

PL-PatchSurfer2: Improved Local Surface Matching-Based Virtual Screening Method That Is Tolerant to Target and Ligand Structure Variation

Woong-Hee Shin,[†] Charles W. Christoffer,[‡] Jibo Wang,[§] and Daisuke Kihara*,^{†,‡}

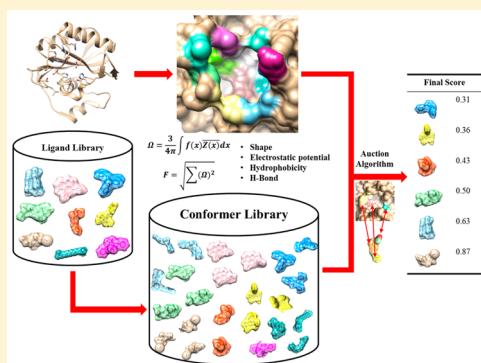
[†]Department of Biological Science, Purdue University, 249 South Martin Jischke Street, West Lafayette, Indiana 47907, United States

[‡]Department of Computer Science, Purdue University, 305 North University Street, West Lafayette, Indiana 47907, United States

[§]Discovery Chemistry Research and Technologies, Eli Lilly and Company, 893 South Delaware Street, Indianapolis, Indiana 46225, United States

Supporting Information

ABSTRACT: Virtual screening has become an indispensable procedure in drug discovery. Virtual screening methods can be classified into two categories: ligand-based and structure-based. While the former have advantages, including being quick to compute, in general they are relatively weak at discovering novel active compounds because they use known actives as references. On the other hand, structure-based methods have higher potential to find novel compounds because they directly predict the binding affinity of a ligand in a target binding pocket, albeit with substantially lower speed than ligand-based methods. Here we report a novel structure-based virtual screening method, PL-PatchSurfer2. In PL-PatchSurfer2, protein and ligand surfaces are represented by a set of overlapping local patches, each of which is represented by three-dimensional Zernike descriptors (3DZDs). By means of 3DZDs, the shapes and physicochemical complementarities of local surface regions of a pocket surface and a ligand molecule can be concisely and effectively computed. Compared with the previous version of the program, the performance of PL-PatchSurfer2 is substantially improved by the addition of two more features, atom-based hydrophobicity and hydrogen-bond acceptors and donors. Benchmark studies showed that PL-PatchSurfer2 performed better than or comparable to popular existing methods. Particularly, PL-PatchSurfer2 significantly outperformed existing methods when apo-form or template-based protein models were used for queries. The computational time of PL-PatchSurfer2 is about 20 times shorter than those of conventional structure-based methods. The PL-PatchSurfer2 program is available at <http://www.kiharalab.org/plps2/>.



INTRODUCTION

Virtual screening is a computational technique that searches a molecule library to find candidate active compounds for a target protein.¹ This technique has been widely used in drug discovery.^{2–4} It can be classified into two categories on the basis of the strategy it adopts: ligand-based virtual screening (LBVS) or structure-based virtual screening (SBVS). LBVS methods rank ligands in a library by calculating similarities between the ligands and a template active compound. Depending on how compounds are represented, LBVS methods can be further classified into those that use one-dimensional,^{5,6} two-dimensional (2D),^{7–9} or three-dimensional (3D) representations.^{10–12} LBVS methods are generally fast and require only template compounds. The fact that these methods do not require receptor structures can be considered among their advantages; however, this disregard of receptor structure weakens their ability to discover novel scaffold compounds.¹³ On the other hand, SBVS methods predict a binding pose of a ligand in a target receptor pocket and calculate the binding affinity from that pose. Thus, protein–

ligand docking^{14–17} and receptor pharmacophore search^{18,19} are common strategies for SBVS. Since it uses a receptor structure and does not require a priori knowledge of known drugs, hits found from a ligand library could be more diverse than those from LBVS. However, SBVS methods usually take more time than LBVS methods because they search conformational space and calculate binding energy. Thus, an ideal VS method would have advantages from both types of methods: it would find diverse active compounds quickly and would be applicable to target receptors that do not have experimentally solved structures.

Here we report a novel SBVS program, PL-PatchSurfer2, that is a substantial upgrade from the prototype of the program we developed earlier.²⁰ Unlike conventional SBVS methods that use atomic-detailed descriptions of molecules, PL-PatchSurfer2 uses a molecular surface representation, which has several notable advantages, as we will show in the benchmark results

Received: March 21, 2016

Published: August 8, 2016

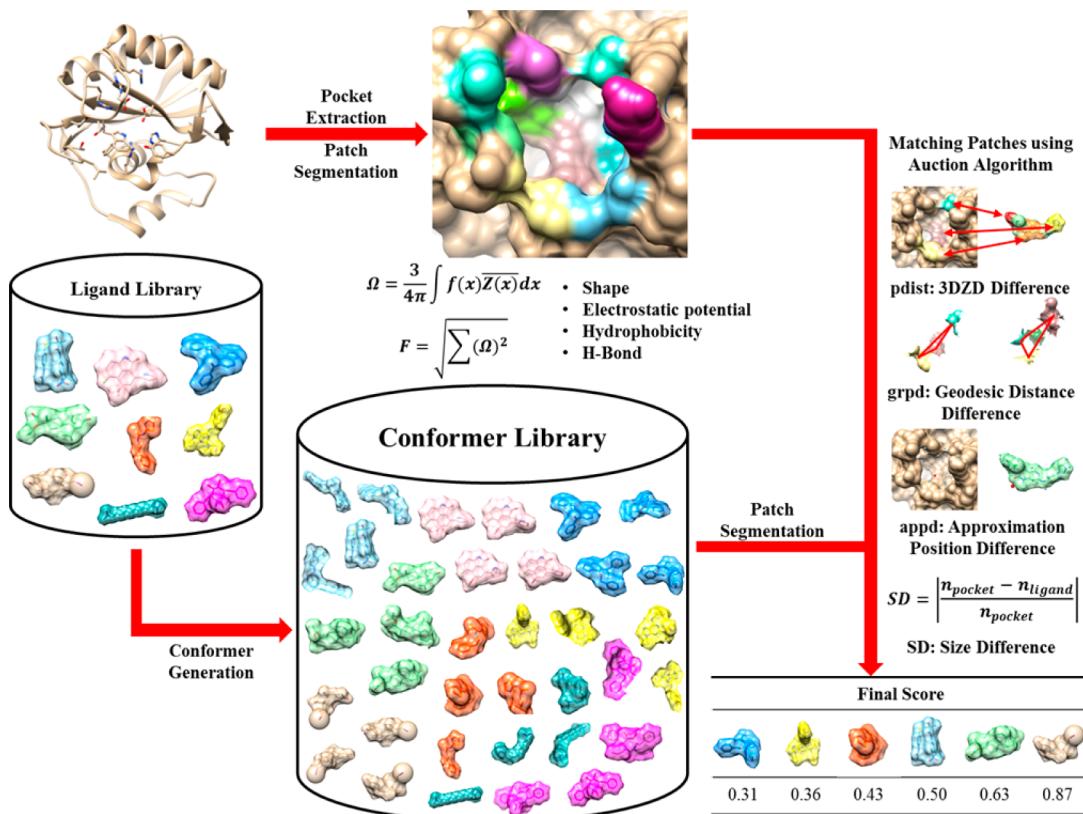


Figure 1. Workflow of PL-PatchSurfer2.

later in this report. Surfaces of a target binding pocket and a ligand are segmented into overlapping local surface patches to enable evaluation of the local compatibility of the pocket and the ligand. In addition to the surface shape, physicochemical properties are mapped onto the surface and represented with three-dimensional Zernike descriptors (3DZDs). A 3DZD is a rotation-invariant vector representation of a 3D scalar-valued function in Euclidean space (e.g., shape, properties mapped in the space) derived from a series expansion of the 3D function based on the 3D Zernike moments.^{21,22} The 3DZD representation of molecular surfaces has been applied to solve many structural biology problems, including ligand–ligand comparison,¹² pocket–pocket comparison,^{23–26} electron density maps from electron microscopy,²⁷ and protein structure similarity search.²⁸ It has been shown that representing local surface patches with multiple 3DZDs is effective for capturing conformational changes of pockets and ligands in pocket–pocket comparison and ligand–ligand comparison.^{12,25}

The largest difference between PL-PatchSurfer2 and its earlier version is that atom-based hydrophobicity and hydrogen-bond donor/acceptor positions, which are important chemical characteristics in protein–ligand interactions, have been newly implemented to describe the molecular surface in addition to the shape and surface electrostatics used in the earlier version. Moreover, weights that combine scoring terms have now been trained using the BEDROC score²⁹ as the target function rather than the area under the receiver operating characteristic (ROC) curve (AUC) used in the previous version, as AUC is known to fail to promote early recognition of active compounds.^{29,30} The performance of PL-PatchSurfer2 was tested extensively against four benchmark data sets in comparison with three popular virtual screening methods,

AutoDock Vina,¹⁶ DOCK6,¹⁷ and ROCS.¹¹ In the first test on the DUD data set,³¹ PL-PatchSurfer2 performed better than or comparably to the three other methods compared. Remarkably, when tested on a data set with apo (ligand-free) binding pocket structures and another data set with template-based computational models of receptor proteins, PL-PatchSurfer2 significantly outperformed the existing methods, with almost twice the enrichment factor at 1% (EF_{1%}) on the former set and over 5 times the EF_{1%} on the latter data set compared with AutodockVina. PL-PatchSurfer2 also showed results that were better than or comparable to those of the existing methods on the WOMBAT data set,³² which contains more diverse active compounds than the DUD set. PL-PatchSurfer2 runs about 20 times faster than AutoDock Vina and DOCK6. On the basis of the benchmark results, unique characteristics of PL-PatchSurfer2 are discussed with a focus on the possibility of expanding target proteins that can be handled by applying computational structure modeling.

METHODS

Description of PL-PatchSurfer2. Overview. PL-PatchSurfer2 is a virtual screening program that ranks ligands in a library for a target pocket by identifying complementarities between pocket and ligand local surface regions. The overall workflow of PL-PatchSurfer2 is shown in Figure 1. The first step of PL-PatchSurfer2 is to generate multiple conformations of each ligand using OMEGA³³ to consider the conformational flexibility of the ligand. Then the molecular surface of a target protein and the ligands in different conformations are generated by APBS.³⁴ Four features are assigned to the surfaces of the molecules: shape, electrostatic potential, hydrophobicity, and hydrogen-bond acceptor/donor positions. The surfaces of the

protein pocket and ligand are segmented into a number of overlapping patches, and the four physicochemical features of the patches are represented with 3DZDs. For a target pocket and a ligand molecule, surface patches are compared to identify compatible patch pairs and evaluated by a score. Then ligands in the library are ranked according to their scores. Compared with the previous version of PL-PatchSurfer (PL-PatchSurfer1), PL-PatchSurfer2 has been improved in five aspects: (1) atom-based hydrophobicity and hydrogen-bonding acceptors/donor information have been added for features of molecules; (2) the maximum number of conformations to be generated for ligands has been increased from 20 to 50; (3) weighting factors for scoring terms in the scoring function, named the *Conformer Score*, which are important for summarizing individual scoring terms for assessing the compatibility of a ligand with a target pocket, have been more thoroughly optimized; (4) the target function for optimization of the *Conformer Score* is BEDROC instead of AUC to put more emphasis on early recognition; and (5) we have devised a Boltzmann averaging method to compute the score for a ligand instead of taking the average of the top 10 scoring conformations as in the previous version. The PL-PatchSurfer2 program is available at <http://www.kiharalab.org/plps2/>. We explain each step in greater detail in the following sections.

