Assessing the Performance of Mixed Strategies to Combine Lipophilic Molecular Similarity and Docking in Virtual Screening

Javier Vázquez,[†,‡] Alessandro Deplano,[†] Albert Herrero,[†] Enric Gibert,[†] Enric Herrero,[†] and F. Javier Luque[‡]

[†] *Pharmacelera, Plaça Pau Vila, 1, Sector C 2a, Edificio Palau de Mar, Barcelona 08039, Spain*

[‡] *Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Sciences, Institute of Biomedicine (IBUB), and Institute of Theoretical and Computational Chemistry (IQTC-UB), University of Barcelona, Av. Prat de la Riba 171, Santa Coloma de Gramanet E-08921, Spain*

Abstract

The accuracy of structure-based (SB) virtual screening (VS) is heavily affected by the scoring function used to rank a library of screened compounds. Even in cases where the docked pose agrees with the experimental binding mode of the ligand, the limitations of current scoring functions may lead to sensible inaccuracies in the ability to discriminate between actives and inactives. In this context, the combination of SB and ligand-based (LB) molecular similarity may be a promising strategy to increase the hit rates in VS. This study explores different strategies that aim to exploit the synergy between LB and SB methods in order to mitigate the limitations of these techniques, and to enhance the performance of VS studies by means of a balanced combination between docking scores and 3D similarity. Particularly, attention is focused to the use of measurements of molecular similarity with PharmScreen, which exploits the 3D distribution of atomic lipophilicity determined from quantum mechanical-based continuum solvation calculations performed with the MST model, in conjunction with three docking programs: Glide, rDock and GOLD. Different strategies have been explored to combine the information provided by docking and similarity measurements for reranking the screened ligands. For a benchmarking of 44 data sets, including 41 targets, the hybrid methods increase the identification of active compounds according to the early (ROCe%) and total (AUC) enrichment metrics of VS compared to pure LB and SB methods. Finally, the hybrid approaches are also more effective in enhancing the chemical diversity of active compounds. The datasets employed in this work are available in https://github.com/Pharmacelera/Molecular-Similarity-and-Docking.

Keywords

Virtual screening, molecular docking, molecular similarity, hydrophobicity, binding mode, protein–ligand interactions, drug design

INTRODUCTION

Structure-based (SB) and ligand-based (LB) methods have been widely used to perform virtual screening (VS) of chemical libraries in computer-aided drug design.[1,2] SB techniques exploit the three-dimensional (3D) structure of the macromolecular target to provide a putative binding mode of fragment-like and drug-like compounds and estimate the strength of the ligand-target interaction, often involving simplified approximations of the enthalpic and entropic components of the binding affinity.[3,4] The most widely used SB technique is molecular docking, which predicts the preferred pose of the ligand in the binding pocket through the use of scoring functions, often supplemented with pharmacophoric constraints, and is exploited to identify hits in VS studies. On the other hand, LB techniques encompass a diverse group of strategies, which primarily disclose similarity relationships between molecular descriptors of the ligand. Under the framework of the similarity principle property,[5] which assumes that similar compounds should have similar properties, a plethora of methods have been developed to disclose structure-activity relationships, derive pharmacophores to rationalize the biological activity, and perform similarity measurements in the search of novel chemical scaffolds.[6–8]

The accuracy and predictive power of both LB and SB methods are limited by several challenges. Besides the lack of structural information of the target, LB methods are limited by the quality of the descriptors used to characterize the chemical features of the compounds, the consistency and chemical diversity of the training set, and the similarity metrics used in the comparison of molecules. On the other hand, SB methods may be affected by the accuracy of the 3D structural data, the involvement of different conformational states of the target protein, often arising from ligand-induced effects, and the assistance of structural waters in mediating ligand binding. Even in the case of well-defined structural models of the target protein, the predictive power of SB techniques may be affected by the simplifications introduced in the

scoring function, and the exhaustiveness of the sampling search, which may lead to a substantial computational cost for VS applications.[9-15]

