JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

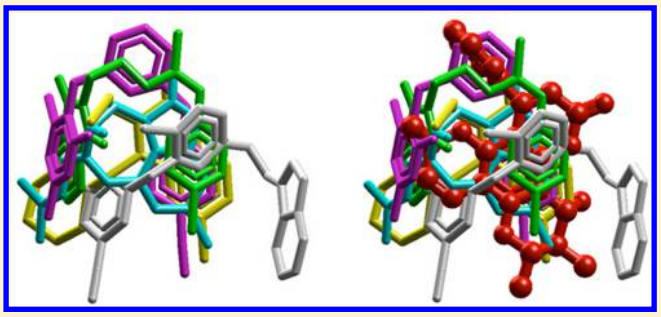
pubs.acs.org/jcim

VS-APPLE: A Virtual Screening Algorithm Using Promiscuous Protein−Ligand Complexes

Tatsuya Okuno,[†,∥,§] Koya Kato,[‡,§] Tomoki P. Terada,[‡] Masaki Sasai,[‡] and George Chikenji*[,‡]

[†]Department of Applied Physics and [‡]Department of Computational Science and Engineering, Nagoya University, Nagoya, Aichi 464-8603, Japan

ABSTRACT: As the number of structurally resolved protein−ligand complexes increases, the ligand-binding pockets of many proteins have been found to accommodate multiple different compounds. Effective use of these structural data is important for developing virtual screening (VS) methods that identify bioactive compounds. Here, we introduce a VS method, VS-APPLE (Virtual Screening Algorithm using Promiscuous Protein−Ligand complExes), based on promiscuous protein−ligand binding structures. In VS-APPLE, multiple ligands bound to a pocket are combined into a query template for screening. Both the structural match between a test compound and the multiple-ligand template and the possible collisions between the test compound and the target protein are evaluated by an efficient geometric hashing method. The performance of VS-APPLE was examined on a filtered, clustered version of the Directory of Useful Decoys data set. In Area Under the Curve analyses of this data set, VS-APPLE outperformed several popular screening programs. Judging from the performance of VS-APPLE, the structural data of promiscuous protein−ligand bindings could be further analyzed and exploited for developing VS methods.



INTRODUCTION

Quite commonly, a single protein promiscuously binds various natural or artificial compounds.[1,2] Our understanding of protein promiscuity has considerably advanced as more crystal structures of protein−ligand complexes have become available.[2−5] Gao and Skolnick[2] compared the structures of a nonredundant set of approximately 20,000 ligand-binding protein pockets and showed that many pockets promiscuously bind ligands with diverse scaffolds. Indeed, the Protein Data Bank (PDB) contains the structures of more than 75 complexes of ligand-binding HIV reverse transcriptase and 260 complexes of ligand-binding HIV protease.[5] Figure 1 illustrates the promiscuous binding of HIV reverse transcriptase to several different ligands. Other compounds bound to this protein should share common features with these distributed ligands, namely, their structures, chemical properties, and binding positions on the protein surface. This example suggests that, by investigating the data of distributed multiple ligands bound to the same protein pocket, we could search for new active compounds. Therefore, a deeper analysis of protein promiscuity may lead to new virtual screening (VS) methods for drug discovery.

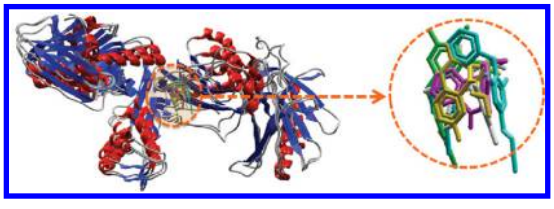


Figure 1. Promiscuous binding of HIV reverse transcriptase to multiple ligands. Structures of five protein−ligand complexes (PDB ID: 1c1b, 1fkp, 1sv5, 2opp, 2zd1) are superposed by maximizing the overlap among the protein structures. Ligands superposed in this manner are enlarged in the right circle.

In conventional ligand-based VS methods, a single template ligand is selected for each target protein as a query for searching new active compounds. However, as increasing numbers of ligands are reported to bind to promiscuous target proteins, the effective use of multiple ligands has attracted significant interest in VS research.[6−10] In particular, multiple compounds bound to a single target can be superposed by maximizing their mutual structural overlap, thereby revealing common structural features of the ligands. The consensus molecular-shape pattern of ligands found by this technique has already proven effective for screening candidate compounds.[8−10] However, a fundamental issue remains: because a promiscuous protein can bind different ligands at its different sites, how these different ligands should be superposed to build an appropriate query shape pattern remains unclear. To resolve this difficulty, we could superpose only the ligands with similar structures and therefore similar binding mode;[8] such mutually similar templates can be chosen by structurally clustering the ligands. However, if the templates are obtained by clustering, novel active compounds with scaffolds different from the templates are unlikely to be found. Therefore, when seeking active compounds with novel

Received: March 11, 2015
Published: June 9, 2015

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

scaffolds, we require further analyses on protein−ligand interactions.

Gao and Skolnick[2] analyzed a large number of pocket-ligand interactions by examining the atomic contacts between hydrophilic, hydrophobic, and aromatic groups. They reported that approximately 60% of the interactions are shared among multiple ligands that bind to the same or similar promiscuous pockets. This observation implies that pocket-ligand interactions are neither extremely specific at the atomic level (i.e., overlapping by almost 100%) nor random with very low overlap. Although different ligands interact with different surfaces of the pocket, these surfaces considerably overlap and the pocket flexibly changes its atomic arrangement, especially at the residue side chains, to accommodate different ligands. Consequently, the interactions among different ligands bound to a pocket overlap to an intermediate degree, enabling us to identify novel active compounds.

This paper proposes a new VS method called VS-APPLE (Virtual Screening Algorithm using Promiscuous Protein−Ligand complExes). In VS-APPLE, protein−ligand complexes are superposed on the target protein by maximizing the structural overlap between each protein in the complex and the target protein. The ensemble of ligands superposed in this manner should reflect the flexible pocket environment, providing a template for compound screening. During this screening, we check whether the candidate compounds avoid collision against the pocket. In this manner, the target protein structure serves two purposes, i.e., building the multiple-ligand template and checking collisions. VS-APPLE does not compute the individual atomic interactions between the test compounds and the pocket; therefore, its computational speed is comparable to that of other ligand-based VS methods, facilitating the screening of large compound databases. As shown later, VS-APPLE outperforms several ligand-based methods, even when identifying active compounds that are dissimilar from the template ligands, suggesting its utility in scaffold hopping. In the example in Figure 2, five ligands bound to the same pocket form a multiple-ligand template, but the selected compound is dissimilar to any of the ligands included in this template.

In this paper, we introduce VS-APPLE as a new VS method using a protein-based multiple-ligand template. The Methods section defines the protein-based multiple-ligand template and the score function used in VS-APPLE. The Results and Discussion section compares the performances of VS-APPLE and other widely used ligand-based VS methods (ROCS[15,16] and BABEL[17]) and structure-based or docking-simulation-based VS methods (Glide,[18−20] DOCK,[21] and GOLD[22]). The performances are evaluated on a filtered and clustered version[11] of the Directory of Useful Decoys (DUD)[13] data set. The comparison is quantified by the Receiver Operator Characteristic (ROC) curves.[26] The quality of each ROC curve is characterized by the area under the curve (AUC).[27,30] The scaffold-hopping ability of VS-APPLE is examined by imposing small similarity between the template ligands and the actives to be distinguished in the test data set. To analyze the effects of the multiple-ligand template, we replace the multi- with a single-ligand template. In addition, we analyze the effects of the protein-based spatial arrangement of multiple ligands in the template. To this end, we compare VS-APPLE with the multiple ligands-based method, in which ligands are superposed to maximize their mutual structural overlap. We show that the consideration of the effects of collision of the test compound against the target protein is effective particularly in the early phase of screening for excluding compounds whose structures are inconsistent with the pocket surface shape. The last section is devoted to conclusions.

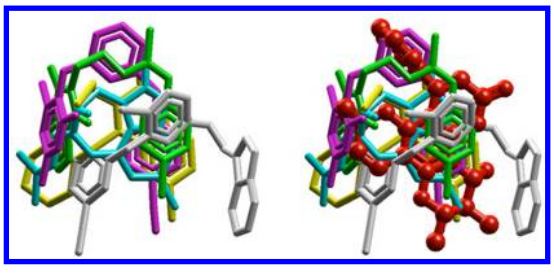


Figure 2. Selection of an active compound with different scaffold from the ligands comprising the multiple-ligand template. In this example, the multiple-ligand template is the superposition of five ligands bound to the promiscuous pocket of HIV reverse transcriptase. Structures of the template ligands are highlighted in different colors (left). The compound selected in the screening process (red) is dissimilar to any of the ligands used for screening (right).

METHODS

In this section, we first construct the multiple-ligand template Q^{multi} of VS-APPLE. Using this template, we define the score $S(k, P^t)$ that estimates the likeliness of the compound k being active to the target protein P^t. In addition, we explain the data set used for benchmarking VS-APPLE.

Multiple-Ligand Template. For each target protein structure P^t, we built a multiple-ligand template Q^{multi}. Q^{multi} should include enough number of diverse ligands to precisely represent the pocket environment. However, because the structural comparison between Q^{multi} and the test compound is relatively time-consuming, excessive number of ligands in Q^{multi} will degrade the efficiency of the calculation. Here, we restricted the number of ligands to 10 by clustering them. Thus, the template Q^{multi} unifies the structures of 10 different ligands, denoted as L_1^*, L_2^*, ..., L_{10}^*, where the superscript * indicates a representative of a particular ligand cluster. The procedure for building Q^{multi} is schematically explained in Figure 3.