Assigning Physicochemical Properties on the Molecule Surface. Four features—geometric shape, electrostatic potential, hydrophobicity, and hydrogen-bond acceptor/donor positions—are used to characterize molecular surfaces. The molecular surface of a protein or a ligand is computed with APBS.³⁴ It also computes the electrostatic potential on the molecular surface by solving the Poisson–Boltzmann equation. The computed shape and electrostatic potential information are mapped to 3D voxels. A protein structure is mapped onto a 3D grid such that a voxel point is marked with 1 if it overlaps with the protein surface and 0 otherwise. Likewise, the computed electrostatic potential, hydrophobicity, and hydrogen-bond acceptors/donors at each position on molecular surface are mapped at corresponding voxels.

For the newly introduced hydrophobicity in PL-PatchSurfer2, a hydrophobic field is calculated and assigned to the surface in a similar way as for the electrostatic potential. Similar to a work by Kellogg et al.,³⁵ who calculated a molecular hydrophobic field as a distance-weighted atomic logP value, we calculate the hydrophobicity of a surface grid point using a Fermi-like distance-based weight as follows:³⁶

$$\text{MHP}_i = \frac{\sum_{j=1}^N f_j [1 + e^{(\alpha r_{ij} - 4)}]^{-1}}{\sum_{j=1}^N [1 + e^{(\alpha r_{ij} - 4)}]^{-1}} \quad (1)$$

where i , j , and N are a voxel point, an atom, and the number of atoms in the molecule, respectively, f_j is the atomic hydrophobicity, and r_{ij} is the distance between voxel point i and atom j . The parameter α is set to 1.0 \AA^{-1} , so the effective distance between a voxel point and a ligand atom is 4.0 \AA . The atomic hydrophobicity f_j is computed by the XlogP3³⁷ program.

The hydrogen-bond property, another new feature in PL-PatchSurfer2, is assigned to a voxel point on the basis of character of the closest atom. If the closest atom is a hydrogen-bond donor (or acceptor), 1 (or -1) is assigned to the voxel point. If the closest atom is neither a donor nor an acceptor, 0 is assigned.

Generation of Multiple Conformations of Ligands, Protein Pocket Extraction, and Surface Segmentation. Conformations of a ligand are generated with OMEGA.³³ The maximum number of conformations is set to 50, and the maximum energy difference between the highest-energy and the lowest-energy conformations is set to 15.0 kcal/mol. The root-mean-square deviation (RMSD) cutoffs between generated conformers are set to 0.5, 0.8, and 1.0 Å for ligands with <6, 6–10, and >10 rotatable bonds, respectively.

A protein pocket surface is extracted after all four features are assigned to the protein surface. The pocket region is determined by casting rays from the center of the pocket to the pocket wall.²⁴ The surfaces of the pocket and ligands are segmented into a number of overlapping patches as follows: First, a set of centers of patches, called seed points, are spread on the surface. Seed points are selected iteratively so that points are separated by at least 3.0 Å. Then spheres of 5.0 Å radius centered at each seed point are placed, and the surface regions within the spheres are segmented. These segmented local surface regions, called patches, are represented with 3DZDs as described in the next section.

3D Zernike Descriptors. A 3DZD represents a 3D function using a series expansion based on Zernike–Canterakis basis functions.^{21,22} Here a brief introduction to the 3DZD is given. More detailed explanations of the descriptor are found in previous works.^{21,22} The Zernike–Canterakis basis function, $Z_{nl}^m(r, \theta, \varphi)$, is shown in eq 2:

$$Z_{nl}^m(r, \theta, \varphi) = R_{nl}(r)Y_l^m(\theta, \varphi) \quad (2)$$

in which $R_{nl}(r)$ is a radial function, where r is the distance from the origin, while $Y_l^m(\theta, \varphi)$ is a spherical harmonic. The integers, n , l , and m are called the order, degree, and repetition, respectively. Conditions for these numbers are $-l \leq m \leq l$, $0 \leq l \leq n$, and $(n - l)$ is an even number. Thus, the ranges of l and m are determined by the order n .

To represent the 3D molecular surface shape or a physicochemical property on the surface with a 3DZD, 3D voxels with the mapped property values are considered as a 3D function $f(x, y, z)$. The 3D Zernike moments of this function, Ω_{nl}^m , are obtained as follows (eq 3):

$$\Omega_{nl}^m = \frac{3}{4\pi} \int_{|\mathbf{x}| \leq 1} f(\mathbf{x}) \overline{Z_{nl}^m(\mathbf{x})} \, d\mathbf{x} \quad (3)$$

where $\mathbf{x} = (x, y, z)$. The rotation-invariant 3DZD, F_{nb} , can be obtained by taking the norm of the moments Ω_{nl}^m (eq 4):

$$F_{nl} = \sqrt{\sum_{m=-l}^{m=l} (\Omega_{nl}^m)^2} \quad (4)$$

The dimension of F_{nl} depends on the order n , which determines the “resolution” of the 3DZD representation. In PL-PatchSurfer, the order n is set to 15, so that n ranges from zero to 15, which makes each 3DZD a 72-dimensional vector. Thus, 3DZDs provide a compact and rotation-invariant description of 3D shapes and numerical features mapped on a 3D shape.

Matching Patches between a Target Pocket and a Ligand. To search optimal complementary patch pairs between a receptor pocket and a ligand in a certain conformation, a modified auction algorithm is used.²⁵ The algorithm minimizes the *Patch Score* (eq 5) by selecting proper patch pairs in an iterative fashion. The *Patch Score* between protein patch p and ligand patch l is computed as

$$\begin{aligned} \text{Patch Score}(p, l) = & \omega_p \times \text{pdist}(p, l) + \omega_g \times \text{grpd}(p, l) \\ & + \omega_a \times \text{appd}(p, l) \end{aligned} \quad (5)$$

In eq 5, *pdist* is a weighted sum of the Euclidean distances between the 3DZDs for *p* and *l* (eq 6):

$$\text{pdist}(p, l) = \sum_i \nu_i |\mathbf{3DZD}_{p,i} - \mathbf{3DZD}_{l,i}| \quad (6)$$

where *i* represents a physicochemical feature (shape, electrostatic potential, hydrophobicity, or hydrogen-bond acceptor/donor) and ν_i is a relative weight assigned to the corresponding feature. The weights ν were trained so that the score *pdist* can correctly identify the interacting protein patch for each query ligand patch in the benchmark data set of 195 protein–ligand complex structures from the PDBBind set.³⁸ A protein–ligand patch pair is defined as physically interacting if the distance between their patch centers is less than 3.0 Å in the crystal structure. The optimization yielded weight values of 1.0, 0.4, 0.1, and 0.2 for shape, electrostatic potential, hydrophobicity, and hydrogen-bond acceptor/donor, respectively.

The second term, *grpd* (geodesic relative position difference), is defined as the sum of geodesic distance differences of the matched pair to the other patch pairs (eq 7):

$$\text{grpd}(p, l) = \frac{1}{n_{\text{pairs}}} \sum_{(p', l') \in M} |G(p, p') - G(l, l')| \quad (7)$$

where *M* is the set of matched patch pairs in the previous iteration of the modified auction algorithm, n_{pairs} is the number of pairs, and $G(a, a')$ is a geodesic distance (i.e., the distance along the molecular surface) between the patch centers of *a* and *a'*. In the initial round of the modified auction algorithm, *grpd* is set to zero because there are no matched pairs from a previous round.

The last term in eq 5, *appd*, represents the approximate patch position difference (eq 8):

$$\text{appd}(p, l) = |\mathbf{APP}_p - \mathbf{APP}_l| \quad (8)$$

in which **APP** is a vectorized histogram of the geodesic distances from a patch center to other patch centers in the pocket or the ligand. The bin size was set to 1.0 Å. An approximate position of a patch in a pocket or a ligand, e.g., whether a patch is located in the middle or at the edge of the molecule, can be represented by this vector.

The weights in *Patch Score* (eq 5) were optimized to maximize the success rate of finding correct patch pairs from a target pocket and a ligand. The weights were trained on the PDBBind set³⁸ and tested on the ASTEX diverse set,³⁹ which consists of 85 protein–ligand complex structures. The results are shown in Table S1 in the Supporting Information. We trained weights for large pockets (whose surfaces were segmented into 40 or more patches) and small pockets (less than 40 surface patches) separately because that was found to improve the overall success rate compared with using a single set of weights for the entire pocket in the data sets.

Scoring Function To Rank Ligands for a Target Pocket. As described in the previous section, for a target pocket *P* and a conformation *C* of a ligand, complementary patches are matched using eq 5. Then the overall fitness of the ligand conformation *C* to the pocket *P*, denoted as the *Conformer Score*, is calculated as follows (eq 9):

$$\text{Conformer Score}(P, C)$$

$$\begin{aligned} = & \omega_p^c \times \text{avg}(\text{pdist}) + \omega_g^c \times \text{avg}(\text{grpd}) + \omega_a^c \times \text{avg}(\text{appd}) \\ & + \omega_s^c \times SD \end{aligned} \quad (9)$$

where $\text{avg}(x)$ is the average of a term *x* over all matched patch pairs. In the last term, the quantity *SD* is the size difference between the pocket and the ligand conformer, which is quantified by the difference between the numbers of patches covering the pocket and the conformer (eq 10):

$$SD = \begin{cases} \frac{n_{\text{pocket}} - n_{\text{ligand}}}{n_{\text{pocket}}} & \text{if } n_{\text{pocket}} \geq n_{\text{ligand}} \\ \frac{n_{\text{ligand}} - n_{\text{pocket}}}{n_{\text{ligand}}} & \text{if } n_{\text{ligand}} > n_{\text{pocket}} \end{cases} \quad (10)$$

where n_{pocket} and n_{ligand} are the numbers of patches in the pocket and the ligand conformer, respectively. This term captures the similarity of the approximate surface areas of the pocket and the conformer.

All of the weight parameters in eq 9 were determined to maximize a virtual screening performance measured by BEDROC.²⁹ The DUD set,³¹ which is composed of 40 proteins, was used to train and evaluate the virtual screening performance. As before, a grid search was used to find the optimum weight set. The weights were changed from 0.0 to 1.0 with an interval of 0.1. For each ligand molecule, up to 50 conformations were generated, each of which obtained a *Conformer Score* for a target pocket. From the scores given to the different conformers of the ligand, the final score for the ligand for that target pocket was computed in two different ways. First, the lowest *Conformer Score* (LCS) was taken as the final score for the ligand. Second, we computed the Boltzmann-weighted ligand score (BS), a weighted average of the *Conformer Scores* inspired by the Boltzmann distribution:^{40,41}

$$\text{BS}(P, L) = \frac{\sum_{C=1}^{N_{\text{conf}}} CS(P, C) \times \exp[-\beta \times CS(P, C)]}{\sum_{C=1}^{N_{\text{conf}}} \exp[-\beta \times CS(P, C)]} \quad (11)$$

where $CS(P, C)$ is the *Conformer Score* of pocket *P* and conformation *C*, N_{conf} is the number of top-scoring conformations considered in computing the score, and β is a parameter. As shown in the Results and Discussion, we tried different values of β and N_{conf} .

Benchmark Data Sets. The performance of PL-PatchSurfer2 was evaluated on four data sets: the DUD data set, a set of apo structures of targets in DUD, a set of template-based structure models of targets in DUD, and the WOMBAT data set. In this section we briefly describe these data sets.

