JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

POSIT: Flexible Shape-Guided Docking For Pose Prediction

Brian P. Kelley,[†,‡] Scott P. Brown,[⊥,§] Gregory L. Warren,[*,†] and Steven W. Muchmore[⊥]

[†]OpenEye Scientific Software, 9 Bisbee Court, Santa Fe, New Mexico 87508, United States

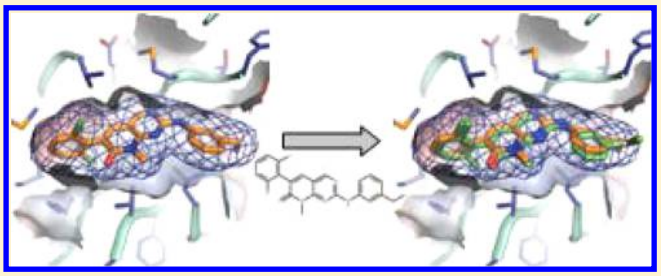
[‡]Novartis Institutes for BioMedical Research, Inc, 250 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

[§]Sunovion Pharmaceuticals Inc, 84 Waterford Drive, Marlborough, Massachusetts 01752, United States

[⊥]Global Pharmaceutical Research Division, AbbVie, 10 Waukegan Road, North Chicago, Illinois 60064, United States

Supporting Information

ABSTRACT: We present a new approach to structure-based drug design (POSIT) rigorously built on the simple concept that pose prediction is intimately coupled to the quality and availability of experimental structural data. We demonstrate the feasibility of the approach by performing retrospective analyses on three data sets designed to explore the strengths and weaknesses of POSIT relative to existing methods. We then present results documenting 2.5 years of prospective use of POSIT across a variety of structure-based industrial drug-discovery research projects. We find that POSIT is well-suited to guiding research decision making for structure-based design and, in particular, excels at enabling lead-optimization campaigns. We show that the POSIT framework can drive superior pose-prediction performance and generate results that naturally lend themselves to prospective decision making during lead optimization. We believe the results presented here are (1) the largest prospective validation of a pose prediction method reported to date (71 crystal structures); (2) provide an unprecedented look at the scope of impact of a computational tool; and (3) represent a first-of-its-kind analysis. We hope that this work inspires additional studies that look at the real impact and performance of computational research tools on prospective drug design.



I. INTRODUCTION

To effectively execute an industrial drug-discovery campaign, typically hundreds to thousands of small-molecule inhibitors are made by medicinal chemists in the course of optimizing molecular properties to produce the desired pharmacological effects. The workflow for optimization typically involves some measurement of activity against a selection of protein targets thought to be relevant to therapeutic goals of the research. For structure-based drug-design projects, research investigators will have available the three-dimensional (3D) coordinates of a number of protein targets, which ideally contain bound inhibitors with direct relevance to the project. Knowledge of 3D structural information allows researchers to employ both theoretical and experimental tools to help guide medicinal-chemistry efforts, with the potential to positively impact drug-design cycle times. 3D structures with sufficiently high resolution can provide detailed insight into the specific interactions between bound small-molecule inhibitors and the protein binding site. Thus, experimentally derived structural information can be highly prized as it contains knowledge directly pertinent to interactions between protein and inhibitory ligand, the improvement of which are one of the primary goals of lead-optimization (LO).

Generally, only a small fraction (typically much less than 1%) of the inhibitors made for a given project will ever have X-ray crystallographic (XRC) determination of the 3D coordinates for the bound protein−inhibitor complex. This disparity presents a great opportunity for computational methods capable of predicting binding modes; however, it also presents a substantial challenge for computational prediction. To generate a useful prediction, a computational method should (1) incorporate knowledge of all available structural information, ideally both existing protein and ligand structures; (2) deliver results in a timely manner consistent with research project timelines; and (3) define a level of confidence in the prediction. Without incorporation of some combination of all these aspects, it is difficult for predictions to realize their full impact on decision making and consistently guide chemistry in a rational manner.

One of the most frequently used pose-prediction methods in drug discovery is the approach known as molecular docking. Molecular docking attempts to predict bound ligand conformations based solely on knowledge of the 3D structure and arrangement of residues in the protein active site. Docking approaches attempt to approximate the inherently complex process of in vivo molecular recognition, and hence, have traditionally produced results containing substantial uncertainty. In fact, the inaccuracies and inconsistent performance of computational pose prediction in docking programs has limited their ability to positively impact drug-discovery efficiency.[1−4]

Received: March 13, 2015
Published: July 7, 2015

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

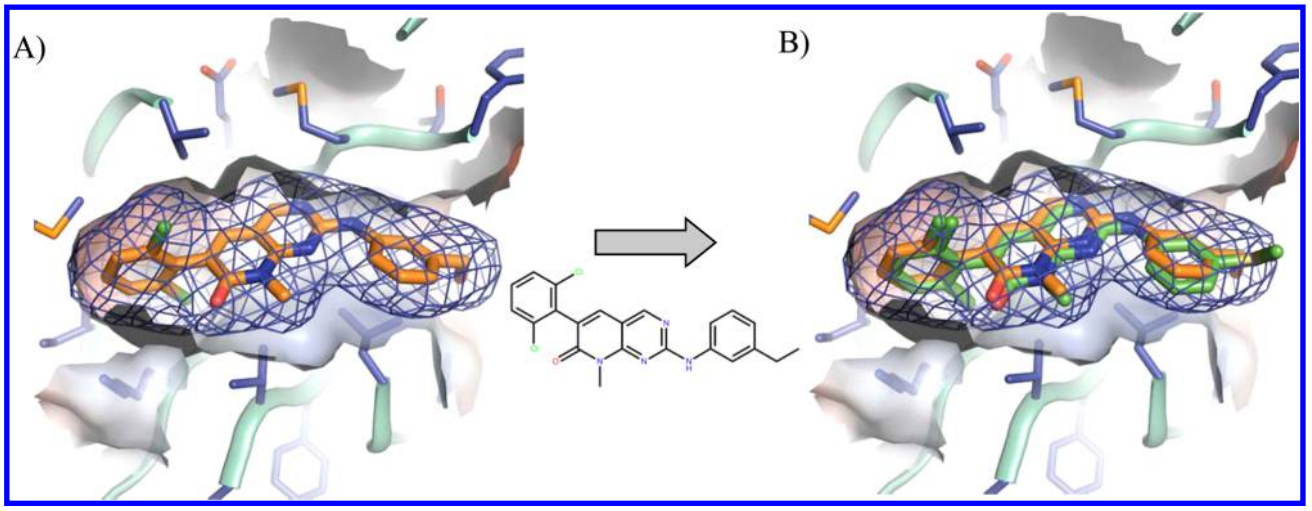


Figure 1. Depiction of the overlay procedure used by POSIT. A known bound pose (A) is used as a template to bring molecules into common alignment (B). This is an example of self-docking where we are docking the known molecule back into its receptor. POSIT initially generates a Gaussian volume (blue mesh) for the known inhibitor, which is used for pose placement and optimization. POSIT then generates an initial set of 3D conformations, rigidly aligns these conformation to the Gaussian volume, and selects a small set of best alignments for the optimization. Poses from the optimization are scored based on shape and chemistry alignment to the known bound pose.