In the first step of template building, the PDB is searched for structures of the target protein and structurally similar proteins, which are collected into a library of protein−ligand complexes. This search was performed by the structure comparison method MICAN,[31] which compares two protein structures by the spatial arrangements of their secondary structure elements. Although MICAN can identify structure similarity by both nonsequential and sequential alignment procedures,[31,32] here we adopt the sequential alignment mode only. Structural similarity was evaluated by the popularly used TM-score,[33] by which homologous structures are collected if they satisfy the threshold similarity to the target structure (TM-score $\geq$0.6).[34]

The resulting structure-data files of protein−ligand complexes $\{C\}$ contain the coordinates of ions and water molecules; these were deleted such that each C_i in $\{C\}$ comprised a protein P_i bound to a ligand L_i. At this stage, structure-data files containing no ligands were also eliminated. Next, it was decided whether the proteins in $\{C\}$ bind ligands in the same pocket. For this purpose, we defined a reference complex C^t comprising the target P^t and the reference ligand L^t. If all the atoms of L_i and L^t were separated by more than 2 Å, L_i was assumed to bind to a different pocket from L^t, and C_i was eliminated from

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

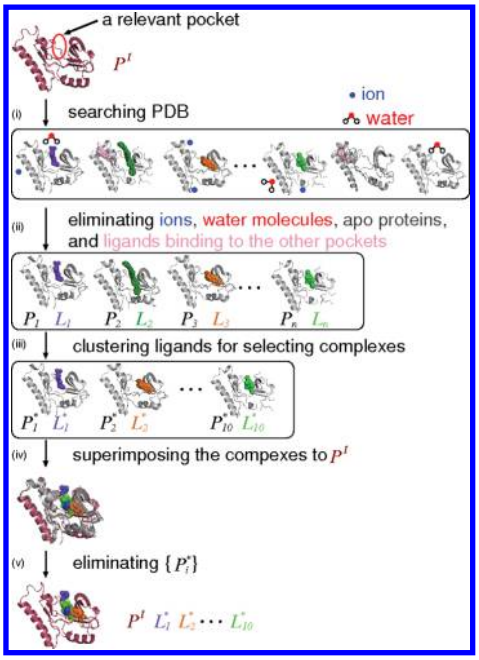


Figure 3. Building a multiple-ligand template Q^{multi} for the target protein P^t. (i) First, the Protein Data Bank is searched for the structures of P^t and its homologues. (ii) Ions and water molecules are eliminated from the structure data files of the obtained structures. Structures that bind ligands on different pockets as well as apo proteins (which do not bind any ligand) are also eliminated. (iii) Ligand clustering is based on the Tanimoto coefficient for the 2D-fingerprint measure, and the 10 largest clusters are selected. From these 10 clusters, we obtain 10 representative ligands L_1^*, L_2^*, ..., L_{10}^* and their corresponding proteins, P_1^*, P_2^*, ..., P_{10}^*. (iv) To establish the spatial position of L_i^*, the complex $C_i^* = L_i^* + P_i^*$ is superposed on P^t by maximizing the structural overlap between P_i^* and P^t. (v) Eliminating P_i^* from the structure data, we obtain the spatially arranged P^t and multiple ligands L_1^*, L_2^*, ..., L_{10}^*. Q^{multi} is stored as an array of L_1^*, L_2^*, ..., L_{10}^*, which are spatially arranged by maintaining their positions relative to P^t.

$\{C\}$. Here, C^t should be selected from high-resolution structures; here, we used the complexes chosen by Huang et al.[13] and Kinnings and Jackson[7] (see Table 1 in the subsection *Active and Decoy library for the Benchmark Test*). For each i, the structures of C_i and C^t were superposed by maximizing the structural overlap of P_i and P^t using the structure alignment measure of MICAN. The L_i and L^t in this superposed complex were then compared, and the complex C_i was eliminated from $\{C\}$ if the least atomic distance between L_i and L^t exceeded 2 Å. In this manner, we obtained an ensemble of structures of protein−ligand complexes $\{C\}$ for a given target protein. In the benchmark test of the present paper, a large ensemble of structures were gathered for each target protein (typically exceeding 100; see Table 2 in the subsection *Active and Decoy library for the Benchmark Test*).

Structures of the multiple-ligand template and a test compound were compared by a variant of the geometric hashing method. Because structure comparison by geometric hashing is relatively time-consuming, the number of comparisons was reduced by clustering the ligands. In the clustering process, the features of ligand L_i were evaluated by a MACCS key 2D-fingerprint obtained by open BABEL software,[17] and the similarity among the features was measured by the Tanimoto coefficient. From this similarity measure, we performed a hierarchical clustering. Clusters of ligands were

Table 1. Dataset Used for Benchmarking VS-APPLE

target protein (abbrev)	reference complex[a] (ligand[b])	# actives/ # decoys	average # of conformers[c]
angiotensin converting enzyme (ace)	1o86 (LPR)	46/1797	69
acetylcholinesterase (ache)	1eve (E20)	100/3892	46
cyclin-dependent kinase 2 (cdk2)	1ckp (PVB)	47/2074	36
cyclooxygenase 2 (cox2)	1cx2 (558)	212/13289	33
epidermal growth factor receptor (egfr)	1m17 (AQ4)	365/15996	42
factor Xa (fxa)	1f0r (815)	64/5745	80
HIV reverse transcriptase (hivrt)	1rt1 (MKC)	34/1519	29
enoyl ACP reductase InhA (inha)	1p44 (GEQ)	57/3266	35
p38 mitogen activated protein (p38)	1kv2 (B96)	137/9141	26
phosphodiesterase (pde5)	1xp0 (VDN)	26/1978	66
platelet derived growth factor receptor kinase (pdgfrb)	1t46 (STI)	124/5980	90
tyrosine kinase Src (src)	2src (ANP)	98/6319	69
vascular endothelial growth factor receptor (vegfr2)	1fgi (SU1)	48/2906	57

[a]PDB code. [b]Ligand code. [c]Average number of conformers generated for each compound.

Table 2. Number of Protein−Ligand Complexes Used for Constructing Multiple-Ligand Templates

target protein	number of complexes[a]	number of clusters[b]
ace	37	14
ache	116	52
cdk2	2526	425
cox2	49	20
egfr	2532	420
fxa	851	189
hivrt	158	55
inha	387	85
p38	1804	346
pde5	180	59
pdgfrb	2564	430
src	2522	426
vegfr2	2544	409

[a]Complexes in $\{C\}$ obtained after removing the structures that bind the same ligands as any actives in the data set. [b]Clusters of complexes defined by the similarity between ligands with a threshold Tanimoto coefficient of 0.7.

defined by cutting off the Tanimoto coefficient at 0.7. The ten largest clusters were then selected, and their cluster centroids, L_1^*, L_2^*, ..., L_{10}^*, were identified. These ten ligands were accumulated into a set $\{L^*\} = (L_i^*$ with $i = 1-10)$, from which we defined the corresponding set of ten complexes $\{C^*\}$ and the set of proteins $\{P^*\}$. The $\{C^*\}$ were superposed on C^t by maximizing the individual structural overlap of each P_i^* and P^t using the structure alignment measure of MICAN, which determined the spatial coordinates of the ligand set L_i^*. Finally, fixing P^t on its superposed position, proteins P_i were eliminated to obtain the structure data file representing the spatially arranged complexes of the target protein P^t and the multiple-ligand template $Q^{multi} = L_1^* + L_2^* + \cdots + L_{10}^*$. It should be noted that ligand L^t in the reference complex C^t served a single

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

purpose, i.e., to select ligands bound to the same pocket. As such, L^t was not necessarily a member of Q^{multi}.

Scoring with Geometric Hashing. When evaluating the kth test compound, N_k^{conf} conformers were generated using OMEGA.[38] The structure of the lth conformer is denoted as $\Gamma_k(l)$ with $1 \le l \le N_k^{\mathrm{conf}}$. The 3D orientation and position of $\Gamma_k(l)$ was determined by superposing the conformer l to Q^{multi} using a variant of the geometric hashing method, as explained in the next subsection. Expressing the operation, i.e., the rotation and displacement of $\Gamma_k(l)$ for this superposition as R, the 3D coordinates of the superposed conformer are given by $R\Gamma_k(l)$. To measure the atomic overlap between $R\Gamma_k(l)$ and Q^{multi}, the 3D space was divided into a grid with a cell size of 0.8 Å. If the centers of two atoms occupied the same cell in the grid, the atoms were regarded as overlapping. Suppose that test compound k has N_k^{atom} atoms classified into six types: C, N, O, S, P, and others. We evaluated $N^{\mathrm{lig}}(i, R\Gamma_k(l), Q^{\mathrm{multi}})$, the number of atoms of Q^{multi} of the same type as atom i that overlap atom i of $R\Gamma_k(l)$. The evaluation of $N^{\mathrm{lig}}(i, R\Gamma_k(l), Q^{\mathrm{multi}})$ is illustrated in Figure 4. Summing over all atoms in the test compound, we have

$$S^{\mathrm{match}}(R\Gamma_k(l), Q^{\mathrm{multi}}) = \sum_{i=1}^{N_k^{\mathrm{atom}}} N^{\mathrm{lig}}(i, R\Gamma_k(l), Q^{\mathrm{multi}}) \quad (1)$$

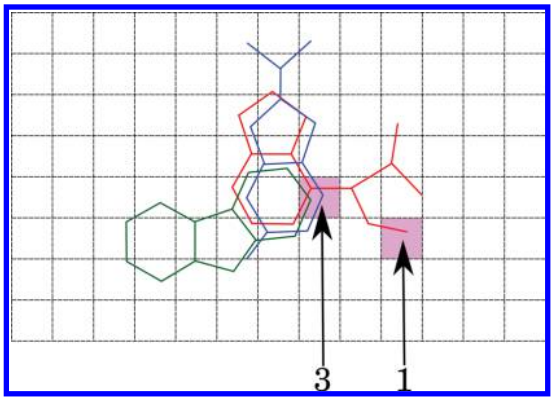


Figure 4. Evaluation of $N^{\mathrm{lig}}(i, R\Gamma_k(l), Q^{\mathrm{multi}})$ on a 2D grid. In this example, the multiple-ligand template Q^{multi} comprises three ligands (red, blue, and green line drawings). When carbon atom i of conformer l of the test compound k with configuration $R\Gamma_k(l)$ is superposed on the position designated by Arrow 3, three carbon atoms of Q^{multi} share the same cell. Therefore, $N^{\mathrm{lig}}(i, R\Gamma_k(l), Q^{\mathrm{multi}}) = 3$. When carbon atom i with configuration $R'\Gamma_k(l)$ sits at the position designated by Arrow 1, then $N^{\mathrm{lig}}(i, R'\Gamma_k(l), Q^{\mathrm{multi}}) = 1$. VS-APPLE performs this calculation on a 3D grid.