The DUD Benchmark Set. The DUD set³¹ is composed of 40 important pharmaceutical targets. The ligand library was constructed for individual protein targets, a combination of known active compounds and decoy compounds that have physicochemical properties similar to those of the actives. Fifteen targets were removed from the original DUD set because they have cofactors, ions, or disordered regions and thus cannot be handled by APBS to compute the molecular surface and the surface electrostatic potential. They are angiotensin-converting enzyme, adenosine deaminase, catechol O-methyltransferase, phosphodiesterase 5, dihydrofolate reductase, glycinnamide ribonucleotide transformylase, aldose reduc-

tase, enoyl ACP reductase, glycogen phosphorylase β , purine nucleoside phosphorylase, S-adenosyl-homocysteine hydrolase, hydroxymethylglutaryl-CoA reductase, trypsin, human shock protein 90, and thymidine kinase. The remaining 25 targets were divided into two sets to perform a twofold cross-validation to train and test the weighting parameters. The full list of training and test set targets can be found in Table S2. To speed up the virtual screening process, the ratio of the number of actives to the number of decoys was changed to 1:29 from the original 1:36 by randomly selecting decoy compounds from the library. If the number of molecules in the whole library was higher than 3000, the library was reconstructed by selecting 60 active compounds and 1740 decoy compounds randomly from the original DUD library.

Apo Structure Data Set. In addition to the holo (target-bound) forms of the target proteins in DUD, we also tested PL-PatchSurfer2 on the apo (ligand-free) forms of the targets. Using the apo form of a target pocket is generally difficult for virtual screening because pockets change their shape upon binding.⁴² Out of the 25 targets, 19 have available crystal structures in the apo form. The Protein Data Bank (PDB) IDs of the apo forms are listed in Table S2.

Template-Based Structure Model Data Set. We also tested the performance of PL-PatchSurfer2 on computational models of target proteins. The applicability of SBVS can be dramatically enlarged if a SBVS method can show sufficiently good performance with computational protein structure models.^{42–44}

Particularly, because of the expansion of the PDB,⁴⁵ an increasing number of protein structures can be modeled using template-based modeling (TBM) methods. For each of the 25 targets in the DUD set, we used PSI-BLAST⁴⁶ to find proteins homologous to the targets in the PDB. For 18 targets, a template with a sequence identity of over 30% and an alignment coverage of over 90% relative to the target was found. Moreover, among them, for six targets, three templates, one each at sequence identities of 30–50%, 50–70%, and 70–90% with coverages of over 90% were found. The identified templates are listed in Table S2. From these templates, template-based models were constructed using Modeller9v11.⁴⁷

The WOMBAT Data Set. The last data set used was the WOMBAT data set. This data set is intended to collect more diverse active compounds than the DUD data set.³² In the WOMBAT data set, active compounds that have lead-like conditions ($\log P < 4.5$ and molecular weight < 450) were collected mainly from the literature, and similar compounds were removed by clustering using 2D graph comparison.³² Thus, the resulting data set is more diverse than the DUD set and therefore more suitable for testing the scaffold-hopping ability of virtual screening methods.

We used nine targets out of 13 proteins in the WOMBAT set that overlap with the DUD set (Table S3). Since the WOMBAT set provides only active compounds, we supplemented the data set with decoy compounds from the DUD set. The ratio of actives to decoys was set to 1:29, the same as in the DUD data set we used, by random selection of compounds, except for the cases where there were not enough decoy compounds available. In the three such cases, all of the decoy compounds were used.

Existing Virtual Screening Methods Compared. PL-PatchSurfer2 was compared with its original version, PL-PatchSurfer1,²⁰ as well as three popular virtual screening programs: AutoDock Vina,¹⁶ DOCK6,¹⁷ and ROCS.¹¹ The former two programs are SBVS methods, while ROCS is an

LBVS method. Since ROCS performs LBVS, which does not use target protein structures, its performance was tested only on the DUD and WOMBAT data sets.

To run the four SBVS programs, including PL-PatchSurfer1 and -2, the binding pocket of a target was defined from the geometric center of the cocrystallized ligand. For the apo and template-based model data sets, the target structure was aligned with the holo-form structure in the DUD set using TM-align,⁴⁸ and the cognate ligand copied to the target structure from the holo structure was used to determine the geometric center of the ligand, from which a binding pocket was defined. For ROCS, which needs a template (query) ligand for screening, we used the same multiple conformers of the native ligand in the holo structure generated by OMEGA for running PL-PatchSurfer2.

Protonation states and atomic charges of proteins and ligands were kept the same as the original assignments in the DUD and WOMBAT data sets. For apo and template-based structures, hydrogen atoms and Gasteiger charges were added using the DockPrep module of Chimera.⁴⁹

To run AutoDock Vina, the docking box was defined as a 25 Å³ cube centered at the geometric center of the cognate ligand. For DOCK6, the docking box was constructed to have a 5.0 Å margin from the space occupied by a set of spheres placed within 10.0 Å from the cognate ligand atoms. The other parameters of AutoDock Vina and DOCK6 were set to their default values.

Evaluation Metrics. The performance of virtual screening was evaluated by the enrichment factor (EF), the area under the ROC curve (AUC), and BEDROC. The EF at the top $x\%$ subset is calculated as follows:

$$EF_{x\%} = \frac{Actives_{x\%}/Compounds_{x\%}}{Actives_{Library}/Compounds_{Library}} \quad (12)$$

where $Actives_{x\%}$ is the number of actives found in the top $x\%$ score-ranked compounds, $Compounds_{x\%}$ is the total number of compounds within the top $x\%$ subset, and $Actives_{Library}$ and $Compounds_{Library}$ are the number of active compounds and total number of compounds in the library, respectively. Random retrieval would yield $EF_{x\%} = 1$. We examined EFs at 1%, 5%, and 10%.

We also used the AUC, which shows the performance of a virtual screening program by plotting the true positive rate (rate of finding active compounds) relative to the false positive rate (rate of finding decoy compounds). The AUC is 0.5 for random retrieval and has a maximum possible value of 1.0, which is obtained when the program finds all of the active compounds before ranking any decoy compounds.

BEDROC assigns larger weights to actives detected earlier in a rank to reward early detection of actives:^{29,30}

$$\text{BEDROC} = \frac{\sum_{i=1}^N e^{-\alpha r_i/N}}{R_a \left(\frac{1-e^{-\alpha}}{e^{\alpha/N}-1} \right)} \times \frac{R_a \sinh\left(\frac{\alpha}{2}\right)}{\cosh\left(\frac{\alpha}{2}\right) - \cosh\left(\frac{\alpha}{2} - \alpha R_a\right)} + \frac{1}{1 - e^{\alpha(1-R_a)}} \quad (13)$$

where N is the number of compounds, n is the number of hits, $R_a = n/N$, r_i is the ranking of the i th hit, and α is a weight parameter, which was set to 20 as suggested by Truchon and Bayly.²⁹

RESULTS AND DISCUSSION

Improvements in PL-PatchSurfer2 over PL-PatchSurfer1. First, we compare PL-PatchSurfer2 (PL-PS2) with its previous version, PL-PatchSurfer1 (PL-PS1). As shown in Table 1, we investigated the effect of each of the following three

Table 1. Comparison between PL-PatchSurfer1 and PL-PatchSurfer2^a

method	EF _{1%}	EF _{5%}	EF _{10%}	AUC	BEDROC
PL-PS1	10.85	3.65	2.39	0.476	0.225
PL-PS1 + HP	12.72	4.54	2.67	0.530	0.266
PL-PS1 + HB	13.24	4.65	2.79	0.535	0.276
PL-PS1 + HB + HP	13.41	4.68	2.80	0.537	0.280
PL-PS2 (LCS)	13.74	4.71	2.84	0.539	0.280

^aA hydrophobicity term (HP) and a hydrogen-bond term (HB) were added one by one to the original PL-PatchSurfer1 (PL-PS1). The last row, PL-PatchSurfer2 (PL-PS2) further uses a maximum of 50 ligand conformations rather than 20, which was originally used in PL-PS1. The *p* values for differences between the methods are provided in Table S4.

new main developments of PL-PatchSurfer2 relative to PL-PatchSurfer1: (1) addition of a hydrophobicity feature; (2) addition of a hydrogen-bond feature; and (3) an increase in the maximum number of ligand conformations generated from 20 to 50. The results were obtained for the 25 targets in the DUD set. The lowest *Conformer Score* (LCS) (eq 9) was used for PL-PatchSurfer2.

All four new schemes showed improvements over PL-PatchSurfer1 in all five metrics. Adding the hydrogen-bond term (HB) was more effective than adding the hydrophobicity feature (HP). Increasing the number of ligand conformations showed improvement on average (PL-PS2 over PL-PS1 + HP + HB). Table S4 shows the *p* values of pairwise comparisons. On the basis of the Bonferroni correction for 10 comparisons among five methods, which takes the *p* < 0.1 level to *p* < 0.01 (=0.1/10 comparisons), the improvement by PL-PatchSurfer2

over the original PL-PatchSurfer1 was significant by all of the other variations for EF_{5%}, AUC, and BEDROC. In terms of EF_{1%}, the performances of PL-PS1 + HB + HP and PL-PS2 were significantly different compared with PL-PS1.

Among the 25 targets, the largest improvement by PL-PatchSurfer2 was observed for RXR α in terms of BEDROC (PL-PS1, 0.029; PL-PS1 + HB, 0.291; PL-PS1 + HP, 0.291; PL-PS1 + HB + HP, 0.301; PL-PS2, 0.312). This is probably due to the new implementation of hydrophobicity and hydrogen-bond terms in PL-PatchSurfer2 because these two are the main interactions between RXR α and its ligands (Figure S1 in the Supporting Information).

Performance on the DUD Data Set. Twenty-five targets of the DUD benchmark set were divided into two subsets (sets 1 and 2) to perform a twofold cross-validation (Table 2). Training of weights for the *Conformer Score* (eq 9) using sets 1 and 2 yielded ($\omega_p^c, \omega_g^c, \omega_w^c, \omega_s^c$) = (0.6, 0.4, 0.7, 0.6) and (0.4, 0.3, 0.5, 0.3), respectively. Table 2 summarizes the results for PL-PatchSurfer2 on set 1, set 2, and the whole set in comparison with Autodock Vina, DOCK6, and ROCS. For PL-PatchSurfer2, two scoring functions were used: LCS (eq 9) and the *Boltzmann-Weighted Ligand Score* (BS) (eq 11). Results for individual targets are provided in Table S5.

Comparing the two scoring schemes of PL-PatchSurfer2, BS showed better performance than LCS in all of the metrics used. This improvement is consistent with the results of a previous study that used the Boltzmann weighting scheme for scoring.⁴⁰ The performance of BS with different values of the parameters β and N_{conf} is provided in Table S6. We used $\beta = 1$ and $N_{\text{conf}} = 50$ because this combination gave the largest BEDROC and EF_{1%} among the parameter combinations tested. For a fixed value of β , the performance improved as N_{conf} increased. For a fixed value of N_{conf} using the largest β value ($\beta = 100$) gave the worst results except for one case (BEDROC, $N_{\text{conf}} = 5$).

