JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

Cite This: J. Chem. Inf. Model. 2019, 59, 2600−2616

pubs.acs.org/jcim

Ligity: A Non-Superpositional, Knowledge-Based Approach to Virtual Screening

Jean-Paul Ebejer,[†] Paul W. Finn,[‡,¶] Wing Ki Wong,[§] Charlotte M. Deane,[§] and Garrett M. Morris*,[§]

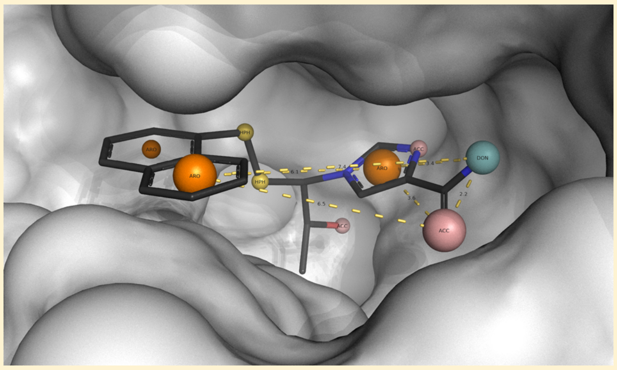
[†]Centre for Molecular Medicine and Biobanking, University of Malta, Msida, MSD 2080, Malta

[‡]Oxford Drug Design Limited, Oxford Centre for Innovation, New Road, Oxford OX1 1BY, U.K.

[¶]The School of Computing, University of Buckingham, Hunter Street, Buckingham, MK18 1EG, U.K.

[§]Oxford Protein Informatics Group, Department of Statistics, University of Oxford, 24-29 St. Giles', Oxford, OX1 3LB, U.K.

S Supporting Information



ABSTRACT: We present Ligity, a hybrid ligand-structure-based, non-superpositional method for virtual screening of large databases of small molecules. Ligity uses the relative spatial distribution of pharmacophoric interaction points (PIPs) derived from the conformations of small molecules. These are compared with the PIPs derived from key interaction features found in protein−ligand complexes and are used to prioritize likely binders. We investigated the effect of generating PIPs using the single lowest energy conformer versus an ensemble of conformers for each screened ligand, using different bin sizes for the distance between two features, utilizing triangular sets of pharmacophoric features (3-PIPs) versus chiral tetrahedral sets (4-PIPs), fusing data for targets with multiple protein−ligand complex structures, and applying different similarity measures. Ligity was benchmarked using the Directory of Useful Decoys-Enhanced (DUD-E). Optimal results were obtained using the tetrahedral PIPs derived from an ensemble of bound ligand conformers and a bin size of 1.5 Å, which are used as the default settings for Ligity. The high-throughput screening mode of Ligity, using only the lowest-energy conformer of each ligand, was used for benchmarking against the whole of the DUD-E, and a more resource-intensive, "information-rich" mode of Ligity, using a conformational ensemble of each ligand, were used for a representative subset of 10 targets. Against the full DUD-E database, mean area under the receiver operating characteristic curve (AUC) values ranged from 0.44 to 0.99, while for the representative subset they ranged from 0.61 to 0.86. Data fusion further improved Ligity's performance, with mean AUC values ranging from 0.64 to 0.95. Ligity is very efficient compared to a protein−ligand docking method such as AutoDock Vina: if the time taken for the precalculation of Ligity descriptors is included in the comparason, then Ligity is about 20 times faster than docking. A direct comparison of the virtual screening steps shows Ligity to be over 5000 times faster. Ligity highly ranks the lowest-energy conformers of DUD-E actives, in a statistically significant manner, behavior that is not observed for DUD-E decoys. Thus, our results suggest that active compounds tend to bind in relatively low-energy conformations compared to decoys. This may be because actives—and thus their lowest-energy conformations—have been optimized for conformational complementarity with their cognate binding sites.

■ INTRODUCTION

Ligand-based virtual screening (LBVS) is underpinned by the hypothesis that compounds with similar chemical structures tend to have similar biological activities.[1] LBVS methods use various representations of a small molecule, such as fingerprints, chemical topology, 3D shape, pharmacophoric features, physicochemical properties, or some combination of these.[2] These are often captured in a descriptor that is effectively a feature vector representing the molecule. Such descriptors are then compared to that of a known biologically active molecule—a "query"—using a similarity measure or metric, yielding a quantitative score of the similarity of the two molecules. Many similarity measures and metrics, such as Tanimoto, Cosine, Dice, and Tversky, have been reported in the literature (see Supporting Information, Schema S1).[3−9] Consensus scoring combines the results of multiple LBVS searches, typically using data fusion methods, and has been shown to improve the accuracy of virtual screening.[10,11] LBVS can perform well, especially when finding new hits with the same chemotype as known actives. Chemical topology-based descriptors do not take ligand 3D information into account

Received: November 2, 2018
Published: May 22, 2019

DOI: 10.1021/acs.jcim.8b00779

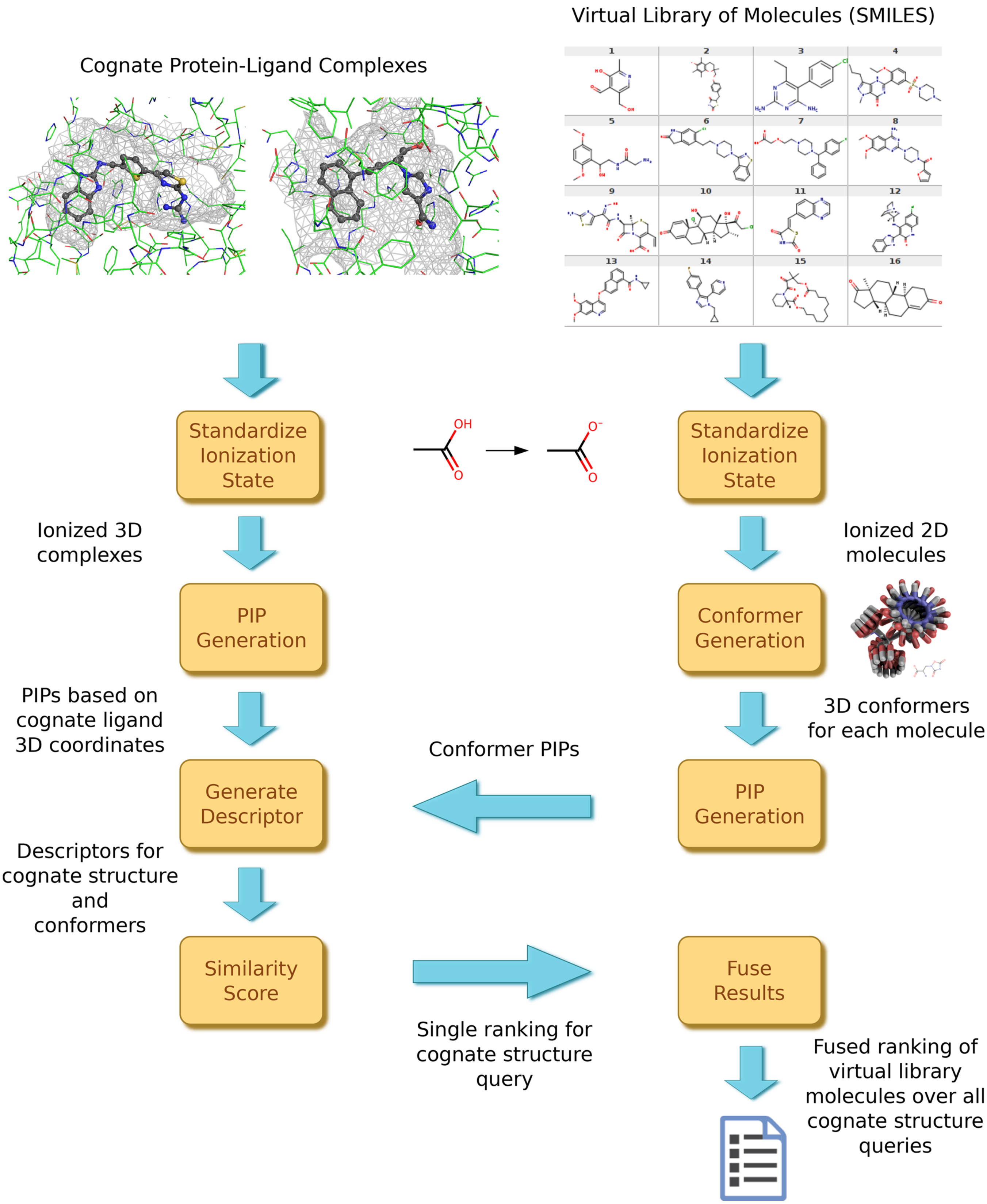

Virtual Library of Molecules (SMILES)
Cognate Protein-Ligand Complexes
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Standardize Ionization State
OH
O
O⁻
O
Standardize Ionization State
Ionized 3D complexes
Ionized 2D molecules
PIP Generation
Conformer Generation
PIPs based on cognate ligand 3D coordinates
3D conformers for each molecule
Conformer PIPs
Generate Descriptor
PIP Generation
Descriptors for cognate structure and conformers
Similarity Score
Fuse Results
Single ranking for cognate structure query
Fused ranking of virtual library molecules over all cognate structure queries

Figure 1. Ligity algorithm. Note that some processes, ionization and generate PIPs, are repeated because they have different implementations depending on whether their input is a 3D protein−ligand complex or a 2D SMILES molecule. Conformer generation uses the protocol published by Ebejer et al.[49]

and tend to be worse at scaffold hopping than 3D approaches. Many popular shape-based ligand-based methods,[12,13] on the other hand, require an optimal structural superposition before comparing the ligands, which can be slow.

In contrast, structure-based virtual screening (SBVS) methods use information from the 3D structure of the target protein.[14−16] Explicit SBVS methods propose a structural hypothesis for how a putative inhibitor binds to a target, by searching for the binding site and optimal binding mode. Candidate docked solutions are ranked by a scoring function, which is based on one of four conceptual approaches: statistical, knowledge-based, force-field-based;[17] or, more recently, machine-learning-based.[18] Despite incremental advances, current SBVS scoring functions tend to correlate poorly with experimental protein−ligand binding affinity.[17] SBVS is a widely used computational method for hit identification, and molecular docking often identifies active submicromolar compounds.[19] The success of homology-model-based SBVS is directly related to the quality of the model,[20] and can be improved by the use of multiple models instead of one.[21] Implicit SBVS methods rely on one or more 3D pharmacophores of active compounds, but they usually require superposition on a query molecule, which can be expensive. Implicit SBVS methods include Discovery Studio's Catalyst and HipHop,[22] LigandScout,[23] LS-Align,[24] PHASE,[25,26] MOE,[27] ROCS,[12] SHAFTS,[28] and WEGA.[29] Thus, both LBVS and SBVS have limitations restricting their predictive abilities.

Hybrid approaches attempt to improve the performance of virtual screening by combining multiple approaches to overcome the weaknesses of the separate methods.[30,31] Virtual screening methods can be combined in three ways: (i) in sequence,[32] (ii) in parallel,[33] or (iii) as pure hybrid methodologies.[34] In the sequential approach, different VS methods are used one after another in a workflow or pipeline, with the output of one method serving as the input of another. The idea is to use faster, coarser methods first as filters to reduce the initial very large numbers of molecules to a subset, which are then used in more accurate but time-consuming methods, such as free energy perturbation (FEP).[35] In the parallel approach, several different VS methods are run simultaneously, with the top results of each method being combined and taken to the next phase. In hybrid approaches, such as the one presented in this paper, a single method makes use of information about both the target protein structure and the active ligand(s) to discriminate between actives and inactives.

Most hybrid approaches use information such as shape similarity, volume overlap, or pharmacophoric similarity to one or more known ligands to guide, or filter, the output of a docking program. The use of active ligand information can improve performance; however, a relatively computationally expensive docking run is still required.

We were motivated to develop an alignment-free, hybrid virtual screening method to utilize both ligand and protein information in a computationally efficient manner. Previous work has explored individual aspects of this approach. FuzCav[36] uses a 4833-integer vector to describe a protein−ligand binding site that counts pharmacophoric triplets assigned to the Cα atomic coordinates of residues lining the binding site. While FuzCav is a non-superpositional method, it is mainly used for binding site comparison. The FLAP approach of Baroni et al.[37] describes a common framework for comparing proteins and ligands, based on GRID[38−41] molecular interaction fields. Local minima in the interaction fields for the protein of interest are sampled to create four-point pharmacophores, against which conformations of ligands can be searched, on the basis of the GRID atom-type pharmacophores they contain. FLAP enables fast searching but does not utilize knowledge about bound ligands. Both LigandScout[23] and Discovery Studio[22] can be run as hybrid VS methods, but they do not develop descriptors and, furthermore, are dependent on structural superposition, which can be a relatively expensive process. We hypothesized that bound ligands, especially if structurally diverse, contain key information about the energetically most-important interactions that would be useful in guiding the creation of pharmacophore-based database queries.