One attempt to improve the inaccuracies and inconsistent performance of protein-only molecule docking algorithms was to add known ligand binding information. Historically this information has been added in two ways. The more prevalent approach is through atom−atom constraints where particular ligand-protein interaction must be present and if not present the potential pose is rejected. This method tends to be less robust because the addition of too many atomic constraints results in few successful poses being generated. The second approach is the use of the molecule shape of the ligand as either a driving function for pose generation, e.g., SG-DOCK and SDOCKER,[5,6] or filter or scoring function for initial pose generation, e.g., SP-DOCK and HYBRID.[3−5] Both methods have been shown to improve pose prediction performance.[5,6] To date, none of these methods have been shown to improve inconsistency of pose prediction performance across different protein target classes.[3−6]

Ligand-based approaches are often used as an alternative to structure-based docking and can produce complementary results. One very well-established type of ligand-based approach uses a target's known active ligands to produce a 3D pharmacophore for the target.[7] A pharmacophore attempts to produce a common 3D alignment for a set of inhibitors by finding conformational and electrostatic elements that are common across the set (structure-based pharmacophores can also be generated using protein active-site alignments).[8] Once built, the pharmacophore is used as a first-pass test for new ligands, which are placed within the pharmacophore and ranked by quality of match to the pharmacophoric features.

Two dimensional (2D) ligand-based methods are also used that are based on extensions to the concept of maximum common substructure.[9] Maximum Overlapping Set[10] is one such attempt to account for common (not necessarily connected) fragments between molecules. A similar approach, termed Graph-based Molecular Alignment,[11] attempts to use more sophisticated optimization procedures to boost retrieval rates and widen the robustness and scope of the alignments.

Finally, another class of ligand-based pose-prediction strategies uses a hybrid approach to combine both 2D and 3D information. The 2D aspects of the molecules are captured using elements of molecular graphs in combination with 3D conformations, which are used to align databases of molecules onto common frameworks. An example of this type of approach is Complexes Restricted by Experimental Structures (CORES).[12] CORES breaks a crystallographic ligand conformation into bound fragments, which are used to guide the posing of conformations for new compounds.

We present here an alternative to the above methods, which represents a fundamental departure from traditional approaches used in structure-based and ligand-based design. The focus of this new method, termed POSIT, is on rigorous extrapolations from experimental structural data to drive pose prediction and confidence measures. This is accomplished, in part, by using ligand representations based on 3D molecular shape (depicted in Figure 1). Molecular shape can be defined as a volume function (inside or out) based on atom-centered functions (e.g., spheres or other radial functions such as Gaussians) or directly from the electron density or theoretical estimates of such. POSIT uses shape overlap and chemical similarity to quantify an extrapolation distance from experimental structures, which provides for more robust estimates of prediction confidence. These confidence estimates lend themselves to more effective use of structural data by informing better critical assessments. Because the predictions are inherently more reliable, each experimentally determined protein−ligand structure has increased research value.

II. DATA AND METHODS

In this section we present the underlying algorithm used in the POSIT approach and describe the data sets that will be used to critically investigate its pose-prediction performance. In total we utilize four different data sets. The first two data sets are designed to facilitate comparison to previously published structural docking studies. The third data set explores the effect of structure selection on pose-prediction probabilities, which will be used to establish expectation values of confidence estimates for prospective POSIT predictions. Lastly, we describe the acquisition of a prospective data set, the sine qua non for the objective validation of a conceptual approach such as POSIT.

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

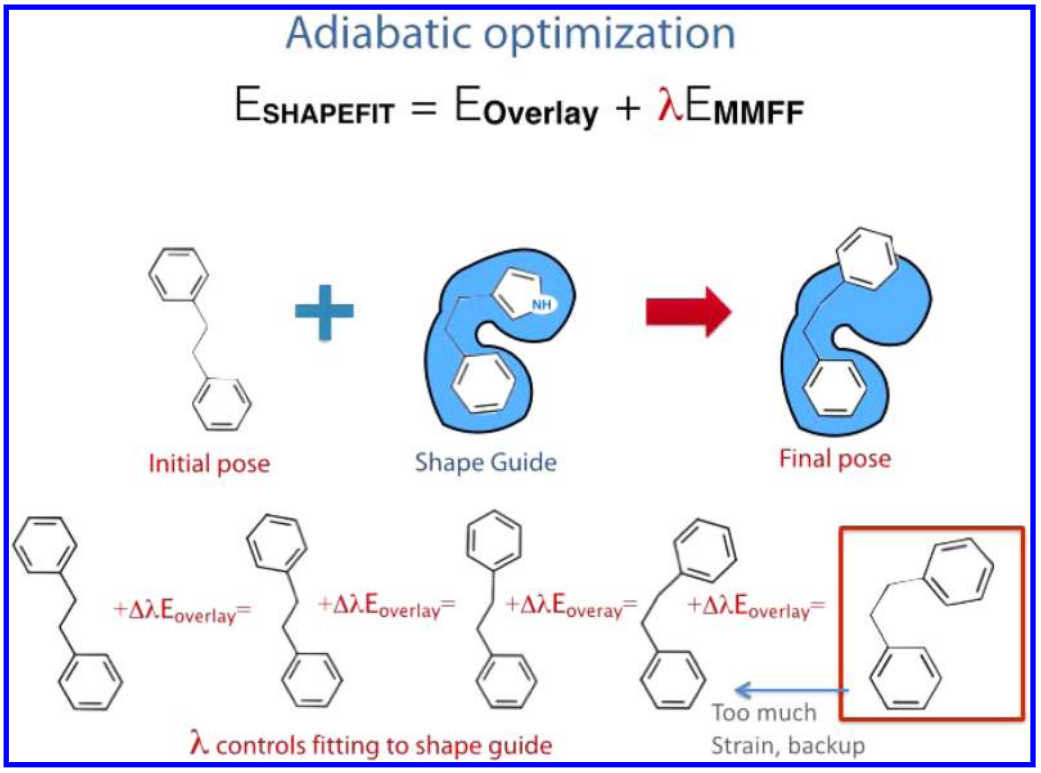


Figure 2. Schematic illustration of the POSIT process showing the SHAPEFIT potential used during the adiabatic optimization. Molecular poses are chosen using rigid alignment to the known binding modes. A shape guide (Gaussian volume) generated from the known bound ligand is used to drive an all-atom optimization of the initial poses. The optimization is halted when either convergence is achieved, or when the induced strain of the overlay becomes greater than 10 kcal/mol relative to its starting conformation.