In addition to the above calculation of S^{match}, $R\Gamma_k(l)$ was evaluated by checking whether the atoms in $R\Gamma_k(l)$ collide with atoms in P^t. This checking was performed on the same grid of 0.8 Å cells. If an atom i of $R\Gamma_k(l)$ occupied a cell whose center was within 1.1 Å, of a heavy atom (i.e., an atom other than hydrogen) of P^t, a collision between the two atoms was assumed. In the event of a collision, $N^{\mathrm{prot}}(i, R\Gamma_k(l), P^t)$ was set to 1; otherwise, it was set to 0. Therefore, the total number of atoms in $R\Gamma_k(l)$ colliding with any heavy atom of P^t is

$$S^{\mathrm{coll}}(R\Gamma_k(l), P^t) = \sum_{i=1}^{N_k^{\mathrm{atom}}} N^{\mathrm{prot}}(i, R\Gamma_k(l), P^t) \quad (2)$$

Because S^{coll} is calculated using a hash table, the calculation is significantly faster than that of many docking simulations, which repetitively calculate the pairwise distances.

S^{match} and S^{coll} are summarized as

$$S^{\mathrm{config}}(R\Gamma_k(l), P^t, Q^{\mathrm{multi}}) = S^{\mathrm{match}}(R\Gamma_k(l), Q^{\mathrm{multi}}) - \omega S^{\mathrm{coll}}(R\Gamma_k(l), P^t) \quad (3)$$

When the weight factor vanishes ($\omega = 0$), collisions between the test compound and the target protein are disregarded, and the test compound is evaluated solely by the overlap between the multiple-ligand template and the test compound. On the other hand, when $\omega \gg 1$, only those test compounds that do not collide with the target protein are selected and evaluated according to their overlap with the multiple-ligand template. Because $S^{\mathrm{match}} \approx 10$ for typical compounds, a moderate weight of $\omega = O(1)$ should allow compounds to collide with the target protein to certain extent. Such weak collisions could be absorbed by the flexible atomic rearrangement of the target protein and increase the possibility of finding active compounds. Therefore, we may reasonably set a moderate ω in S^{config}. In subsequent analysis, we set $\omega = 2$, unless otherwise stated.

To calculate the score of a given compound k (representing the likeliness of the compound being active to the target protein P^t), we generated all possible Rs by the geometric hashing method, selected the R that maximizes $S^{\mathrm{config}}(R\Gamma_k(l), P^t, Q^{\mathrm{multi}})$, and averaged over the conformers l as

$$S(k, P^t) = \frac{1}{N_k^{\mathrm{conf}}} \sum_{l=1}^{N_k^{\mathrm{conf}}} [\max_R S^{\mathrm{config}}(R\Gamma_k(l), P^t, Q^{\mathrm{multi}})] \quad (4)$$

Compounds in the library were then evaluated by their scores $S(k, P^t)$.

Superposition Based on a Geometric Hashing Method. The operation R that superposes a conformer onto the template was generated by a variant of the geometric hashing method. This subsection explains how R is generated. The procedure basically follows the conventional geometric hashing method:[35] (i) Select three atoms in $R\Gamma_k(l)$ and three atoms in Q^{multi}. (ii) For an atom triplet, generate a set of basis vectors representing its 3D coordinate system, $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$, and its origin $\boldsymbol{r}_0$, where $\boldsymbol{r}_0$ is a vector representing the position of one atom in the triplet. This step defines $(\boldsymbol{r}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)_T$ in Q^{multi} and $(\boldsymbol{r}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)_\Gamma$ in $\Gamma_k(l)$. (iii) Define R as the operation that superposes $(\boldsymbol{r}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)_\Gamma$ on $(\boldsymbol{r}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)_T$.

This superposition procedure was further refined to highlight the chemical groups. Because each ligand has several characteristic chemical groups which are important for binding activity, superpositions that neglected the spatial distribution of the atoms within these chemical groups were excluded from the R-generating procedure: Because atom triplets comprising atoms in different ligands of Q^{multi} do not represent local chemical groups of the ligands, these interligand atom triplets were discarded, and the coordinates in Q^{multi} were computed only from the intraligand atom triplets. Further, to enhance sampling of local chemical groups, atom triplets were chosen only when the atoms in each triplet were separated by less than 2.5 Å. Under these restrictions, the generated coordinates properly represent the structures of local chemical groups.

To reduce the computation time and increase the calculation efficiency, we must reduce the number of operations R. In the

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

most naive algorithm, the complexity of all possible comparisons is $O(N_T \times N_\Gamma \times N_k^{\mathrm{atom}})$. Here, N_T is the number of generated $(\boldsymbol{r}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)_T$, N_Γ is the number of generated $(\boldsymbol{r}_0, \boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)_\Gamma$, and N_k^{atom} is the number of atoms in the kth test compound. The standard geometric hashing technique described in Eidhammer et al.[36] reduces this complexity to $O(N_T \times N_{Q_{\mathrm{multi}}}^{\mathrm{atom}} + N_\Gamma \times N_k^{\mathrm{atom}})$, where $N_{Q_{\mathrm{multi}}}^{\mathrm{atom}}$ is the number of atoms in the multiple ligand template, and $O(N_T \times N_{Q_{\mathrm{multi}}}^{\mathrm{atom}})$ and $O(N_\Gamma \times N_k^{\mathrm{atom}})$ are the complexities of the preprocessing and recognition steps, respectively. Because $N_T \times N_{Q_{\mathrm{multi}}}^{\mathrm{atom}}$ is usually significantly larger than $N_\Gamma \times N_k^{\mathrm{atom}}$, fast computation requires a reduction in $N_T \times N_{Q_{\mathrm{multi}}}^{\mathrm{atom}}$. Here, we propose a method that further reduces the complexity by reducing N_T. We note that $N^{\mathrm{lig}}(i, R\Gamma_k(l), Q^{\mathrm{multi}})$ in eq 1 is large only when atom i of $R\Gamma_k(l)$ is superposed on a crowded atomic environment of Q^{multi}. This scenario is most likely when the coordinate origin of $R\Gamma_k(l)$ is superposed on acoordinate origin in Q^{multi} that is surrounded by the origins of many other possible coordinates. Therefore, only the coordinates in the dense region of Q^{multi} should be used in the N^{lig} calculation. To reduce N_T, we define the degree of crowdedness of the coordinates generated in Q^{multi}. For instance, suppose that the coordinate of $\Gamma_k(l)$ is superposed on the coordinate p of Q^{multi}. The crowdedness around coordinate p is defined as

$$D^{\mathrm{crowd}}(p) = \frac{1}{N_T} \sum_{q=1}^{N_T} \exp(-d_{pq}/2) \tag{5}$$

where d_{pq} is the distance between the two coordinates $(\boldsymbol{r}_0^p, \boldsymbol{e}_1^p, \boldsymbol{e}_2^p, \boldsymbol{e}_3^p)_T$ and $(\boldsymbol{r}_0^q, \boldsymbol{e}_1^q, \boldsymbol{e}_2^q, \boldsymbol{e}_3^q)_T$

$$d_{pq} = \sqrt{(\boldsymbol{r}_0^p - \boldsymbol{r}_0^q)^2 + \sum_{k=1}^{3} (\boldsymbol{r}_0^p - \boldsymbol{r}_0^q + \boldsymbol{e}_k^p - \boldsymbol{e}_k^q)^2} \tag{6}$$

The coordinates in Q^{multi} were ranked by their D^{crowd} values, and R was defined using the 10% most-crowded coordinates; the remaining 90% of coordinates were discarded. We confirmed that this reduction of N_T did not significantly change the calculated $S(k, P^t)$. Meanwhile, the calculation time was dramatically reduced. The complexity of the proposed algorithm is $O(0.1 \times N_T \times N_{Q_{\mathrm{multi}}}^{\mathrm{atom}} + N_\Gamma \times N_k^{\mathrm{atom}})$, corresponding to a 10-fold increase in computational speed (relative to the standard method). This improvement is clarified in the following estimation. Suppose that the test compound and all template ligands are composed of n atoms. The estimated complexity of N_T is then $10 \times n^3$. Similarly, $N_{Q_{\mathrm{multi}}}^{\mathrm{atom}} \sim 10 \times n$, $N_\Gamma \sim n^3$, and $N_k^{\mathrm{atom}} = n$. From these estimates, the complexities of the standard and proposed methods are $O(100 \times n^4 + n^4)$ and $O(10 \times n^4 + n^4)$, respectively. It follows that the complexity of the proposed method (relative to the standard method) linearly reduces with decreasing N_T. The actual run time needed for computation by VS-APPLE is reasonably short. Using the data set listed in Table 1, it took about 0.096 s to evaluate one conformer of a test compound on 2.4 GHz Opteron processor. Since a test compound in the data set has about 50 conformers on average, the computation time needed for evaluation of one test compound was about 4.8 s. This computational time is about three times shorter than that of Glide: According to the support page of Schrödinger company (http://www.schrodinger.com/kb/1012), the required computation time for one test compound by "Standard Precision" (SP) mode of Glide is 15 s on a 2.2 GHz Opteron processor.