Therefore, we present "Ligity", a hybrid method that derives interaction features from experimentally determined structures of protein−ligand complexes. These interaction "hot-spots" are mapped to pharmacophoric interaction points, or "PIPs" in the ligand space. Ligity descriptors are built from triangular or chiral tetrahedral combinations of these PIPs. Typically, only a portion of the cognate ligand will be in contact with the protein structure and Ligity considers only these PIPs when constructing a query. Candidate ligands (database compounds) are also converted into Ligity descriptors on the basis of the 3D distribution of pharmacophoric features in their enumerated low-energy conformational ensembles. In this case, all PIPs are used in descriptor construction. Thus, Ligity descriptors for queries will generally be smaller than the Ligity descriptors of database molecules. We have therefore included consideration of the asymmetric Tversky similarity measure to compare descriptors, as this has the useful property that it can detect similarities even if the number of the features in the two descriptors differs significantly. If there are multiple structures for the target protein with different query (cognate) active molecules, Ligity fuses the results of each query. Unlike the hybrid methods in LigandScout and Discovery Studio, Ligity has the advantage that it does not require structural superposition to compare two sets of pharmacophores. Our results show that Ligity performs virtual screening as well as or better than docking methods, but it is much more computationally efficient.

■ METHODS

Ligity is a non-superpositional, knowledge-based, virtual screening method. It uses one or more existing protein−ligand complexes to construct a query. This is used to find biologically active molecules within a large database of small molecule conformational ensembles. An overview of how Ligity works is shown in Figure 1. The Ligity suite of programs is written in C++ and Python,[42,43] and it uses the open source cheminformatics toolkit RDKit.[44] When used for virtual screening, the main steps of Ligity are as follows:

(1) For each 3D protein−ligand complex used as (part of) a query, (a) standardize the ionization state of the protein and cognate ligand in the complex and (b) find protein−ligand interactions based on the maximal distance per interaction type and corresponding SMARTS[45] (SMILES[46−48] *ar*bitrary *t*arget *s*pecification) definitions for pharmacophoric features, and (c) for each protein−ligand interaction, find the set of PIPs on the cognate ligand.

(2) For each molecule in the compound database represented as a SMILES string, (a) standardize the ionization state of the molecule, using the same rules as for the query protein−ligand complex; (b) generate low-energy conformers using the protocol published by Ebejer et al.[49] (Briefly, the number of conformers generated depends on the number of

rotatable bonds in the ligand, with 50 for 7 or fewer rotatable bonds, 200 for 8−12 rotatable bonds, and 300 conformers for more flexible ligands. An initial conformer is generated using distance geometry and energy minimized using the UFF force field,[50] and conformers more similar than 0.35 Å RMSD are eliminated); and (c) find the set of PIPs corresponding to each conformer.

(3) For the query ligand(s) and compound database PIPs, generate the corresponding Ligity descriptors. These encode the distances and pharmacophoric features between all triangular or chiral tetrahedral combinations of PIPs. The Ligity descriptor is the multidimensional array that stores the counts of the component pharmacophorically labeled triangles or tetrahedra. Each axis in the multidimensional array corresponds to the distance between a pair of PIPs.

(4) Calculate the similarity between the descriptors for the query ligand and each conformer in the compound database.

(5) Rank the compound database for each protein−ligand query by decreasing similarity.

(6) If multiple protein−ligand queries are used, use data fusion to combine the ranked lists into one final list.

Steps 2 and 3, generation of the conformational ensembles and Ligity descriptors for the database compounds, are preprocessing steps that need only be carried out once.

Ligity's Input and Output. There are two main inputs to the Ligity algorithm: (1) one or more holo protein−ligand complex structures for the target of interest, the "query", and (2) a database of small molecules in SMILES format to be screened.

When more than one protein−ligand structure is used for the same target, this represents the "information-rich" mode of Ligity. We used the screening-PDB, or sc-PDB,[51−53] and its hierarchically clustered binding sites to define these information-rich queries; however, the user is free to select multiple structure queries by other methods. The clustering in sc-PDB is based on both the Enzyme Commission (EC) number and the structural similarity of their binding sites, as defined by SiteAlign.[54,55] SiteAlign quantifies the topological and pharmacophoric features of binding sites using 1D-fingerprints and attempts to find the best alignment of a target site with the largest query site. Thus, any given cluster of similar binding sites could be used as input for Ligity. It should be emphasized here that Ligity does not require that either the proteins or their holo binding sites should be superimposed, since Ligity is a non-superpositional method.

The input virtual screening library can consist of simply a list of SMILES strings describing the small molecules to be screened. The output of Ligity is a ranked list of all the molecules in the virtual library, ordered by decreasing similarity to the triangular (or chiral tetrahedral) Ligity descriptor(s) of the active cognate ligand(s).

Standardization of the Ionization State. In order to make chemically meaningful comparisons, we standardize the ionization states of both the protein-bound ligands and the small molecules in the compound database by first adding hydrogens using RDKit and then applying in-house rules (see Table 3.3 in the work by Ebejer[56]). These transformations ensure that ionizable groups are consistently charged, e.g., deprotonating COOH into COO^-. Although this ionization happens in 3D for the protein-bound cognate ligands and in 2D for the virtual screening library, the same chemical transformation rules are applied in both cases. The SMARTS patterns used to treat ionization do not require explicit hydrogens to be added to the protein.

Conformer Generation. The list of standardized 2D molecules in the virtual screening database are subjected to our RDKit-based[44] conformer generation protocol[49] to generate an ensemble of low-energy 3D atomic coordinates. As described earlier in step 2b, the number of low-energy UFF conformers generated for a given molecule is determined by its number of rotatable bonds, with at most 300 conformers for the most flexible compounds.[49] Normally, Ligity considers all conformers in a molecule's ensemble, but there is also a "high-throughput" mode where only the lowest-energy conformer of the small molecule is used. This makes subsequent steps in the algorithm take a fraction of the time taken for the full conformer set (see "Computational Efficiency").

Generation of PIPs. Pharmacophoric models are automatically generated for the standardized 3D protein−ligand query or queries and the conformers for the molecules in the virtual library or database. These pharmacophoric models are represented by a collection of PIPs. These highlight the interesting parts of the molecule or key features necessary for molecular recognition of a ligand by the receptor. Ligity defines six types of pharmacophoric features[57] in the ligands: hydrophobic, aromatic, hydrogen bond donor (HBD), hydrogen bond acceptor (HBA), anion (−), and cation (+) using SMARTS patterns.

The SMARTS patterns representing pharmacophoric features are specified in a feature definition file in RDKit (in `BaseFeatures.fdef`). These were modified to specify the six "PIP types" that are used by Ligity (listed in Table S1 in the Supporting Information). These rules occasionally generate duplicate PIPs with the same PIP type and (x, y, z)-coordinates; in these cases, the duplicates are filtered out to give a set of unique PIPs describing the molecule's shape and pharmacophoric features.

PIP Generation for the Query Protein−Ligand Complex. In the case of a query protein−ligand complex, we first identify PIPs in the ligand by applying the SMARTS patterns. PIPs are then discarded if there is no complementary receptor PIP within a defined distance of the ligand PIP. For example, for a hydrogen bond acceptor in the bound ligand, there must a hydrogen bond donor in the protein within 3.9 Å for this ligand feature for the PIP to be retained. This culling of features is determined by the feature pairs and distance criteria shown in Table 1, which are partly assembled from the literature.[58−60] Ligity does not distinguish between weak, moderate, and strong hydrogen bonds, which are typically defined using different distance intervals.[61] Also, Ligity does not consider less common and weaker π-interactions, such as π donor−acceptor interactions.[62] PIPs are placed at either the center of an atom or a pseudoatom, as appropriate for the PIP

Table 1. Pharmacophoric Features and Distance Thresholds Used To Define Queries in Ligity

interacting receptor−ligand PIP pairs	distance threshold (Å)
(hydrophobic, hydrophobic)	4.5
(acceptor, donor)	3.9
(cation, anion)	4.0
(aromatic, aromatic)	4.5
(cation, aromatic)	4.0

DOI: 10.1021/acs.jcim.8b00779

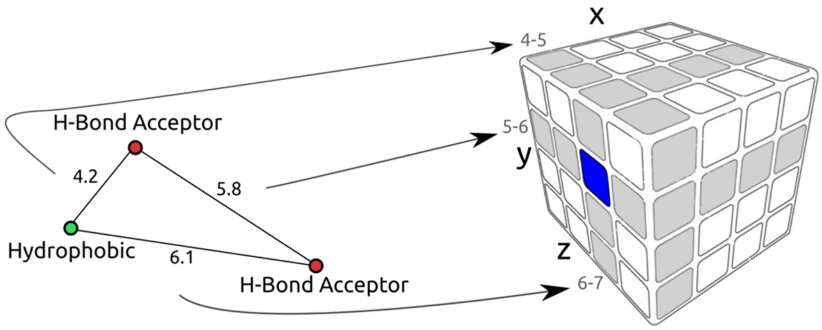


Figure 2. Relationship between PIPs and the Ligity descriptor. The lengths of the edges of each triangle are mapped to their corresponding indices in the array of triangle counts. Shown here is a 3-PIP triangle, with features ⟨HBA, HBA, hydrophobic⟩ and edge lengths of ⟨5.8, 6.1, 4.2⟩ mapping onto a Ligity descriptor that uses an edge-length bin size of 1.0 Å. The blue-bin count will be incremented.

type; e.g., a benzene ring would have an aromatic PIP positioned at the center of the ring.

For hydrogen bond acceptor and donor interactions, we also require that the angle between the hydrogen bond donor heavy atom, its bonded hydrogen atom, and the hydrogen bond acceptor atom, should form an angle greater than 90°.

This helps to both prioritize and limit the number of interaction points.

The query protein−ligand complex PIPs are therefore defined as those ligand PIPs that interact with the protein, filtering out parts of the ligand that do not contribute directly to binding to the target, such as solvent-exposed parts of the ligand. These are typically a subset of all the possible PIPs in a ligand. This has a key influence when selecting a similarity measure. An example of PIP generation for a cognate ligand of adenosine deaminase is shown in the Supporting Information (Figure S1).

PIP Generation for Database Compounds. PIPs are generated for every conformer of each database compound using the same SMARTS patterns. Database compounds do not go through the culling step, so for every conformer, all PIPs are retained. Thus, the size of the set of PIPs for database compounds tends to be larger than for binding site ligands from the query protein−ligand complex. While every conformer of a given molecule has the same number and types of PIPs, their positions will depend on its atomic coordinates.

Ligity Descriptor Generation. The list of PIPs generated either for the query or database compound is used to generate a Ligity descriptor. A descriptor consists of all possible combinations of sets of 3- or 4-PIPs and is assembled as shown in Figure 2.

3-PIP Combinations. In the case of 3-PIP combinations, each triplet describes a triangle whose vertices are the PIP types (e.g., ⟨donor, acceptor, acceptor⟩), and whose edges correspond to the distances between those PIPs. Each triangle is then used to populate a 3D "array", mapping each of the three lengths of the triangle's edges to corresponding indices that determine which bin should accumulate the count of that particular PIP triplet combination.

4-PIP Combinations. In the case of 4-PIP combinations, each quadruplet of PIPs describes a tetrahedron whose vertices are the four PIP types (e.g., ⟨donor, acceptor, cation, acceptor⟩) and whose six edges again correspond to the distances between pairs of PIPs. Tetrahedron counts are stored in a six-dimensional array, hereafter referred to as a *hypercube*, using the lengths of each of the six edges to map to the indices corresponding to the appropriate bin that accumulates the counts of PIP types for that specific tetrahedron.

Ligity Descriptor Bin Sizes. The location of the bin in the multidimensional array corresponds to the lengths of the edges of the triangle (3-PIP) or tetrahedron (4-PIP) geometries. Each bin stores the number of occurrences of every possible combination of 3-PIP triangle or 4-PIP tetrahedron. The bin size determines the "resolution" of the Ligity descriptor. If the maximum distance between two PIPs of a molecule is 15 Å and a bin size of 1 Å is used, there would be 15 equally sized bins in any dimension of the descriptor. A very fine bin size, e.g., 0.1 Å, would give a descriptor with many bins with most bin counters set to either 0 or 1. On the other hand, a large bin size, of say 30 Å, would result in only one bin with all counters set to the total number of occurrences of each PIP type combination.