POSIT Algorithm. The POSIT approach to docking represents a modification and extension of the molecular-shape-based ligand-alignment algorithm used in ROCS.[13] The shape-matching algorithm drives the alignment of test compounds to an experimental ligand electron density, or alternatively a simulated ligand electron density, which is used to produce the initial overlays for the POSIT method. The 3D similarity of the overlays is assessed by both shape and chemistry similarity between the known ligand and the posed (test) ligand. Since chemistry is labeled or "colored" on the molecular graph it is generically referred to here as "color". The combined measure, the TanimotoCombo (TC), is the sum of two separate Tanimoto measures (shape similarity and color similarity), its values span the range from 2.0 (identical) to 0.0 (completely dissimilar). As there is always some shape overlap, the lower bound on TC is always greater than zero.

All POSIT analyses in this work follow the same initial sequence of steps. This begins by first converting the test molecule into a connection table, or SMILES string to ensure removal of existing and potentially biasing 3D information. This same step is performed by all methods presented here. The SMILES are input directly into POSIT, which contains its own internal 3D conformational generation routines (in contrast to the other methods used here that require a separate conformation-generation step). POSIT performs rigid-body 3D alignments on the test-molecule conformations using shape- and chemistry-driven optimization against the shape and color of the known (cognate) ligand. A final set of conformations is obtained by sorting and filtering using TC. To make this procedure more efficient, candidate poses are removed if closer than 0.8 Å root-mean-square deviation (RMSD) to a superior candidate poses. By default, ten starting poses are chosen, which was found to balance quality of result and computational cost.

Each pose from the previous step is then flexibly optimized using a protocol that attempts to maximize overlap with the conformation of the known binding mode while avoiding inducing too much strain. This optimization protocol is based on a combined force-field and shape potential (the SHAPEFIT potential), which has been used previously to fit ligands into modeled electron densities.[14] The formal definition of the SHAPEFIT potential is shown in eq 1:

$$E_{\mathrm{SHAPEFIT}} = E_{\mathrm{MMFF}} + \lambda E_{\mathrm{Overlay}} \quad (1)$$

E_{MMFF} is the potential energy calculated using the Merck Molecular Force Field (MMFF94s);[15−18] $E_{Overlay}$ the shape "potential" is proportional to the shape overlap; and λ is a variable weighting factor that slowly increases during optimization. The concept here is that we do not want the sudden addition of the shape potential to the internal energy of the molecule to cause "barrier" crossing to new conformational states. Instead, λ is slowly increased to "adiabatically" shift the ligand to a better matching overlay. If, for an increment in λ, either the improvement of fit is below a given threshold or the strain energy in E_{MMFF} exceeds a different threshold, typically 10 kcal/mol, then the process is terminated. The choice of 10 kcal/mol is somewhat arbitrary but is consistent with the strain-energy work of Perola and Charifson.[19] Monitoring conformational stain is important for producing reasonable pose geometries, which is particularly relevant for test ligands with significant chemical differences from the cognate ligand. A graphical illustration of this process is depicted in Figure 2.

Once the adiabatic optimization is complete, the posed ligand undergoes an all-atom restrained refinement. Atoms that

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

lie inside the density of the cognate ligand, or "shape guide," are restrained using a harmonic-well potential. Atoms lying outside the shape-guide density, and thus more "reactive" to interactions with the protein, are allowed to adapt more to local environmental conditions, e.g., by optimizing their positions relative to hydrogen-bond interactions, or clashes with atoms of the protein, etc. The poses resulting from this process are ranked a final time from highest to lowest TC relative to the cognate ligand.

Retrospective Data I: Self-Docking Study. We use the Iridium data set,[20] summarized in the Supporting Information, to perform the self-docking analysis. Iridium is a valuable source for this test because all structures have complete ligand electron density. In addition, its annotations allowed us to remove structures with alternate conformations of ligands or proteins. The justification for this is that POSIT drives ligand posing using either simulated or experimental ligand densities. Hence, it is desirable to remove complexes where pose placement might be influenced by the presence of alternate conformations for the bound ligand or residues in the protein active site. The primary purpose of this data set is to rigorously test predictions of ligand placement in protein binding sites when the cognate ligand is well-defined and free from confounding factors. For this reason the structures are very highly curated to ensure a quality for the electron density that is sufficient to assign well-defined positions for every ligand heavy atom. That is, for each structure the ligand conformation and active-site placement are known both precisely and accurately. The subset so chosen consisted of 178 proteins with crystal structures for 281 protein−ligand complexes (some structures have more than one protein−ligand instance).

We compared the results of POSIT against the rigid alignment algorithm of ROCS.[13] The initial step in the ROCS self-docking procedure generates the isomeric canonical SMILES for each ligand using OEChem 1.6.[21] The SMILES are then input into Omega (v2.3.2)[22] to generate 3D conformations. These conformations are then subsequently passed to ROCS for rigid-body alignment using shape- and color-driven optimization against the cognate ligand. The poses are ranked by TC with only the highest scoring pose retained. A heavy-atom RMSD value is then calculated between the best pose and the cognate ligand using the OERMSD algorithm implemented in OEChem 1.6.

The procedure for the POSIT self-docking analysis also begins by generating isomeric canonical SMILES for each ligand using OEChem 1.6. The SMILES representation is then passed as a command-line argument to POSIT. Only the highest scoring pose is retained from the calculation, followed by a calculation of heavy-atom RMSD using the same procedure as described for ROCS.

The expectation of this study was that POSIT will drive structures closer to their cognate forms, whereas ROCS will, inevitably, be limited to the quality of the input structures. Our goal was to assess the degree to which this was, in fact, true.

Retrospective Data II: Cross-Docking Study. Curated data for 421 XRC structures was kindly provided by Tuccinardi et al. (summarized in the Supporting Information).[23] This data set was used in a previous study to critically examine docking performance and allowed us to extend our analysis and compare POSIT to the methods examined by Tuccinardi et al. in their extensive cross-docking study.

In the work of Tuccinardi, each structure file was obtained directly from the RSCB database, and hydrogen atoms were added to the protein and ligand using Babel v3.3[24] and Szybki v1.3.2,[25] respectively. Each ligand was then extracted from the structure and analyzed for atom-typing errors by comparison to results from Macromodel minimization[23] as well as the original XRC structure. The presence of any potential alternate conformations was checked by visual inspection. Where available, deposited structure factors were used to minimize the ligand in situ using a combination of electron density, as calculated by EDS (http://eds.bmc.uu.se/)[26] and the MMFF94s force field. The resulting ligand conformation was visually compared to the deposited ligand conformation in order to screen for highly conformationally strained ligands (strain energies greater than 10 kcal/mol). There were no overstrained ligands present in any of 421 complexes.