Active and Decoy Library for the Benchmark Test. The performance of VS-APPLE was evaluated on a subset of the DUD data set. Each of the 40 target proteins in DUD is provided with a set of active compounds and a large number of decoy compounds. The decoys have similar chemical properties to actives but are presumed to be inactive, with low binding affinity to the targets.[13] The DUD data set has been widely used to determine whether VS methods can distinguish a small number of active compounds from many decoy compounds. However, it was pointed out that some of the targets in the DUD data set are not suitable for evaluating 3D molecular similarity methods because nearly all active molecules recorded for these targets are trivial analogues of each other.[12] To resolve this problem, Good and Oprea[11] selected mutually dissimilar actives from the DUD data set by filtering and clustering them. The diversity of Good and Oprea's actives was examined by Cheeseright et al.,[21] and the resulting subset of DUD comprising 13 targets and the corresponding actives and decoys have since been used to test various VS methods.[7,16,37] Since actives for these 13 DUD targets are largely different from each other, they were suitable for evaluating the 3D molecular similarity methods. Therefore, we used this 13-target subset of DUD in the benchmark test of VS-APPLE. This subdatabase of DUD is summarized in Table 1. For each target protein in Table 1, a protein−ligand complex (comprising the target protein P^t and the ligand L^t) was chosen as a reference complex C^t. Huang et al.[13] and Kinnings and Jackson[7] assumed the 13 ligands L^t as single-ligand templates. In contrast, we used the reference complexes C^t to collect multiple ligands that bind to the same pocket. To this end, we checked whether a pocket binds both L^t and other ligands (see subsection *Multiple-Ligand Template*).

For each active or decoy molecule k, N_k^{conf} conformers were generated by OMEGA[38] under the constraint $N_k^{\mathrm{conf}} \leq 100$. Kirchmair et al.[39] examined the energy threshold in OMEGA that optimizes a set of candidate conformers for VS; an overly low threshold excludes the high-energy active conformers as candidate conformers, whereas an overly high threshold yields too many high-energy conformers, reducing the efficiency of the procedure. Here, we adopted Kirchmair et al. value of 25 kcal mol^{-1} when generating conformers. Table 1 summarizes the average number of conformers generated for each compound. To eliminate similar conformers, the diversity threshold was set to 1 Å root-mean-square deviation. The same energy and diversity threshold parameters were adopted by Sperandio et al.[40] and Kinnings and Jackson.[7]

For each target protein in Table 1, an ensemble of structures of protein−ligand complexes $\{C\}$ was selected from the PDB. From this ensemble, we constructed each multiple-ligand template in VS-APPLE. Complexes binding the same ligand as any active ligand in the test data set were removed from $\{C\}$. The numbers of remaining complexes are summarized in Table 2. For most of the targets, numerous complexes were retrieved from the PDB. The diversity of bound ligands in these complexes was quantified by the Tanimoto coefficient of the MACCS key 2D-fingerprint, which measures the similarity among the ligands. Table 2 lists the numbers of complex clusters obtained with a similarity threshold (Tanimoto coefficient) of 0.7. The large number of these clusters shows that the target proteins are sufficiently promiscuous to bind many distinctly different ligands.

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

Table 3. AUC Values of 13 Target Proteins Obtained by Various VS Methods[a]

	protein-based arrangement		ligand based[b]			docking-simulation based[c]		
target protein	VS-APPLE (ω = 0)	VS-APPLE (ω = 2)	ROCS1	ROCS2	BABEL[d]	Glide	DOCK	GOLD[e]
ace	0.91	0.92	0.70	0.82	**0.94**	0.73	0.68	0.44
ache	0.55	0.53	**0.77**	**0.77**	0.70	0.68	0.68	0.69
cdk2	**0.76**	0.75	0.68	0.69	0.57	**0.76**	0.57	0.63
cox2	0.78	0.79	0.93	0.95	0.63	**0.97**	0.82	0.80
egfr	0.93	0.94	**0.95**	**0.95**	0.73	0.84	0.57	0.46
fxa	0.84	**0.86**	0.39	0.66	0.56	0.75	0.73	0.72
hivrt	**0.84**	0.83	0.66	0.72	0.70	0.83	0.68	0.59
inha	0.53	0.62	0.72	0.78	**0.79**	0.63	0.27	0.70
p38	0.33	0.40	0.52	0.48	0.42	**0.61**	0.42	0.63
pde5	**0.82**	0.77	0.53	0.59	0.34	0.74	0.56	0.73
pdgfrb	0.92	**0.95**	0.34	0.30	0.73	–[f]	0.36	0.63
src	0.84	**0.90**	0.38	0.39	0.60	0.89	0.48	0.67
vegfr2	**0.86**	0.78	0.43	0.44	0.54	0.66	0.38	0.70
mean	0.76	**0.77**	0.62	0.66	0.63	0.76	0.55	0.65

[a]The highest AUC of each target is highlighted in bold. [b]ROCS1 and ROCS2 are taken from Kirchmair et al.[15] and Venkatraman et al.,[16] respectively. [c]Results for Glide are taken from Repasky et al.[20] Results for DOCK and GOLD are taken from Meier et al.[22] [d]Results for BABEL were obtained by the present authors using the open BABEL software,[17] with the ligands listed in Table 1 as templates. [e]Results for GOLD were calculated using ChemScore.[41] [f]Data not recorded in Repasky et al.[20]

RESULTS AND DISCUSSION

In this section, the performance of VS-APPLE is examined using the data set of Table 1. The results are analyzed by drawing the ROC curves[26] of each target protein, which relate the true positive rate (TPR) to the false positive rate (FPR). The TPR and FPR define the fraction of actives and decoys, respectively, in a group of compounds with S scores (see eq 4) larger than a given value. The overall quality of the ROC curve is represented by the AUC,[27,30] defined as

$$\mathrm{AUC} = \frac{1}{N^{\mathrm{active}}} \sum_{n=1}^{N^{\mathrm{active}}} (1 - f_n) \tag{7}$$

where N^{active} is the number of actives for a given target, and f_n is the fraction of decoys with larger S value than the nth-ranked active. The range of AUC is $0 \leq \mathrm{AUC} \leq 1$. Another widely used metric for assessing VS performance is the enrichment factor (EF),[30] defined by

$$EF_{x\%} = \frac{N_{x\%}^{\mathrm{active}}/N_{x\%}}{N^{\mathrm{active}}/N} \tag{8}$$

where N is the total number of compounds (both actives and decoys) for a target protein, $N_{x\%}$ is the number of compounds scoring the top x%, and $N_{x\%}^{\mathrm{active}}$ is the number of actives among the top x%-scoring compounds. Because only a small fraction of a database is tested experimentally, it is desirable to recognize active molecules as early as possible. The EF is suitable for evaluating this ability. In the following analysis, we first compare VS-APPLE with other widely used ligand-based and structure-based (i.e., docking-simulation-based) VS methods and then analyze the performance of VS-APPLE with several modified calculations.

Comparison with Other Ligand-Based and Structure-Based Methods. The AUC values calculated by VS-APPLE and several existing VS methods are listed in Table 3. In Table 3, results obtained by two variants of VS-APPLE, one taking into account collisions of test compounds against the target proteins in the AUC (i.e., setting ω = 2 in eq 3) and the other ignoring the collisions (ω = 0), are shown, but the comparison of these two variants will be discussed in detail in the subsection *Effects of Checking Collisions against Target Proteins*, and in this subsection, the VS-APPLE with ω = 2 is focused on. In the ligand-based method ROCS, two molecules are superposed by maximizing the volume overlap, and their 3D similarity is expressed by a score similar to the Tanimoto coefficient.[14] Here, the AUC values of ROCS reported in two previous publications are labeled as ROCS1[15] and ROCS2,[16] whose difference is in the way of conformer generation: both methods used the program OMEGA, but ROCS1 used the default setting of OMEGA and ROCS2 generated 100 conformers per ligand. BABEL is based on the 2D-fingerprints of ligands.[17] Also compared are the popular structure-based methods Glide,[18−20] DOCK,[21] and GOLD.[22]

Comparing the mean AUC values in Table 3, we observe that VS-APPLE performs comparably to the structure-based method Glide and significantly better than the other structure-based methods (DOCK and GOLD) and the ligand-based methods (ROCS and BABEL). Since each of the 13 DUD targets we used here possesses at least 15 different clusters of actives, the results in Table 3 show that VS-APPLE can pick up a chemically wide variety of active compounds although the search algorithm of VS-APPLE is based on the 3D similarity to known reference compounds.