In both 3-PIP and 4-PIP combinations, index determinism is ensured by sorting the vertex labels into a canonical order, which ensures that the same 3-PIP triangles or 4-PIP tetrahedra contribute to the same bin counter. For example, with 1 Å sized bins, (⟨HBA, HBD, +⟩, ⟨2.6, 3.7, 3.2⟩) and (⟨HBA, +, HBD⟩, ⟨3.2, 3.7, 2.6⟩) would contribute to the same bin, ⟨+ , HBA, HBD⟩ with indices ⟨3, 2, 3⟩. When a geometry, i.e., a triangle or tetrahedron, contains more than one of the same PIP type (e.g., 3-PIP ⟨HBA, HBD, HBD⟩), the edges for the pairs containing those identical PIP types are sorted by length. This guarantees that identical geometries are also assigned to the same bin counter. For example, using 1 Å sized bins, the 3-PIP triangles {⟨HBA, HBD, HBD⟩, ⟨3.7, 4.2, 2.1⟩} and {⟨HBA, HBD, HBD⟩, ⟨2.1, 4.2, 3.7⟩}, would be counted in the same bin, ⟨HBA, HBD, HBD⟩ with indices ⟨2, 4, 3⟩. Triangles or tetrahedra having any side shorter than 1.5 Å or longer than 15 Å are filtered out. The lower-bound filtering eliminates large numbers of repetitive geometries (such as those found in a six-membered aliphatic rings). We also filter the descriptor to remove 3-PIP and 4-PIP geometries that appear infrequently (e.g., that occur only once) or those that are very common (e.g., >10 times). PIP geometries that occur frequently in all molecules tend not to be useful for discriminating between actives and decoys.

Chirality Detection. A 4-PIP tetrahedron is chiral if all four PIP types are different. We capture this chirality by calculating the chiral volume of the tetrahedron after the four PIPs have been deterministically sorted by their type. Equation 1 gives a positive or negative volume, which distinguishes between the chiral forms of the tetrahedron

$$V = \frac{(\vec{a} - \vec{d})\cdot[(\vec{b} - \vec{d}) \times (\vec{c} - \vec{d})]}{6} \qquad (1)$$

where $\vec{a}$, $\vec{b}$, $\vec{c}$, and $\vec{d}$ are the Cartesian coordinates of the vertices of the PIP tetrahedron. As a performance optimization, we drop the denominator from the volume calculation, as we only need the sign of the result to distinguish chirality.

Similarity Score. The PIP descriptors for the query and for a database compound are used to calculate their similarity. We tested the similarity measures Tanimoto, Dice, Cosine, Common Counts, and Tversky. The similarity score of a database compound is taken to be the highest score from all of its conformers in the conformational ensemble.

The Tversky similarity, $S_{\alpha}\beta(A,B)$, of the Ligity descriptor counts is shown in Equation 2

$$S_{\alpha\beta}(A,B) = \frac{\sum_{i=1}^{n} \min(A_i,B_i)}{\alpha \sum_{i=1}^{n} A_i + \beta \sum_{i=1}^{n} B_i + (1 - \alpha - \beta) \sum_{i=1}^{n} \min(A_i,B_i)} \qquad (2)$$

where A is the Ligity descriptor of a protein–ligand query, B is the Ligity descriptor of a ligand in the virtual library, n is the number of distance bins in the descriptor, and A_i and B_i are the counts of PIP geometries in the ith bin. This is the nonbinary implementation of the Tversky index, designed to work with integer vectors as opposed to bit vectors. We tested the Tversky similarity with weights of $\alpha = 1$, $\beta = 0$; $\alpha = 0.95$, $\beta = 0.05$; $\alpha = 0.9$, $\beta = 0.1$; and $\alpha = 0.85$, $\beta = 0.15$. We hypothesized that the asymmetric Tversky measure would work well because, unlike symmetric measures such as the Tanimoto (or Jaccard[63]) index, it can be used to emphasize substructural features of a query and would thus help to focus on only those parts of the query ligand that interact with the protein.

Fusion of Ranked Results. When multiple queries are available, we fuse the results from the different queries into a single list. This is achieved using the maximum similarity (MAX-SIM) data fusion method as described by Nasr et al.[64] and is shown in Equation 3

$$S(A,B_1,...,B_{|\vec{B}|}) = \max_{j} S(\vec{A},\vec{B_j}) \qquad (3)$$

where S is the similarity score, A is the query vector, B is the vector of the ligand in the multiple lists, B_j is the vector of the ligand in list j, and $|\vec{B}|$ is the number of structures being compared. The highest scoring instance of a molecule across all ranking lists is used in a final, single ranking. It should be noted that other fusion methods, such as exponential Tanimoto discriminant, have been reported to do slightly better[64] than MAX-SIM, but they are much more complex and require parameters to be fitted.

Performance Measurement. All receiver operating characteristic (ROC) curves and area under the ROC curve (AUC) calculations were performed in the R software environment for statistical computing and graphics (version 3.0.0)[65] using the package ROCR (version 1.0–4).[66,67] BEDROC values were calculated using the R package enrichvs (version 0.0.5).[68] For BEDROC, we used an α value of 20, as suggested by the authors of this method;[69] this means that "80% of the maximum contribution to the BEDROC comes from the first 8% of the list".

Ligity Parameter Refinement Data Set. There are several parameters in the Ligity method. To identify optimal values we selected three targets that are present in both sc-PDB and DUD-E,[70] which is a widely used benchmark set for evaluating virtual screening methods. DUD-E contains 50 decoy compounds chosen to have similar physicochemical properties but dissimilar 2D topology, for every active compound in the set of 22 886 actives, grouped into 102 diverse target sets, with an average of 224 actives per target. These three targets were adenosine deaminase (ADA), cyclin-dependent kinase 2 (CDK2), and trypsin I (TRY1), as shown in Table 2. For the CDK2 and TRY1 receptor sets, we randomly selected 100 actives and 100 decoys from the DUD-E sets. For ADA this was not possible, as there are only 93 actives in the DUD-E data set, so we selected all of these and 100 ADA decoys.

Table 2. Targets Used To Identify Optimal Parameter Values and Data Structures for Ligity, along with Their sc-PDB Cluster IDs, PDB IDs, and the Number of Conformers, n_c, for the Active and Decoy Sets[a]

		PDB ID			n_c	
target	sc-PDB cluster ID	site 1	site 2	site 3	for actives	for decoys
adenosine deaminase (ADA)	0085	1ndv	2e1w	3km8	6002	6394
cyclin-dependent kinase 2 (CDK2)	1424	1pxm	2bts	2c6m	6839	6697
trypsin 1 (TRY1)	1463	1bjv	1o3o	3m35	4647	4956

[a] Three structures of each target were chosen so as to capture the protein flexibility. Each receptor has a randomly selected subset of 100 active and 100 decoy molecules taken from DUD-E, except ADA, which has only 93 actives.

Validation Data Sets. To benchmark the two modes of Ligity, "HTS mode" and "information-rich mode", we used DUD-E to create two data sets. These consisted of (i) all of the targets in DUD-E and (ii) a randomly selected subset of 10 targets from the DUD-E set that were also present in sc-PDB (2011 release). The focused subset consisted of angiotensin-converting enzyme (ACE), adenosine deaminase (ADA), cyclin-dependent kinase 2 (CDK2), coagulation factor X (FA10), coagulation factor VII (FA7), glucocorticoid receptor (GCR), human immunodeficiency virus type 1 integrase (HIVINT), human immunodeficiency virus type 1 protease (HIVPR), thrombin (THRB), and trypsin I (TRY1). These included the three targets described in the previous section used for Ligity parameter refinement. For each target, we used the sc-PDB cluster of cognate protein–ligand complexes to construct information-rich Ligity queries. We removed the few structures whose ligands caused errors when read by RDKit and then standardized the ionization state of the remaining bound ligands. We also standardized the ionization state and generated conformers for the corresponding actives and decoys of each target taken from DUD-E using the same standardization rules. When present, we removed the cognate ligand in the sc-PDB structure from the corresponding DUD-E actives set, to eliminate any bias in the method. As might be expected, a conformer of the cognate ligand tends to score very highly with the cognate ligand's Ligity descriptor. This effect would be accentuated when using the MAX-SIM data fusion method, which takes the highest score for each virtual library molecule across all binding pockets for a target. The final benchmarking data set is shown in Table 3. We list the number

Table 3. DUD-E Target Datasets Used To Benchmark Ligity, along with Their sc-PDB Clusters, and the Number of Actives and Decoys, with the Number of Conformers, n_c, Given in Parentheses

target	sc-PDB cluster ID	number of sc-PDB structures	number of actives (n_c)	number of decoys (n_c)
angiotensin-converting enzyme (ACE)	0132	9	282 (27 346)	16 900 (1 307 531)
adenosine deaminase (ADA)	0085	20	93 (6 720)	5 450 (371 990)
cyclin-dependent kinase 2 (CDK2)	1424	109	474 (20 480)	27 850 (1 360 619)
coagulation factor X (FA10)	0224	81	537 (39 732)	28 325 (1 799 269)
coagulation factor VII (FA7)	0223	15	114 (9 759)	6 250 (398 145)
glucocorticoid receptor (GCR)	0367	7	258 (8 682)	14 999 (640 882)
human immunodeficiency virus type 1 integrase (HIVINT)	1167	3	98 (5 096)	6 650 (327 474)
human immunodeficiency virus type 1 protease (HIVPR)	0654	166	535 (27 975)	35 750 (2 189 091)
thrombin (THRB)	0830	113	461 (34 936)	27 004 (2 020 395)
trypsin I (TRY1)	0850	74	449 (30 311)	25 980 (1 706 265)

Table 4. Effect of Different Similarity Measures on Ligity's ROC AUC Performance Using the Parameter Optimizaion Dataset[a]

		AUC						
		Tversky (1)	Tversky (2)	Tversky (3)	Tversky (4)			
	query	$\alpha = 1$	$\alpha = 0.95$	$\alpha = 0.9$	$\alpha = 0.85$			
target	PDB ID	$\beta = 0$	$\beta = 0.05$	$\beta = 0.1$	$\beta = 0.15$	Tanimoto	Cosine	Counts
ADA	1ndv	0.791	0.814	0.831	**0.839**	0.799	0.810	0.791
	2e1w	0.904	0.911	0.919	**0.927**	0.909	0.912	0.904
	3km8	**0.796**	0.790	0.784	0.770	0.672	0.714	**0.796**
mean		0.830	0.838	**0.845**	**0.845**	0.793	0.822	0.830
fusion		0.893	0.913	0.925	**0.935**	0.912	0.926	0.913
CDK2	1pxm	**0.636**	0.581	0.541	0.516	0.508	0.530	**0.636**
	2bts	0.700	**0.701**	0.669	0.634	0.539	0.593	0.700
	2c6m	**0.575**	0.553	0.539	0.522	0.513	0.481	**0.575**
mean		**0.637**	0.611	0.583	0.557	0.520	0.535	**0.637**
fusion		**0.634**	0.575	0.532	0.508	0.488	0.503	0.576
TRY1	1bjv	**0.724**	0.511	0.479	0.467	0.412	0.557	**0.724**
	1o3o	**0.671**	0.456	0.394	0.363	0.300	0.417	**0.671**
	3m35	**0.728**	0.709	0.663	0.615	0.483	0.687	**0.728**
mean		**0.708**	0.559	0.512	0.482	0.398	0.554	**0.708**
fusion		**0.729**	0.712	0.662	0.611	0.472	0.687	0.728

[a] The best AUC in each row is shown in bold; it can be seen that Tversky with $\alpha = 1$ and $\beta = 0$ performed best. Note that Dice similarity calculations (not shown) gave identical results to Tanimoto. The mean AUC across all three cognate structure queries for each target is shown, as well as the AUC of the MAX−SIM fused results over the three individual queries.

of sc-PDB protein−ligand structures used as queries, the number of active and decoy compounds, and how many conformers they have.