The cross-docking experiment attempts to dock each ligand from a given kinase family into all other structures within that family. To facilitate post hoc analysis, all of the proteins within each family were brought into a common frame of reference. This was accomplished by backbone alignment using the Mustang software package.[27] To spot check for sequence-alignment bias, proteins were also aligned using Magic Fit[28,29] (http://chemistry.umeche.maine.edu/CHY431/Swiss.html). The Magic Fit results were consistent with the Mustang algorithm, producing variations of less than 0.1 Å RMSD, which fall within the coordinate error of the resolution for the XRC structures.

Docking results for the following software packages were taken directly from Tuccinardi et al.: AUTODOCK 4.0;[30] FRED 2.1;[3] GLIDE 5.0;[31−33] and GOLD 4.1.[34] We reproduced the results for the docking program FRED to check that our procedure was sufficiently equivalent to that used by Tuccinardi et al. in the original study. It should be noted that none of these methods incorporate protein flexibility.

As with the other analyses, the first step in POSIT cross docking is to generate the isomeric canonical SMILES for each of the 421 ligands using OEChem 1.6. The SMILES string for each ligand was then passed to POSIT and docking was attempted against all protein structures residing within each ligand's kinase family, e.g., the ephrin type-B receptor 4 kinase family contains 8 complexes and thus required 64 separate cross-docking runs. RMSD values for each cross-docked pose were calculated using the cognate ligand coordinates from the coaligned structure. Although each ligand gets docked into multiple structures within its family, the RMSD value evaluated is always relative to its own (aligned) cognate pose, hence the importance of having all structures aligned into a consistent reference frame. A pose prediction is considered successful if its calculated RMSD is less than 2.0 Å, a value commonly purported to be sufficient for the purposes of LO.

Retrospective Data III: Ensemble-Docking Study. For the purposes of this study we used FRED and POSIT, to evaluate the consequences of structure selection on the performance of the docking predictions.

The ensemble-docking study uses the same set of 421 structures as described in the previous section; however, the structures are sampled differently. For each kinase family in the cross-docking data set, five structures were randomly selected for use in docking runs. The ligands within each kinase family were then docked into the structures from this randomly selected set in one of two ways: (1) a structure from the set was chosen at random for each ligand; or (2) a structure was chosen with the highest TC between the two cognate ligands. After docking, RMSD values were then assessed for each pose

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

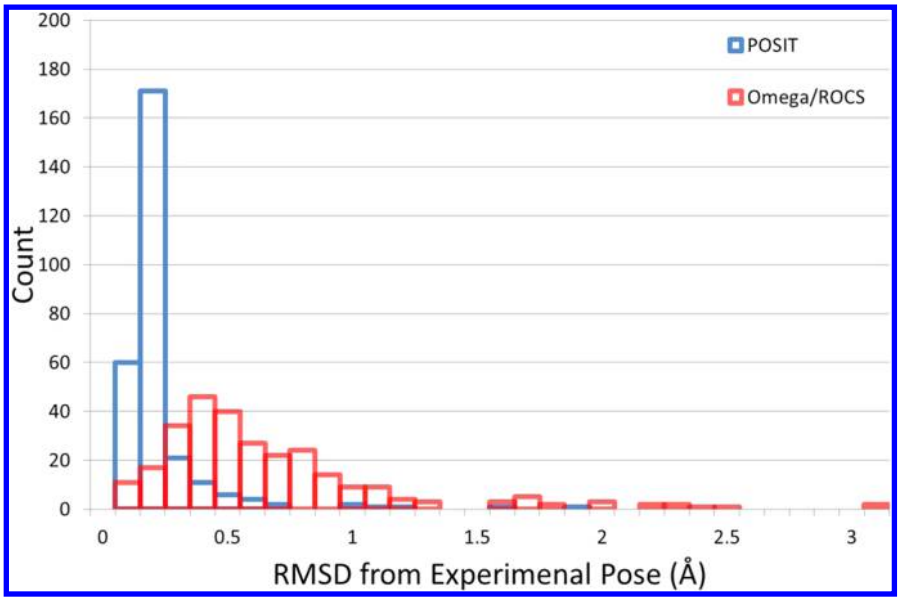


Figure 3. Self-docking results for two shape-based methods. The results from POSIT are shown in blue for the 281 protein−ligand complexes from the Iridium data set,[20] which contains complexes for which the ligand electron density is complete. In red are the Iridium results for the ligand-based shape-fitting method ROCS. The comparison shows that for self-docking the rigid alignment alone is insufficient for precise pose reproduction. POSIT produces precise self-docking pose predictions because it includes flexible coordinate optimization and also checks for clashes with the protein active site. For the POSIT generated poses, 94% are less than 0.4 Å RMSD compared to 38% for the OMEGA/ROCS approach.

relative to its cognate ligand. Pose predictions were considered successful for RMSD values less than 2.0 Å. The steps above were then repeated 1000 times, each time with a new set of randomly selected five structures.

The goal of generating a priori probabilities in this way is that it allows us to assess the consequences of different structure selection rules on pose-prediction success probabilities. A better understanding of how to select the most appropriate structure for docking has the potential benefit of adding a systematic positive bias to improve docking results. If such a selection method is robust and computationally tractable, it should be incorporated into docking protocols and used routinely to select the most appropriate structure for each docking run.

We used an identical procedure to generate the POSIT predictions as described in the previous section for cross docking. For FRED, the first step converts the test molecule into SMILES, which is then input into Omega (v2.3.2) and 3D conformations are generated. FRED requires some additional preparatory steps before docking can begin. The protein must be converted into an annotated data structure. This process was automated and default settings used to process each protein structure in the data set. The 3D ligand conformations were then input into FRED, which was then run using only the protein structure to generate docking scores. For each conformation, FRED exhaustively searches through all possible positions and orientations using a search grid placed within the active site (all of which are defined in the annotated protein data object). Each of the millions of candidate poses is scored and kept in a ranked list. At the end of the run only the highest scoring pose is retained.

Prospective Data: Discovery Research. In this section we describe the technical aspects involved in collecting all of the data for analysis of a large number of prospective predictions acquired during discovery-research operations over a significant period of time. We attempt to provide sufficient detail to elucidate the processes in place that allowed the predictive method POSIT to become widely accessible to discovery researchers, which was ultimately responsible for its routine use as a prospective research tool for structure-based design. A full description is presented in the Supporting Information.

All of the prospective predictions reported here were made during a span of roughly 2.5 y of discovery research at Abbott Laboratories, beginning in late 2008 and extending into early 2011. The total number of POSIT predictions recorded during this period, across all discovery therapeutic areas and projects, was in excess of 30 000. Note that this number represents the total number of submissions, and not a count of the unique compounds as some compounds were submitted more than once. The majority of the predictions were requests submitted by project-team medicinal chemists. In analyzing the POSIT submissions, 71 instances were identified for which the compound in the original prediction was subsequently synthesized and selected for followed up XRC structural determination by Abbott Laboratories' Structural Biology group. We will refer to this set as the Abbott Prospective Set (APS).