Examining the AUC results of individual target proteins, we find that VS-APPLE is superior to other methods for most of the targets; the exceptions are ache, inha, and p38. p38 is well-recognized as a difficult target[19] because its binding pocket is structurally flexible.[23] The low AUC value for p38 by VS-APPLE likely arises from the combined use of multiple ligands that bind to the flexible ligand-binding site of p38, suggesting that the structures of the flexible pocket require further analyses and classification before useful p38 templates can be built into VS-APPLE. The poor performance of VS-APPLE compared with the other ligand-based methods for ache and inha can be attributed to the high dissimilarity of template ligands to the DUD active compounds. Shown in Figure 5 is the box-and-whisker plot of the Tanimoto coefficient for the MACCS key 2D-fingerprint measure between the DUD active compounds and the template ligands used in ROCS and BABEL (colored in

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

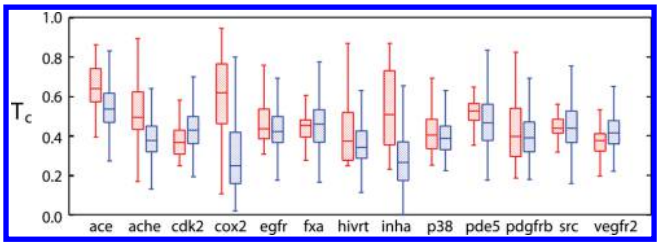


Figure 5. Box-and-whisker plots of the Tanimoto coefficient for the MACCS key 2D-fingerprint measure. The red bar represents the distribution of the Tanimoto coefficient between the DUD actives and the template ligands for ROCS and BABEL used in the literatures,[15,16] and the blue one represents that between the DUD actives and all ten template ligands used in VS-APPLE.

red) and that between the DUD actives and all the template ligands used in VS-APPLE (colored in blue). The distributions of the Tanimoto coefficient of VS-APPLE for ache and inha are largely shifted to the small values compared with those of the other ligand-based methods, and templates with high similarity to the actives ($T_c \sim 0.9$) were included in the templates of ROCS ans BABEL, implying that VS-APPPLE is at a disadvantage in terms of the quality of template ligands for these targets. Interestingly, this disadvantage does not always lead to a poor performance of VS-APPLE. As shown in Figure 5, the template ligands of VS-APPLE for hivrt and pdgfrb are less similar to the actives than those of ROCS and BABEL. The screening performance for these targets by VS-APPLE, however, is much higher than those of ROCS and BABEL, as indicated by AUC values in Table 3. This result demonstrates one of the advantages of VS-APPLE: It can search for active compounds using template ligands that are dissimilar to the actives at least for some targets.

Next, we compare $EF_{1\%}$ values obtained by the methods compared here. Listed in Table 4 are the $EF_{1\%}$ values obtained by VS-APPLE, and those of other methods which were taken from the literature.[15,22,41,44] Unlike the AUC, the average value of $EF_{1\%}$ by VS-APPLE is smaller than the others except for DOCK. This result indicates that although VS-APPLE can pick up a wide variety of active ligands as shown by the AUC analysis, they are not included in the early phase (~1%) of screening for many targets.

Although the average value of $EF_{1\%}$ by VS-APPLE was lower than the other methods, it was superior to the others for some targets; VS-APPLE demonstrated the best performance for two targets (pde5 and src) and a similar performance to the best methods for other two targets (ace and hivrt). In addition, the performance of VS-APPLE is largely different from the others, and correlation coefficients of $EF_{1\%}$s between VS-APPLE and the others are small; −0.24, −0.39, −0.28, 0.09, and −0.06 for ROCS, BABEL, Glide, DOCK, and GOLD, respectively. Therefore, VS-APPLE can compensate the weakness of the other methods, and the combined[24] or the consensus virtual screening strategies[25] should benefit from including VS-APPLE for accurate virtual screening.

Ability for Scaffold Hopping. The screening score in VS-APPLE was calculated from $N^{\mathrm{lig}}(i, R\Gamma_k(l), Q^{\mathrm{multi}})$ in eq 1, which depends on the atomic density in the region of the template Q^{multi} superposed by atom i of the conformer $R\Gamma_k(l)$. Note that N^{lig} is large when i is superposed on a densely populated region of Q^{multi} (See Figure 4). However, too much emphasis on the dense atomic regions in the existing structures of Q^{multi} would hinder the search for novel scaffolds that are dissimilar to the existing structures. To weaken the emphasis on dense regions, we replace N^{lig} in eq 1 with $N^{\mathrm{lig}}_{\mathrm{step}}$, defined as

$$N^{\mathrm{lig}}_{\mathrm{step}}(i, R\Gamma_k(l), Q^{\mathrm{multi}}) = \theta(N^{\mathrm{lig}}(i, R\Gamma_k(l), Q^{\mathrm{multi}})) \quad (9)$$

where $\theta(x)$ is a step function; namely, $\theta(x) = 1$ for $x \geq 1$ and $\theta(x) = 0$ for $x = 0$. The version of VS-APPLE based on $N^{\mathrm{lig}}_{\mathrm{step}}$ is referred to as VS-APPLE$^{\mathrm{step}}$ to distinguish it from the original VS-APPLE based on N^{lig}.

The ability of VS-APPLE$^{\mathrm{step}}$ and VS-APPLE to find novel scaffolds can be quantified by examining how the methods distinguish active compounds when the template ligands and actives are dissimilar. The similarity between the template ligands and the actives in the test data set was measured by the Tanimoto coefficient of the MACCS key 2D-fingerprint. Ligands were incorporated into a template Q^{multi} only when

Table 4. Comparison of VS Methods in Terms of $EF_{1\%}$ Values Calculated for 13 Target Proteins[a]

	protein-based arrangement		ligand based[b]		docking-simulation based[c]		
target protein	VS-APPLE ($\omega = 0$)	VS-APPLE ($\omega = 2$)	ROCS	BABEL[d]	Glide	DOCK	GOLD[e]
ace	29.5	29.5	27.7	19.5	**30.6**	8.7	2.1
ache	2.0	2.0	**26.8**	23.0	3.8	0.0	4.0
cdk2	4.1	4.1	26.4	17.0	**57.1**	10.6	16.4
cox2	2.3	1.9	**35.7**	31.6	29.4	16.9	21.9
egfr	7.1	7.9	33.0	**39.1**	5.0	4.1	0.8
fxa	10.8	7.7	4.3	3.1	**54.2**	9.5	3.7
hivrt	17.1	20.0	**21.1**	5.8	10.0	6.2	5.5
inha	0.0	3.4	36.7	**45.6**	31.8	0.0	40.5
p38	0.0	0.0	**10.2**	5.8	0.4	0.0	3.3
pde5	11.0	**18.4**	16.2	0.0	2.0	7.7	16.4
pdgfrb	5.6	6.4	7.7	**20.9**	–[f]	0.0	4.9
src	7.1	**16.1**	2.0	14.2	8.4	1.0	5.8
vegfr2	12.3	20.5	0.0	4.1	–[f]	2.1	**35.7**
mean	8.3	10.6	19.0	17.6	**22.3**	5.1	12.3

[a]The largest $EF_{1\%}$ value for each target is highlighted in bold. [b]Results for ROCS were taken from Kirchmair et al.[15] [c]Results for Glide are taken from Svensson et al.[44] Results for DOCK and GOLD are taken from Meier et al.[22] [d]Results for BABEL were obtained by the present authors, using the open BABEL software,[17] where ligands listed in Table 1 were used as templates. [e]Results for GOLD were calculated with ChemScore.[41] [f]Data was not available from Svensson et al.[44]

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

their Tanimoto coefficient was below the threshold T_c. The previous subsection assumed the maximal Tanimoto coefficient $T_c = 1$; by reducing the T_c, we selected ligands that were less similar to the actives. More specifically, protein−ligand complexes were eliminated from the ensemble of complexes C if their similarity between ligand and active exceeded T_c. The remaining complexes were clustered, and the ten top-ranking clusters were selected from C to generate ten ligands in Q^{multi}. The dependence of performance on T_c was similarly calculated for other VS methods; the template ligands in these methods were restricted to a smaller Tanimoto coefficient than T_c.

Figure 6 plots the AUC value averaged over 13 targets, $\overline{\text{AUC}}(T_c)$, as a function of T_c for different VS methods, namely,

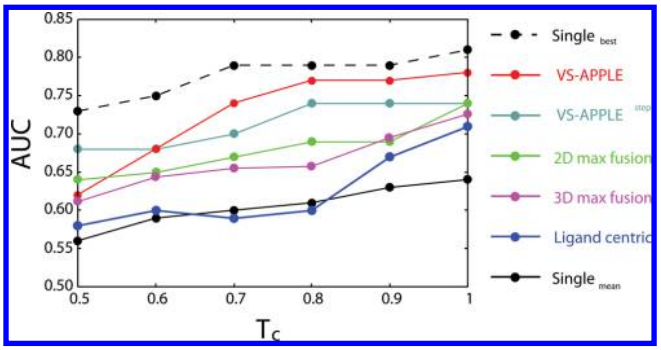


Figure 6. AUC values as functions of the similarity cutoff T_c, which defines the maximum similarity between the template ligands and actives in the test data set. The mean AUC values $\overline{\text{AUC}}(T_c)$ were averaged over 13 targets. The mean AUCs calculated by VS-APPLE (red) and VS-APPLE$^{\text{step}}$ (dark green) with $\omega = 0$ are compared with those of other methods based on multiple templates, i.e., the ligand-centric method (blue), the 2D max fusion method (light green), and the 3D max fusion method (pink), along with methods based on single-ligand templates, i.e., Single$_{\text{best}}$ (black dashed line) and Single$_{\text{mean}}$ (black solid line). Single$_{\text{best}}$, or the upper limit of the single-ligand template method, was obtained *ad hoc* from knowledge of the answer actives. Single$_{\text{mean}}$ denotes the average performance of the single-ligand template method.

VS-APPLE, its variant VS-APPLE$^{\text{step}}$, methods using multiple ligands for a single target (the ligand-centric method and the 2D and 3D max fusion methods[7,28,29]), and the single-ligand template methods. The ligand-centric methods will be discussed in the subsection *Effects of Protein-Based Ligand Arrangements*; other methods will be discussed in *Comparison between Multiple and Single Template Methods*.