In order to compare Ligity to another non-superpositional ligand-based virtual screening method, we also tested Oxford Drug Design's proprietary implementation of ElectroShape 4D (version 2.0.2).[71] The virtual screening study was performed by using the 3D cognate ligand as an input to ElectroShape 4D to generate a query descriptor that was used to rank the active and decoy molecules of each receptor. Similar to Ligity, the ionization states of the query molecule and the active and decoy compounds were standardized at physiological pH. This was particularly important for ElectroShape, as it uses partial charges in its similarity computation. All Ligity comparisons were carried out using a single receptor structure, as defined in DUD-E.

Unless otherwise stated, the following computational experiments were carried out using the validation data set described in Table 2. We tested the effects of (1) using 3-PIP versus 4-PIP combinations for descriptor generation, (2) varying the descriptor length bin sizes used to count similar multi-PIP geometries, (3) using different similarity measures, (4) ranking using either single or MAX−SIM fused results, and (5) using the lowest-energy conformer versus multiple conformers for each molecule in a virtual library.

RESULTS AND DISCUSSION

We present the results in two parts: the first discusses our investigation to identify the optimal parameters for Ligity. The second part presents the validation of Ligity for virtual screening using DUD-E.[70] The validation was carried out at two levels: (i) using the HTS mode of Ligity on all of DUD-E and (ii) using the information-rich mode of Ligity on a randomly chosen subset of 10 DUD-E targets (ACE, ADA, CDK2, FA10, FA7, GCR, HIVINT, HIVPR, THRB, and TRY1). Of these 10, 3 were selected, ADA, CDK2, and TRY1, for Ligity parameter optimization.

Ligity Parameter Optimization. *3-PIP versus 4-PIP Descriptors.* We found that, in terms of the method's accuracy as measured by ROC AUC, the 4-PIP descriptors performed

DOI: 10.1021/acs.jcim.8b00779

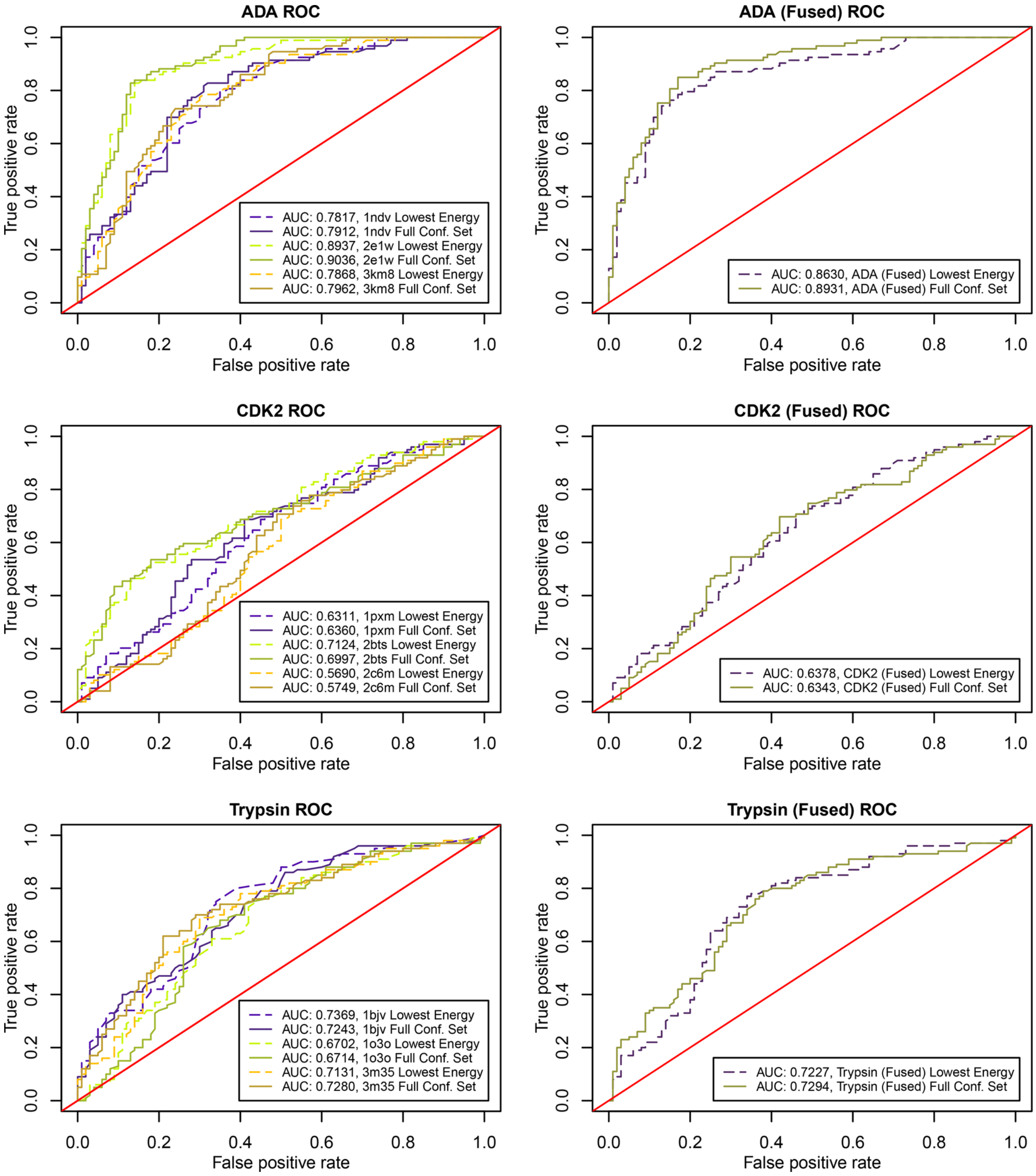


Figure 3. Ligity showed little difference when using only the lowest-energy conformer of the active or decoy molecule (left column). Using the lowest-energy conformer when fusing the results also exhibited a minimal effect (right column). The performance of the method using only the lowest-energy conformer is shown with dashed lines, while the performance using the full conformer set is shown with solid lines.

marginally better than 3-PIPs on our parameter optimization data set (Supporting Information, Figure S2). However, in terms of storage space, the 3-PIP descriptors are on average 90.75% smaller than the 4-PIP descriptors. This is an important consideration in virtual screening experiments, where it is not uncommon to have millions of molecules to test.

4-PIP descriptors perform better than 3-PIPs for ADA (average AUC improvement of 0.018) and TRY1 (average AUC improvement of 0.020). However, there is a decrease in performance for CDK2, where the 3-PIP descriptor performs better (average 3-PIP AUC improvement over 4-PIP of 0.028). The same pattern is observed in the fused results scoring. It is possible that CDK2's ligands, which tend to be flat, could be better captured by the triangular 3-PIP descriptors. However, given the three-dimensional and chiral nature of ligand binding sites, Ligity uses 4-PIP descriptors by default.

Descriptor Bin Size. We evaluated the performance of 4-PIP Ligity descriptors with hypercube bin sizes of 0.5, 1.0, 1.5, and 2.0 Å. The number of query-matching tetrahedra increases with the bin size. The results are shown in Table S2 of the Supporting Information. Results indicate that the optimal bin size may be dependent on the receptor; Ligity performs slightly better for CDK2 with a bin size of 0.5 Å, while it performs better with a bin size of 1.5 Å for ADA and TRY1. It is possible that this observation could be related to the flexibility of the molecules. CDK2 actives and decoys sets are less flexible, as measured by the number of rotatable bonds, than the ADA and

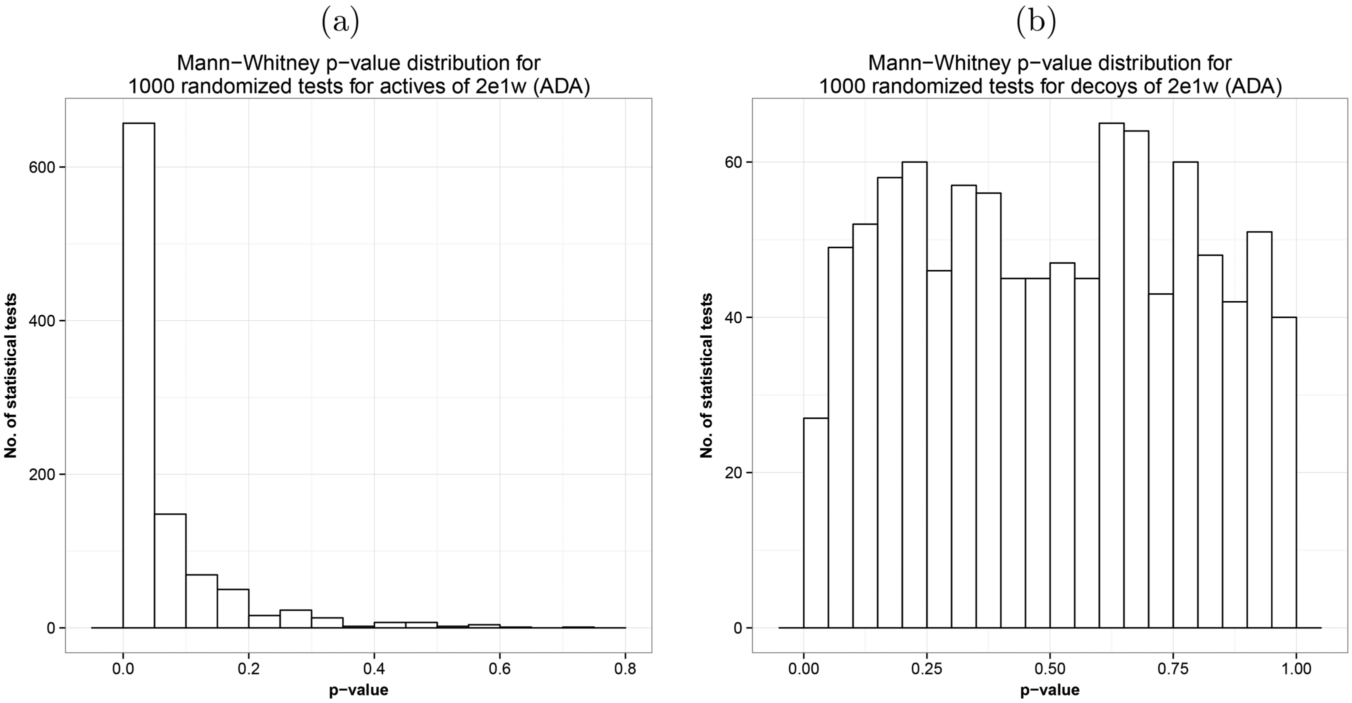


Figure 4. Ligity preferentially selects lower-energy conformers for actives but not for decoys: (a) lower conformer identifiers are selected for actives, rather than what one would expect at random, and (b) there is no difference between conformer identifiers selected for decoys and ones selected at random.

TRY1 sets (Supporting Information, Figure S3). Larger, more flexible molecules have many degrees of freedom and a smaller bin size may not capture similarities between the query and database molecules adequately. On the other hand, smaller, less flexible molecules would match a large number of tetrahedrons if large bins are used because many molecules would populate the same bins, increasing the false positive rate and resulting in a decrease in performance. The default value of the bin size in Ligity was chosen to be 1.5 Å.

Similarity Measures. Using 4-PIP descriptors and a bin size of 1.5 Å, we investigated the performance of the following similarity measures, Tanimoto, Cosine, Dice, Counts (simple counts of common bins), and Tversky (including different values for α and β). In each case, the ability to rank actives over decoys was assessed by computing the area under the ROC curve for each optimization data set protein. The results are presented in Table 4.

The Tanimoto and Dice similarity measures give identical ROC AUCs (see eqs 2 and 3 in Schema S1 in the Supporting Information). These two measures are similar in spirit, the only difference is that Tanimoto is the ratio of the number of common features between the two data structures over the total number of features, while the Dice measure is the ratio of the total number of common features over the average size of the features in the two data structures.

Tversky, with $\alpha = 1$ and $\beta = 0$, and Counts give identical results. This is because Tversky with $\beta = 0$ is a special case of Counts, where the number of common features are divided by the number of features in the query, which does not vary, therefore giving the same rankings of actives and decoys. Note that results are different when we consider fused rankings results: this is because when using the Tversky measure, the different protein−ligand queries will normalize the common counts between query and database ligand by a different amount.

Table 4 shows that the different similarity measures perform more consistently for ADA than for CDK2 and TRY1. For CDK2 and TRY1, similarity measures that quantify global similarity between the query and virtual library molecule, such as Tanimoto, perform poorly. This can be explained by considering the number of PIPs in the query compared to the number of database compound PIPs (Supporting Information, Figure S4). The number of query PIPs for CDK2 and TRY1 is smaller than the number of database compound PIPs. In order to capture this asymmetry, we need a similarity measure that captures the substructure nature of the query. By setting the α parameter to 1.0 in the Tversky measure, we place all the importance on the query PIPs (and effectively ignore ligand's PIPs that do not match).

