilizes correlation with SAR to guide final model selection.[40] Small molecule conformational shape fitting can also be achieved with transfer NOE as demonstrated in the design of a flexible macrocyclic inhibitor.[41]

Small molecule-based pharmacophore models are an extension of molecular shape fitting by identifying chemical commonalities among binders of a given receptor. Known ligand binders of a receptor are aligned as flexible conformers and common features are identified in 3D space to identify the pharmacophore. This approach can also be performed with protein side chains to create receptor side pharmacophores.[42] One such method PharmDock converts the receptor and small molecule to hydrogen bonding and hydrophobic pharmacophores before using an alignment algorithm to match pairs. On a test of the PDBBind Core Set, PharmDock identified a native-like top model 56% of the time when native conformations were used but only 37% when using Omega generated conformers, signifying the importance of input pharmacophore conformations.[43] Pharmacophores can be generated in combination with molecular dynamics to generate an ensemble of binding pocket models as shown in a development study of small molecules for a highly flexible sulfotransferase binding pocket.[44] Yang et al. further discuss both ligand- and structure-based pharmacophore modeling along with potential challenges

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233

such as data set construction, molecular alignment, and feature selection.[42] Use of SARs and multiple active small molecules in docking simulations is an exciting area of research. In particular, activity cliffs, highly similar compounds with orders-of-magnitude differences in potency, provide powerful SAR information. New methodologies that utilize the experimental knowledge provided by SARs around activity cliffs can guide the creation of additional structural analogues.[45] The current release of the ChEMBL database contains over 13 million activities recorded against over 10 000 targets.[46] Furthermore, there are system specific small molecule SAR databases for common drug discovery targets such as GLASS for GPCRs[47] and KLIFS for kinases.[48] One potential way to incorporate this data is to use an ensemble docking method that can simultaneously optimize multiple protein−small molecule complexes and correlate their calculated scores. MLSD is an extension of AutoDock4 that allows the simultaneous optimization of multiple small molecule fragments, though it is restricted to molecules that concurrently bind to the same target.[49] Mass spectrometry of protein unfolding can also be used to examine multiple small molecule bindings and their combined interactions on protein stability in the gas phase, though further methods are necessary to translate this to binding modes.[50]

In a reverse modality, small molecule information can also be used to generate distance-dependent pair potentials for protein comparative modeling. The MOBILE program docks small molecules into a starting ensemble average of homology models, and then generates restraints for subsequent rounds of model refinement. When used in combination with MODELLER, MOBILE produced native-like binding pocket geometries in 70% of test cases, improving results in 60% of cases compared to restraint free comparative modeling.[51]

INTERFACE-BASED DATA FOR DOCKING

Protein−small molecule interactions are the most powerful combination of the protein and small molecule information as they can identify specific contact points that can be used in determining both location and orientation. In testing small molecule docking into G-protein coupled receptors, Nguyen et al. demonstrated that sampling efficiency can increase by an order of magnitude for every ten known protein−small molecule contacts. The gain was even greater when utilizing more detailed information such as a specific ionic interaction translated as a 3.0 Å distance restraint.[52]

A protein−small molecule double mutant cycle analysis identifies interactions by comparing ITC binding data of a single protein mutation, a single small molecule functional group substitution, and both simultaneously. A substantial nonadditive interaction energy change is evidence for a direct interaction. A collection of these pairwise interactions can be used to derive protein-small molecule distance constraints to incorporate into the energy gradient docking grid. Roisman et al. probed 100 potential pairwise interactions in an interferon-receptor complex and identified five significant interactions. Docking simulations were run until a converged model was generated satisfying the five restraints.[53] Blum et al. utilized double mutant cycle analysis with a nicotine analogue and an acetylcholine receptor backbone amide substitution to show a hydrogen bond interaction.[54] Similar success has been obtained with double mutant cycle alanine scanning for a yeast Ste2p GPCR[55] and with an allosteric binding site on a hM1 muscarinic receptor complex.[56] Compared to the traditional single site-directed mutagenesis method, this approach has the potential to differentiate the impact on small molecule binding from disrupting favorable interactions vs protein stability.

Another method of directly determining intermolecular distance restraints is with protein−small molecule NOEs. However, this is generally limited by the need to assign resonances for the protein-small molecule complex. One alternative that relies on matching only small molecule resonance assignments was developed by Constantine et al.[57] NMR NOE experiments can also be used to derive relative orientations between two weakly binding, competitive small molecules. The INPHARMA technique relies on the transfer of NOEs between the two small molecules and a common receptor target. This in combination with a crystal structure of one protein−small molecule complex can be used in determining the binding of a related small molecule series without assuming binding in a similar fashion.[58]

Small molecule similarity is a type of interface restraint based on the known binding mode between a related small molecule and the given protein target. The binding mode of the related molecule can be used as a guide in placement and orientation during docking simulations. In the 2015 D3R Grand Challenge involving blinded docking of HSP90 and MAP4K4 binders, the most successful workflows superpositioned targets onto similar small molecules instead of sampling large, unrestricted binding space.[59] HybridDock is one approach that augments docking with molecular similarity by generating possible binding modes using existing cocrystallized molecules. This significantly improved both the binding energy correlation and native binding mode recovery in a CSAR 2013−2014 test.[60] LigBEnD is a similar approach that uses the cocrystallized molecule to generate an atomic property force field. Scoring models with this small molecule-based force field correctly predicted 30 out of 36 compounds in the D3R docking challenge.[61] Related binding pockets may be found even among proteins of distinct global folds and evolutionary history. An analysis of potential enzyme drug targets and evolutionarily distant proteins in the PDB found similar binding pockets with different global folds in 61%, 10%, and 61% of kinases, phosphatases, and proteases, respectively.[62]