According to Figure 6, the $\overline{\text{AUC}}(T_c)$ values of all methods decrease as T_c decreases. Especially, the performance of VS-APPLE, which showed the best performance at $T_c = 1.0$, was significantly degraded by reducing the T_c. On the contrary, the performance of VS-APPLE$^{\text{step}}$ was relatively robust against the change of T_c, although the same multiple ligand templates are used as in VS-APPLE. Since the only difference between VS-APPLE and VS-APPLE$^{\text{step}}$ is the scoring function, a steep decline in AUC of VS-APPLE can be attributed to the emphasis on the dense atomic regions of Q^{multi} in the scoring function. Furthermore, apart from Single$_{\text{best}}$ representing the upper limit of single-ligand template methods, VS-APPLE$^{\text{step}}$ provides the best VS method for identifying actives that are dissimilar ($T_c < 0.6$) from the template ligands. These results show that although the emphasis on the dense atomic regions of Q^{multi} is useful for finding the relatively similar ($T_c \geq 0.7$) compounds to the references, too much emphasis hinders the search for novel scaffolds. Thus, either VS-APPLE or VS-APPLE$^{\text{step}}$ should be chosen depending on the problem whether compounds with highly novel scaffolds are needed or not.

Comparison between Multiple and Single Template Methods. A pronounced feature of VS-APPLE is its use of multiple ligands L_1^*, L_2^*, ..., L_{10}^* to build a template Q^{multi}. To analyze the effects of such multiple ligands, we compare the performances of VS-APPLE and a single-ligand version of VS-APPLE, named VS-APPLE$^{\text{single}}(L_i^*)$, in which the multiligand template Q^{multi} is replaced with single-ligand template L_i^*. There are two major differences between VS-APPLE and VS-APPLE$^{\text{single}}(L_i^*)$. First, in VS-APPLE, N^{lig} is calculated only from the coordinates in the dense region of Q^{multi}, whereas all coordinates defined by three atoms are included in the L_i^* of VS-APPLE$^{\text{single}}(L_i^*)$. The second difference is the score by which compounds are evaluated. The score used in VS-APPLE$^{\text{single}}(L_i^*)$ is defined as

$$S(k, P^t) = \frac{1}{N_k^{\text{conf}}} \sum_{l=1}^{N_k^{\text{conf}}} [\max_R S^{\text{config}}(R\Gamma_k(l), P^t, L_i^*)] \quad (10)$$

When $\omega = 0$, VS-APPLE$^{\text{single}}(L_i^*)$ resembles LigMatch developed by Kinnings et al.[7]

The overall behaviors of VS-APPLE$^{\text{single}}(L_i^*)$ are examined by averaging the AUC values over the 13 target proteins in the DUD-derived database. The AUC calculated by VS-APPLE$^{\text{single}}(L_i^*)$ and averaged over the 13 targets is written as $\overline{\text{AUC}}_i^{\text{single}}$. The largest $\overline{\text{AUC}}_i^{\text{single}}$, where $i = 1-10$, is denoted $\overline{\text{AUC}}_{\text{best}}^{\text{single}}$, and the average of the ten $\overline{\text{AUC}}_i^{\text{single}}$ values is expressed as $\overline{\text{AUC}}_{\text{mean}}^{\text{single}}$. In Figure 6, $\overline{\text{AUC}}_{\text{mean}}^{\text{single}}$ and $\overline{\text{AUC}}_{\text{best}}^{\text{single}}$ are plotted as Single$_{\text{mean}}$ and Single$_{\text{best}}$, respectively, and are compared with $\overline{\text{AUC}}$ obtained by VS-APPLE and VS-APPLE$^{\text{step}}$ at different similarity cutoffs T_c. The $\overline{\text{AUC}}$s of VS-APPLE and VS-APPLE$^{\text{step}}$ are smaller than Single$_{\text{best}}$ but considerably larger than Single$_{\text{mean}}$ for all examined values of T_c. Because we cannot know *a priori* which ligand achieves Single$_{\text{best}}$, the best ligand is identified only in an *ad hoc* manner by using the results obtained with each single-ligand template. In contrast, even though VS-APPLE and VS-APPLE$^{\text{step}}$ do not use knowledge of the performance of single-ligand templates, VS-APPLE and VS-APPLE$^{\text{step}}$ provide much better results than the average performance of the single-ligand template method.

Figure 6 also compares the results of methods based on max fusion rules.[7,28,29] In these methods, the similarity scores between each test compound and all the available single-ligand templates are computed. The score of the test compound is defined as the highest computed score. In this work, the similarity between the test compound and the ligand L_i^* was measured in two ways: by the Tanimoto coefficient of the MACCS key 2D-fingerprint and by 3D structural similarity expressed by the score of VS-APPLE$^{\text{single}}$ normalized by the number of atoms present in a test compound. The former and latter similarity measures are referred to as the 2D and 3D max fusion methods, respectively. The $\overline{\text{AUC}}$s calculated by the 2D and 3D max fusion methods are plotted in Figure 6. Both the methods yield $\overline{\text{AUC}}$ values between Single$_{\text{mean}}$ and $\overline{\text{AUC}}$ of VS-APPLE or VS-APPLE$^{\text{step}}$, indicating that VS-APPLE based on multiple-ligand templates achieves superior results than methods using a single ligand for each template.

These results highlight the advantage of VS-APPLE as a method to utilize the reference ligand data. Recently, it has become evident that more and more proteins are promiscuous to interact with multiple distinct ligands, and the number of

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

experimentally determined promiscuous protein−ligand complexes has been rapidly increasing.[2−5] Accordingly, a large number of ligands are currently available as reference ligands in VS methods. This situation raises an important problem on the way to choose a ligand molecule as a reference or on the way to use multiple ligands for high performance screening. This problem is critically important for ligand-based approaches, because, as shown in Figure 7, the performance of a single-ligand template method strongly depends on the choice of a reference ligand, and the best single-reference ligand cannot be known *a priori*. VS-APPLE provides a solution to this problem, since it is free from the problem of choosing a reference ligand, and it shows a similar performance to the best single-ligand template method at least for $T_c \geq 0.8$ as shown in Figure 6.

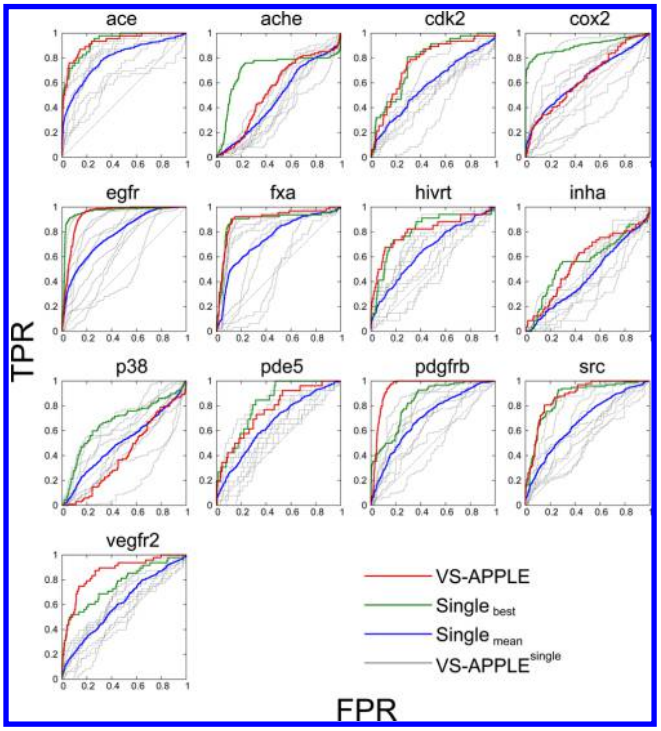


Figure 7. ROC curves of 13 target proteins obtained by VS-APPLE (red), $\text{Single}_{\text{mean}}$ (blue), and $\text{Single}_{\text{best}}$ (green). The nine gray ROC curves were obtained by $\text{VS-APPLE}^{\text{single}}(L_i^*)$. The similarity cutoff was $T_c = 1$, and collisions were neglected in the score calculation ($\omega = 0$).

For more detailed analyses, we plot the ROC curves. We denote the ROC curve of the ligand yielding $\text{AUC}^{\text{single}}{}_{\text{best}}$ by $\text{Single}_{\text{best}}$ and the mean curve averaged over ten ROC curves of $\text{VS-APPLE}^{\text{single}}(L_i^*)$ by $\text{Single}_{\text{mean}}$. Figure 7 plots the ROC curves of VS-APPLE, $\text{VS-APPLE}^{\text{single}}(L_i^*)$, $\text{Single}_{\text{best}}$, and $\text{Single}_{\text{mean}}$ for the 13 target proteins. In these plots, we set $\omega = 0$ and $T_c = 1$. VS-APPLE and $\text{Single}_{\text{mean}}$ yielded similar ROC curves for cox2, but $\text{Single}_{\text{mean}}$ yielded a higher AUC than VS-APPLE for p38. For the remaining 11 targets, the ROC curves of VS-APPLE were superior to those of $\text{Single}_{\text{mean}}$ in terms of AUC values. Overall, VS-APPLE demonstrated better performance than the single-ligand template method. For six of the targets, namely, ace, cdk2, fxa, hivrt, inha, and src, the ROC of VS-APPLE approached $\text{Single}_{\text{best}}$; for five targets (ache, cox2, egfr, p38, and pde5), it was below $\text{Single}_{\text{best}}$, and for the remaining two targets, pdgfrb and vegfr2, it clearly exceeded $\text{Single}_{\text{best}}$.

Summarizing these results, by adopting multiple ligands for each template in VS-APPLE, we can better represent the atomic features of the binding pocket in many cases and enhance the screening performance. However, our multiligand template cannot capture the flexibility of the p38 binding pocket. To solve this problem, we require further analyses of pocket-ligand interactions.