The APS compounds all originate from structure-enabled projects and represent requested predictions from 20 distinct discovery research programs. They are representative of the variety of active research projects, and capture the challenges of real drug-design scenarios faced by discovery researchers. More importantly, the APS represents predictions made in a truly blind manner, and hence provide direct insight into the predictive research impact of POSIT.

To assess the accuracy of the predictions for the APS compounds, the protein structure that was used in the cognate docking was aligned to the new structure using the program Mustang.[27] This brings the two ligands into a common frame of reference and allows direct evaluation of the accuracy of the binding-mode prediction made by POSIT. Once aligned, the heavy-atom RMSD values for each ligand were calculated using the OERMSD algorithm in OEChem 1.6.

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

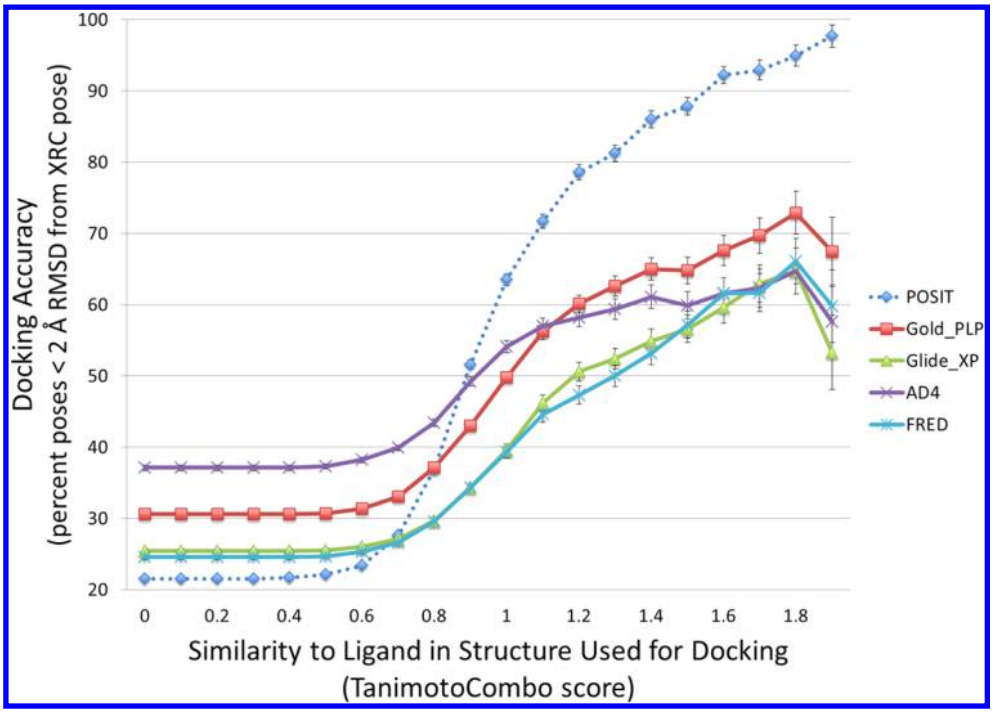


Figure 4. Cross-docking pose-prediction performance for POSIT and several other structure-based docking techniques. Plotted is similarity between the ligand in the complex and the ligand being docked (as measured by TanimotoCombo) versus percent of top-ranked poses within 2 Å RMSD of the X-ray crystal-structure pose. For POSIT, cross-docking pose reproduction is substantially better than any of the docking methods displayed when the TanimotoCombo ligand similarity is greater than around 0.9. Gold_PLP indicates Gold v4.1 with the PLP scoring function. Glide_XP is Glide v5.0 with the XP scoring function. AD4 represents AutoDock v4.0. FRED indicates FRED v2.1.[23]

III. RESULTS AND DISCUSSION

Evaluation of the pose-prediction performance of POSIT will be presented in four sections. The first three sections represent all of the retrospective analyses: the results for self-docking; the results for cross docking; and the results for the comparison of random selection of structures versus using the best by ligand similarity. These retrospective analyses establish a baseline for POSIT performance but also demonstrate its performance relative to the minimum requirements for a predictive structure-based method. In the final section, we present results for the most realistic assessment possible for a predictive method: the direct evaluation of POSIT performance during prospective drug-discovery research.

Retrospective Results I: Self-Docking Validation. Self-docking is a commonly employed validation technique in the literature,[4,35−41] but by itself is not sufficient to gauge a method's prospective value as it is not representative of the types of scenarios encountered in research. However, self-docking results have value in that they represent a minimum requirement standard.[42]

Shown in Figure 3 is a comparison of ROCS and POSIT, both 3D-shape-based posing methods. The distributions of heavy-atom RMSD values in Figure 3 reveal two important aspects: (1) the use of a rigid method by itself (ROCS) is not sufficient to reproduce the high-precision atom positions in the Iridium structures; and (2) the addition of the flexible optimization in POSIT produces a significant improvement in performance. Statistical comparison of the results from Figure 3 produces a one-sided *p-value* for the paired *t*-test[43] of less than 0.0001, and a Cohen's *d* value[44] of 1.64. Cohen's *d* assesses effect size, i.e. whether a difference is both statistically significant and of a size that is relevant compared to the underlying dispersion of results. A value greater than one is considered a large effect.

On the basis of the results of the self-docking, we find that POSIT is able to generate pose guesses with a median RMSD of 0.13 Å, as compared to a median value of 0.48 Å for the rigid ROCS approach. POSIT and ROCS use the same procedure to generate their initial posing, with the main difference between the two being the additional optimization employed by POSIT. Without flexible optimization the quality of pose guessing can be seen to be significantly compromised; however, this improvement does incur an additional non-negligible increase in the computational cost.

Retrospective Results II: Cross-Docking Validation. In this section we focus on the analysis of cross docking. Cross docking is more representative of scenarios likely to be encountered in discovery research, and thus can provide better insight into potential prospective performance. We compare the cross-docking pose-prediction performance of POSIT to the results reported by Tuccinardi et al., using the same data set used in their published analysis.[23]

Figure 4 shows a comparison of the results of POSIT relative to four other docking methods. Plotted in Figure 4 is pose-prediction performance as a function of ligand similarity, which is measured as TC between the docked (test) compound and the compound used as the "shape guide" (the known ligand). Below a similarity value of around 0.9 it can be seen that all of the methods perform roughly equivalently. However, for TC values above 1.0 POSIT begins to excel relative to the other approaches. TC similarity values above 1.0 lie in the range of what is typically explored in LO programs, with a historical average of roughly 1.2 for LO campaigns at Abbott Laboratories. Thus, the range of values with the most relevance to LO programs lies solidly in the region where POSIT

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

outperforms all other docking methods. POSIT appears to be uniquely applicable to enabling LO-stage research. The authors also provide a single example comparing POSIT cross-docking performance to the performance for a ligand binding information docking method (HYBRID) found in Supporting Information Figure S1.