SYSTEM SPECIFIC EXPERIMENTAL DATA CAN IMPROVE BOTH SAMPLING AND SCORING

Small molecule docking can be divided into two codependent challenges: sampling and scoring. Efficient sampling requires the program to rapidly generate near-native binding poses, while accurate scoring needs the program to distinguish correct binding interactions from decoys. Algorithms incorporating system specific data offer improvements to both sampling and scoring.

Ross et al. demonstrated that a universal scoring function, trained across a diverse data set, is less accurate than a target-specific scoring function when applied to a single system. A score function that captures experimental affinities across the entire data set has varying accuracy on single protein data sets.[63] Thus, protein target dependence remains a significant challenge in choosing diverse benchmark targets for scoring function development, and in selecting a score function for use in a particular application system. Current approaches to overcome this bias are to test a multitude of scoring protocols before selecting one, or to utilize rescoring algorithms such as NNScore that can be suited to a specific receptor for screening applications.[64] Trained scoring functions such as the random

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233

forest based SFCscore or the support vector machine based SVRR may be fitted to a given target by careful choice of training data and/or descriptors.[65,66] One major advantage of tuned scoring functions is the ability to reproduce experimental activity data. AutoShim, for example, uses provided IC50 activity data and partial least-squares regression to parametrize the scoring function. This empirical correction improved experimental SAR correlation from an all-purpose scoring function best of 0.32 to 0.5 across a GSK docking set.[67] However, care must be taken to avoid overfitting as having the same protein families in training and validation can produce unrealistically high prediction accuracy.[68]

MOLECULAR SIMILARITY AS AN INTERFACE TYPE DOCKING RESTRAINT

One common assumption in small molecule docking and screening is that chemically similar small molecules bind a given receptor in similar fashions. The common molecular scaffold is presumed to make similar interactions with the binding pocket, while peripheral functional group modifications create different contacts that explain SARs. Whether or not this assumption holds up across a broad set of diverse protein-small molecule complexes is critical in accessing the applicability of molecular similarity restraints.

Previous analysis of 206 protein−small molecule structure pairs observed similar small molecule binding modes in 90 percent of related structures. Binding similarity was defined as having an optimized small molecule shape Tanimoto of greater than 0.8. The receptors in those cases exhibited very similar backbone structures with the primary differences due to side chain conformation or water architecture.[69] An examination of scaffold building pairs found that in 41 out of 297 cases the binding mode changed upon chemical elaboration of a scaffold.[70]

The current work expands the previous surveys through the 2014 version of the PDBBind refined set,[71] which contains 3446 structures controlled for protein structure quality, accurate binding data, small molecule properties, and nonsurface interactions.[13] Each protein−small molecule cocrystal structure is paired with experimentally measured K_d, K_i, or IC50.

Selecting systems with at least two crystal structures produces 2443 structures across 441 targets. All complexes within a system were then aligned based on binding pocket residues within 15 Å of the small molecule. The in-house BioChemicalLibrary (BCL) software suite, available at http://www.meilerlab.org/bclcommons, is used to calculate small molecule properties and make all possible intrasystem pairwise comparisons. The pairwise normalized RMSD (nRMSD) is calculated based on the heavy atom RMSD of the largest common connected substructure and normalized to ten heavy atoms using a small molecule analog of RMSD-100.[72] The pairwise Tanimoto similarity coefficient is computed as the number of atoms in the largest common connected substructure divided by the total number of unique atoms.

All 34 461 pairwise comparisons were then filtered to eliminate identical small molecule pairs (Tanimoto < 1), trivial common scaffolds (scaffold heavy atoms ≥ 6), and different binding pockets (small molecule center distance > 3.0). A large number of the remaining comparisons were from the HIV-1-Protease system, which dominates the PDBBind refined set. In order to avoid biasing the results toward any particular system, only a randomly chosen small subset of HIV-1-Protease comparisons are included. Small molecule symmetry was factored in when calculating nRMSDs for pairs where symmetrical molecules are flipped. The final analysis included a total of 7298 comparisons across 366 targets. The median Tanimoto similarity across the data set was 0.333, and the median nRMSD of the common scaffold was 1.071 Å.