Subsequently, we compared the performance of PL-PatchSurfer2 with those of two SBVS methods, AutoDock Vina and DOCK6, and an LBVS method, ROCS (see the All

Table 2. Results on the DUD Benchmark Data Set^a

	EF _{1%}	EF _{5%}	EF _{10%}	AUC	BEDROC
Set 1 (12 Targets)					
PL-PS2 (LCS)	12.18 (13.22)	4.27 (4.33)	2.66 (2.74)	0.551 (0.541)	0.254 (0.260)
PL-PS2 (BS)	12.51 (13.93)	4.51 (4.67)	2.76 (2.81)	0.573 (0.564)	0.266 (0.277)
AutoDock Vina	9.54	5.78	3.84	0.659	0.317
DOCK6	13.10	4.15	2.50	0.518	0.253
ROCS	14.74	6.86	4.37	0.690	0.393
Set 2 (13 Targets)					
PL-PS2 (LCS)	15.18 (15.33)	5.12 (5.30)	3.00 (3.04)	0.527 (0.546)	0.303 (0.307)
PL-PS2 (BS)	16.90 (16.58)	5.80 (5.73)	3.39 (3.54)	0.581 (0.599)	0.346 (0.345)
AutoDock Vina	6.42	4.38	2.95	0.606	0.238
DOCK6	9.96	3.90	2.45	0.486	0.227
ROCS	9.01	4.33	2.75	0.529	0.246
All Proteins (25 Targets)					
PL-PS2 (LCS)	13.74	4.71	2.84	0.539	0.280
PL-PS2 (BS)	14.79	5.18	3.09	0.577	0.308
AutoDock Vina	7.92	5.05	3.37	0.632	0.276
DOCK6	11.47	4.02	2.47	0.502	0.239
ROCS	11.76	5.54	3.52	0.606	0.316

^aFor results of PL-PatchSurfer2 on sets 1 and 2, values obtained when the data set was used as the testing set are primarily shown, while values obtained when the data set was used for training are reported in parentheses. The overall set results reported for PL-PatchSurfer2 are averages of the 25 targets when they were handled as test data.

Proteins results in Table 2). Table S7 shows the statistical significance of the performance differences of the methods observed in Table 2 (i.e., p values of pairwise Student's t test for EF_{1%}, AUC, and BEDROC). PL-PatchSurfer2 showed the best performance at EF_{1%}, while ROCS performed best on three metrics, EF_{5%}, EF_{10%}, and BEDROC. Statistically speaking, PL-PatchSurfer2's performance at EF_{1%} was significantly better than that of AutoDock Vina at the $p < 0.1$ level (which corresponds to $p < 1.67 \times 10^{-2}$ when the Bonferroni correction is applied, where $1.67 \times 10^{-2} = 0.1/6$, in which 6 is the number of pairwise comparisons between four methods). The distinction of PL-PatchSurfer2 is large at EF_{1%}, and the values for all of the programs became similar as the cutoff for the enrichment increased. In terms of AUC, AutoDock Vina showed the best performance. These results indicate that PL-PatchSurfer2 has an advantage in the early detection of actives.

We further benchmarked the methods by dividing the DUD data set on the basis of the average dissimilarity between the active compounds in the library and the cognate ligands in the crystal structures of the targets. The dissimilarity was calculated by SIMCOMP,⁹ which compares 2D structures of compounds as graphs. With a dissimilarity threshold of 0.75, the DUD data set was split into two groups containing 12 and 13 targets having average dissimilarity values of <0.75 or ≥ 0.75 , respectively. Furthermore, five targets having dissimilarity values of ≥ 0.8 were selected. The box plots of EF_{1%}, AUC, and BEDROC on the three subsets are reported in Figure 2.

Comparing the results on the three subsets (Figure 2), the performance of AutoDock Vina, DOCK6, and ROCS declined sharply as the dissimilarity of the active compounds increased to 0.8 to 1.0 in all three metrics. The drop in performance of ROCS was expected because LBVS programs identify compounds similar to the template (query) active compound.⁵⁰ In contrast, notably the performance of PL-PatchSurfer2 did not deteriorate; rather, it showed even better results as the dissimilarity increased on these data sets. The stable performance of PL-PatchSurfer2 would be due to its local surface matching between a target pocket and a ligand, which is more tolerant to global shape differences than other virtual screening programs.

An example where PL-PatchSurfer2 showed better performance than the other three methods is shown in Figure 3. PL-PatchSurfer2 demonstrated a higher earlier active compound detection rate than the other three methods for this target, HIVPR (Figure 3A), which was also reflected by larger BEDROC scores: the BEDROC values for PL-PatchSurfer2 (LCS), PL-PatchSurfer2 (BS), AutoDock Vina, DOCK6, and ROCS were 0.503, 0.559, 0.286, 0.121, and 0.064, respectively. The three active compounds shown in Figure 3C are highly dissimilar to the target's template compounds (Figure 3B), and thus, AutoDock Vina, DOCK6, and ROCS could not retrieve them at an earlier rank. ZINC03815630 was ranked by AutoDock Vina, DOCK6, and ROCS at very low ranks (152nd, 1535th, and 614th, respectively). Similarly, ZINC03833855 was ranked 1587th, 1549th, and 1547th and ZINC03815642 was ranked 405th, 290th, and 525th by AutoDock Vina, DOCK6, and ROCS, respectively. In contrast, unlike the three methods, PL-PatchSurfer2 ranked them very high: ZINC03815630, third by LCS and 29th by BS; ZINC03833855, eighth by LCS and fourth by BS; ZINC03815642, 54th by LCS and 28th by BS.

Benchmark Results on the Apo Structure Data Set.

Considering protein structure flexibility is important for

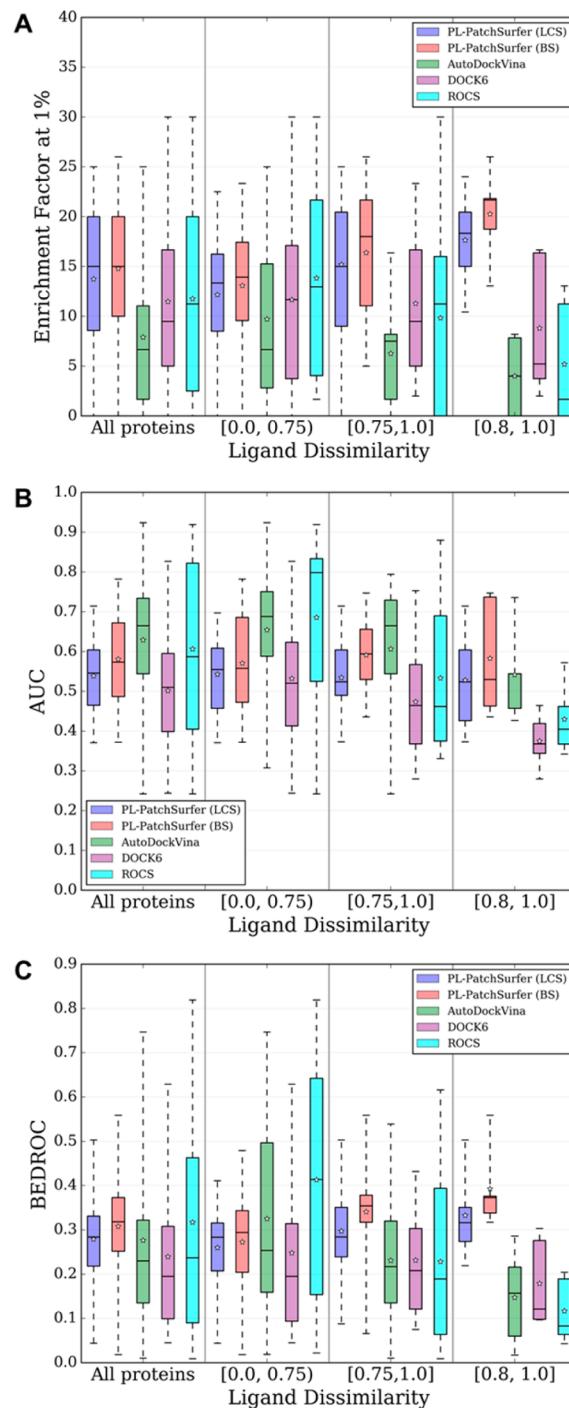


Figure 2. Performance depending on the dissimilarity between the active compounds and the cocrystallized ligand of a target, as measured by (A) EF_{1%}, (B) AUC, and (C) BEDROC. Three subsets from the DUD targets were used, those with dissimilarities of <0.75 , ≥ 0.75 , and ≥ 0.8 . The values were taken from the All Proteins section in Table 2. Average values are marked as white stars on the boxes.

protein–ligand docking, not only for predicting binding poses⁵¹ but also for virtual screening.⁵² Since a binding pocket changes its conformation upon docking of a ligand, it is more difficult in general to identify active compounds if a target is in its apo form. To examine how receptor structure changes affect the virtual screening performance, we benchmarked using 19 apo target structures. The results are summarized in Table 3,

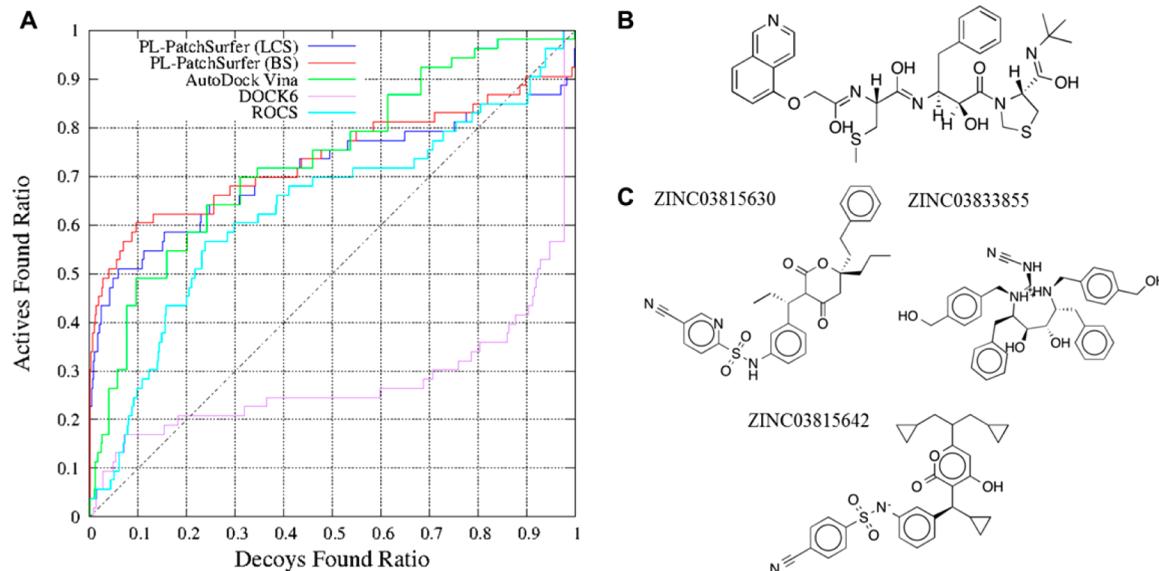


Figure 3. Screening results for HIVPR. (A) ROC curves for the five methods. (B) Structure of the cognate ligand of the target (PDB ID 1HPX). (C) Structures of three active compounds, ZINC03815630, ZINC03833855, and ZINC03815642. The dissimilarities of these three compounds from the template ligand are 0.805, 0.872, and 0.852, respectively.

Table 3. Virtual Screening Benchmark Results on the Apo Structure Data Set (19 Targets); The Results for the Same Targets in Their Holo Forms Are Also Shown for Comparison^a

	EF 1%	EF 5%	EF 10%	AUC	BEDROC
Apo Structures					
PL-PS2 (LCS)	10.51 (14.11)	3.98 (5.26)	2.58 (3.60)	0.532 (0.596)	0.245 (0.340)
PL-PS2 (BS)	11.67 (17.78)	4.42 (6.55)	2.84 (4.18)	0.559 (0.654)	0.267 (0.399)
AutoDock Vina	3.86 (1.33)	2.55 (3.41)	2.00 (2.67)	0.549 (0.604)	0.155 (0.180)
DOCK6	7.69 (1.67)	3.58 (1.60)	2.48 (1.32)	0.511 (0.402)	0.212 (0.103)
Holo Structures					
PL-PS2 (LCS)	12.90 (19.67)	4.74 (7.13)	2.88 (4.20)	0.557 (0.642)	0.278 (0.419)
PL-PS2 (BS)	14.00 (19.22)	5.07 (7.74)	3.02 (4.89)	0.577 (0.692)	0.299 (0.451)
AutoDock Vina	7.35 (10.78)	4.95 (8.22)	3.39 (5.11)	0.653 (0.780)	0.272 (0.417)
DOCK6	12.49 (18.44)	4.60 (6.19)	2.86 (3.84)	0.548 (0.516)	0.270 (0.367)

^aThe numbers in parentheses are averages for three apo-form targets whose binding sites have a larger than 2.0 Å Cα RMSD from their holo forms.

and the detailed results are shown in Table S8. The *p* values for the pairwise method comparison results are shown in Table S9.