In this context, strategies that combine LB and SB methods may be valuable to reinforce the success of VS campaigns by minimizing the intrinsic weaknesses of these techniques.[16–19] Following Drwal and Griffith,[2] combined LB+SB strategies can be grouped into three categories: sequential, parallel, and hybrid approaches. The former splits the screening process into various consecutive steps to overcome the computational expensiveness of the SB approach. Accordingly, a prefiltering step is performed at the beginning of the VS using the less expensive LB techniques, and the retrieved hits are subsequently evaluated using molecular docking.[20–22] In the parallel approach, LB and SB methods are run independently, and then the results are merged to obtain a mixed ranking.[16,23–25] Finally, hybrid approaches integrate LB and SB techniques, generally through the translation of key protein-ligand interactions into pharmacophoric constraints for guiding the screening of compounds[26] and for profiling purposes,[27] or alternatively by exploiting the similarity of molecules to a known crystallographic ligand to rerank the docked poses.[18,28–30]

The success of these strategies depends not only on the robustness of LB and SB techniques, but also on the synergy and complementarity of the molecular descriptors encoded in these methods, and the mathematical formalism adopted to combine them into a ranking function that alleviates the potential deficiencies of these methods.[31,32] Here we address these questions by examining the performance of parallel and hybrid approaches to discriminate between active and inactive compounds in VS of chemical libraries. To this end, attention is paid to the use of lipophilic (Hyphar) descriptors,[33–35] which have been examined in the context of 3D-QSAR studies[36,37] and similarity measurements of molecular overlays.[38] The lipophilic descriptors are based on the atomic decomposition of the global lipophilicity of a given compound estimated from quantum mechanical (QM) continuum solvation calculations in

water and *n*-octanol, yielding a 3D lipophilicity distribution of the molecule.[39,40] In this work the Hyphar-based similarity measurements have been exploited in conjunction with three docking programs: rDock,[41] GOLD[42] and Glide.[43-45] In particular, we have evaluated the synergy between these LB and SB techniques to enhance the performance of VS through the analysis of new rankings obtained by using different parallel and hybrid combination strategies.

METHODS

Test dataset. Since the performance of VS methods is sensitive to the set of compounds, a diverse number of receptors should be considered to calibrate the performance of LB and SB techniques. For our purposes here, a subset of the Directory of Useful Decoys (DUD; http://dud.docking.org/)[46,47] has been used. Although DUD is suitable to address the weaknesses of docking methods, LB methods can easily account for differences between actives and inactives.[46] Accordingly, the subset of DUD proposed by Good and Oprea (DUD_LIB_VS_1.0; http://dud.docking.org/jahn/) has been chosen,[48] as this specific dataset was conceived with the aim of avoiding an overestimation of the performance of LB methods. In particular, a lead-like filtering and clustering algorithm was applied to eliminate large molecules with inappropriate physicochemical properties and to reduce the artificial bias between structural analogs and actives during the enrichment test.[49–51] This set contains known actives and mimetic decoys for 40 targets downloaded from the DUD website. Additionally, four sets (DHFR, GR, HIV1PR, and VEGFR2) taken from DEKOIS V2.0 (http://www.dekois.com),[52] which were used to evaluate the combination of LB and SB methods,[53] were also considered. In this latter benchmark, each different set has the same size and the same number of active ligands, selected from BindingDB.[54] For ease of reading, these sets will be denoted as BS1 (benchmarking set 1; 40 targets) and BS2 (benchmarking set 2; 4

targets), respectively. A detailed description of the datasets is provided in the Supporting Information (SI; Table S1).

Ligand Preparation. In this study, two complementary aspects of ligand-receptor interactions were analyzed: the MST-derived lipophilicity (Hyphar) descriptors, and the ligand docking into the binding site, which was examined using rDock, GOLD, and Glide. Hyphar descriptors were obtained using the MST solvation model[55] parametrized at the semiempirical RM1 level.[56] All the compounds were modeled considering a neutral state in order to avoid artefactual effects in similarity measurements arising from the large contribution of charged chemical groups to the 3D lipophilicity distribution. MolVS,[57] a standardization tool in the RDKit[58] chemistry framework, was applied to neutralize the compounds. The geometry of all ligands was minimized using the RM1 Hamiltonian using a locally modified version of MOPAC.[59] To take into account the conformational flexibility of compounds, up to 100 conformations were calculated for each ligand using the Distance Geometry method implemented in RDKit. With regard to the docking calculations, the ionization and tautomeric states assigned to compounds in the original sets were adopted in this study.