Figure 3 shows the relationship between the nRMSD and Tanimoto similarity for all pairwise comparisons. The tabular

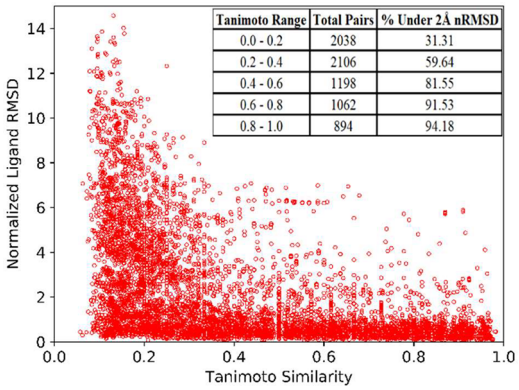


Figure 3. Pairwise small molecule scaffold nRMSD vs Tanimoto similarity. (inset) Number of pairs and percentage under nRMSD cutoff for each Tanimoto range.

breakdown shows an increasing percentage of pairs below 2 Å nRMSD with increasing Tanimoto similarity. The data set included 548 pairs where the common substructure is equivalent in size to the smaller molecule. A small molecule binding mode change (nRMSD > 2.0 Å) was observed in 52 cases (9.5%), slightly lower than the percent changed based on volume overlap comparison used by Malhotra and Karanicolas.[70]

A Tanimoto similarity of 0.85 is often regarded as a cutoff for two molecules having similar biological behavior.[73] However, recent studies suggest that different metrics and different systems will often produce distinct limits.[74] The pairwise comparisons were decomposed as distribution plots in ten bins of Tanimoto similarity shown in Figure 4. Although the median nRMSD is below 2.0 Å for Tanimoto values above 0.2, the

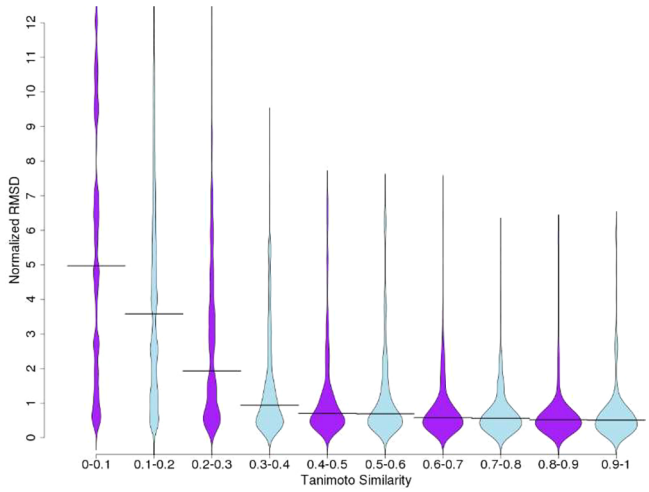


Figure 4. nRMSD distributions for each of ten bins of Tanimoto similarity. The median of each distribution is shown in black.

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233

nRMSD distribution continues to shift downward. In particular, the maximum nRMSD observed decreases significantly.

The distribution appears fairly stable beyond a Tanimoto value of 0.7, suggesting a small molecule similarity cutoff in this neighborhood.

This suggests molecular similarity as a simple metric for determining if docking positioning can rely on known binding modes, though predictive power may be improved by factoring additional properties such as potency, size, and lipophilicity.[70]

EXCEPTIONS TO MOLECULAR SIMILARITY INTERFACE RESTRAINTS

In 11 small molecule pairs, highly similar small molecules (Tanimoto > 0.7) exhibited significantly different binding modes (nRMSD > 5.0 Å). In one particular example, a series of diflunisal derivatives exhibited two opposite orientations when bound to transthyretin, a protein involved in amyloidogenesis. The lead compound diflunisal was found to bind in a forward and a reverse orientation. More interestingly, a *meta*-difluoro derivative (Figure 5, left) was found exclusively in the reverse binding mode while an *ortho*-difluoro derivative (Figure 5, right) was found exclusively in the forward binding mode.[75]

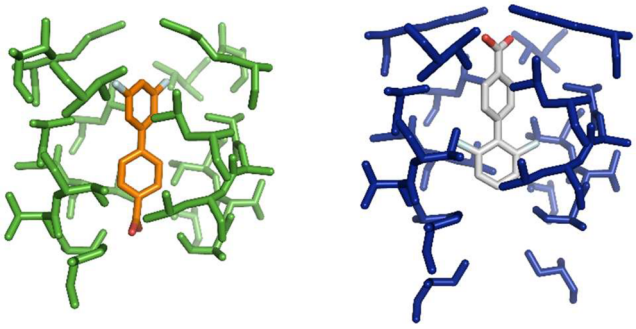


Figure 5. *meta*-Difluoro diflunisal derivative (left, PDB: 2B9A) and *ortho*-difluoro diflunisal derivative (right, PDB: 2F7I) bound to transthyretin.

Another counterexample in the data set is the binding of bile acids, taurocholate and cholate, to *Campylobacter jejuni* CmeR regulator protein. The two compounds differ by a distal anionic group but are found bound in antiparallel orientations as shown in Figure 6. The two compounds share the same volume of the binding pocket and is suggested to interact similarly with a previously identified glycerol binding site. Furthermore, the pocket is highlighted by a large hydrophobic tunnel with numerous mini-pockets, suggesting a reason as to why it is capable of binding diverse small molecules in diverse fashions.[76]

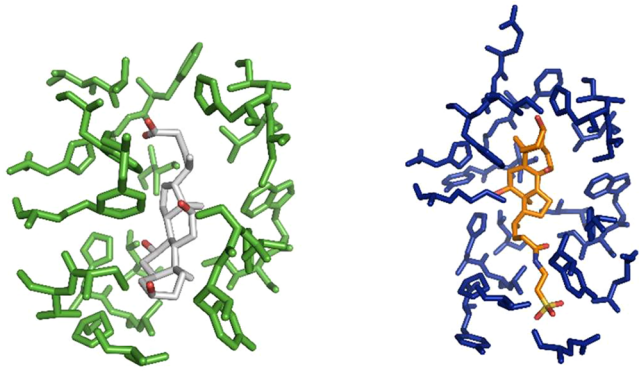


Figure 6. Cholate (left, PDB: 3QPS) and taurocholate (right, PDB: 3QQA) bound to CmeR regulator protein.