Effects of Protein-Based Ligand Arrangements. Another important feature of VS-APPLE is the spatial arrangement of multiple ligands in the template. Each of the protein−ligand complexes is superposed on the target protein to maximize the structural overlap of the proteins. This protein-based ligand arrangement should reflect the flexible atomic environment of the pocket of the target. To explore the effects of this protein-based arrangement, we compared the performances of VS-APPLE and the ligand-centric method, in which multiple ligands are superposed to maximize their mutual structural overlap. The ligand-centric template was constructed by calculating the mutual structural similarity among the ten ligands L_1^*, L_2^*, ..., L_{10}^* using the geometric hashing method. Here, the similarity between ligands i and j was measured by the score S_{ij}, which was calculated similarly to S^{match} in eq 1. The centroid ligand i with the highest total score $\sum_j S_{ij}$ was selected from the obtained S_{ij}. The remaining nine ligands were then superposed on the selected ligand by maximizing structural overlap. In the ligand-centric method, the protein-based multiple-ligand template of VS-APPLE was replaced by this constructed ten-ligand template.

Figure 8 plots the ROC curves for the 13 target proteins obtained by the original VS-APPLE and the ligand-centric method. Again, we set $T_c = 1$ and $\omega = 0$. Moreover, we compared the results of the 2D and 3D max fusion methods. VS-APPLE demonstrated the best performance for five targets (ace, fxa, hivrt, pdgfrb, and vegfr2) and a similar performance to the best of the other methods for other four targets (ache, cdk2,

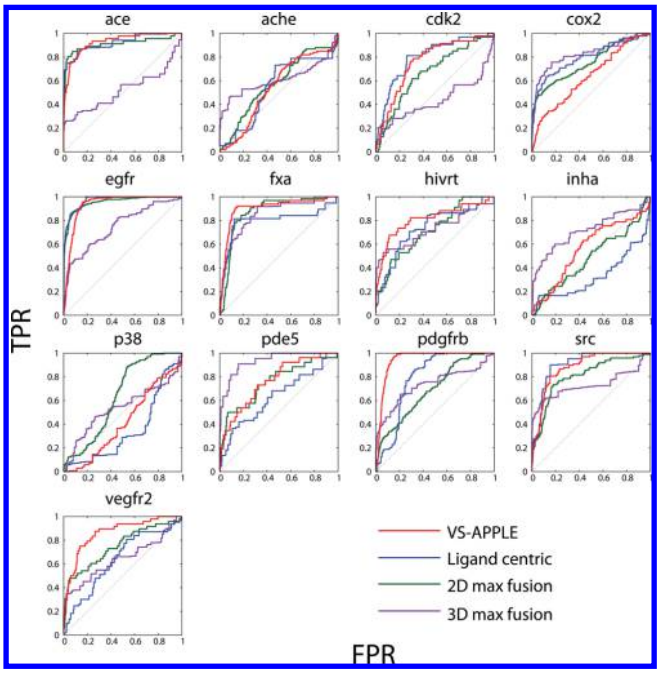


Figure 8. ROC curves for 13 target proteins obtained by VS-APPLE (red), the ligand-centric method (blue), the 2D max fusion method (green), and the 3D max fusion method (purple). $T_c = 1$ and $\omega = 0$.

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

egfr, and src). For the remaining four targets (cox2, inha, pde5, and p38), VS-APPLE was outperformed by the other methods. The ligand-centric method yielded better results than VS-APPLE for cox2 but poorer results than VS-APPLE for six targets (fxa, inha, p38, pde5, pdgfrb, and vegfr2). Overall, VS-APPLE is superior to the ligand-centric method. We conclude that the protein-based arrangement of multiple ligands is necessary for high performance because it reflects the flexible atomic environment of the target pocket. It should be noted that the ligand-centric method provides the best results for cox2 but the poorest results for inha and p38 (see Figure 8). This observation suggests that the spatial arrangement of ligands drastically changes the results and is important for handling difficult targets.

Effects of Collisions against Target Proteins. The score calculation in VS-APPLE not only matches the test compound to the template but also accounts for collisions between the test compound and the target protein (see eq 3). The effects of the collisions were not evident in the AUC values of Table 3 but become significant when the $EF_{1\%}$ is examined. The $EF_{1\%}$ was calculated by VS-APPLE in the presence and absence of collisions (ω = 2 and ω = 0 respectively), and the results are compared in Table 4. For six of the target proteins (egfr, hivrt, inha, pde5, pdgfrb, src, and vegfr2), the $EF_{1\%}$ is larger for ω = 2 than for ω = 0, thus increasing the mean $EF_{1\%}$ in the ω = 2 case. The collisions especially enhanced the performance of VS-APPLE on pde5, src, and vegfr2. Thus, by assigning a nonzero weight factor ω for collision checking, we can considerably enhance the performance of VS-APPLE during the early phase of screening.

This performance enhancement can be further examined by plotting the ω = 2 scores versus the ω = 0 scores of each compound for a given target protein. As an example, the scores $S(k, P^t)$ (see eq 4) calculated with ω = 2 and ω = 0 for the target protein P^t = src are plotted in the two-dimensional plane in Figure 9. In this figure, a group of decoys score highly when ω = 0 but significantly lower when collision effects are considered (ω = 2). Such discrimination of decoys explains the high $EF_{1\%}$ in the ω = 2 calculation for src. The structure of an example decoy in this group of compounds is shown in Figure 10. Note that this decoy largely overlaps the ligands in the multiple-ligand template but also strongly collides with the target protein, which was given a low score by VS-APPLE with ω = 2.

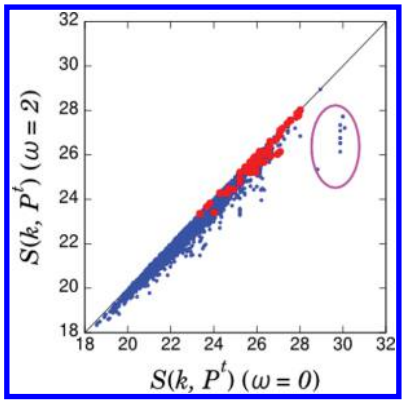


Figure 9. Scores $S(k, P^t)$ calculated by VS-APPLE with ω = 2 and ω = 0 are plotted in the 2-dimensional plane for each compound k and the target protein P^t = src. Scores of decoys and actives are presented in blue and red, respectively. The pink circle encloses the decoy scores that are high when ω = 0 but low when ω = 2, i.e., decoys that are influenced by considering the collisions.

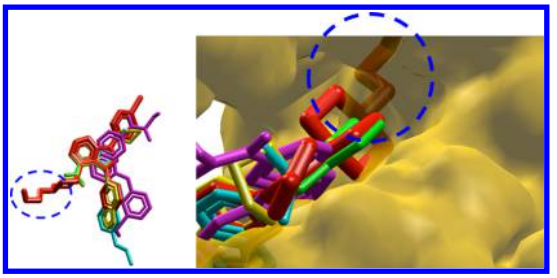


Figure 10. Structure of a decoy in the group of compounds enclosed by the pink circle in Figure 9. (Left) The red decoy largely overlaps with ligands in the multiple-ligand template (purple, yellow, green, and light blue). (Right) The decoy and ligands on the left are displayed along with the target protein surface. The blue-dashed circle encloses the colliding part of the decoy.

CONCLUSION

We have developed a new VS method, VS-APPLE, which utilizes the accumulated structure data of multiple ligands bound to promiscuous protein pockets. VS-APPLE exploits the correspondence between the configurations of the bound ligands and the flexible atomic environment of the pocket. A multiple-ligand template for screening candidate compounds was constructed from spatially arranged multiple ligands, which were obtained by protein-based superposition of protein–ligand complexes. Structural overlap between the multiple-ligand template and the test compound was evaluated by a geometric hashing method. By this approach, we could detect active compounds with scaffolds different from the template ligands.

By reducing the number of coordinates in the geometric hashing comparison between the multiple-ligand template and the test compound, we largely accelerated the computation speed, rendering VS-APPLE an efficient computation tool. The performance of VS-APPLE was examined in comparisons with methods based on 2D-fingerprints or the 3D-shape of ligands and with methods based on docking simulations. Comparisons were made on a filtered and clustered DUD subset, comprising 13 target proteins and their corresponding actives and decoys. Performance was quantified by the mean AUC averaged over 13 targets. Judging from the AUC scores, VS-APPLE performed similarly to the structure-based (docking-simulation-based method) Glide and outperformed the ligand-based methods (ROCS and BABEL) and the other structure-based methods (DOCK and GOLD). The performance of VS-APPLE remained high even when the template was constructed from ligands dissimilar to the actives in the test compound data set; in this case, VS-APPLE distinguished the actives from decoys with no obvious reduction in AUC. This implies that VS-APPLE can effectively search for active compounds with novel scaffolds.

To examine the roles of the multiple-ligand template, we compared the performances of VS-APPLE based on multiple- and single-ligand templates. The average AUC values were consistently larger in the multiligand case than when single-ligand templates were used. For two of the 13 target proteins in the reduced DUD database, the ROC curves calculated by VS-APPLE were superior to those of the best single-ligand template selected from 10 ligands. However, when averaged over 13 proteins, the best single-ligand template method gave

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108–1119

better results than VS-APPLE and VS-APPLE^{step}. However, because the best single-ligand template cannot be known *a priori*, we expect that VS-APPLE and its derivative will deliver better results than the single-ligand template method in systematic virtual screening. VS-APPLE outperformed the 2D or 3D max fusion rules. To examine the roles of the spatial arrangement of the multiple ligands in the template, we compared VS-APPLE with a method that superposes multiple ligands in the template to maximize their mutual structural overlap. In most cases, the ligand-centric method yielded smaller AUC values than VS-APPLE, indicating that the protein-based arrangement of multiple ligands in the template is important for the high performance of VS-APPLE. In addition, we examined the effects of collisions of the test compound against the target protein. The consideration of collision was particularly effective in the early phase of screening, as evidenced by the large $EF_{1\%}$ values.