It is important to note that in this retrospective data set the majority of cross-docking ligand pairs have TC values lying below 1.0 (see histogram of scores per TC bin in Supporting Information Figure S2). For the similarity range below 0.9, which includes (apo) proteins with no bound ligands, structure-based docking is likely the best choice for pose prediction. However, this limitation may be mitigated by multiple crystal structures of different ligands bound to the same protein, i.e. as long as one structure has a TC > 1.0 ligand information may be superior. We will explore this idea further in the next section.

As a final note, it is worth mentioning that the results in Figure 4 reveal a possible criterion that could be used to assist in deployment of crystallography resources, and perhaps to justify both the timing of further crystallography and prioritize the best choice of compounds to pursue. New compounds would be prioritized based on their TC similarity to the existing ligands in current XRC structures. That is to say, the most structurally informative new compound is one with a shape that is significantly different from all of the other existing ligands. This difference can be quantified by TC similarity. A new candidate compound for crystallography should have TC values less than 0.9 compared to all existing ligands. It would not be difficult to adopt this principle into an experimental design and solve those structures that are most informative for future medicinal chemistry. This could present new opportunities for synergies between crystallography and computational drug design.

Retrospective Results III: Ensemble-Docking Validation. We explore in this section an approach to the problem of selecting the most appropriate structure for use in docking runs. In particular, we wish to investigate whether a simple TC similarity criterion is sufficient. To assess this, the a priori success probabilities of POSIT and FRED were compared using two very different selection rules. The first rule models the purely random selection of docking structure, whereas the second rule models a selection bias that uses TC value between test and known ligands.

Figure 5 shows the results for this process. What is immediately clear from the plot in Figure 5 is that selection of a docking structure using TC has a dramatic effect on the pose-prediction performance of both POSIT and FRED. In can also be seen that for either structure selection scenario POSIT shows statistically superior performance relative to FRED. This is perhaps not surprising as FRED is using a rigid structure-driven approach. However, FRED also shows a highly significant increase in performance.

While difficult to prove, the results of this section, and those of cross docking, appear to suggest that POSIT is fundamentally more efficient in its use of structural information, with its results showing a greater fidelity than any purely structure-driven docking approach. The most interesting aspect is the evidence that a simple similarity metric can be used to select the docking structure that will produce the greatest likelihood for success. The best candidate structure for docking is the one that has a cognate ligand with the highest TC to the compound to be docked. This presumably is because, as two ligands become more similar, their induced-fit effects begin to converge and, as in the Tuccinardi work, illustrate the underappreciated importance of this for docking.

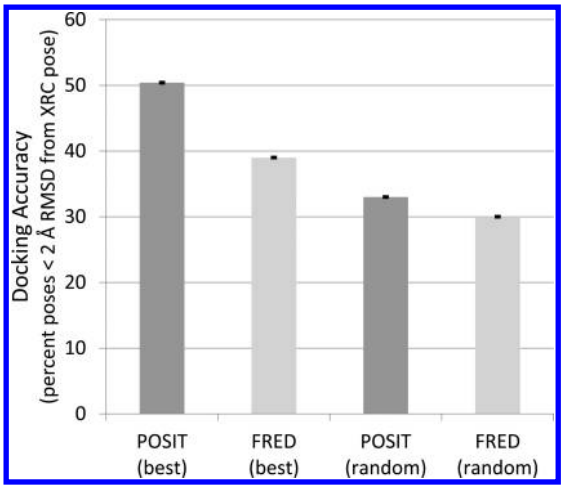


Figure 5. Ensemble-docking results using the methods POSIT and FRED. A statistical bootstrapping procedure was used to assess each method under two different rules for structure selection: (best) a single structure whose known ligand has the highest TanimotoCombo similarity to the test ligand and (random) the purely random choice of structure. It can be seen that structure selection by ligand similarity produces statistically and substantially significant improvement to the docking performance of both methods relative to random. Also evident from the graph is that POSIT excels over a traditional pose prediction methods (FRED) when structure selection is based on ligand similarity.

Prospective Results: Discovery Research. In this section we present the results for the prospective predictions, which were made using POSIT while providing structure-based computational support for discovery research over approximately 2.5 y of research at Abbott Laboratories. This data represents real predictions made for discovery research, averaged over time, scientists, and research projects. As such it reflects the real challenges present in prosecuting industrial drug discovery using structure-based design. Further detail and analysis on the properties of these compounds is provided as Supporting Information.

Shown in Figure 6 is the probability of predicting a binding mode within 2.0 Å RMSD of the actual binding mode for the seventy-one APS compounds as a function of TC similarity to the shape guide ligand used to drive the pose prediction. It can be seen from the plot that POSIT approaches 100% success for TC values above 1.5 At a TC value of around 1.2, which is approximately the historical average TC for Abbott LO projects, POSIT is providing results with approximately 75% confidence, compared to 50−60% expected for protein-structure-based docking programs (Figure 4).

Figure 7 shows the actual RMSD values for the APS. The line in the plot represents the expected probability based on the retrospective analysis performed by ensemble docking. This shows the prospective results to be consistent with the probabilities expected from retrospective bootstrap sampling.[45] These results also support the idea that one can use a simple TC assessment to select a priori the best structure to use in a prospective docking study.

The research value of a predictive computational method is most directly assessed by empirical examination of its performance in situations where the answer is *not* known (i.e., predictions made *prospectively*). To the best of the authors'

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

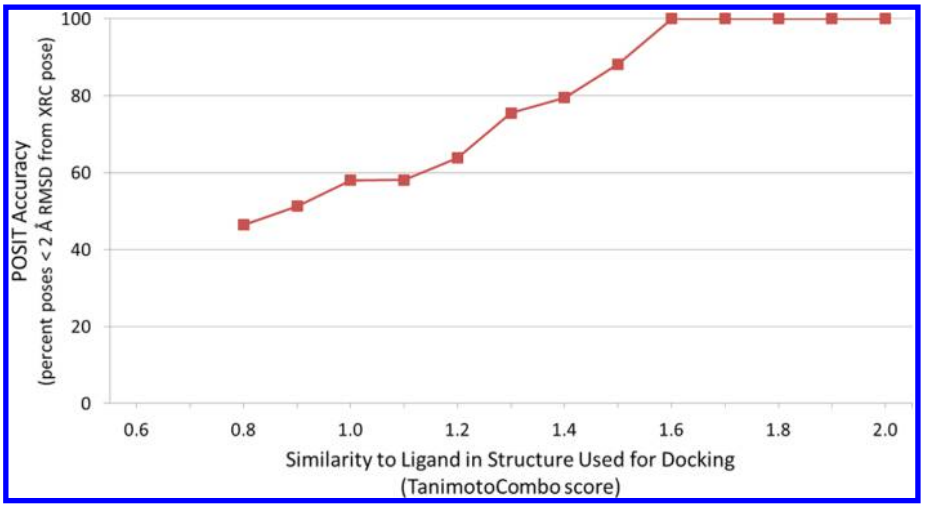


Figure 6. Successful prospective POSIT pose predictions for 71 compounds versus TanimotoCombo similarity. A pose prediction is considered successful if its RMSD to the experimental binding mode is less than 2 Å. The experimental binding modes were all determined by X-ray crystallography subsequent to the POSIT predictions. The TanimotoCombo similarity values are measured against the "shape guide" ligand used by POSIT to initially dock the compound.