Protein Preparation. For each ligand a protein target was selected to perform the docking calculations. For the BS1 set, all targets were directly taken from the DUD Web site (DUD release 2). Let us note, however, that the original X-ray structure for the target ADA is not correct, and hence it was replaced by the X-ray structure 1NDW taken from the Protein Data Band (PDB).[60,61] In addition, the PDB codes for COX-1 and SRC (entries 1P4G and 2SRC, respectively) in DUD do not correspond to the protein structures compiled in the set, which are 1QAG and 1Y57, respectively. On the other hand, only water molecules present in the X-ray structures prepared by the DUD's authors were maintained in the binding pocket. Although this may be a relevant factor in determining the optimal performance of the SB

methods for specific sets, this decision was motivated for the sake of comparison with other tests.

For the BS2 set, the targets were obtained from the PDB, and water molecules and cofactors were processed following the approach described by Anighoro and Bajorath.[53] For calculations performed with Glide and rDock, the protein targets were prepared using the Protein Preparation Wizard module in Maestro.[62] For GOLD, the protein structures were prepared using the CSD Python API following the standard wizard workflow. A detailed description of the targets and the changes made to certain protein residues are provided in the Supporting Information Tables S1 and S2, respectively.

Query Preparation. The query structures chosen to perform the similarity search in the LB analysis were adopted from previous works.[46,47,53] With regard to BS1, the queries proposed by Huang *et al*.,[46] which were used in the validation of LB tools,[49,63] were selected and the structures downloaded from the DUD Web site (DUD release 2). For BS2, the same co-crystallized ligands used in ref. 53 were extracted from the PDB (Supporting Information Table S1).

LB and SB VS tools. PharmScreen[38,64] was used as the LB VS tool using the default configuration. PharmScreen exploits the MST-derived atomic lipophilic contributions in conjunction with a hydrogen-bond (HB) descriptor, which accounts for the HB donor/acceptor properties of atoms in a molecule. These lipophilic contributions were determined from quantum mechanical computations of the ligand immersed in water and *n*-octanol through the use of RM1 semiempirical version of the MST model,[55] and the subsequent partition of the *n*-octanol/water partition coefficient into 3D maps of atomic contributions, which were further decomposed into electrostatic and cavitation components (the reader is addressed to refs. 38 and 39 for details). As an example, Supporting Information Figure S1 provides a graphical representation of the comparison between two pairs of molecules according to these

hydrophobicity components.

The SB VS was performed using rDock, GOLD and Glide, as these docking programs involve distinct sampling algorithms and scoring functions. rDock is an open source molecular docking package, while Glide and GOLD are widely distributed and validated docking tools. In all cases, the receptor grids were centered on the query molecule selected in the LB method. Grid dimensions were defined using the default parameters, except for Glide in BS2, where the grid was adjusted to fit the size defined in ref. 23, and docking calculations were performed using the default procedure (representative files for docking calculations are provided in the Supporting Information), which is briefly detailed as follows:

i) The rDock calculations were performed using a rigid model for the protein target, keeping nevertheless the flexibility of the hydrogen-bond donors, the scoring function "dock.prm", and performing a total of 50 docking runs per ligand.[41] rDock uses a combination of stochastic and deterministic search techniques to generate low-energy ligand poses. A genetic algorithm search was applied as a first stage followed by low temperature Monte Carlo and Simplex minimization stages. Generally, rDock is used in conjunction with pharmacophoric constraints, which have not been applied in this study.

ii) GOLD was run using the default options and the GoldScore fitness function.[42] GOLD employs a genetic algorithm to explore the full range of conformational flexibility of the ligand in the protein binding site. Free corners of ligand rings were allowed to flip during docking. Also, GOLD was allowed to vary ligand non-fused ring conformations during docking based on a library of ring conformations extracted from the Cambridge Structural Database.[65]

iii) Finally, Glide was used in both HTVS and SP modes.[44,45] The general van der Waals radius-scaling factor was reduced by a factor of 0.9 in order to decrease the number of rejected

molecules. In addition, the cutoff of Coulomb-van der Waals energy and HB score were disallowed (the default value of 0.0 was modified to 1000). Even with these changes, some molecules were rejected in docking calculations, and they were relocated at the end of BS1 and BS2 rankings, first hits and then decoys. A list of relocated molecules is reported in the Supporting Information Table S3. A systematic search of the ligand's conformation and orientation in the binding pocket was performed through a hierarchical series of filters.