The input structures to VS-APPLE are multiple resolved crystal structures of protein−ligand complexes. This strategy should become more effective as more crystal structures of protein−ligand complexes become available. Indeed, ligand soaking techniques[42,43] with automated high-throughput crystallization have increased the feasibility of crystallizing protein−ligand complexes, and the number of available structures is expected to increase. Therefore, more extensive use of multiple structures is strongly desired. VS-APPLE, a new method based on multiple-ligand templates, should promote further exploitation of the growing library of promiscuous protein−ligand complex structures.

AUTHOR INFORMATION

Corresponding Author

*Phone: 81 (52)789 3718. Fax: 81 (52)789 3719. E-mail: chikenji@tbp.cse.nagoya-u.ac.jp.

Present Address

[‖]Division of Neurogenetics, Center for Neurological Diseases and Cancer, Nagoya University Graduate School of Medicine, Nagoya, Aichi 466-8550, Japan.

Author Contributions

[§]Contributed equally to this work.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Platform for Drug Discovery, Informatics, and Structural Life Science from the Ministry of Education, Culture, Sports, Science and Technology, Japan. We thank OpenEye Scientific Software for providing us with an academic license for use of its software.

REFERENCES

(1) Nobeli, I.; Favia, A. D.; Thornton, J. M. Protein Promiscuity and its Implications for Biotechnology. *Nat. Biotechnol.* **2009**, *2*, 157−167.

(2) Gao, M.; Skolnick, J. A Comprehensive Survey of Small-Molecule Binding Pockets in Proteins. *PLoS Comput. Biol.* **2013**, *9*, e1003302.

(3) Gao, M.; Skolnick, J. APoc: Large-Scale Identification of Similar Protein Pockets. *Bioinformatics* **2013**, *29*, 597−604.

(4) Kufareva, I.; Ilatovskiy, A. V.; Abagyan, R. Pocketome: an Encyclopedia of Small-Molecule Binding Sites in 4D. *Nucleic Acids Res.* **2012**, *40*, D535−D540.

(5) Berman, H. M. Creating a Community Resource for Protein Science. *Protein Sci.* **2012**, *2*, 1587−1596.

(6) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of Topological Descriptors for Similarity-Based Virtual Screening Using Multiple Bioactive Reference Structures. *Org. Biomol. Chem.* **2004**, *2*, 3256−3266.

(7) Kinnings, S. L.; Jackson, R. M. LigMatch: A Multiple Structure-Based Ligand Matching Method for 3D Virtual Screening. *J. Chem. Inf. Model.* **2009**, *49*, 2056−2066.

(8) Pérez-Nueno, V. I.; Ritchie, D. W. Using Consensus-Shape Clustering to Identify Promiscuous Ligands and Protein Targets and to Choose the Right Query for Shape-Based Virtual Screening. *J. Chem. Inf. Model.* **2011**, *51*, 1233−1248.

(9) Wei, N.-N.; Hamza, A. SABRE: Ligand/Structure-Based Virtual Screening Approach Using Consensus Molecular-Shape Pattern Recognition. *J. Chem. Inf. Model.* **2014**, *54*, 338−346.

(10) Hamza, A.; Wei, N.-N.; Zhan, C.-G. Ligand-Based Virtual Screening Approach Using a New Scoring Function. *J. Chem. Inf. Model.* **2012**, *52*, 963−974.

(11) Good, A. C.; Oprea, T. I. Optimization of CAMD Techniques 3. Virtual Screening Enrichment Studies: a Help or Hindrance in Tool Selection. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 169−178.

(12) Irwin, J. J. Community Benchmarks for Virtual Screening. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 193−199.

(13) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789−6801.

(14) *ROCS - Rapid Overlay of Chemical Structures. 2.2*; OpenEye Scientific Software, Inc.:2006. http://www.eyesopen.com/ (accessed June 1, 2015).

(15) Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How to Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information. *J. Chem. Inf. Model.* **2009**, *49*, 678−692.

(16) Venkatraman, V.; Perez-Nueno, V. I.; Mavridis, L.; Ritchie, D. W. Comprehensive Comparison of Ligand-Based Virtual Screening Tools Against the DUD Dataset Reveals Limitations of Current 3D Methods. *J. Chem. Inf. Model.* **2010**, *50*, 2079−2093.

(17) *The Open Babel Package, ver.2.3.1*. http://openbabel.org/wiki/ (accessed November 2011) .

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

(19) 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*, 1750−1759.

(20) 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*, 787−799.

(21) Cheeseright, T. J.; Mackey, M. D.; Melville, J. L.; Vinter, J. G. FieldScreen: Virtual screening Using Molecular Fields. Application to the DUD Data Set. *J. Chem. Inf. Model.* **2008**, *48*, 2108−2117.

(22) Meier, R.; Pippel, M.; Brandt, F.; Sippl, W.; Baldauf, C. PARADOCKS: A Framework for Molecular Docking with Population-Based Metaheuristics. *J. Chem. Inf. Model.* **2010**, *50*, 879−889.

(23) Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P. F.; Gilmore, T.; Graham, A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP Kinase by Utilizing a Novel Allosteric Binding Site. *Nat. Struct. Biol.* **2002**, *9*, 268−272.

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

(25) Yang, J. M.; Chen, Y. F.; Shen, T. W.; Kristal, B. S.; Hsu, D. F. Consensus Scoring Criteria for Improving Enrichment in Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 1134−1146.

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119

(26) Hanley, J. A.; McNeil, B. J. The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve. *Radiology* **1982**, *143*, 29−36.
(27) Hawkins, P. C. D.; Warren, G. L.; Skillman, A. G.; Nicholls, A. How to Do an Evaluation: Pitfalls and Traps. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 179−190.
(28) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of Fingerprint-Based Methods for Virtual Screening Using Multiple Bioactive Reference Structures. *J. Chem. Inf. Model.* **2004**, *44*, 1177−1185.
(29) Whittle, M.; Gillet, V. J.; Willett, P.; Alex, A.; Loesel, J. Enhancing the Effectiveness of Virtual Screening by Fusing Nearest Neighbor Lists: A Comparison of Similarity Coefficients. *J. Chem. Inf. Model.* **2004**, *44*, 1840−1848.
(30) Mackey, M. D.; Melville, J. L. Better Than Random? The Chemotype Enrichment Problem. *J. Chem. Inf. Model.* **2009**, *49*, 1154−1162.
(31) Minami, S.; Sawada, K.; Chikenji, G. MICAN: a Protein Structure Alignment Algorithm that Can Handle Multiple-Chains, Inverse Alignments, Cα Only Models, Alternative Alignments, and Non-Sequential Alignments. *BMC Bioinf.* **2013**, *14*, 24.
(32) Minami, S.; Sawada, K.; Chikenji, G. How a Spatial Arrangement of Secondary Structure Elements is Dispersed in the Universe of Protein Folds. *PLoS One* **2014**, *9*, e107959.
(33) Zhang, Y.; Skolnick, J. Scoring Function for Automated Assessment of Protein Structure Template Quality. *Proteins* **2004**, *57*, 702−710.
(34) Xu, J.; Zhang, Y. How Significant is a Protein Structure Similarity with TM-score = 0.5? *Bioinformatics* **2010**, *26*, 889−895.
(35) Wolfson, H. J.; Rigoutsos, I. Geometric Hashing: An Overview. *Comput. Sci. Eng.* **1997**, *4*, 10−21.
(36) Eidhammer, I.; Jonassen, I.; Taylor, W. R. *Protein Bioinformatics*; John Wiley & Sons, Ltd.: Chichester, UK, 2001.
(37) Jahn, A.; Hinselmann, G.; Fechner, N.; Zell, A. Optimal Assignment Methods for Ligand-Based virtual Screening. *J. Cheminf.* **2009**, *1*, 14.
(38) Bostrom, J.; Greenwood, J. R.; Gottfries, J. Assessing the Performance of OMEGA with Respect to Retrieving Bioactive Conformations. *J. Mol. Graphics Modell.* **2003**, *2*, 449−462.
(39) Kirchmair, J.; Wolber, G.; Laggner, C.; Langer, T. Comparative Performance Assessment of the Conformational Model Generators Omega and Catalyst: a Large-Scale Survey on the Retrieval of Protein-Bound Ligand Conformations. *J. Chem. Inf. Model.* **2006**, *46*, 1848−1861.
(40) Sperandio, O.; Andrieu, O.; Miteva, M. A.; Vo, M. Q.; Souaille, M.; Delfaud, F.; Villoutreix, B. O. MED-SuMoLig: a New Ligand-Based Screening Tool for Efficient Scaffold Hopping. *J. Chem. Inf. Model.* **2007**, *47*, 1097−1110.
(41) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical Scoring Functions 1: The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425−445.
(42) Hassell, A. M.; An, G.; Bledsoe, R. K.; Bynum, J. M.; Carter, H. L., III; Deng, S.-J. J.; Gampe, R. T.; Grisard, T. E.; Madauss, K. P.; Nolte, R. T.; Rocque, W. J.; Wang, L.; Weaver, K. L.; Williams, S. P.; Wisely, G. B.; Xu, R.; Shewchuka, L. M. Crystallization of Protein−Ligand Complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2007**, *63*, 72−79.
(43) le Maire, A.; Gelin, M.; Pochet, S.; Hoh, F.; Pirocchi, M.; Guichou, J. F.; Ferrer, J. L.; Labesse, G. In-Plate Protein Crystallization, in Situ Ligand Soaking and X-ray Diffraction. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2011**, *67*, 747−755.
(44) Svensson, F.; Karlen, A.; Skold, A. Virtual Screening Data Fusion Using Both Structure- and Ligand-Based Methods. *J. Chem. Inf. Model.* **2012**, *52*, 225−232.

DOI: 10.1021/acs.jcim.5b00134
J. Chem. Inf. Model. 2015, 55, 1108−1119