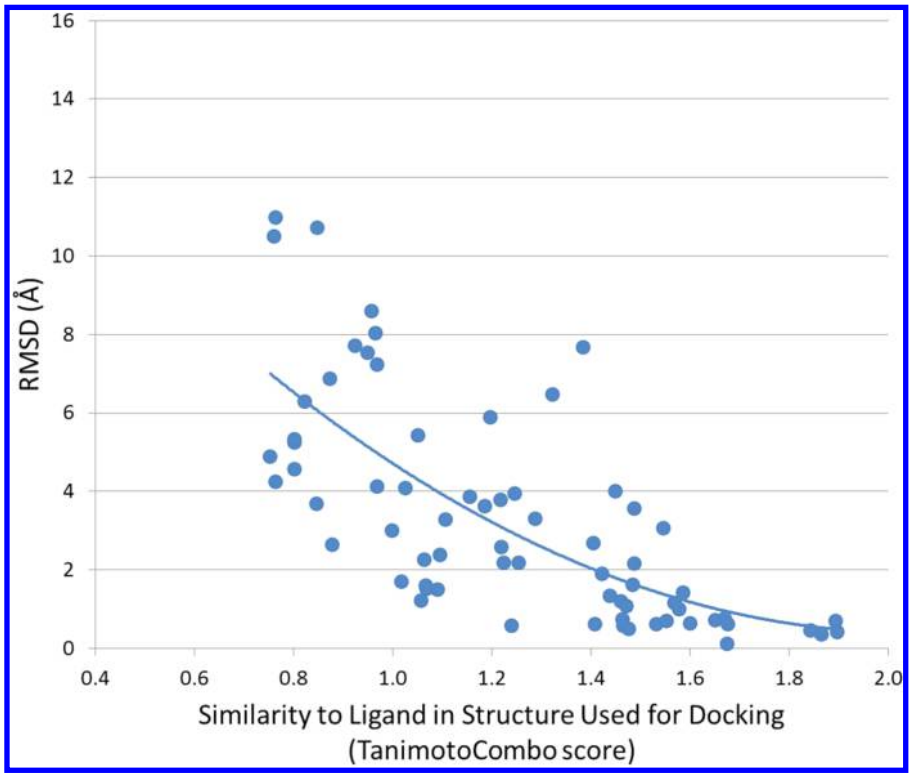


Figure 7. RMSD of POSIT predictions for 71 compounds versus TanimotoCombo similarity. The RMSD values are calculated relative to the experimental binding mode as determined by X-ray crystallography subsequent to the POSIT prediction. The TanimotoCombo values shown are measured against the ligand used as the shape guide in the POSIT docking. The circles show actual prospective results for the 71 compounds. The line represents a median fit to the data and has been added for visual guidance.

knowledge, this analysis represents the most comprehensive set of truly prospective results published to date. The results show that POSIT can provide structure-based predictions with quantitative confidence, and allow simple TC similarity to be used for selection of the best structure for docking. These aspects allow POSIT to be more easily incorporated into research decision making processes by allowing researchers to critically assess its predictive value.

IV. CONCLUSIONS

Structure-based docking attempts to solve the particularly difficult problem of pose prediction. Historically, structure-based methods have performed poorly when used to support decisions in LO campaigns. A contributing factor to the poor performance has been the reliance on using only protein information to generate pose predictions. This has an inherent limitation as it does not incorporate all of the available knowledge contained in the structure, i.e., the bound ligand

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

conformations. We have illustrated this by comparing structure-based performance to a more robust method (POSIT), which uses an adiabatic algorithm to couple force-field terms to experimentally derived ligand shape. POSIT also incorporates shape and chemistry similarity to leverage knowledge of existing bound conformations, which allows for superior selection of the protein receptor most likely to generate the best pose prediction.

POSIT cannot generate reliable pose predictions for very different ligands. However, the data presented in Figure 4 shows that traditional docking programs do not generate very reliable pose predictions for different ligands either. One question the reader might be asking is "what about those cases where a small chemical change results in a very different binding mode?" POSIT is unlikely to generate a correct binding mode for these cases but these cases are rare. A quick examination of the Figures 4 and 6 shows that when the TC similarity is 1.8 or higher the chance of a new binding mode and/or of POSIT generating an incorrect binding mode is historically only 5% and for the prospective data presented here 0% of the time. More importantly the POSIT probability scoring function reports exactly what the probability of these rare events is expected to be. Our data analysis shows that these rare events are more likely when the ligand is small and binding is driven by electrostatic interactions. Development is underway to see if this information might be exploited to further improve POSIT's pose prediction reliability.

POSIT outperforms protein-only methods in retrospective analyses examining self-docking and cross-docking performance. In addition, the POSIT methodology is shown to be systematically superior in the context of LO pose prediction. It is interesting to note that our evaluation shows that the pose-prediction performance of POSIT for LO is very similar in both retrospective and prospective contexts (81% versus 79% for TC similarities above 1.4). In summary, POSIT represents a fundamental improvement in pose-prediction methodology to support LO-stage drug discovery.

■ ASSOCIATED CONTENT

Supporting Information

Tables that summarize the structures contained in the Iridium data set, the targets used by Tuccinardi et al., and a list of targets for the prospective data set. Additionally, a more detailed description of the prospective data is provided. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.5b00142.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: greg@eyesopen.com. Phone: 505-473-7385 × 50. Fax 505-473-0833.

Notes

G.L.W. is employed by and has a financial interest in OpenEye Scientific Software, Inc., the developer and distributer of programs POSIT, ROCS, OMEGA, FRED, HYBRID, AFITT, BABEL, SZYBKI, and OEChem used and/or mentioned in this paper. B.P.K. was employed by, during the time period covered by this paper, and has a financial interest in OpenEye Scientific Software, Inc. S.P.B. and S.W.M. have no competing financial interest.

■ REFERENCES

(1) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.* **2006**, *49*, 5912−31.

(2) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: Pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455−74.

(3) McGann, M. Fred pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.* **2011**, *51*, 578−96.

(4) McGann, M. Fred and hybrid docking performance on standardized datasets. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 897−906.

(5) Fradera, X.; Knegtel, R. M.; Mestres, J. Similarity-driven flexible ligand docking. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 623−36.