As expected, the performance of all of the methods deteriorated for the apo structures compared with the holo structure results. Remarkably, however, PL-PatchSurfer2 did not deteriorate greatly when compared with the other two conventional SBVS programs. For example, the EF_{1%} values of AutoDock Vina and DOCK6 decreased over 40% (AutoDock Vina from 7.35 to 3.86; DOCK6 from 12.49 to 7.69), while PL-PatchSurfer2 with the two scoring schemes dropped only about 20% (LCS from 12.90 to 10.51; BS from 14.00 to 11.67). In terms of EF_{1%}, PL-PatchSurfer2 is significantly better than the other two programs at the *p* < 0.1 level, even after Bonferroni correction is applied (the *p* value becomes $3.33 \times 10^{-2} = 0.1/3$ when the correction for three comparisons is applied). For AUC, the three programs are distinguishable, but for BEDROC, PL-PatchSurfer2 is significantly better than AutoDock Vina (Table S9). The differences in EF_{1%}, BEDROC, and AUC observed for the holo and apo forms of the individual targets are shown in Figure S2. Overall, the performance difference between the apo and holo forms tends to increase as the binding site RMSD between the two forms becomes larger for AutoDock Vina and DOCK6. In contrast,

PL-PatchSurfer2 is more tolerant of structural changes in the binding pockets. When the three methods were compared, PL-PatchSurfer2 clearly showed the highest performance for all of the metrics.

In Table 3, we have also reported results for three targets (RXR α , HIVPR, and SRC) that underwent large conformational changes of the binding site in going from the apo form to the holo form (>2 Å C α RMSD). The average values for the three proteins are shown in the parentheses in Table 3. For AutoDock Vina and DOCK6, EF_{1%} decreased by more than 90% (AutoDock Vina from 10.78 to 1.33; DOCK6 from 18.44 to 1.67) and BEDROC by over 50% (AutoDock Vina from 0.417 to 0.180; DOCK6 from 0.367 to 0.103). In contrast, PL-PatchSurfer2 with the LCS and BS scoring schemes retained over 70% and 90% of the performance, respectively.

Figure 4 illustrates a typical situation that highlights the performance difference between PL-PatchSurfer2 and the other two methods on the apo form of a target binding site. It shows SRC kinase, which exhibits a large conformational change between the two forms (2.14 Å C α RMSD). This large conformational change takes place in two loops, the glycine-rich flap and the activation loop, in the same way as for other kinases.^{52,53} When the two structures are superimposed, the

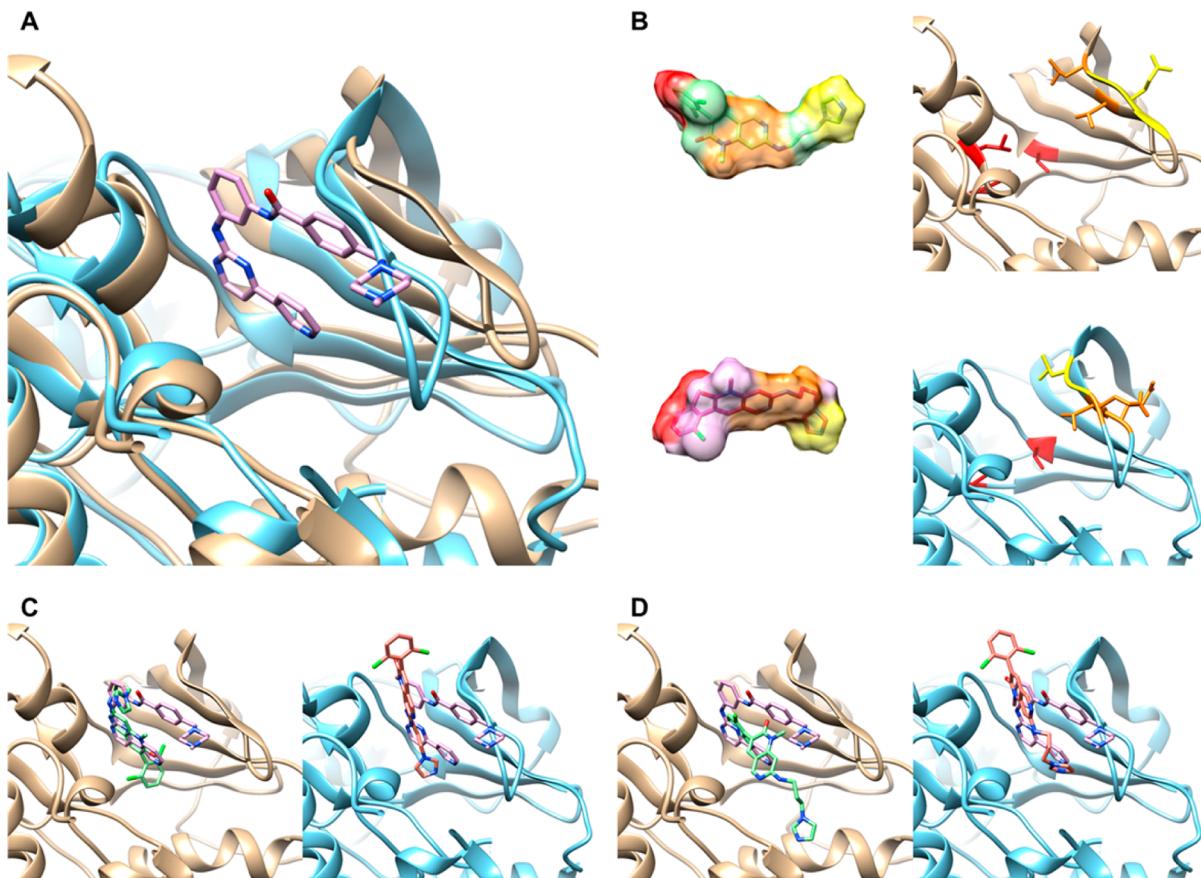


Figure 4. (A) Structure superimposition of SRC holo (gold) and apo (cyan) structures. The cognate ligand from the holo structure is colored in pink. (B) Top-scored conformation of an active compound, ZINC03815530, by PL-PatchSurfer2 with the LCS scoring scheme. The two upper panels are from the holo results, while the two lower panels are from the apo results. Matching local regions between the compound and the binding pocket in the receptor are indicated in the matching colors. (C) Binding pose predicted by AutoDock Vina on the holo (left panel, the predicted pose is in green) and the apo (right panel, the predicted pose is in magenta) structures. The compound in pink is the cognate ligand of the holo structure. (D) Binding pose predicted by DOCK6 on the holo (left) and apo (right) structures. The color codes are the same as in (C).

Table 4. Virtual Screening Benchmark on the Template-Based Models of 18 DUD Targets; The Results for the Holo Form of the Same Targets Are Shown in the Bottom Half of the Table^a

	EF 1%	EF 5%	EF 10%	AUC	BEDROC
Template-Based Models					
PL-PS2 (LCS)	11.48 (9.16)	4.40 (4.56)	2.70 (2.95)	0.540 (0.610)	0.255 (0.255)
PL-PS2 (BS)	12.02 (10.69)	4.72 (5.11)	2.90 (3.28)	0.563 (0.642)	0.280 (0.297)
AutoDock Vina	1.86 (1.68)	1.39 (1.30)	1.21 (1.30)	0.478 (0.556)	0.089 (0.092)
DOCK6	6.91 (2.58)	3.07 (1.88)	2.06 (1.58)	0.484 (0.494)	0.180 (0.099)
Holo Structures					
PL-PS2 (LCS)	12.51 (10.91)	4.70 (4.78)	2.85 (3.10)	0.554 (0.606)	0.274 (0.278)
PL-PS2 (BS)	13.67 (11.97)	4.96 (5.24)	2.97 (3.29)	0.573 (0.634)	0.295 (0.306)
AutoDock Vina	7.34 (8.63)	5.00 (6.14)	3.42 (4.09)	0.654 (0.730)	0.273 (0.331)
DOCK6	12.90 (11.70)	4.69 (4.40)	2.92 (2.98)	0.554 (0.569)	0.277 (0.266)

^aFor targets with multiple templates in Table S2, the template with the lowest sequence identity to the target was used for modeling. The numbers in parentheses are the averages for models of the nine targets that were built using a template with 30% sequence identity.

cognate ligand in the holo structure causes steric crashes with the glycine-rich flap of the apo structure (Figure 4A). This makes it difficult for AutoDock Vina and DOCK6 to find active compounds with the target's apo form. Substantial drops of EF_{1%} and BEDROC were observed for AutoDock Vina and DOCK6 on this target between the holo form and the apo form: EF_{1%} dropped from 3.33 to 0.0 (-3.33) and 23.33 to 5.00 (-18.33) for AutoDock Vina and DOCK6, respectively, and BEDROC dropped from 0.217 to 0.069 (-0.149) and 0.432 to

0.196 (-0.236) for AutoDock Vina and DOCK6, respectively. In contrast, PL-PatchSurfer2 did not exhibit much performance deterioration and in fact often somehow exhibited better performance for the apo form. With the LCS/BS scoring scheme, the EF_{1%} values for the holo and apo forms were 25.00/21.67 and 23.33/23.33, respectively, and the BEDROC values were 0.441/0.437 and 0.447/0.440 (LCS/BS) for the holo and apo forms, respectively.