The number of PIPs in the query and the Tversky score AUCs have a moderate Pearson correlation coefficient (r) of 0.510, $p < 0.05$. If we remove the 2c6m query, which seems to be an outlier, from the CDK2 set, the Pearson's correlation coefficient improves to $r = 0.741$, $p < 0.05$.

Single Versus Fused Results Rankings. As expected, Table 4 also shows that consensus scoring using MAX-SIM data fusion over the results from multiple queries improves the performance of Ligity. For example, for ADA using Tversky with $\alpha = 1$ and $\beta = 0$, the mean AUC was 0.830, while the fusion AUC was 0.893. Ligity therefore uses MAX-SIM data fusion methods across multiple protein−ligand complexes by default.

Using Lowest Energy Conformer versus Conformational Ensemble. We were interested in testing two different ligand conformer representations, one where we used only the lowest energy conformer of the molecule and the other where we used a conformational ensemble of up to a maximum of 300 conformers per molecule. Using a single conformer greatly increase the speed and reduces the storage requirements of the method. We tested Ligity using chiral tetrahedral 4-PIPs, a bin

Table 5. Performance of Ligity in HTS Mode against the Ligity-Compatible DUD-E Targets[a]

			ROC AUC		BEDROC		$EF_{1\%}$	
target	no. of actives	no. of decoys	Tanimoto	Tversky	Tanimoto	Tversky	Tanimoto	Tversky
ABL1	182	10 750	0.563	0.473	0.077	0.077	1.653	2.204
ACE	281	16 877	0.787	0.787	0.336	0.401	12.425	19.525
ACES	453	26 242	0.634	0.645	0.077	0.155	1.766	5.518
ADA	93	5 450	0.724	0.660	0.149	0.147	3.251	3.251
ADA17	532	35 898	0.638	0.728	0.103	0.283	1.317	9.030
ADRB1	247	15 850	0.523	0.647	0.065	0.129	1.619	5.262
ADRB2	231	14 999	0.523	0.589	0.052	0.040	1.735	0.000
AKT1	293	16 450	0.386	0.548	0.039	0.107	2.737	3.080
AKT2	117	6 900	0.511	0.685	0.140	0.194	8.568	8.568
ALDR	159	8 988	0.574	0.610	0.202	0.172	10.747	6.322
AMPC	48	2 845	0.521	0.541	0.049	0.023	0.000	0.000
ANDR	269	14 349	0.722	0.742	0.194	0.354	4.839	24.938
AOFB	121	6 875	0.422	0.464	0.045	0.027	1.652	0.000
BACE1	283	18 100	0.441	0.775	0.017	0.310	0.000	13.062
BRAF	152	9 950	0.612	0.639	0.208	0.165	12.502	5.264
CASP3	199	10 694	0.600	0.734	0.068	0.258	0.502	7.031
CDK2	474	27 838	0.467	0.507	0.021	0.048	0.000	1.055
COMT	41	3 846	0.789	0.889	0.338	0.665	19.447	58.341
CP2C9	120	7 449	0.518	0.634	0.058	0.186	1.660	8.299
CP3A4	170	11 787	0.450	0.493	0.022	0.057	0.000	2.345
CSF1R	166	12 149	0.526	0.542	0.136	0.152	6.031	7.238
CXCR4	40	3 405	0.575	0.722	0.217	0.134	12.665	0.000
DEF	102	5 699	0.732	0.833	0.212	0.379	10.786	15.689
DHI1	330	19 348	0.481	0.595	0.089	0.062	2.422	1.211
DPP4	533	40 941	0.586	0.591	0.154	0.157	4.312	3.937
DRD3	480	34 048	0.484	0.441	0.043	0.046	1.251	0.626
DYR	231	17 196	0.694	0.758	0.210	0.230	6.504	7.371
EGFR	542	35 047	0.593	0.491	0.054	0.037	0.922	0.000
ESR1	383	20 683	0.838	0.861	0.527	0.594	31.281	39.101
ESR2	367	20 199	0.844	0.870	0.563	0.644	20.130	32.644
FA10	537	28 324	0.564	0.674	0.058	0.118	0.930	2.232
FA7	114	6 249	0.762	0.859	0.210	0.332	6.105	8.721
FABP4	47	2 749	0.786	0.744	0.191	0.276	0.000	10.623
FAK1	100	5 350	0.642	0.531	0.111	0.065	2.019	0.000
FGFR1	139	8 698	0.511	0.522	0.036	0.088	0.722	1.445
FKB1A	111	5 799	0.605	0.751	0.162	0.164	8.122	3.610
FNTA	592	51 493	0.411	0.625	0.012	0.132	0.000	4.053
FPPS	85	8 842	0.917	0.985	0.323	0.776	2.360	36.581
GCR	258	14 998	0.805	0.834	0.244	0.324	3.092	8.116
GLCM	54	3 790	0.667	0.685	0.182	0.279	1.873	11.240
GRIA2	158	11 842	0.662	0.684	0.248	0.154	11.392	5.696
GRIK1	101	6 547	0.656	0.668	0.203	0.102	7.978	1.995
HDAC2	185	10 300	0.676	0.734	0.187	0.201	4.318	4.318
HDAC8	170	10 449	0.640	0.819	0.120	0.377	2.946	8.250
HIVINT	100	6 640	0.390	0.554	0.030	0.116	0.000	3.018
HIVPR	535	35 724	0.663	0.872	0.072	0.490	0.187	23.898
HIVRT	338	18 884	0.495	0.475	0.124	0.085	4.443	1.777
HMDH	170	8 750	0.480	0.906	0.068	0.652	2.358	35.963
HS90A	88	4 850	0.635	0.506	0.096	0.083	0.000	3.436
HXK4	92	4 700	0.662	0.803	0.206	0.307	15.192	9.766
IGF1R	148	9 300	0.502	0.575	0.057	0.189	2.037	14.941
INHA	43	2 300	0.493	0.575	0.031	0.045	0.000	0.000
ITAL	138	8 500	0.619	0.465	0.037	0.065	0.000	0.728
JAK2	107	6 500	0.472	0.475	0.073	0.118	2.807	6.549
KIF11	116	6 850	0.755	0.781	0.149	0.219	4.289	2.574
KIT	166	10 449	0.463	0.437	0.045	0.030	0.000	0.000
KITH	57	2 850	0.649	0.838	0.228	0.709	14.069	47.483
KPCB	135	8 699	0.753	0.813	0.220	0.338	8.923	12.641
LCK	419	27 391	0.471	0.437	0.031	0.043	0.000	1.910

DOI: 10.1021/acs.jcim.8b00779

Table 5. continued

			ROC AUC		BEDROC		EF$_{1\%}$	
target	no. of actives	no. of decoys	Tanimoto	Tversky	Tanimoto	Tversky	Tanimoto	Tversky
LKHA4	171	9 448	0.718	0.694	0.238	0.150	8.203	1.758
MAPK2	101	6 148	0.660	0.670	0.174	0.199	5.988	3.992
MCR	94	5 149	0.816	0.888	0.215	0.454	6.436	19.307
MET	166	11 249	0.566	0.531	0.130	0.065	6.032	0.603
MK01	79	4 550	0.518	0.602	0.121	0.206	5.095	3.821
MK10	104	6 600	0.488	0.489	0.020	0.031	0.962	0.962
MK14	578	35 847	0.511	0.589	0.040	0.064	0.173	0.519
MMP13	572	37 199	0.648	0.753	0.134	0.268	2.446	9.957
MP2K1	121	8 146	0.669	0.569	0.187	0.058	3.293	0.823
NOS1	98	8 028	0.483	0.451	0.109	0.041	3.071	0.000
NRAM	98	6 200	0.853	0.859	0.342	0.290	11.221	3.060
PA2GA	99	5 150	0.793	0.756	0.225	0.153	1.020	3.059
PARP1	508	30 029	0.635	0.692	0.215	0.231	11.234	7.884
PGH1	195	10 798	0.645	0.637	0.077	0.100	0.000	2.050
PGH2	435	23 139	0.716	0.780	0.166	0.291	3.444	9.874
PLK1	107	6 800	0.658	0.531	0.123	0.048	1.871	0.000
PNPH	103	6 946	0.575	0.578	0.161	0.181	4.888	8.799
PPARA	373	19 399	0.783	0.778	0.262	0.280	6.693	7.764
PPARD	240	12 250	0.547	0.544	0.078	0.098	1.665	2.498
PPARG	484	25 299	0.515	0.605	0.055	0.118	0.619	4.955
PRGR	293	15 648	0.740	0.793	0.142	0.318	2.053	14.714
PTN1	130	7 249	0.398	0.538	0.055	0.090	0.000	3.068
PUR2	50	2 700	0.851	0.837	0.281	0.255	7.857	1.964
PYGM	77	3 944	0.403	0.492	0.016	0.137	0.000	3.917
PYRD	111	6 449	0.682	0.710	0.462	0.413	34.027	16.118
RENI	104	6 956	0.720	0.789	0.043	0.138	0.000	0.000
ROCK1	100	6 300	0.347	0.449	0.020	0.084	1.000	4.000
RXRA	131	6 950	0.788	0.900	0.219	0.596	6.091	27.407
SAHH	63	3 450	0.874	0.852	0.598	0.542	35.050	27.084
SRC	524	34 500	0.565	0.477	0.065	0.050	0.382	0.573
TGFR1	133	8 499	0.609	0.639	0.147	0.154	10.565	4.528
THB	103	7 450	0.794	0.762	0.238	0.150	10.614	0.965
THRB	461	27 000	0.605	0.706	0.063	0.166	2.166	5.632
TRY1	449	25 975	0.711	0.815	0.147	0.280	2.898	6.688
TRYB1	148	7 650	0.670	0.670	0.153	0.132	3.378	3.378
TYSY	109	6 745	0.594	0.725	0.071	0.226	0.911	5.468
UROK	162	9 850	0.525	0.650	0.036	0.120	0.000	1.854
VGFR2	409	24 948	0.632	0.578	0.083	0.093	1.465	1.465
WEE1	102	6 150	0.934	0.929	0.789	0.797	59.348	61.294
XIAP	100	5 150	0.752	0.974	0.190	0.897	8.077	51.490

[a] The mean (and standard deviation in parentheses) values of ROC AUC using Tanimoto is 0.622 (±0.132), while for Tversky it is 0.671 (±0.142); the mean EF$_{1\%}$ using Tanimoto is 5.648 (±8.668), while for EF$_{1\%}$ using Tversky it is 9.047 (±12.713).

size of 1.5 Å, and Tversky similarity with $\alpha = 1$ and $\beta = 0$. Surprisingly, we found that using the lowest-energy conformer rather than the full ensemble had a small effect on the AUC performance of Ligity (Figure 3). The average difference in AUC between all nine individual binding site structure queries for ADA, CDK2, and TRY1 is 0.003. For fused results, the average difference in AUC between the full conformer ensemble and the lowest energy conformers is 0.011, with the full conformational ensemble model doing just slightly better.

Since Ligity performed equally well when using only the lowest energy conformer, we decided to test whether the lowest-energy conformers of a molecule scored best when the full conformational ensemble is used. For any molecule, each conformer is assigned a unique integer identifier assigned sequentially to the conformers when sorted by energy. We investigated whether Ligity picked lower conformer identifiers more often than would be expected by random chance. Not all molecules have the same number of conformers in their ensembles, and a random selection is therefore more likely to pick conformer identifier 1 than conformer identifier 300, because every molecule will have at least one conformer but very few will have the maximum possible of 300. Therefore, for the active and decoy sets for ADA, CDK2, and TRY1, we built a theoretical probability function for the selection of conformer identifiers (as identifiers are generated according to the energy rank of the conformer).