(6) Wu, G.; Vieth, M. Sdocker: A method utilizing existing x-ray structures to improve docking accuracy. *J. Med. Chem.* **2004**, *47*, 3142−8.

(7) Barillari, C.; Marcou, G.; Rognan, D. Hot-spots-guided receptor-based pharmacophores (hs-pharm): A knowledge-based approach to identify ligand-anchoring atoms in protein cavities and prioritize structure-based pharmacophores. *J. Chem. Inf. Model.* **2008**, *48*, 1396−410.

(8) Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. Molecule-pharmacophore superpositioning and pattern matching in computational drug design. *Drug Discovery Today* **2008**, *13*, 23−9.

(9) Raymond, J. W.; Willett, P. Maximum common subgraph isomorphism algorithms for the matching of chemical structures. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 521−33.

(10) Stahl, M.; Mauser, H.; Tsui, M.; Taylor, N. R. A robust clustering method for chemical structures. *J. Med. Chem.* **2005**, *48*, 4358−66.

(11) Marialke, J.; Korner, R.; Tietze, S.; Apostolakis, J. Graph-based molecular alignment (gma). *J. Chem. Inf. Model.* **2007**, *47*, 591−601.

(12) Hare, B. J.; Walters, W. P.; Caron, P. R.; Bemis, G. W. Cores: An automated method for generating three-dimensional models of protein/ligand complexes. *J. Med. Chem.* **2004**, *47*, 4731−40.

(13) Hawkins, P. C.; Skillman, A. G.; Nicholls, A. Comparison of shape-matching and docking as virtual screening tools. *J. Med. Chem.* **2007**, *50*, 74−82.

(14) Wlodek, S.; Skillman, A. G.; Nicholls, A. Automated ligand placement and refinement with a combined force field and shape potential. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2006**, *62*, 741−9.

(15) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of mmff94. *J. Comput. Chem.* **1996**, *17*, 490−519.

(16) Halgren, T. A. Merck molecular force field. Ii. Mmff94 van der waals and electrostatic parameters for intermolecular interactions. *J. Comput. Chem.* **1996**, *17*, 520−552.

(17) Halgren, T. A. Merck molecular force field. Iii. Molecular geometries and vibrational frequencies for mmff94. *J. Comput. Chem.* **1996**, *17*, 553−586.

(18) Halgren, T. A.; Nachbar, R. B. Merck molecular force field. Iv. Conformational energies and geometries for mmff94. *J. Comput. Chem.* **1996**, *17*, 587−615.

(19) Perola, E.; Charifson, P. S. Conformational analysis of drug-like molecules bound to proteins: An extensive study of ligand reorganization upon binding. *J. Med. Chem.* **2004**, *47*, 2499−2510.

(20) Warren, G. L.; Do, T. D.; Kelley, B. P.; Nicholls, A.; Warren, S. D. Essential considerations for using protein-ligand structures in drug discovery. *Drug Discovery Today* **2012**, *17*, 1270−81.

(21) *Openeye toolkits v1.6.1*; Openeye Scientific Software: Santa Fe, NM, 2008; http://www.eyesopen.com.

(22) Hawkins, P. C.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer generation with omega: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. *J. Chem. Inf. Model.* **2010**, *50*, 572−84.

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780

(23) Tuccinardi, T.; Botta, M.; Giordano, A.; Martinelli, A. Protein kinases: Docking and homology modeling reliability. *J. Chem. Inf. Model.* **2010**, *50*, 1432−41.
(24) *Babel v3.3*; Openeye Scientific Software: Santa Fe, NM, 2007; http://www.eyesopen.com.
(25) *Szybki v1.3.2*; Openeye Scientific Software, Santa Fe, NM, 2008; http://www.eyesopen.com.
(26) Kleywegt, G. J.; Harris, M. R.; Zou, J. Y.; Taylor, T. C.; Wahlby, A.; Jones, T. A. The uppsala electron-density server. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2240−9.
(27) Konagurthu, A. S.; Whisstock, J. C.; Stuckey, P. J.; Lesk, A. M. Mustang: A multiple structural alignment algorithm. *Proteins: Struct., Funct., Genet.* **2006**, *64*, 559−74.
(28) Guex, N.; Peitsch, M. C. Swiss-model and the swiss-pdbviewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714−23.
(29) Guex, N.; Peitsch, M. C.; Schwede, T. Automated comparative protein structure modeling with swiss-model and swiss-pdbviewer: A historical perspective. *Electrophoresis* **2009**, *30* (Suppl 1), S162−73.
(30) 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.
(31) 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.
(32) Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177−96.
(33) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750−9.
(34) 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.
(35) Brozell, S. R.; Mukherjee, S.; Balius, 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.
(36) Corbeil, C. R.; Williams, C. I.; Labute, P. Variability in docking success rates due to dataset preparation. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 775−86.
(37) Liebeschuetz, J. W.; Cole, J. C.; Korb, O. Pose prediction and virtual screening performance of gold scoring functions in a standardized test. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 737−48.
(38) Neves, M. A.; Totrov, M.; Abagyan, R. Docking and scoring with icm: The benchmarking results and strategies for improvement. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 675−86.
(39) Novikov, F. N.; Stroylov, V. S.; Zeifman, A. A.; Stroganov, O. V.; Kulkov, V.; Chilov, G. G. Lead finder docking and virtual screening evaluation with astex and dud test sets. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 725−35.
(40) Repasky, M. P.; Murphy, R. B.; Banks, J. L.; Greenwood, J. R.; Tubert-Brohman, I.; Bhat, S.; Friesner, R. A. Docking performance of the glide program as evaluated on the astex and dud datasets: A complete set of glide sp results and selected results for a new scoring function integrating watermap and glide. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 787−99.
(41) Schneider, N.; Hindle, S.; Lange, G.; Klein, R.; Albrecht, J.; Briem, H.; Beyer, K.; Claussen, H.; Gastreich, M.; Lemmen, C.; Rarey, M. Substantial improvements in large-scale redocking and screening using the novel hyde scoring function. *J. Comput.-Aided Mol. Des.* **2012**, *26*, 701−23.
(42) Hawkins, P. C.; Kelley, B. P.; Warren, G. L. The application of statistical methods to cognate docking: A path forward? *J. Chem. Inf. Model.* **2014**, *54*, 1339−55.
(43) Student. The probable error of a mean. *Biometrika* **1908**, *6*, 1−25.
(44) Cohen, J. *Statistical power analysis for the behavioral sciencies*; . Routledge, 1988.
(45) Muchmore, S. W.; Debe, D. A.; Metz, J. T.; Brown, S. P.; Martin, Y. C.; Hajduk, P. J. Application of belief theory to similarity data fusion for use in analog searching and lead hopping. *J. Chem. Inf. Model.* **2008**, *48*, 941−8.

DOI: 10.1021/acs.jcim.5b00142
J. Chem. Inf. Model. 2015, 55, 1771−1780