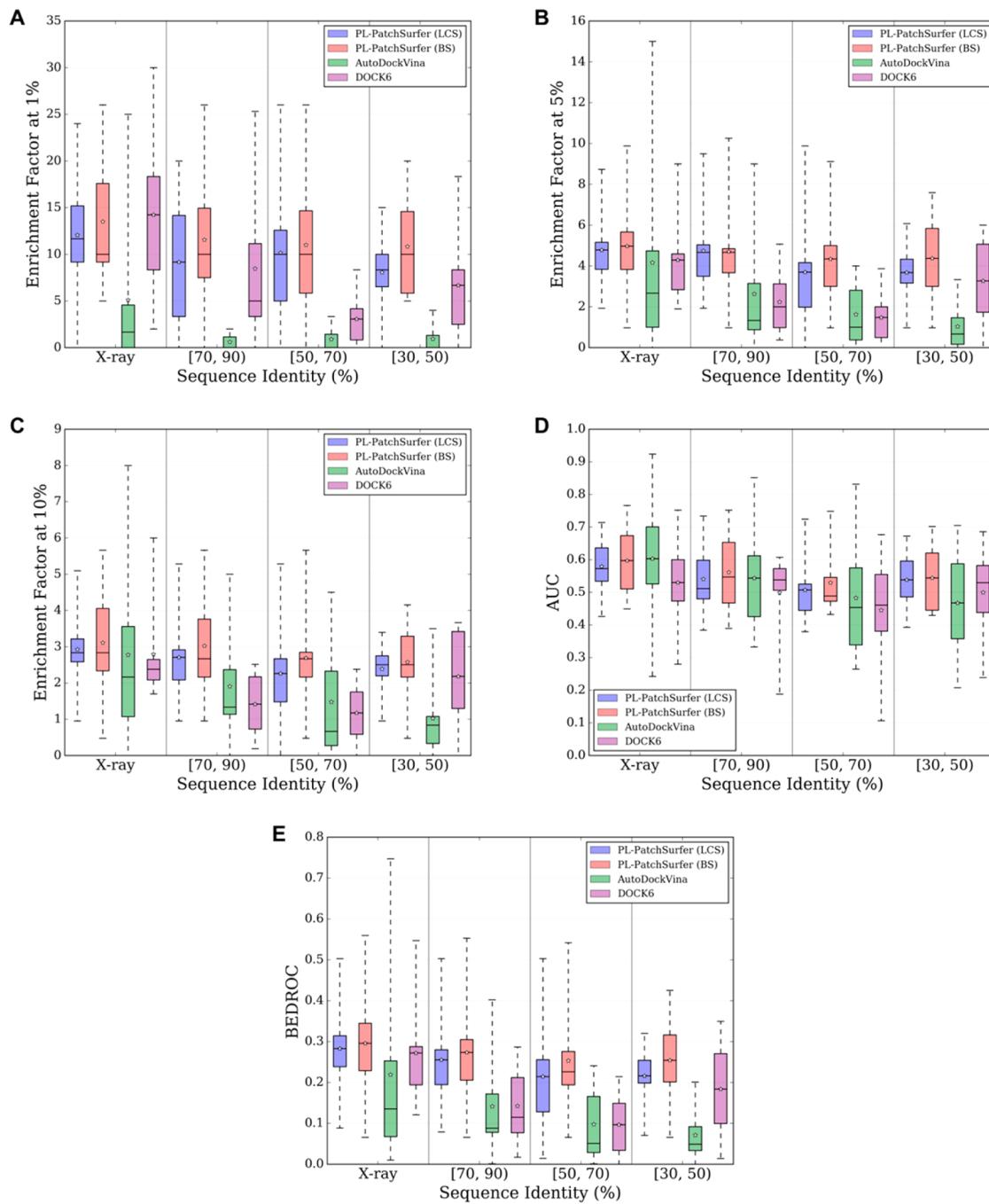


Figure 5. Box plots of virtual screening results for template-based models constructed from templates with varying sequence identity to the targets, as measured by (A) EF_{1%}, (B) EF_{5%}, (C) EF_{10%}, (D) AUC, and (E) BEDROC. Averages over six targets are shown. For comparison, results on the holo structures of the targets are also shown (X-ray, on the left). On each box, the mean value is shown with a white star.

Figure 4B–D shows the docking poses of an active compound, ZINC03815530, predicted by the three methods for the holo and apo forms. This compound was ranked within the top 5% by the three methods when the holo structure was used: it was ranked at 70th, 92nd, 83rd, and 40th by PL-PatchSurfer2 (LCS), PL-PatchSurfer2 (BS), AutoDock Vina, and DOCK6, respectively. However, when the apo structure was used, the ranks changed to 98th, 48th, 1692nd, and 1040th, respectively, showing deterioration with AutoDock Vina and DOCK6 in contrast to PL-PatchSurfer2. PL-PatchSurfer2 found similar compatible surface patch pairs between the compound and the binding site in the holo form (upper panels)

and the apo form (the lower panels) despite a large conformational change between them (Figure 4B). In the cases of AutoDock Vina and DOCK6, although the predicted docking poses for the holo-form binding site were reasonably good (left panels in Figure 4C,D), the predicted poses for the apo form were toward the outside of the binding pocket because of the conformation change in the glycine-rich flap, which blocked the compound from being placed inside the binding pocket (right panels in Figure 4C,D).

Results on the Template-Based Structure Model Data Set. Next we examined the performance of the SBVS methods on template-based models of target proteins. Using computa-

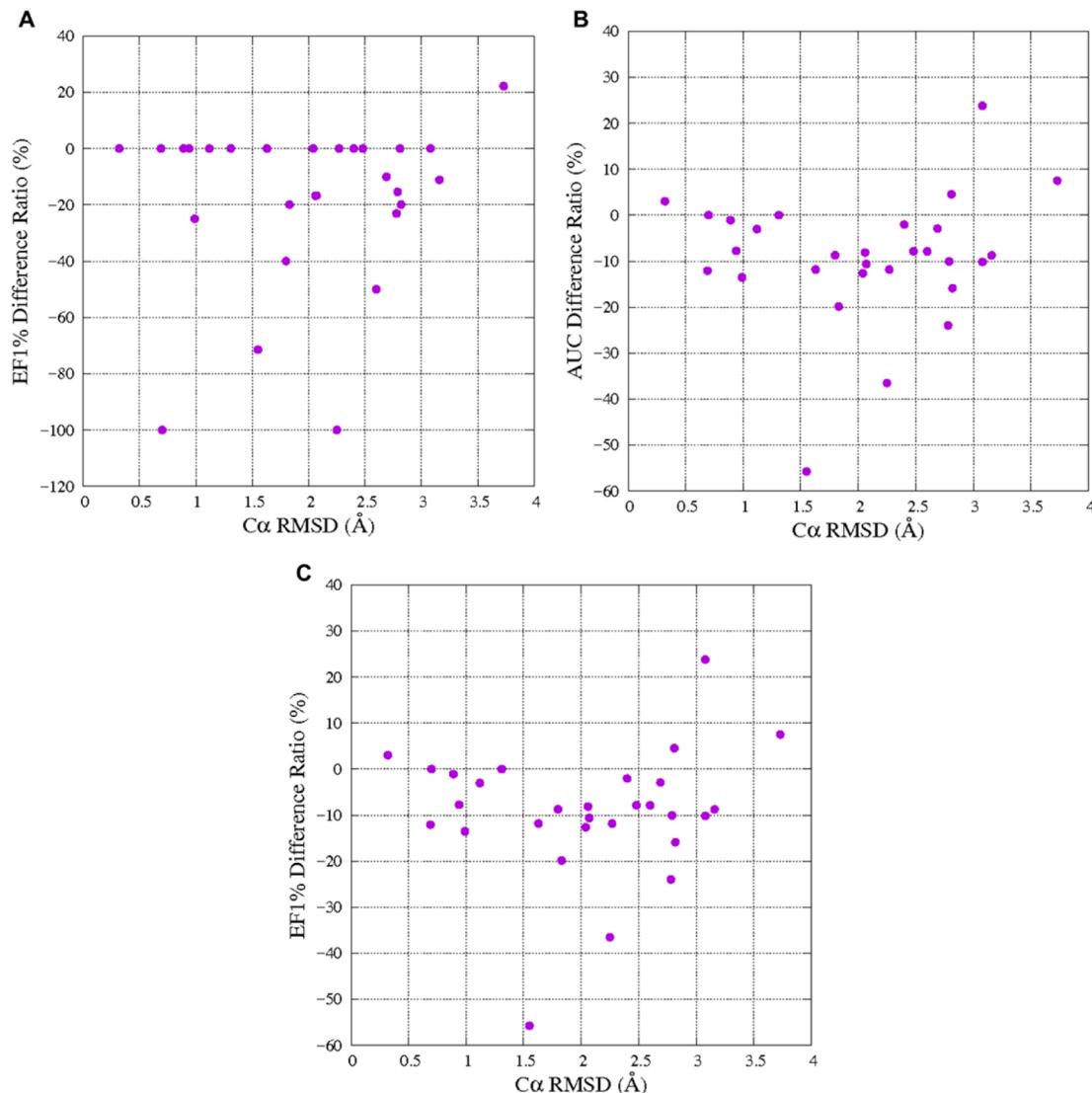


Figure 6. Performance difference between the holo and template-based target models (model – holo)/holo, as measured by (A) EF_{1%}, (B) AUC, and (C) BEDROC. The x axis is the binding-site C_α RMSD between the models and the target in its holo form. PL-PatchSurfer2 (BS) was used.

tional models for target proteins can substantially enlarge the application of SBVS methods, particularly for receptors whose structures are difficult to solve by experimental means,^{42–44} such as G-protein-coupled receptors.⁵⁴ However, using models is a challenge for SBVS methods because the models are not perfectly accurate.⁴³ As shown in Table S2, 18 DUD targets having appropriate template structures available were modeled using Modeller9v11.⁴⁷ The VS results are shown in Table 4, and Table S10 gives results for individual targets. The statistical significance of the performance differences of the three methods is provided in Table S11.

As observed for the previous data sets, the BS scheme showed higher values than the LCS scheme for all of the metrics, and both scoring methods of PL-PatchSurfer2 outperformed AutoDock Vina and DOCK6. Compared with the performance on the holo structures, PL-PatchSurfer2 retained about 90% of its performance on computational models for all of the metrics, while the performance of the other two programs decreased by more than 40%. The better performance of PL-PatchSurfer2 against AutoDock Vina is statistically significant (Table S11) at the $p < 0.05$ level ($p < 0.05$ becomes $p < 1.67 \times 10^{-2}$ when the Bonferroni correction

is applied) for EF_{1%}, AUC, and BEDROC and significant for EF_{1%} and BEDROC against DOCK6.

To further understand how the quality of the structure model influences the virtual screening performance of each program, we constructed three models of different qualities for each of the six target proteins using templates with sequence identities to the target in the ranges of 70–90%, 50–70%, and 30–50%. The results are presented in Figure 5 and Table S12.

Overall, the performance of all of the methods deteriorated as more distant templates were used for target structure modeling. The methods showed relatively stable performance in AUC (Figure 5D). The severity of the performance decline differed among the methods. Compared with AutoDock Vina and Dock6, which exhibited performance drops, the results from PL-PatchSurfer2 did not decline much, demonstrating more tolerance to inaccuracy of structure models of targets.

As shown in Figure 6, we further examined the performance of PL-PatchSurfer2 on template-based models from a different angle, relative to the RMSDs of the models. Thirty template-based models built with template structures in Table S2 were used. The RMSD of the model (x axis) and the drop in performance (y axis) do not show a clear correlation,