Protocols applied to combine LB and SB methods. Three LB+SB protocols were tested, namely, a parallel ranking, and two hybrid approaches: rescoring ranking and consensus ranking.

i) *Parallel ranking (PR)*. In this case the rankings obtained from separate LB and SB screenings are merged to create the final ranking. With the aim to treat both techniques with equal parity, the first molecule of the SB ranking and the first molecule of LB ranking occupies the first and second positions of the PR. Among this pair of molecules, the first is the compound with the lower sum of LB and SB rankings. Accordingly, the molecules ranked second for each method would be reranked third and fourth, and so on until all molecules are reordered.

ii) *Rescoring ranking (RR)*. The final ranking is determined according to the score obtained from the 3D similarity measurement (relative to the reference template) of the best pose generated by the SB docking method.

iii) *Consensus ranking (CR)*. This strategy mixes both parallel and hybrid approaches, as the LB screening used in PR is replaced by the RR ranked compounds. Thus, the final score of the compounds is obtained by combining the rankings provided a) directly from the SB (docking) method and b) from the RR approach, and performing this combination according to the parallel strategy.

Finally, let us note that the similarity measurement performed in RR and CR was estimated

using two metrics: i) a global similarity using the Tanimoto coefficient (Tn), and ii) a local similarity using the Tversky coefficient (Tv).[66] These coefficients can be described as noted in Eq. 1, with α=β=1 for Tn, and α=1 and β=0 for Tv. Accordingly, Tv remarks the asymmetric character in the comparison of two molecules, allowing to give more relevance to the salient features of a specific compound in the similarity measurement relative to the reference template.[67] This will be used here in the comparison of molecules with large differences in their size (see below).

$$Similarity = \frac{C}{\alpha(A-C)+\beta(B-C)+C} \quad (1)$$

where C denotes ~~the common feature~~ the intersection of the fields of the two molecules that are compared, and the terms $A-C$ and $B-C$ stand for the distinctive fields not overlapped ~~features~~ of the two molecules.

Performance Evaluation. Two metrics were used to assess the performance of the three combined (LB+SB) reranking strategies.[68–70] The Receiver Operator Characteristic enrichment factor (ROCe, Eq. 2) was used to measure the performance assuming a given percentage of false positives and to evaluate the quality in the initial positions of the ranking.

$$ROCe\ X\% = \frac{\frac{N^{X\%}_{actives\ selected}}{N\ total\ actives}}{\frac{N^{X\%}_{decoys\ selected}}{N\ total\ decoys}} = \frac{\frac{TP}{TP+FN}}{\frac{FP}{TN+FP}} = \frac{sensitivity}{1-specificity} \quad (2)$$

Following Jan and Nicholls,[71] ROCe values at false positive rates of 0.5%, 1.0%, 2.0%, and 5.0% were determined. Furthermore, the Area Under the ROC curve (AUC) was used as metrics to assess the global performance to identify actives in the VS.

In addition, chemotype clustering analyses was included to examine the chemical diversity in the ranked compounds by means of the awROCe values (Eq. 3).[49]

$$awROCe\ X\% = \frac{\frac{\sum_{j}^{N_{clusters}} \sum_{i}^{N_j} w_{ij} a_{ij}^{X\%}}{\text{N clusters}}}{\frac{\text{N}_{\text{decoys selected}}^{X\%}}{\text{N total decoys}}} \quad (3)$$

where $w_{ij} = \frac{1}{N_j}$ is the weight of the i^{th} structure from the j^{th} cluster, N_j is the number of structures in a given cluster, $a_{ij}^{X\%}$ is 1 or 0 depending on whether the ith structure of the j^{th} cluster already appeared or not in the chosen fraction of the dataset.

By using Eq. 2, the value of a true positive hit is weighted depending on the cluster to which it belongs and on the number of molecules in the cluster. The clusters proposed in the DUD dataset were used to compute this metric. The awROCe parameter was determined taking into account the same percentages adopted for ROCe.

RESULTS AND DISCUSSION

Comparative analysis of LB and SB methods. The VS results obtained by using either LB (PharmScreen) or SB (Glide HTVS, Glide SP, rDock, and GOLD) methods are reported in Table 1, which shows the averaged values of ROCe and AUC for the whole set of targets (results for individual sets are provided in the Supporting Information Tables S4-S7).