A number of notable exceptions can be found in literature as well. Structure-based design of an influenza neuraminidase inhibitor series showed up to 180° variation in the orientation of a central five-member ring. Potent analogues were only found for the congeneric series that bound in the same orientation with consistent SAR.[77] A study of dipeptidyl peptidase IV inhibitors showed chemically similar small molecules with different distal aromatic substitutions and placements bind in distinct orientations. The substituted phenyl ring made $\pi-\pi$ interactions but with distinct residues in the different cases.[78] Another common exception in systems such as HIV-1-Protease involve inhibitors bound in two approximately symmetrical orientations.[79] Kim et al. discusses a number of other exceptions, such as dihydrofolate reductase and cytochrome c peroxidase small molecules, that were identified through examination of outliers left out when constructed QSAR data sets. It should be noted that some of the exceptions involve conformational changes in the distal parts of the small molecule while the main chemical scaffold remains aligned.[80] It may also be possible in these QSAR data sets that similar small molecules reside in different conformations of a flexible protein binding pocket. In such cases, the small molecule orientations remain constant but the protein-small molecule contacts change.[81]

Based on the PDBBind refined set survey, these exceptions are fairly uncommon. Some features frequently seen in these exceptions include nearly symmetrical molecules, large binding pockets allowing multiple orientations, and distal groups capable of making favorable interactions with different residues in the binding pocket. Unfortunately, there are no currently known small molecule or receptor structure factors to distinguish exceptions from regular binders.

Although a similarity based approach toward docking or screening with atoms aligned by identity may not work in these particular cases, there may be remedies using molecular properties. A docking method utilizing pharmacophores with properties such as partial charge or hydrogen bond donor/acceptors can alleviate this problem. Ph4Dock is an example where atoms are represented as electrical charge centers without consideration for identity.[82] Furthermore, similar contact residues are often observed in these situations allowing for productive suggestions of pairwise interaction validation experiments such as double mutant cycles.

CONCLUSION

The introduction of biochemical data in the modeling process is a strong promoter of protein-small molecule modeling success and accuracy. Spectroscopic methods such as NMR and mass spectrometry have been adapted to interrogate receptor−small molecule interactions. One can expect improvements in both sampling and scoring when incorporating experimental contacts or structure activity relationships. Structure activity relationships derive from energetic changes upon modifications to either the receptor or the small molecule. An analysis of the Platinum Database shows that binding affinity changes upon single residue mutagenesis is a moderate predictor of binding contacts. Docking improvements are particularly significant when experimental restraints limit both translational and rotational modes of the small molecule. Such restraints can

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233

be derived from interface-based experiments that identifies specific protein−small molecule interactions, or by relying on binding modes of similar small molecules. These methods would rely on the assumption that similarities in structure translates to similarities in binding. A pairwise comparison across the PDBBind Database shows that this is generally true among congeneric small molecules. Exceptions exist wherein typically highly symmetric molecules bind in inverted orientations. These situations may be addressed by further development of pharmacophore, or molecular property, based alignment methods. In cases where experimental data is lacking, computational ensemble methods, such as molecular dynamics, are presently available to account for protein and small molecule flexibility. However, there is significant opportunity to develop methods utilizing structural ensembles of related proteins and small molecules.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jens.meiler@vanderbilt.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is funded by a Predoctoral Fellowship from the PhRMA foundation awarded to D.Y.F. Work in the J.M. laboratory is supported in part by NIH (R01 GM099842, and R01 DK097376) and the National Science Foundation (CHE 1305874). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors thank Dr. Terry Lybrand, Dr. Andes Hess, and Dr. Corey Hopkins for their discussions regarding this work.

REFERENCES

(1) 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.

(2) Ewing, T. J.; 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.

(3) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(4) Jones, G.; Glen, R. C.; Leach, A. R.; Taylor, R.; Willett, P. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267*, 727−748.

(5) Lemmon, G. H.; Meiler, J. RosettaLigand Docking with Flexible XML Protocols. *Methods Mol. Biol.* **2012**, *819*, 143−155.

(6) Kalyaanamoorthy, S.; Chen, Y.-P. P. Structure-Based Drug Design to Augment Hit Discovery. *Drug Discovery Today* **2011**, *16*, 831−839.

(7) Davis, I. W.; Raha, K.; Head, M. S.; Baker, D. Blind Docking of Pharmaceutically Relevant Compounds Using RosettaLigand. *Protein Sci.* **2009**, *18*, 1998−2002.

(8) Damm-Ganamet, K. L.; Smith, R. D.; Dunbar, J. B.; Stuckey, J. A.; Carlson, H. A. CSAR Benchmark Exercise 2011−2012: Evaluation of Results from Docking and Relative Ranking of Blinded Congeneric Series. *J. Chem. Inf. Model.* **2013**, *53*, 1853−1870.