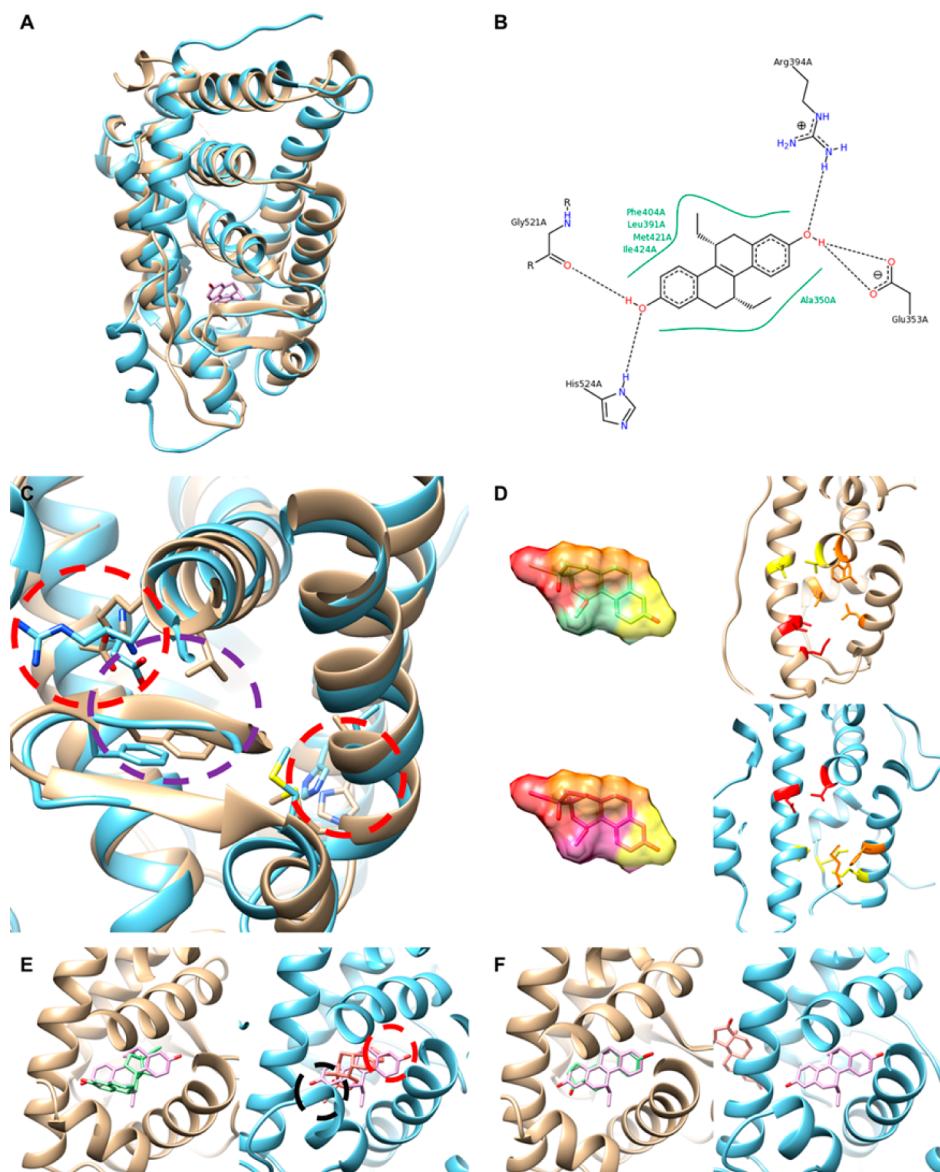


Figure 7. Example of ligand docking to a template-based model. (A) Structure superimposition of ER- α holo (gold) and the template-based model (cyan). The cognate ligand from the holo structure is shown in pink. (B) The cognate ligand, (R,R)-5,11-cis-diethyl-5,6,11,12-tetrahydrochrysene-2,8-diol (PDB Ligand ID ETC) in the holo structure. This diagram was drawn with PoseView.⁵⁶ (C) Side chains of the binding site of ER- α . Two regions that form hydrogen bonds with the cognate ligand are marked by red circles, and a region for hydrophobic interactions is marked by a purple circle. (D) Top-scoring conformations of ZINC01999257 by the PL-PatchSurfer2 LCS scoring scheme. The two upper panels are results using the holo structure, while the lower panels are results using a template-based model of the target. Local regions that match between the compound and the binding pocket are indicated in matching colors. (E) Binding pose predicted for ZINC01686128 by AutoDock Vina in the holo form (left, the compound in the predicted pose is in green) and a template-based model (right, the predicted pose is in magenta). The compound forms hydrogen bonds at one end (black circle) but no bonds were formed on the other end (red circle). (F) Binding pose predicted for ZINC03815416 by DOCK6 on the holo structure (left) and a model (right). The colors of the compounds are the same as in (E).

particularly in terms of EF_{1%} (Figure 6A). The performance of PL-PatchSurfer2 did not deteriorate in terms of EF_{1%} for a substantial number of models (16), which included one with an RMSD of 3.73 Å for which PL-PatchSurfer2 even somehow worked better than for the holo form of the target. If we consider that a 10% decrease in EF_{1%}, AUC, or BEDROC is acceptable when using template-based models, PL-PatchSurfer2's performance was acceptable on 75% of the models (six out of eight) up to an RMSD of 1.5 Å. If up to a 20% decrease in the performance is acceptable, actually overall 76.7%, 90.0%, and 90.0% of the template models (23, 27, and 27 out of 30)

were within the decrease range in terms of EF_{1%}, AUC, and BEDROC, respectively.

Figure 7 is an example of docking to template-based models: a model of ER α with its agonists. The template-based model shown was built from a template (PDB ID 1Z5X) that has a sequence identity of 30% to the target (PDB ID 1L2I) and has a backbone RMSD of 2.69 Å. Superimposition shows that a long loop (at the left bottom) that is a part of the ligand binding site⁵⁵ is misplaced in the model (Figure 7A). The cognate ligand is structurally symmetrical, and its interaction with the binding pocket is mainly governed by hydrophobic

interactions at the middle and hydrogen bonds at the both termini of the molecule (Figure 7B).

The side-chain conformations of the residues at the binding site differ in the model relative to the holo form, particularly for the residues that participate in forming hydrogen bonds (red circles in Figure 7C). Using the model inevitably decreased the VS methods' performance as a result of its structural difference from the holo form. With PL-PatchSurfer2, EF_{1%} decreased from 9.0/15.0 to 7.5/13.5 (using the LCS/BS scheme) when the model was used instead of the holo-form target, compared with 13.5 to 4.5 with AutoDock Vina and 10.5 to 3.0 with DOCK6.

The latter three panels (Figure 7D–F) explain what happened to the predicted binding modes of the top-ranked active compounds when the template-based model was used. ZINC01999257 was ranked second (out of 2010 compounds) in both scoring schemes of PL-PatchSurfer2 with the holo form of the target and first (LCS) and third (BS) with the model, despite the conformational difference between the model and the holo form (Figure 7D). Close inspection found that the compound docked to the pocket in the opposite direction when the model was used (lower panels in Figure 7D) relative to the pose when the holo form was used (upper panels); however, because this active compound is almost symmetric, both termini of the compound can form hydrogen bonds in both orientations between terminal hydroxyl groups and residues in the pocket (patches in red). Also, hydrophobic interactions in the middle of the compound (the patch in orange) can clearly be formed in either orientation. In contrast to PL-PatchSurfer2, AutoDock Vina and DOCK6 suffer from the structural difference of the model. ZINC01686128, which was ranked fifth by AutoDock Vina on the holo structure, was ranked 115th on the model because one terminal of the compound could not form hydrogen bonds (Figure 7E). For DOCK6, ZINC03615416 was the top-scoring active compound when the holo structure was used. However, with the model it was ranked very low (1393th) because a favorable pose was not found within the pocket and the ligand was placed outside the binding site (Figure 7F).

Results on the WOMBAT Data Set. Finally, we tested the four VS methods on the WOMBAT data set. The WOMBAT data set has more diverse active compounds than DUD, addressing observations that VS benchmark results are influenced by the diversity of targets and ligands in a data set.^{32,57,58} The results are summarized in Table 5, and results for individual targets are given in Table S13. Pairwise *p* values are provided in Table S14.

On this data set, PL-PatchSurfer2 (BS) again showed the highest performance among the three SBVS methods in terms of EF_{1%} and BEDROC. The performances of the methods for all of the metrics were worse than on the DUD set, although it is difficult to make direct comparisons because the compound

sets are different from DUD. Interestingly, ROCS performed worst for all of the metrics except for AUC even though it performed best on the DUD benchmark. This may reflect the fact that active compounds in WOMBAT are on average more dissimilar to the cognate ligand than those in DUD (the average dissimilarities are 0.645 in DUD and 0.757 in WOMBAT), which in general makes it more difficult to find actives from a cognate ligand using LBVS methods.^{13,59} It needs to be mentioned, however, that these differences are not statistically significant: after application of the Bonferroni correction, no programs are distinguishable at the *p* < 0.1 level in terms of EF_{1%}. For AUC, AutoDock Vina is statistically significant at *p* < 0.1, and for BEDROC, only AutoDock Vina and ROCS are distinguishable with each other.

Computational Time. Table 6 reports computational times for the four methods. The values shown are the average times

Table 6. Average Computational Times for Screening of the DUD Data Set^a

program	computation time (h)
PL-PatchSurfer2	2.64
AutoDock Vina	58.81
DOCK6	48.03
ROCS	1.38

^aThe values do not include the time for computing compound conformations and the patches of targets and compounds. All of the computations were processed on a Linux machine with an Intel i7-3820 3.60 GHz CPU and 64 GB of RAM.

taken to screen the whole compound library of 25 DUD targets. PL-PatchSurfer2 is 22 and 18 times faster than Autodock Vina and DOCK6, respectively, and on the same scale with ROCS, an LBVS program.

CONCLUSION

We have introduced and benchmarked PL-PatchSurfer2, which evaluates protein–ligand interactions on the basis of compatibility of local molecular surfaces. A molecular surface is divided into a number of overlapping patches, which are described by four physicochemical properties: shape, electrostatic potential, hydrophobicity, and hydrogen bonding. The features are further converted to three-dimensional Zernike descriptors (3DZDs), each of which is a 72-dimensional vector. Representing and comparing the features using 3DZDs has two technical advantages: (1) the descriptors are rotationally invariant and (2) comparisons can be fast because Euclidean distance comparison of vectors is fast.^{12,20,23–28}

The thorough benchmark study performed in this work has indicated that PL-PatchSurfer2 performs better in general than two existing SBVS methods, AutoDock Vina and DOCK6, and particularly outperforms the two methods when apo forms or template-based models of targets are used. ROCS, an LBVS method, showed performance comparable to that of PL-PatchSurfer2 on DUD, but PL-PatchSurfer2 consistently showed a higher value than ROCS on all of the metrics when applied to the WOMBAT data set, which has more diverse active compounds. These results indicate that the unique surface patch representation imparts an interesting strength to PL-PatchSurfer2, i.e., increased tolerance to conformation changes of targets. Taking advantage of this unique strength, PL-PatchSurfer2 can substantially extend the capability of virtual screening by providing more accurate predictions in

Table 5. Virtual Screening Results on the WOMBAT Data Set

	EF _{1%}	EF _{5%}	EF _{10%}	AUC	BEDROC
PL-PS2 (LCS)	6.41	2.70	1.60	0.313	0.177
PL-PS2 (BS)	6.47	2.54	1.61	0.329	0.226
AutoDock Vina	5.37	2.88	2.22	0.558	0.197
DOCK6	5.56	2.39	1.63	0.419	0.155
ROCS	4.70	2.13	1.57	0.434	0.147

difficult cases, such as scaffold hopping or cases where only the apo form or template-based models of a target are available. Sufficient performance on template-based models is worth attention because it can substantially extend the target space where virtual screening is applicable. To further improve the performance, it may be effective to perform ensemble docking, where multiple conformations are used also for the target protein in addition to the ligand conformation ensemble, to cover receptor structural flexibility.^{60,61}

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.jcim.6b00163](https://doi.org/10.1021/acs.jcim.6b00163).

Results of training and testing of weights in eq 5; targets in the DUD data set with their apo structures and templates used for building their computational models; targets from the WOMBAT data set used in this study; pairwise *p* values for comparison between PL-PatchSurfer1 and PL-PatchSurfer2; results for individual targets in the DUD benchmark data set; performance of the Boltzmann-weighted ligand score with different combinations of parameters; pairwise *p* values from Student's *t* test for the performance on the DUD data set; virtual screening results for individual targets in the apo structure data set; pairwise *p* values from Student's *t* test for the apo-form targets; virtual screening results for individual targets in the template-based model data set; pairwise *p* values for the template-based model data set; virtual screening results of individual template-based models that were constructed with templates of two different ranges of sequence identity; virtual screening results for individual targets in the WOMBAT data set; pairwise *p* values for the WOMBAT set; interaction pattern between RXR α and its cognate ligand; and performance difference between holo- and apo-form targets (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: dkihara@purdue.edu. Tel.: +1-765-496-2284.

Author Contributions

W.-H.S. participated in designing the research, coded the programs, conducted the experiments, and wrote the paper. C.W.C. participated in coding the programs. J.W. participated in designing the research and interpreting the results. D.K. conceived the study, participated in its design and coordination, and wrote the paper.