The model is described in Equation 4

$$P(c_{\mathrm{id}}) = \frac{\sum_{i=1}^{N} \frac{1}{|C_i|}}{N} \qquad (4)$$

Table 6. Ligity Results in Information-Rich Mode Using the DUD-E Validation Subset[a]

receptor	mean AUC (±σ)	mean BEDROC (±σ)	fusion AUC	fusion BEDROC
angiotensin-converting enzyme (ACE)	0.779 (±0.070)	0.424 (±0.181)	*0.948*	0.776
adenosine deaminase (ADA)	0.811 (±0.068)	0.302 (±0.102)	*0.894*	0.557
cyclin-dependent kinase 2 (CDK2)	0.610 (±0.035)	0.081 (±0.047)	*0.643*	0.062
coagulation factor X (FA10)	0.700 (±0.050)	0.195 (±0.079)	0.716	0.208
coagulation factor VII (FA7)	0.750 (±0.026)	0.277 (±0.540)	*0.809*	0.270
glucocorticoid receptor (GCR)	0.790 (±0.094)	0.300 (±0.116)	*0.867*	0.439
human immunodeficiency virus type 1 integrase (HIVINT)	*0.669* (±0.045)	0.173 (±0.068)	0.637	0.139
human immunodeficiency virus type 1 protease (HIVPR)	0.874 (±0.018)	0.584 (±0.057)	*0.876*	0.527
thrombin (THRB)	0.747 (±0.035)	0.220 (±0.079)	*0.752*	0.185
trypsin I (TRY1)	0.725 (±0.060)	0.171 (±0.076)	*0.778*	0.167
mean across all receptors	0.745 (±0.050)	0.273 (±0.135)	*0.792*	0.333

[a] It can be seen that Ligity's mean ROC AUC is moderate to excellent across all targets and generally improves when using data fusion. Standard deviation values, σ, are shown in parentheses. The best ROC AUC for each target is in italic.

where $P(c_{\mathrm{id}})$ is the probability of picking a specific conformer with identifier id for all N molecules, N is the total number of molecules for the receptor (e.g., 93 for ADA actives), and C_i is the conformational ensemble for the ith molecule. The probability of picking a conformer identifier that is larger than the conformer ensemble size for that molecule is zero. In our theoretical model, every conformer in an ensemble is equally likely to be picked, regardless of size. (A simplified example is offered in Table S3 of the Supporting Information.)

We compare the Ligity ranking of all conformers of N molecules with 1000 randomly generated rankings of the same conformers of these N molecules. We generate each of these 1000 random results lists using the probability function in Equation 4. We then run a Mann−Whitney statistical test for the conformer identifier selection produced by Ligity against each of the 1000 random results lists. In Figure 4 we show two histograms, of the 1000 resulting p-values for the active and decoy sets for ADA. There is a clear preference for Ligity to select lower-energy conformers of actives, where most p-values are below our significance level (α) of 0.05 (Figure 4a). In other words, there is a statistically significant difference between the conformer identifiers picked by Ligity for actives and a random set of generated conformer identifiers. We also ran a one-sided test, so we specifically tested for a preference for lower conformer identifier than the random set. For the decoy set (Figure 4b) we saw no significant preference for lower conformer identifiers as opposed to random selection, as there are very few p-values smaller than our chosen α (i.e., smaller than 0.05). This differential behavior was observed for all three targets we used while prototyping Ligity, including CDK2 and TRY1.

Using only the lowest-energy conformers is sufficient for Ligity to distinguish between actives and decoys. Butler et al.[72] found that bioactive conformations are very close to the global energy minimum: two-thirds of their 99-molecule data set were within 0.5 kcal/mol of the global energy minimum. They used "sophisticated QM-based methods to take into account both the internal energy of the ligand and the solvation effect, and the application of physically meaningful constraints to refine the bioactive conformation" and asserted that strain energies larger than 2 kcal/mol tended to be attributable to structural determination inaccuracies. Others have argued that the strain energy of the bound ligand conformation can be higher than the calculated global minimum for their unbound form, by as much as 5−15 kcal/mol.[73−77] Some have claimed that the effect of the bioactive conformation on virtual screening experiments is small.[78,79] Indeed in some of these studies, starting off with a low-energy conformation instead of the bioactive conformation yielded little difference in enrichment.

The importance of any differences between the bioactive and solution conformations of ligands on their binding affinities remains controversial. In this regard, our findings are consistent with those of Butler et al.

Validating Ligity Using DUD-E. Ligity was validated in two ways: (i) using the HTS mode of Ligity against the entire DUD-E benchmark and (ii) using the information-rich mode against the 7 + 3 targets of Table 3. All validation studies used chiral tetrahedral 4-PIP combinations with a bin size of 1.5 Å.

HTS Mode Ligity. When running Ligity against the whole of DUD-E, we used only a single known protein−ligand complex to construct our query Ligity descriptor and only the lowest-energy conformer of the full conformer ensemble of an active or decoy was used for screening. This is done for computational efficiency, to enable the large-scale evaluation against the full DUD-E database.

Ligity descriptor generation failed for three of the 102 DUD-E targets. For CAH2, its very small ligand (trifluoromethane sulfonamide in PDB entry 1BCD) only generated three PIPs, and thus, 4-PIP descriptors could not be generated for this query. RDKit was unable to read in the query molecules of AA2AR and PDE5A. Thus, 99 DUD-E targets were used in these validation experiments.

Performance was evaluated for the Tanimoto and Tversky (α = 1 and β = 0) scoring functions using ROC AUC and enrichment factors at 1%. Table 5 shows the performance of Ligity against the 99 Ligity-compatible DUD-E targets in HTS mode. The mean AUC for the ROC curves using the default Tversky index is 0.671 (slightly better than when using Tanimoto index, which has an AUC of 0.622). Ligity in HTS mode performed very well for some targets, with AUCs greater than 0.85 for COMT, ESR1, ESR2, FA7, FPPS, HIVPR, HMDH, MCR, NRAM, RXRA, WEE1, and XIAP. For early enrichment, as measured by the enrichment factor at 1% ($\mathrm{EF}_{1\%}$), we also found that Tversky performed better than Tanimoto, with an average $\mathrm{EF}_{1\%}$ of 9.048 versus 5.648,

DOI: 10.1021/acs.jcim.8b00779

respectively (Table 5). This indicates good early enrichment behavior using Ligity.

Information-Rich Mode Ligity. We would expect Ligity's virtual screening performance to improve with queries built from multiple cognate ligands. To test this, we built Ligity queries using the cognate ligands taken from the protein−ligand complexes in each target's sc-PDB binding site cluster. This represents the information-rich mode of Ligity, by capturing as much information from known actives as possible. We used data fusion on the results by selecting the maximum score across all separate ranking lists, calculating AUC and BEDROC values.

In Table 6 we show the mean AUC for the information-rich mode of Ligity across each individual query run (i.e., each single sc-PDB ligand−protein complex is a separate query) and the corresponding standard deviation (σ). The ROC AUC of the fusion score based on the ranked results of each individual query is also shown, along with the early enrichment BEDROC score for each individual query and for the fused approach. The BEDROC values were calculated using α = 20. Of the 10 receptors tested using Ligity in the information-rich mode, 9 exhibit early enrichment and only CDK2 shows random early enrichment.

Ligity's performance across all receptors had mean AUC values ranging from 0.6 to 0.9. Data fusion improved the AUC values by more than 0.05 in half the cases and was only marginally worse than the mean in just one case (HIVINT).

Comparison of Ligity to Existing Methods. In Table 7, we compare Ligity's performance to that of two other methods: another non-superpositional, 3D ligand similarity search method, ElectroShape, and the published performance of a protein−ligand docking method, DOCK, for the information-rich data set. For reference, as well as the individual query scores of these PDB structures for Ligity, we also give the data fusion score.

Table 7. ROC AUC Comparison of Methods for a Queries Using Ligity in "Information-Rich Mode"[a])

DUD-E target	PDB ID	ElectroShape	DOCK	Ligity	Ligity fused
ACE	3bkl	0.452	0.716	**0.749**	*0.948*
ADA	2e1w	0.714	0.764	**0.857**	*0.897*
CDK2	1h00	0.433	***0.791***	(0.610)	0.644
FA10	3kl6	0.664	***0.866***	0.716	0.717
FA7	1w7x	0.822	***0.879***	0.762	0.809
GCR	3bqd	0.521	0.439	**0.807**	*0.869*
HIVINT	3nf7	0.578	0.642	***0.717***	0.637
HIVPR	1xl2	0.495	0.596	**0.836**	*0.877*
THRB	1ype	0.646	***0.813***	0.709	0.752
TRY1	2ayw	0.320	***0.934***	(0.725)	0.778
mean AUC		0.565	0.744	**0.749**	*0.793*

[a]For those cases where the specific PDB ID was not present in the corresponding sc-PDB cluster—CDK2 and TRY1—we use all the sc-PDB query descriptors and report the mean AUC in parentheses. Entries with the best AUC among the methods that used only one structure for comparison—DOCK and Ligity—are highlighted in bold. The best AUC for each target is in italic. Ligity with a single active structure does better than all other methods for 4 out of the 10 DUD-E target classes, and 9 times out of 10 Ligity is better than the other non-superpositional method, ElectroShape. Note that when using more than one protein−ligand complex and fusing the results, with the exception of HIVINT, Ligity does even better ("Ligity fused") than when using only one complex ("Ligity").

The AUC values for DOCK in Table 7 were taken from Table S1 of the the Supporting Information reported by Mysinger et al.,[70] where DOCK 3.6 was tested using the same PDB structure (second column of Table 7) and the same actives and decoys as defined in DUD-E. We did not compare performance with that of 2D fingerprints (e.g., Morgan) because, as the authors of DUD-E state, Daylight[80] fingerprints were used to remove any decoys that were similar to actives, and they warned that this may create an artificially favorable enrichment bias for 2D fingerprinting methods.

Finally, it can be seen that Ligity does slightly better, on average, than the other 3D methods in Table 7. The results of Ligity with data-fusion had the highest mean AUC of 0.793, followed by Ligity without fusion at 0.749, closely followed by DOCK at 0.744, and then ElectroShape at 0.565. Ligity fused had slightly worse AUCs than DOCK for CDK2, FA10, FA7, THRB, and TRY1 but had significantly better AUCs for ACE, ADA, GCR, and HIVPR. For this data set, the highest ROC AUC is obtained by DOCK for five targets, Ligity with data fusion for four targets, and Ligity without data fusion for one target.

For CDK2, the query receptor for which Ligity does worst (1h00, AUC = 0.61) was not present in the CDK2 binding-site cluster in the sc-PDB release that we used. These sc-PDB clusters contain binding sites that are similar to one another (above a certain threshold) and only contain structures that pass a stringent quality filter. This implies that the 2003-deposited structure for 1h00 might not be similar enough to the sc-PDB binding sites we used as queries. For CDK2, instead we present the average across all the single CDK2 runs, shown in parentheses. A similar case applied for PDB entry 2ayw of TRY1.

ElectroShape tended to perform poorly for this data set, outperforming Ligity for only one target, FA7.

Computational Efficiency. To determine the computational efficiency of Ligity, we compared it to a popular protein−ligand docking method, AutoDock Vina[81] (refer to Table 8 for speed-up comparison). The benchmarks were carried out using 64-bit GNU/Linux Fedora release 28 running on an Intel Core i7-6700 CPU at 3.40 GHz. A subset of the ADA DUD-E benchmark was constructed using all 93 actives

Table 8. Ligity Is about 4−5 Orders of Magnitude Times Faster than Protein−Ligand Docking, Once Its Descriptors Have Been Precalculated for the Virtual Library Being Screened[a]

method	mode	CPU time (s)	relative speed-up
AutoDock Vina	flexible ligand	12586.8	1.0
Ligity descriptors + VS	information-rich	637.3	19.9
Ligity descriptors	information-rich	585.9	21.5
Ligity virtual screening	information-rich	51.4	244.9
Ligity virtual screening	HTS	1.9	6815.3

[a]This is exemplified by comparing with the already very efficient AutoDock Vina with the ADA target from DUD-E. The "information-rich" mode of Ligity, with multiple conformers per molecule, used a total of 3074 conformers for the 93 actives and 2287 conformers for 100 randomly selected decoys. The "HTS" mode of Ligity used just the lowest-energy conformer in each molecule's conformer ensemble for each active or decoy.

DOI: 10.1021/acs.jcim.8b00779
J. Chem. Inf. Model. 2019, 59, 2600−2616

and 100 randomly selected decoys. Ligity requires two steps to run: (i) precalculation of 4-PIP Ligity descriptors for every conformer of each molecule in the database and (ii) comparison of these descriptors with the 4-PIP Ligity descriptor of the query molecule, in this case, ligand FR233623, residue name FR6, in PDB entry 2e1w. Using our C++ implementation of Ligity with 3074 active conformers and 2287 decoy conformers, step i took 9 min 45.9 s, while step ii took 51.4 s. For the docking run, each active and each decoy from the ADA subset was converted into PDBQT-formatted files. The same compounds were flexibly docked using AutoDock Vina instructed to use a single core of the same machine: this took a total of 209 min 46.8 s. Ligity is therefore about 20 times faster than AutoDock Vina. If the Ligity 4-PIP descriptor precalculation is excluded—this only needs to be done once for the database compounds—Ligity is about 245 times faster than AutoDock Vina. In HTS mode, using only the lowest-energy conformation of the actives and decoys, Ligity is about 6800 times faster than docking.