(9) Li, Y.; Han, L.; Liu, Z.; Wang, R. Comparative Assessment of Scoring Functions on an Updated Benchmark: 2. Evaluation Methods and General Results. *J. Chem. Inf. Model.* **2014**, *54*, 1717−1736.

(10) Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J. A.; Kang, Y. N.; Delproposto, J.; Kubish, G.; Dunbar, J. B.; Carlson, H. A.; Burley, S. K.; Walters, W. P.; Amaro, R. E.; Feher, V. A.; Gilson, M. K. D3R Grand Challenge 2015: Evaluation of Protein−ligand Pose and Affinity Predictions. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 651.

(11) DiMaio, F.; Song, Y.; Li, X.; Brunner, M. J.; Xu, C.; Conticello, V.; Egelman, E.; Marlovits, T. C.; Cheng, Y.; Baker, D. Atomic-Accuracy Models from 4.5-Å Cryo-Electron Microscopy Data with Density-Guided Iterative Local Refinement. *Nat. Methods* **2015**, *12*, 361−365.

(12) Fischer, A. W.; Alexander, N. S.; Woetzel, N.; Karakas, M.; Weiner, B. E.; Meiler, J. BCL::MP-Fold: Membrane Protein Structure Prediction Guided by EPR Restraints. *Proteins: Struct., Funct., Genet.* **2015**, *83*, 1947.

(13) Li, Y.; Han, L.; Li, J.; Liu, Z.; Zhao, Z.; Wang, R.; Liu, J. Comparative Assessment of Scoring Functions on an Updated Benchmark: 1. Compilation of the Test Set. *J. Chem. Inf. Model.* **2014**, *54*, 1700−1716.

(14) Bissantz, C.; Kuhn, B.; Stahl, M. A Medicinal Chemist's Guide to Molecular Interactions. *J. Med. Chem.* **2010**, *53*, 5061−5084.

(15) 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.

(16) Huwe, P. J.; Xu, Q.; Shapovalov, M. V.; Modi, V.; Andrake, M. D.; Dunbrack, R. L. Biological Function Derived from Predicted Structures in CASP11. *Proteins: Struct., Funct., Genet.* **2016**, *84*, 370.

(17) Bordogna, A.; Pandini, A.; Bonati, L. Predicting the Accuracy of Protein-Ligand Docking on Homology Models. *J. Comput. Chem.* **2011**, *32*, 81−98.

(18) 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*, 365−379.

(19) Ferrara, P.; Jacoby, E. Evaluation of the Utility of Homology Models in High Throughput Docking. *J. Mol. Model.* **2007**, *13*, 897−905.

(20) Kaufmann, K. W.; Meiler, J. Using RosettaLigand for Small Molecule Docking into Comparative Models. *PLoS One* **2012**, *7*, 7.

(21) Cooper, D.; et al. X-Ray Crystallography: Assessments and Validation of Protein-Small Molecule Complexes for Drug Discovery. *Expert Opin. Drug Discovery* **2011**, *6*, 771−782.

(22) ten Brink, T.; Aguirre, C.; Exner, T. E.; Krimm, I. Performance of Protein−Ligand Docking with Simulated Chemical Shift Perturbations. *J. Chem. Inf. Model.* **2015**, *55*, 275−283.

(23) Orts, J.; Bartoschek, S.; Griesinger, C.; Monecke, P.; Carlomagno, T. An NMR-Based Scoring Function Improves the Accuracy of Binding Pose Predictions by Docking by Two Orders of Magnitude. *J. Biomol. NMR* **2012**, *52*, 23−30.

(24) Onila, I.; ten Brink, T.; Fredriksson, K.; Codutti, L.; Mazur, A.; Griesinger, C.; Carlomagno, T.; Exner, T. E. On-the-Fly Integration of Data from a Spin-Diffusion-Based NMR Experiment into Protein−Ligand Docking. *J. Chem. Inf. Model.* **2015**, *55*, 1962.

(25) Skjærven, L.; Codutti, L.; Angelini, A.; Grimaldi, M.; Latek, D.; Monecke, P.; Dreyer, M. K.; Carlomagno, T. Accounting for Conformational Variability in Protein-Ligand Docking with Nmr-Guided Rescoring. *J. Am. Chem. Soc.* **2013**, *135*, 5819−5827.

(26) Cala, O.; Guilliere, F.; Krimm, I. NMR-Based Analysis of Protein-Ligand Interactions. *Anal. Bioanal. Chem.* **2014**, *406*, 943−956.

(27) Ishii, K.; Noda, M.; Uchiyama, S. Mass Spectrometric Analysis of Protein−ligand Interactions. *Biophys. Physicobiology* **2016**, *13*, 87−95.

(28) Mouchlis, V. D.; Morisseau, C.; Hammock, B. D.; Li, S.; McCammon, J. A.; Dennis, E. A. Computer-Aided Drug Design Guided by Hydrogen/deuterium Exchange Mass Spectrometry: A Powerful Combination for the Development of Potent and Selective Inhibitors of Group VIA Calcium-Independent Phospholipase A2. *Bioorg. Med. Chem.* **2016**, *24*, 4801−4811.

(29) Leavitt, S.; Freire, E. Direct Measurement of Protein Binding Energetics by Isothermal Titration Calorimetry. *Curr. Opin. Struct. Biol.* **2001**, *11*, 560−566.