Funding

This work was partially supported by a grant from the Lilly Research Award Program, the National Institute of General Medical Sciences of the National Institutes of Health (R01GM097528), and the National Science Foundation (IIS1319551, DBI1262189, and IOS1127027).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Hyunho Chun (Purdue University) for the discussion about statistical analysis of the virtual screening results.

■ ABBREVIATIONS

LBVS, ligand-based virtual screening; SBVS, structure-based virtual screening; 3DZD, three-dimensional Zernike descriptor; AUC, area under the receiver operating characteristic curve; VS, virtual screening; BS, Boltzmann-weighted ligand score; LCS, lowest Conformer Score

■ REFERENCES

- (1) Walters, W. P.; Stahl, M. T.; Murcko, M. A. Virtual Screening—An Overview. *Drug Discovery Today* **1998**, *3*, 160–178.
- (2) Vidović, D.; Busby, S. A.; Griffin, P. R.; Schürer, S. C. A Combined Ligand- and Structure-Based Virtual Screening Protocol Identifies Submicromolar PPAR γ Partial Agonists. *ChemMedChem* **2011**, *6*, 94–103.
- (3) Langmead, C. J.; Andrews, S. P.; Congreve, M.; Errey, J. C.; Hurrell, E.; Marshall, F. H.; Mason, J. S.; Richardson, C. M.; Robertson, N.; Zhukov, A.; Weir, M. Identification of Novel Adenosine A_{2A} Receptor Antagonists by Virtual Screening. *J. Med. Chem.* **2012**, *55*, 1904–1909.
- (4) Fazi, R.; Tintori, C.; Brai, A.; Botta, L.; Selvaraj, M.; Garbelli, A.; Maga, G.; Botta, M. Homology Model-Based Virtual Screening for the Identification of Human Helicase DDX3 Inhibitors. *J. Chem. Inf. Model.* **2015**, *55*, 2443–2454.
- (5) Schwartz, J.; Awale, M.; Reymond, J.-L. SMIfp (SMILES fingerprint) Chemical Space for Virtual Screening and Visualization of Large Databases of Organic Molecules. *J. Chem. Inf. Model.* **2013**, *53*, 1979–1989.
- (6) Durant, J. L.; Leland, B. A.; Henry, D. R.; Nourse, J. G. Reoptimization of MDL Keys for Use in Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1273–1280.
- (7) Raymond, J. W.; Gardiner, E. J.; Willett, P. RASCAL: Calculation of Graph Similarity using Maximum Common Edge Subgraphs. *Comput. J.* **2002**, *45*, 631–644.
- (8) Bender, A.; Mussa, H. Y.; Glen, R. C.; Reiling, S. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708–1718.
- (9) Hattori, M.; Okuno, Y.; Goto, S.; Kanehisa, M. Development of a Chemical Structure Comparison Method for Integrated Analysis of Chemical and Genomic Information in the Metabolic Pathways. *J. Am. Chem. Soc.* **2003**, *125*, 11853–11865.
- (10) Ballester, P. J.; Richards, W. G. Ultrafast Shape Recognition to Search Compound Databases for Similar Molecular Shapes. *J. Comput. Chem.* **2007**, *28*, 1711–1723.
- (11) 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.
- (12) Shin, W.-H.; Zhu, X.; Bures, M. G.; Kihara, D. Three-Dimensional Compound Comparison Methods and Their Application in Drug Discovery. *Molecules* **2015**, *20*, 12841–12962.
- (13) von Korff, M.; Freyss, J.; Sander, T. Comparison of Ligand- and Structure-Based Virtual Screening on the DUD Data Set. *J. Chem. Inf. Model.* **2009**, *49*, 209–231.
- (14) Shin, W.-H.; Kim, J.-K.; Kim, D.-S.; Seok, C. GalaxyDock2: Protein–Ligand Docking Using Beta-Complex and Global Optimization. *J. Comput. Chem.* **2013**, *34*, 2647–2656.
- (15) Jain, A. N. Surflex-Dock 2.1: Robust Performance from Ligand Energetic Modeling, Ring Flexibility, and Knowledge-Based Search. *J. Comput.-Aided Mol. Des.* **2007**, *21*, 281–306.
- (16) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.
- (17) Allen, W. J.; Baliaus, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C. DOCK 6: Impact of New Features and Current Docking Performance. *J. Comput. Chem.* **2015**, *36*, 1132–1156.

- (18) Leach, A. R.; Gillet, V. J.; Lewis, R. A.; Taylor, R. Three-Dimensional Pharmacophore Methods in Drug Discovery. *J. Med. Chem.* **2010**, *53*, 539–558.
- (19) 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.
- (20) Hu, B.; Zhu, X.; Monroe, L.; Bures, M. G.; Kihara, D. PL-PatchSurfer: A Novel Molecular Local Surface-Based Method for Exploring Protein-Ligand Interactions. *Int. J. Mol. Sci.* **2014**, *15*, 15122–15145.
- (21) Canterakis, N. 3D Zernike Moments and Zernike Affine Invariants for 3D Image Analysis and Recognition. In *Proceedings of 11th Scandinavian Conference on Image Analysis*; Dansk Selskab for Automatisk Genkendelse af Mønstre: Lyngby, Denmark, 1999; pp 85–93.
- (22) Novotni, M.; Klein, R. 3D Zernike Descriptors for Content Based Shape Retrieval. *Proceedings of eighth ACM symposium on Solid modeling and applications* **2003**, 216–225.
- (23) Chikhi, R.; Sael, L.; Kihara, D. Real-Time Ligand Binding Pocket Database Search Using Local Surface Descriptors. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2007–2028.
- (24) Sael, L.; Kihara, D. Binding Ligand Prediction for Proteins Using Partial Matching of Local Surface Patches. *Int. J. Mol. Sci.* **2010**, *11*, 5009–5026.
- (25) Sael, L.; Kihara, D. Detecting Local Ligand-Binding Site Similarity in Nonhomologous Proteins by Surface Patch Comparison. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 1177–1185.
- (26) Zhu, X.; Xiong, Y.; Kihara, D. Large-Scale Binding Ligand Prediction by Improved Patch-Based Method Patch-Surfer2.0. *Bioinformatics* **2015**, *31*, 707–713.
- (27) Esquivel-Rodriguez, J.; Xiong, Y.; Han, X.; Guang, S.; Christoffer, C.; Kihara, D. Navigating 3D Electron Microscopy Maps with EM-SURFER. *BMC Bioinf.* **2015**, *16*, 181.
- (28) Xiong, Y.; Esquivel-Rodriguez, J.; Sael, L.; Kihara, D. 3D-SURFER 2.0. Web Platform for Real-Time Search and Characterization of Protein Surfaces. *Methods Mol. Biol.* **2014**, *1137*, 105–117.
- (29) 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.
- (30) Zhao, W.; Hevener, K. E.; White, S. W.; Lee, R. E.; Boyett, J. M. A Statistical Framework to Evaluate Virtual Screening. *BMC Bioinf.* **2009**, *10*, 225.
- (31) Huang, N.; Shoichet, B. K.; Irwin, J. J. Benchmarking Sets for Molecular Docking. *J. Med. Chem.* **2006**, *49*, 6789–6801.
- (32) 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.
- (33) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.
- (34) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037–10041.
- (35) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: A New Method of Empirical Hydrophobic Field Calculation for CoMFA. *J. Comput.-Aided Mol. Des.* **1991**, *5*, 545–552.
- (36) Heiden, W.; Moeckel, G.; Brickmann, J. A New Approach to Analysis and Display of Local Lipophilicity/Hydrophilicity Mapped on Molecular Surfaces. *J. Comput.-Aided Mol. Des.* **1993**, *7*, S03–S14.
- (37) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol–Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J. Chem. Inf. Model.* **2007**, *47*, 2140–2148.
- (38) Li, Y.; Liu, Z.; Li, Z.; Han, L.; Liu, Z.; Zhao, Z.; Wang, L. Comparative Assessment of Scoring Functions on an Updated Benchmark: 1. Compilation of the Test Set. *J. Chem. Inf. Model.* **2014**, *54*, 1700–1716.
- (39) Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. Diverse, High-Quality Test Set for the Validation of Protein–Ligand Docking Performance. *J. Med. Chem.* **2007**, *50*, 726–740.
- (40) Paulsen, J. L.; Anderson, A. C. Scoring Ensembles of Docked Protein:Ligand Interactions for Virtual Lead Optimization. *J. Chem. Inf. Model.* **2009**, *49*, 2813–2819.
- (41) Vosmeer, C. R.; Pool, R.; van Stee, M. F.; Perić-Hassler, L.; Vermeulen, N. P. E.; Geerke, D. P. Towards Automated Binding Affinity Prediction Using an Iterative Linear Interaction Energy Approach. *Int. J. Mol. Sci.* **2014**, *15*, 798–816.
- (42) 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.
- (43) Bordogna, A.; Pandini, A.; Bonati, L. Predicting the Accuracy of Protein–Ligand Docking on Homology Models. *J. Comput. Chem.* **2011**, *32*, 81–98.
- (44) Cavasotto, C. N.; Phatak, S. S. Homology Modeling in Drug Discovery: Current Trends and Applications. *Drug Discovery Today* **2009**, *14*, 676–683.
- (45) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.
- (46) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.
- (47) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
- (48) Zhang, Y.; Skolnick, J. TM-align: A Protein Structure Alignment Algorithm Based on the TM-Score. *Nucleic Acids Res.* **2005**, *33*, 2302–2309.
- (49) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- (50) 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.
- (51) Jain, A. N. Effects of Protein Conformation in Docking: Improved Pose Prediction through Protein Pocket Adaptation. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 355–374.
- (52) Cavasotto, C. N.; Abagyan, R. Protein Flexibility in Ligand Docking and Virtual Screening to Protein Kinases. *J. Mol. Biol.* **2004**, *337*, 209–225.
- (53) Xu, M.; Yu, L.; Wan, B.; Yu, L.; Huang, Q. Predicting Inactive Conformations of Protein Kinases Using Active Structures: Conformational Selection of Type-II Inhibitors. *PLoS One* **2011**, *6*, e22644.
- (54) Kufareva, I.; Katritch, V.; et al. Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. *Structure* **2014**, *22*, 1120–1139.
- (55) Brzozowski, A. M.; Pike, A. C. W.; Dauter, Z.; Hubbard, R. E.; Bonn, T.; Engström, O.; Öhman, L.; Greene, G. L.; Gustafsson, J.-Å.; Carlquist, M. Molecular Basis of Agonism and Antagonism in the Oestrogen Receptor. *Nature* **1997**, *389*, 753–758.
- (56) Stierand, K.; Rarey, M. Drawing PDB: Protein-Ligand Complexes in Two Dimensions. *ACS Med. Chem. Lett.* **2010**, *1*, 540–545.
- (57) 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.
- (58) Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455–1474.
- (59) Venkatraman, V.; Perez-Nueno, V. I.; Mavridis, L.; Ritchie, D. W. Comprehensive Comparison of Ligand-Based Virtual Screening

Tools Against the DUD Data set Reveals Limitations of Current 3D Methods. *J. Chem. Inf. Model.* **2010**, *50*, 2079–2093.

(60) Huang, S. Y.; Zou, X. Ensemble Docking of Multiple Protein Structures: Considering Protein Structural Variations in Molecular Docking. *Proteins: Struct., Funct., Genet.* **2007**, *66*, 399–421.

(61) Campbell, A. J.; Lamb, M. L.; Joseph-McCarthy, D. Ensemble-Based Docking Using Biased Molecular Dynamics. *J. Chem. Inf. Model.* **2014**, *S4*, 2127–2138.