■ CONCLUSIONS

Ligity is a fully automated, non-superpositional, pharmacophore-based method with comparable virtual screening performance to protein−ligand docking methods, while being much faster (some 2−3 orders of magnitude faster than AutoDock Vina). Ligity uses protein−ligand complex structures to build one or more query descriptors based on the geometric arrangement of the pharmacophoric features (PIPs) of the interacting parts of the active ligand(s). Using a subset of DUD-E, we investigated the parameter space of Ligity to maximize enrichment, including using 3-PIP versus 4-PIP pharmacophores, different bin sizes in the descriptor, different similarity measures, data fusion methods to aggregate rankings, and the effect of using the single lowest-energy conformer versus multiple conformers. Optimal results were found using 4-PIP descriptors, a bin size of 1.5 Å, Tversky similarity with $\alpha = 1$, $\beta = 0$, and consensus scoring using MAX−SIM data fusion over the results of the multiple bound ligand conformations. These are used as the default settings. Ligity gave slightly better VS performance with ensembles of conformers for the database molecules, but the improved speed when using a single lowest-energy conformer provides a very competitive high-throughput virtual screening "HTS mode". Ligity also performed better in terms of recovery of known actives when multiple ligands bound to the same protein were combined into an "information-rich" Ligity descriptor. It would be interesting to study how sensitive Ligity's results are to the particular choice of protein−ligand structures. This would ideally require a combinatorial examination of all possible sets of protein−ligand binding cavities, and we believe this would be an interesting follow-up study.

Ligity could be used in conjunction with explicit structure-based virtual screening methods, for example by using docked poses of known actives in either apoprotein X-ray structres or homology models. We suspect that, in general, however, the additional structural noise introduced both by subtle variations in the homology models of the proteins and in the docked binding modes of the ligand might reduce the predictive performance of our method.

The novelty of Ligity lies in its descriptor: a three-dimensional array descriptor for all possible triangular 3-PIP feature sets and a six-dimensional hypercube descriptor for all possible chiral tetrahedral 4-PIP combinations. We found the 4-PIP-based Ligity descriptor to perform better than the 3-PIP variant. Furthermore, receptor flexibility can be incorporated by using alternative conformations of the holo protein−ligand complex where there are multiple experimentally resolved structures or from molecular dynamics simulations. Ligand flexibility is captured by ensembles of low-energy conformers. Ligity is also able to distinguish chiral arrangements of pharmacophoric features using our 4-PIP tetrahedral descriptors. The non-superpositional nature of Ligity's descriptors makes it extremely efficient and particularly well-suited for machine-learning applications.

We found that Ligity preferentially picked low-energy conformers for active molecules as the highest scoring conformers on the targets we used to refine the parameters of the method (CDK2, TRY1, and ADA). We showed that this preference is statistically significant for actives but not for decoys. We postulate that, although this is still an open question in the literature, active compounds bind in a relatively low-energy conformation. This finding is consistent with the process of drug discovery. The structure of the active compounds in the DUD-E data sets will often be the result of medicinal chemistry optimization, where optimizing binding affinity is an important goal. One way to achieve this, while low molecular weight (and thus rule-of-5 compliance) is maintained, is to minimize any energy penalty for obtaining the bound conformation. Artificial decoy molecules have undergone no such optimization and, thus, would not be expected to show this effect. This further suggests that this selection effect will only be observed in retrospective virtual screening experiments (using benchmark sets consisting of true positives and putative negatives) but not in prospective ones.

On the basis of the retrospective analysis presented here, we believe that Ligity has potential as a prospective virtual screening method that is able to efficiently and successfully screen databases consisting of millions of molecules.

■ ASSOCIATED CONTENT

S Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00779.

Further Ligity experiments, as well as implementation details (Schema S1, Figures S1−S4, and Tables S1−S3) (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: garrett.morris@stats.ox.ac.uk. Phone: +44 1865 281770. Fax: +44 1865 282862.

ORCID

Jean-Paul Ebejer: 0000-0003-0888-2637

Wing Ki Wong: 0000-0003-4029-6902

Charlotte M. Deane: 0000-0003-1388-2252

Garrett M. Morris: 0000-0003-1731-8405

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

J.P.E. thanks the European Union Framework Programme 7-funded Marie Curie Initial Training Network STARS, grant agreement PITN-GA-2009-238490. C.D., G.M.M., and

W.K.W. thank the EPSRC and MRC for financial support under grant number EP/L016044/1 and InhibOx Ltd. (now Oxford Drug Design Ltd.) for supporting this research.

■ ABBREVIATIONS

3D, three-dimensional; 4D, four-dimensional; 6D, six-dimensional; AA2AR, adenosine A2a receptor; ABL1, tyrosine-protein kinase ABL; ACE, angiotensin-converting enzyme; ACES, acetylcholinesterase; ADA, adenosine deaminase; ADA17, ADAM17; ADRB1, beta-1 adrenergic receptor; ADRB2, beta-2 adrenergic receptor; AKT1, Serine/threonine-protein kinase AKT; AKT2, serine/threonine-protein kinase AKT2; ALDR, aldose reductase; AMPC, beta-lactamase; ANDR, androgen receptor; AOFB, monoamine oxidase B; AUC, area under the receiver operating characteristic curve; BACE1, beta-secretase 1; BEDROC, Boltzmann-enhanced discrimination of receiver operating characteristic; BRAF, serine/threonine-protein kinase B-raf; CAH2, carbonic anhydrase II; CASP3, caspase-3; CDK2, cyclin-dependent kinase 2; COMT, catechol *O*-methyltransferase; CP2C9, cytochrome P450 2C9; CP3A4, cytochrome P450 3A4; CSF1R, macrophage colony stimulating factor receptor; CXCR4, C-X-C chemokine receptor type 4; DEF, peptide deformylase; DHI1, 11-beta-hydroxysteroid dehydrogenase 1; DPP4, dipeptidyl peptidase IV; DRD3, dopamine D3 receptor; DUD-E, directory of useful decoys enhanced; DYR, dihydrofolate reductase; EF, enrichment factor; EGFR, epidermal growth factor receptor erbB1; ESR1, estrogen receptor alpha; ESR2, estrogen receptor beta; FA10, coagulation factor X; FA7, coagulation factor VII; FABP4, fatty acid binding protein adipocyte; FAK1, focal adhesion kinase 1; FGFR1, fibroblast growth factor receptor 1; FKB1A, FK506-binding protein 1A; FNTA, protein farnesyltransferase/geranylgeranyltransferase type I alpha subunit; FPPS, farnesyl diphosphate synthase; GCR, glucocorticoid receptor; GLCM, beta-glucocerebrosidase; GRIA2, glutamate receptor ionotropic, AMPA 2; GRIK1, glutamate receptor ionotropic kainate 1; HDAC2, histone deacetylase 2; HDAC8, Histone deacetylase 8; HIVINT, human immunodeficiency virus type 1 integrase; HIVPR, human immunodeficiency virus type 1 protease; HIVRT, human immunodeficiency virus type 1 reverse transcriptase; HMDH, HMG-CoA reductase; HS90A, heat shock protein HSP 90-alpha; HXK4, hexokinase type IV; ID, identifier; IGF1R, insulin-like growth factor I receptor; INHA, enoyl-[acyl-carrier-protein] reductase; ITAL, leukocyte adhesion glycoprotein LFA-1 alpha; JAK2, tyrosine-protein kinase JAK2; KIF11, kinesin-like protein 1; KIT, stem cell growth factor receptor; KITH, thymidine kinase; KPCB, protein kinase C beta; LBVS, ligand-based virtual screening; LCK, tyrosine-protein kinase LCK; LEC, lowest-energy conformer; LKHA4, Leukotriene A4 hydrolase; MAPK2, MAP kinase-activated protein kinase 2; MAX-SIM, maximum similarity (data fusion); MCR, Mineralocorticoid receptor; MET, hepatocyte growth factor receptor; MK01, MAP kinase ERK2; MK10, c-Jun N-terminal kinase 3; MK14, MAP kinase p38 alpha; MMP13, matrix metalloproteinase 13; MP2K1, dual specificity mitogen-activated protein kinase kinase 1; NOS1, nitric-oxide synthase, brain; NRAM, neuraminidase; PA2GA, phospholipase A2 group IIA; PARP1, poly[ADP-ribose] polymerase-1; PDB, Protein Data Bank; PDE5A, phosphodiesterase 5A; PGH1, cyclooxygenase-1; PGH2, cyclooxygenase-2; PIP, pharmacophoric Interaction Point; PLK1, serine/threonine-protein kinase PLK1; PNPH, purine nucleoside phosphorylase; PPARA, peroxisome proliferator-activated receptor alpha; PPARD, peroxisome proliferator-activated receptor delta; PPARG, peroxisome proliferator-activated receptor gamma; PRGR, progesterone receptor; PTN1, protein-tyrosine phosphatase 1B; PUR2, GAR transformylase; PYGM, muscle glycogen phosphorylase; PYRD, dihydroorotate dehydrogenase; RENI, renin; RMSD, root-mean-square deviation; ROC, receiver operating characteristic; ROCK1, Rho-associated protein kinase 1; RXRA, retinoid X receptor alpha; SAHH, adenosylhomocysteinase; SBVS, structure-based virtual screening; SMARTS, smiles arbitrary target specification; SMILES, simplified molecular-input line-entry system; SRC, tyrosine-protein kinase SRC; TGFR1, TGF-beta receptor type I; THB, thyroid hormone receptor beta-1; THRB, thrombin; TRY1, trypsin I; TRYB1, tryptase beta-1; TYSY, thymidylate synthase; UROK, urokinase-type plasminogen activator; VGFR2, vascular endothelial growth factor receptor 2; VS, virtual screening; WEE1, serine/threonine-protein kinase WEE1; XIAP, inhibitor of apoptosis protein 3

■ REFERENCES

(1) Johnson, M.; Maggiora, G. *Concepts and Applications of Molecular Similarity*; Wiley-Interscience Publication, 1990.

(2) Ripphausen, P.; Nisius, B.; Bajorath, J. State-of-the-Art in Ligand-Based Virtual Screening. *Drug Discovery Today* **2011**, *16*, 372−376.

(3) Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto Index an Appropriate Choice for Fingerprint-based Similarity Calculations? *J. Cheminf.* **2015**, *7*, 20.

(4) Yan, X.; Liao, C.; Liu, Z.; Hagler, A. T.; Gu, Q.; Xu, J. Chemical Structure Similarity Search for Ligand-based Virtual Screening: Methods and Computational Resources. *Curr. Drug Targets* **2016**, *17*, 1580−1585.

(5) Willett, P. Similarity-Based Virtual Screening Using 2D Fingerprints. *Drug Discovery Today* **2006**, *11*, 1046−1053.

(6) Nikolova, N.; Jaworska, J. Approaches to Measure Chemical Similarity - a Review. *QSAR Comb. Sci.* **2003**, *22*, 1006−1026.

(7) Geppert, H.; Vogt, M.; Bajorath, J. Current Trends in Ligand-Based Virtual Screening: Molecular Representations, Data Mining Methods, New Application Areas, and Performance Evaluation. *J. Chem. Inf. Model.* **2010**, *50*, 205−216.

(8) Kristensen, T. G.; Nielsen, J.; Pedersen, C. N. Methods for Similarity-Based Virtual Screening. *Comput. Struct. Biotechnol. J.* **2013**, *5*, e201302009.

(9) Maggiora, G. M.; Shanmugasundaram, V. In *Chemoinformatics and Computational Chemical Biology*; Bajorath, J., Ed.; Humana Press, 2011; pp 39−100.

(10) Holliday, J. D.; Kanoulas, E.; Malim, N.; Willett, P. Multiple Search Methods for Similarity-Based Virtual Screening: Analysis of Search Overlap and Precision. *J. Cheminf.* **2011**, *3*, 29.

(11) Willett, P. Fusing Similarity Rankings in Ligand-Based Virtual Screening. *Comput. Struct. Biotechnol. J.* **2013**, *5*, e201302002.

(12) Rush, T. S., 3rd; Grant, J. A.; Mosyak, L.; Nicholls, A. A Shape-Based 3-D Scaffold Hopping Method and its Application to a Bacterial Protein-Protein Interaction. *J. Med. Chem.* **2005**, *48*, 1489−1495.