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233

(30) Pires, D. E. V.; Blundell, T. L.; Ascher, D. B. Platinum: A Database of Experimentally Measured Effects of Mutations on Structurally Defined Protein-Ligand Complexes. *Nucleic Acids Res.* **2015**, *43*, D387−D391.
(31) Kufareva, I.; Rueda, M.; Katritch, V.; et al. Status of GPCR Modeling and Docking as Reflected by Community-Wide GPCR Dock 2010 Assessment. *Structure* **2011**, *19*, 1108−1126.
(32) Kufareva, I.; Katritch, V.; Stevens, R. C.; Abagyan, R. Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure* **2014**, *22*, 1120−1139.
(33) Rueda, M.; Bottegoni, G.; Abagyan, R. Consistent Improvement of Cross-Docking Results Using Binding Site Ensembles Generated with Elastic Network Normal Modes. *J. Chem. Inf. Model.* **2009**, *49*, 716−725.
(34) Korb, O.; Olsson, T. S. G.; Bowden, S. J.; Hall, R. J.; Verdonk, M. L.; Liebeschuetz, J. W.; Cole, J. C. Potential and Limitations of Ensemble Docking. *J. Chem. Inf. Model.* **2012**, *52*, 1262−1274.
(35) Amaro, R. E.; 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.
(36) Xu, M.; Lill, M. A. Utilizing Experimental Data for Reducing Ensemble Size in Flexible-Protein Docking. *J. Chem. Inf. Model.* **2012**, *52*, 187−198.
(37) Sinko, W.; Lindert, S.; McCammon, J. A. Accounting for Receptor Flexibility and Enhanced Sampling Methods in Computer-Aided Drug Design. *Chem. Biol. Drug Des.* **2013**, *81*, 41−49.
(38) Feixas, F.; Lindert, S.; Sinko, W.; McCammon, J. A. Exploring the Role of Receptor Flexibility in Structure-Based Drug Discovery. *Biophys. Chem.* **2014**, *186*, 31−45.
(39) Wei, N.-N.; Hamza, A. SABRE: Ligand/structure-Based Virtual Screening Approach Using Consensus Molecular-Shape Pattern Recognition. *J. Chem. Inf. Model.* **2014**, *54*, 338−346.
(40) Vieth, M.; Cummins, D. J. DoMCoSAR: A Novel Approach for Establishing the Docking Mode That Is Consistent with the Structure-Activity Relationship. Application to HIV-1 Protease Inhibitors and VEGF Receptor Tyrosine Kinase Inhibitors. *J. Med. Chem.* **2000**, *43*, 3020−3032.
(41) Dinsmore, C. J.; Bogusky, M. J.; Culberson, J. C.; Bergman, J. M.; Homnick, C. F.; Zartman, C. B.; Mosser, S. D.; Schaber, M. D.; Robinson, R. G.; Koblan, K. S.; Huber, H. E.; Graham, S. L.; Hartman, G. D.; Huff, J. R.; Williams, T. M. Conformational Restriction of Flexible Ligands Guided by the Transferred NOE Experiment: Potent Macrocyclic Inhibitors of Farnesyltransferase [25]. *J. Am. Chem. Soc.* **2001**, *123*, 2107−2108.
(42) Yang, S.-Y. Pharmacophore Modeling and Applications in Drug Discovery: Challenges and Recent Advances. *Drug Discovery Today* **2010**, *15*, 444−450.
(43) Hu, B.; Lill, M. A. PharmDock: A Pharmacophore-Based Docking Program. *J. Cheminf.* **2014**, *6*, 14.
(44) Rakers, C.; Schumacher, F.; Meinl, W.; Glatt, H.; Kleuser, B.; Wolber, G. In Silico Prediction of Human Sulfotransferase 1E1 Activity Guided by Pharmacophores from Molecular Dynamics Simulations. *J. Biol. Chem.* **2016**, *291*, 58−71.
(45) Dimova, D.; Heikamp, K.; Stumpfe, D.; Bajorath, J. Do Medicinal Chemists Learn from Activity Cliffs? A Systematic Evaluation of Cliff Progression in Evolving Compound Data Sets. *J. Med. Chem.* **2013**, *56*, 3339−3345.
(46) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42*, D1083−D1090.
(47) Chan, W. K. B.; Zhang, H.; Yang, J.; Brender, J. R.; Hur, J.; Özgür, A.; Zhang, Y. GLASS: A Comprehensive Database for Experimentally-Validated GPCR-Ligand Associations. *Bioinformatics* **2015**, *31*, btv302.
(48) Van Linden, O. P. J.; Kooistra, A. J.; Leurs, R.; De Esch, I. J. P.; De Graaf, C. KLIFS: A Knowledge-Based Structural Database to Navigate Kinase-Ligand Interaction Space. *J. Med. Chem.* **2014**, *57*, 249−277.
(49) Li, H.; Li, C. Multiple Ligand Simultaneous Docking: Orchestrated Dancing of Ligands in Binding Sites of Protein. *J. Comput. Chem.* **2010**, *31*, 2014−2022.
(50) Allison, T. M.; Reading, E.; Liko, I.; Baldwin, A. J.; Laganowsky, A.; Robinson, C. V. Quantifying the Stabilizing Effects of Protein-Ligand Interactions in the Gas Phase. *Nat. Commun.* **2015**, *6*, 8551.
(51) Evers, A.; Gohlke, H.; Klebe, G. Ligand-Supported Homology Modelling of Protein Binding-Sites Using Knowledge-Based Potentials. *J. Mol. Biol.* **2003**, *334*, 327−345.
(52) Nguyen, E. D.; Norn, C.; Frimurer, T. M.; Meiler, J. Assessment and Challenges of Ligand Docking into Comparative Models of G-Protein Coupled Receptors. *PLoS One* **2013**, *8*, e67302.
(53) Roisman, L. C.; Piehler, J.; Trosset, J.-Y.; Scheraga, H. A.; Schreiber, G. Structure of the Interferon-Receptor Complex Determined by Distance Constraints from Double-Mutant Cycles and Flexible Docking. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 13231−13236.
(54) Blum, A. P.; Lester, H. A.; Dougherty, D. A. Nicotinic Pharmacophore: The Pyridine N of Nicotine and Carbonyl of Acetylcholine Hydrogen Bond across a Subunit Interface to a Backbone NH. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 13206−13211.
(55) Naider, F.; Becker, J. M.; Lee, Y. H.; Horovitz, A. Double-Mutant Cycle Scanning of the Interaction of a Peptide Ligand and Its G Protein-Coupled Receptor. *Biochemistry* **2007**, *46*, 3476−3481.
(56) Marquer, C.; Fruchart-Gaillard, C.; Letellier, G.; Marcon, E.; Mourier, G.; Zinn-Justin, S.; Ménez, A.; Servent, D.; Gilquin, B. Structural Model of Ligand-G Protein-Coupled Receptor (GPCR) Complex Based on Experimental Double Mutant Cycle Data: MT7 Snake Toxin Bound to Dimeric hM1Muscarinic Receptor. *J. Biol. Chem.* **2011**, *286*, 31661−31675.
(57) Constantine, K. L.; Davis, M. E.; Metzler, W. J.; Mueller, L.; Claus, B. L. Protein-Ligand NOE Matching: A High-Throughput Method for Binding Pose Evaluation That Does Not Require Protein NMR Resonance Assignments. *J. Am. Chem. Soc.* **2006**, *128*, 7252−7263.
(58) Orts, J.; Tuma, J.; Reese, M.; Grimm, S. K.; Monecke, P.; Bartoschek, S.; Schiffer, A.; Wendt, K. U.; Griesinger, C.; Carlomagno, T. Crystallography-Independent Determination of Ligand Binding Modes. *Angew. Chem., Int. Ed.* **2008**, *47*, 7736−7740.
(59) Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J. A.; Kang, Y. N.; Delproposto, J.; Kubish, G.; Dunbar, J. B.; Carlson, H. A.; Burley, S. K.; Walters, W. P.; Amaro, R. E.; Feher, V. A.; Gilson, M. K. D3R Grand Challenge 2015: Evaluation of Protein−ligand Pose and Affinity Predictions. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 651−668.
(60) Huang, S.-Y.; Li, M.; Wang, J.; Pan, Y. HybridDock: A Hybrid Protein-Ligand Docking Protocol Integrating Protein- and Ligand-Based Approaches. *J. Chem. Inf. Model.* **2016**, *56*, 1078.
(61) Lam, P. C.-H.; Abagyan, R.; Totrov, M. Ligand-Biased Ensemble Receptor Docking (LigBEnD): A Hybrid Ligand/receptor Structure-Based Approach. *J. Comput.-Aided Mol. Des.* **2017**, DOI: 10.1007/s10822-017-0058-x.
(62) Skolnick, J.; Gao, M.; Roy, A.; Srinivasan, B.; Zhou, H. Implications of the Small Number of Distinct Ligand Binding Pockets in Proteins for Drug Discovery, Evolution and Biochemical Function. *Bioorg. Med. Chem. Lett.* **2015**, *25*, 1163−1170.
(63) Ross, G. a; Morris, G. M.; Biggin, P. C. One Size Does Not Fit All: The Limits of Structure-Based Models in Drug Discovery. *J. Chem. Theory Comput.* **2013**, *9*, 4266−4274.
(64) Durrant, J. D.; Friedman, A. J.; Rogers, K. E.; McCammon, J. A. Comparing Neural-Network Scoring Functions and the State of the Art: Applications to Common Library Screening. *J. Chem. Inf. Model.* **2013**, *53*, 1726−1735.
(65) Zilian, D.; Sotriffer, C. a. SFCscore(RF): A Random Forest-Based Scoring Function for Improved Affinity Prediction of Protein-Ligand Complexes. *J. Chem. Inf. Model.* **2013**, *53*, 1923−1933.