(13) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50*, 74−82.

(14) Lyne, P. D. Structure-Based Virtual Screening: An Overview. *Drug Discovery Today* **2002**, *7*, 1047−1055.

(15) Lionta, E.; Spyrou, G.; Vassilatis, D. K.; Cournia, Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr. Top. Med. Chem.* **2014**, *14*, 1923−1938.

(16) Li, Q.; Shah, S. In *Protein Bioinformatics: From Protein Modifications and Networks to Proteomics*; Wu, C. H., Arighi, C. N., Ross, K. E., Eds.; Springer: New York, 2017; pp 111−124.

(17) 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.

(18) Wojcikowski, M.; Ballester, P. J.; Siedlecki, P. Performance of Machine-Learning Scoring Functions in Structure-Based Virtual Screening. *Sci. Rep.* **2017**, *7*, 46710.

(19) Ripphausen, P.; Stumpfe, D.; Bajorath, J. Analysis of Structure-Based Virtual Screening Studies and Characterization of Identified Active Compounds. *Future Med. Chem.* **2012**, *4*, 603−613.

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

(21) Fan, H.; Irwin, J. J.; Webb, B. M.; Klebe, G.; Shoichet, B. K.; Sali, A. Molecular Docking Screens Using Comparative Models of Proteins. *J. Chem. Inf. Model.* **2009**, *49*, 2512−2527.

(22) Dassault Systémes BIOVIA. *Discovery Studio Modeling Environment*, Release 2019. https://3dsbiovia.com/products/collaborative-science/biovia-discovery-studio/ (accessed January 2, 2019).

(23) Wolber, G.; Langer, T. LigandScout: 3-D Pharmacophores Derived from Protein-Bound Ligands and Their Use as Virtual Screening Filters. *J. Chem. Inf. Model.* **2005**, *45*, 160−169.

(24) Hu, J.; Liu, Z.; Yu, D. J.; Zhang, Y. LS-Align: An Atom-Level, Flexible Ligand Structural Alignment Algorithm for High-Throughput Virtual Screening. *Bioinformatics* **2018**, *34*, 2209−2218.

(25) Dixon, S. L.; Smondyrev, A. M.; Rao, S. N. PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching. *Chem. Biol. Drug Des.* **2006**, *67*, 370−2.

(26) Dixon, S. L.; Smondyrev, A. M.; Rao, S. N. PHASE: ANovel Approach to Pharmacophore Modeling and 3D Database Searching. *Chem. Biol. Drug Des.* **2006**, *67*, 370−2.

(27) Chemical Computing Group. *MOE 2019.01*. https://www.chemcomp.com/index.htm (accessed January 2, 2019).

(28) Liu, X.; Jiang, H.; Li, H. SHAFTS: A Hybrid Approach for 3D Molecular Similarity Calculation. 1. Method and Assessment of Virtual Screening. *J. Chem. Inf. Model.* **2011**, *51*, 2372−85.

(29) Yan, X.; Li, J.; Liu, Z.; Zheng, M.; Ge, H.; Xu, J. Enhancing Molecular Shape Comparison by Weighted Gaussian Functions. *J. Chem. Inf. Model.* **2013**, *53*, 1967−78.

(30) Tanrikulu, Y.; Krüger, B.; Proschak, E. The Holistic Integration of Virtual Screening in Drug Discovery. *Drug Discovery Today* **2013**, *18*, 358−364.

(31) Drwal, M. N.; Griffith, R. Combination of Ligand- and Structure-Based Methods in Virtual Screening. *Drug Discovery Today: Technol.* **2013**, *10*, e395−e401.

(32) Muddassar, M.; Jang, J. W.; Gon, H. S.; Cho, Y. S.; Kim, E. E.; Keum, K. C.; Oh, T.; Cho, S.-N.; Pae, A. N. Identification of Novel Antitubercular Compounds Through Hybrid Virtual Screening Approach. *Bioorg. Med. Chem.* **2010**, *18*, 6914−6921.

(33) Sastry, G. M.; Inakollu, V. S. S.; Sherman, W. Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking. *J. Chem. Inf. Model.* **2013**, *53*, 1531−1542.

(34) 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*, 1078−1087.

(35) Aldeghi, M.; Heifetz, A.; Bodkin, M. J.; Knapp, S.; Biggin, P. C. Accurate Calculation of the Absolute Free Energy of Binding for Drug Molecules. *Chem. Sci.* **2016**, *7*, 207−218.

(36) Weill, N.; Rognan, D. Alignment-Free Ultra-High-Throughput Comparison of Druggable Protein-Ligand Binding Sites. *J. Chem. Inf. Model.* **2010**, *50*, 123−35.

(37) Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J. S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands and Proteins (FLAP): Theory and Application. *J. Chem. Inf. Model.* **2007**, *47*, 279−294.

(38) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(39) Wade, R. C.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 2. Ligand Probe Groups with the Ability to Form More Than Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36*, 148−156.

(40) Wade, R. C.; Clark, K. J.; Goodford, P. J. Further Development of Hydrogen Bond Functions for Use in Determining Energetically Favorable Binding Sites on Molecules of Known Structure. 1. Ligand Probe Groups with The Ability to Form Two Hydrogen Bonds. *J. Med. Chem.* **1993**, *36*, 140−147.

(41) Carosati, E.; Sciabola, S.; Cruciani, G. Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field. *J. Med. Chem.* **2004**, *47*, 5114−5125.

(42) van Rossum, G.; de Boer, J. Linking a Stub Generator (AIL) to a Prototyping Language (Python). *Spring 1991 EurOpen Conference Proceedings*; 1991; pp 229−247.

(43) van Rossum, G. *Python Tutorial*, Technical Report CS-R9526; Centrum voor Wiskunde en Informatica (CWI): Amsterdam. 1995.

(44) Landrum, G. *RDKit: Open-Source Cheminformatics*. http://www.rdkit.org (accessed January 2, 2019).

(45) Daylight Chemical Information Systems, Inc. *SMARTS—A Language for Describing Molecular Patterns*; 2018. http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html (accessed October 26, 2018).

(46) Weininger, D. SMILES, a Chemical Language and Information-System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Model.* **1988**, *28*, 31−36.

(47) Weininger, D.; Weininger, A.; Weininger, J. L. SMILES. 2. Algorithm for Generation of Unique SMILES Notation. *J. Chem. Inf. Model.* **1989**, *29*, 97−101.

(48) Weininger, D. SMILES. 3. Depict. Graphical Depiction of Chemical Structures. *J. Chem. Inf. Model.* **1990**, *30*, 237−243.

(49) Ebejer, J.; Morris, G.; Deane, C. Freely Available Conformer Generation Methods: How Good Are They? *J. Chem. Inf. Model.* **2012**, *52*, 1146−1158.

(50) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. Uff, a Full Periodic-Table Force-Field for Molecular Mechanics and Molecular-Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024−10035.

(51) Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D. sc-PDB: an Annotated Database of Druggable Binding Sites from the Protein Data Bank. *J. Chem. Inf. Model.* **2006**, *46*, 717−727.

(52) Meslamani, J.; Rognan, D.; Kellenberger, E. sc-PDB: A Database for Identifying Variations and Multiplicity of 'Druggable' Binding Sites in Proteins. *Bioinformatics* **2011**, *27*, 1324−1326.

(53) Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E. sc-PDB: a 3D-database of Ligandable Binding Sites-10 Years On. *Nucleic Acids Res.* **2015**, *43*, D399−D404.

(54) Meslamani, J.; Rognan, D.; Kellenberger, E. sc-PDB: A Database for Identifying Variations and Multiplicity of 'Druggable' Binding Sites in Proteins. *Bioinformatics* **2011**, *27*, 1324−1326.

(55) Schalon, C.; Surgand, J.-S.; Kellenberger, E.; Rognan, D. A Simple and Fuzzy Method to Align and Compare Druggable Ligand-Binding Sites. *Proteins: Struct., Funct., Genet.* **2008**, *71*, 1755−1778.

(56) Ebejer, J. P. Data Driven Approaches to Improve the Drug Discovery Process: A Virtual Screening Quest in Drug Discovery. Ph.D. thesis, Oxford University, 2014.

(57) Langer, T.; Wolber, G. Pharmacophore Definition and 3D Searches. *Drug Discovery Today: Technol.* **2004**, *1*, 203−207.

(58) Marcou, G.; Rognan, D. Optimizing Fragment and Scaffold Docking by Use of Molecular Interaction Fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195−207.

(59) Schreyer, A.; Blundell, T. CREDO: A Protein-Ligand Interaction Database for Drug Discovery. *Chem. Biol. Drug Des.* **2009**, *73*, 157−167.

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

(61) Jeffrey, G. *An Introduction to Hydrogen Bonding*; Topics in Physical Chemistry Series; Oxford University Press, 1997.

(62) Meyer, E. A.; Castellano, R. K.; Diederich, F. Interactions with Aromatic Rings in Chemical and Biological Recognition. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210−1250.

(63) Jaccard, P. The Distribution of the Flora in the Alpine Zone. *New Phytol.* **1912**, *11*, 37−50.

(64) Nasr, R.; Swamidass, S. J.; Baldi, P. Large Scale Study of Multiple-Molecule Queries. *J. Cheminf.* **2009**, *1*, 7.

(65) R Core Team. *R: A Language and Environment for Statistical Computing*, R version 3.0.0; R Foundation for Statistical Computing: Vienna, Austria, 2013.

(66) Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. ROCR: Visualizing Classifier Performance in R. *Bioinformatics* **2005**, *21*, 3940−3941.

(67) Sing, T.; Sander, O.; Beerenwinkel, N.; Lengauer, T. *ROCR: Visualizing the Performance of Scoring Classifiers*, R package version 1.0-4; 2012.

(68) Yabuuchi, H. *enrichvs: Enrichment Assessment of Virtual Screening Approaches*, R package version 0.0.5; 2011.

(69) Truchon, J.-F.; Bayly, C. I. Evaluating Virtual Screening Methods: Good and Bad Metrics for the "Early Recognition" Problem. *J. Chem. Inf. Model.* **2007**, *47*, 488−508.

(70) Mysinger, M. M.; Carchia, M.; Irwin, J. J.; Shoichet, B. K. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* **2012**, *55*, 6582−6594.

(71) Armstrong, S. M.; Morris, G. M.; Finn, P. W.; Sharma, R.; Moretti, L.; Cooper, R. I.; Richards, W. G. ElectroShape: Fast Molecular Similarity Calculations Incorporating Shape, Chirality and Electrostatics. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 789−801.

(72) Butler, K. T.; Luque, F. J.; Barril, X. Toward Accurate Relative Energy Predictions of the Bioactive Conformation of Drugs. *J. Comput. Chem.* **2009**, *30*, 601−610.

(73) Nicklaus, M. C.; Wang, S.; Driscoll, J. S.; Milne, G. W. Conformational Changes of Small Molecules Binding to Proteins. *Bioorg. Med. Chem.* **1995**, *3*, 411−428.

(74) Boström, J.; Norrby, P. O.; Liljefors, T. Conformational Energy Penalties of Protein-Bound Ligands. *J. Comput.-Aided Mol. Des.* **1998**, *12*, 383−396.

(75) 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.

(76) Sitzmann, M.; Weidlich, I. E.; Filippov, I. V.; Liao, C.; Peach, M. L.; Ihlenfeldt, W.-D.; Karki, R. G.; Borodina, Y. V.; Cachau, R. E.; Nicklaus, M. C. PDB Ligand Conformational Energies Calculated Quantum-Mechanically. *J. Chem. Inf. Model.* **2012**, *52*, 739−756.

(77) He, M. W.; Lee, P. S.; Sweeney, Z. K. Promiscuity and the Conformational Rearrangement of Drug-like Molecules: Insight from the Protein Data Bank. *ChemMedChem* **2015**, *10*, 238−244.

(78) Renner, S.; Schwab, C. H.; Gasteiger, J.; Schneider, G. Impact of Conformational Flexibility on Three-Dimensional Similarity Searching U Correlation Vectors. *J. Chem. Inf. Model.* **2006**, *46*, 2324−2332.

(79) Zhang, Q.; Muegge, I. Scaffold Hopping through Virtual Screening Using 2D and 3D Similarity Descriptors: Ranking, Voting, and Consensus Scoring. *J. Med. Chem.* **2006**, *49*, 1536−1548.

(80) *Daylight Theory Manual*, version 4.9; 2011. http://www.daylight.com/dayhtml/doc/theory/ (accessed October 18, 2017).

(81) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with A New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2009**, *31*, 455−461.