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233

(66) Wang, W.; He, W.; Zhou, X.; Chen, X. Optimization of Molecular Docking Scores with Support Vector Rank Regression. *Proteins: Struct., Funct., Genet.* **2013**, *81*, 1386−1398.
(67) Martin, E.; Sullivan, D. AutoShim: Empirically Corrected Scoring Functions for Quantitative Docking with a Crystal Structure and IC50 Training Data. *J. Chem. Inf. Model.* **2008**, *48*, 861−872.
(68) Kramer, C.; Gedeck, P. Leave-Cluster-out Cross-Validation Is Appropriate for Scoring Functions Derived from Diverse Protein Data Sets. *J. Chem. Inf. Model.* **2010**, *50*, 1961−1969.
(69) Boström, J.; Hogner, A.; Schmitt, S. Do Structurally Similar Ligands Bind in a Similar Fashion? *J. Med. Chem.* **2006**, *49*, 6716−6725.
(70) Malhotra, S.; Karanicolas, J. When Does Chemical Elaboration Induce a Ligand To Change Its Binding Mode? *J. Med. Chem.* **2017**, *60*, 128−145.
(71) Liu, Z.; Li, Y.; Han, L.; Li, J.; Liu, J.; Zhao, Z.; Nie, W.; Liu, Y.; Wang, R. PDB-Wide Collection of Binding Data: Current Status of the PDBbind Database. *Bioinformatics* **2015**, *31*, 405−412.
(72) Gregory, K. J.; Nguyen, E. D.; Malosh, C.; Mendenhall, J. L.; Zic, J. Z.; Bates, B. S.; Noetzel, M. J.; Squire, E. F.; Turner, E. M.; Rook, J. M.; Emmitte, K. a.; Stauffer, S. R.; Lindsley, C. W.; Meiler, J.; Conn, P. J. Identification of Specific Ligand-Receptor Interactions That Govern Binding and Cooperativity of Diverse Modulators to a Common Metabotropic Glutamate Receptor 5 Allosteric Site. *ACS Chem. Neurosci.* **2014**, *5*, 282−295.
(73) Patterson, D. E.; Cramer, R. D.; Ferguson, a M.; Clark, R. D.; Weinberger, L. E. Neighborhood Behavior: A Useful Concept for Validation Of "molecular Diversity" descriptors. *J. Med. Chem.* **1996**, *39*, 3049−3059.
(74) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in Virtual Screening: Foundations, Limitations and Novel Approaches. *Drug Discovery Today* **2007**, *12*, 225−233.
(75) Adamski-Werner, S. L.; Palaninathan, S. K.; Sacchettini, J. C.; Kelly, J. W. Diflunisal Analogues Stabilize the Native State of Transthyretin. Potent Inhibition of Amyloidogenesis. *J. Med. Chem.* **2004**, *47*, 355−374.
(76) Lei, H. T.; Shen, Z.; Surana, P.; Routh, M. D.; Su, C. C.; Zhang, Q.; Yu, E. W. Crystal Structures of CmeR-Bile Acid Complexes from Campylobacter Jejuni. *Protein Sci.* **2011**, *20*, 712−723.
(77) Stoll, V.; Stewart, K. D.; Maring, C. J.; Muchmore, S.; Giranda, V.; Gu, Y. Y.; Wang, G.; Chen, Y.; Sun, M.; Zhao, C.; Kennedy, A. L.; Madigan, D. L.; Xu, Y.; Saldivar, A.; Kati, W.; Laver, G.; Sowin, T.; Sham, H. L.; Greer, J.; Kempf, D. Influenza Neuraminidase Inhibitors: Structure-Based Design of a Novel Inhibitor Series. *Biochemistry* **2003**, *42*, 718−727.
(78) Pei, Z.; Li, X.; Longenecker, K.; von Geldern, T. W.; Wiedeman, P. E.; Lubben, T. H.; Zinker, B. A.; Stewart, K.; Ballaron, S. J.; Stashko, M. a; Mika, A. K.; Beno, D. W. A.; Long, M.; Wells, H.; Kempf-Grote, A. J.; Madar, D. J.; McDermott, T. S.; Bhagavatula, L.; Fickes, M. G.; Pireh, D.; Solomon, L. R.; Lake, M. R.; Edalji, R.; Fry, E. H.; Sham, H. L.; Trevillyan, J. M. Discovery, Structure−Activity Relationship, and Pharmacological Evaluation of (5-Substituted-Pyrrolidinyl-2-Carbonyl)-2-Cyanopyrrolidines as Potent Dipeptidyl Peptidase IV Inhibitors. *J. Med. Chem.* **2006**, *49*, 3520−3535.
(79) Fitzgerald, P. M.; McKeever, B. M.; VanMiddlesworth, J. F.; Springer, J. P.; Heimbach, J. C.; Leu, C. T.; Herber, W. K.; Dixon, R. A.; Darke, P. L. Crystallographic Analysis of a Complex between Human Immunodeficiency Virus Type 1 Protease and Acetyl-Pepstatin at 2.0-A Resolution. *J. Biol. Chem.* **1990**, *265*, 14209−14219.
(80) Kim, K. H. Outliers in SAR and QSAR: Is Unusual Binding Mode a Possible Source of Outliers? *J. Comput.-Aided Mol. Des.* **2007**, *21*, 63−86.
(81) Kim, K. H. Outliers in SAR and QSAR: 2. Is a Flexible Binding Site a Possible Source of Outliers? *J. Comput.-Aided Mol. Des.* **2007**, *21*, 421−435.
(82) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: Pharmacophore-Based Protein - Ligand Docking. *J. Med. Chem.* **2004**, *47*, 6804−6811.

DOI: 10.1021/acs.jcim.7b00418
J. Chem. Inf. Model. 2018, 58, 225−233