The results obtained with Glide HTVS and Glide SP are highly similar for the BS1 dataset (DUD), as noted in the similar values determined for ROCe (ranging from 27.3 to 6.8 at 0.5% and 5%, respectively, for Glide HTVS, and from 28.3 to 6.8 for Glide SP), and AUC (0.70 and 0.74 for Glide HTVS and Glide SP, respectively). This resemblance is remarkable when

one takes into account the lower computational cost required for the VS with Glide HTVS compared to Glide SP. On the other hand, they compare with the values obtained with PharmScreen, as the ROCe values vary from 33.2 to 6.6, and the AUC is 0.66. In contrast, the results obtained for the BS2 dataset reveal larger differences between Glide HTVS and Glide SP, as the ROCe values obtained for HTVS (ranging from 6.3 to 3.4) are drastically improved when Glide SP is used, leading to ROCe values varying from 35.0 to 6.3. Similarly, the AUC increases from 0.52 for Glide HTVS to 0.69 for Glide SP. For this set of compounds, the results obtained with PharmScreen yields ROCe values ranging from 15.0 to 5.4, and an AUC of 0.70.

Table 1. ROCe and AUC for PharmScreen, Glide HTVS, Glide SP, rDock and GOLD.

Metric	**PharmScreen**	**Glide HTVS**	**Glide SP**	**rDock**	**GOLD**
BS1 dataset					
ROCE 0.5	33.2	27.3	28.3	4.1	17.4
ROCE 1	21.0	19.7	18.8	4.9	12.2
ROCE 2	12.5	12.2	12.4	4.6	8.1
ROCE 5	6.6	6.4	6.8	3.4	4.4
AUC	0.66	0.70	0.74	0.62	0.55
BS2 dataset					
ROCE 0.5	15.0	6.3	35.0	5.0	16.3
ROCE 1	8.1	5.6	18.1	5.6	11.3
ROCE 2	6.3	4.1	11.9	3.1	5.9
ROCE 5	5.4	3.4	6.3	3.4	2.6
AUC	0.70	0.52	0.69	0.64	0.58

[a] GS: Global (GS) and local (LS) similarity measurement performed with Tanimoto and Tversky coefficients.

Compared to PharmScreen and Glide, rDock exhibits a lower performance, especially regarding the early recovery of actives. This trend agrees with the results reported by Ruiz-Carmona *et al*.,[41] as the VS performance of rDock was reported to be inferior to Glide for most systems unless pharmacophore constraints were used, whereas the inclusion of this

information improved significantly the performance of rDock, which was similar to Glide. In contrast, whereas the results obtained in the VS of the BS1 dataset with GOLD point out a lower performance compared to PharmScreen and Glide, it compares well with PharmScreen and outperforms Glide HTVS in the VS for the BS2 dataset. On the other hand, it is worth noting that the two docking methods show consistent trends for the two compound datasets. Thus, for rDock the ROCe values range from 4.1 to 3.4 for BS1, and from 5 to 3.4 for BS2, and the AUC values are 0.62 and 0.64 for BS1 and BS2, respectively. In the case of GOLD, the ROCe values range from 17.4 to 4.4 for BS1, and from 16.3 to 2.6 for BS2, whereas the AUC amounts to 0.55 and 0.58 for BS1 and BS2, respectively.

It is worth noting that the active compounds recovered from the VS performed with LB and SB techniques exhibit notorious differences. This can be stated in Figure 1, which shows the position (normalized from 1% to 100%) occupied by all active compound as ranked by PharmScreen and by the four docking methods (Glide HTVS, Glide SP, rDock and GOLD). The plots show that most of the compounds recovered in the first 10% with PharmScreen are in most cases scattered up to 60% according to Glide HTVS and Glide SP, and can be found up to 80-90% according to rDock and GOLD. Similarly, the hits in the first 10% obtained with the different docking programs are scattered up to 60-100% with PharmScreen. This behavior is also observed in the early recovered actives, as most of the compounds lying in the first 1% with docking methods are largely scattered according to the PharmScreen ranking, and vice versa.

Keeping in mind the intrinsic features that underlie similarity measurements and docking calculations, the different behavior exhibited by PharmScreen and the three docking programs regarding the recovery of actives is not unexpected. However, this also suggests that a combined strategy between LB and SB techniques may be valuable not only for enhancing the recovery of actives, but also for the enrichment in the diversity of chemical scaffolds.

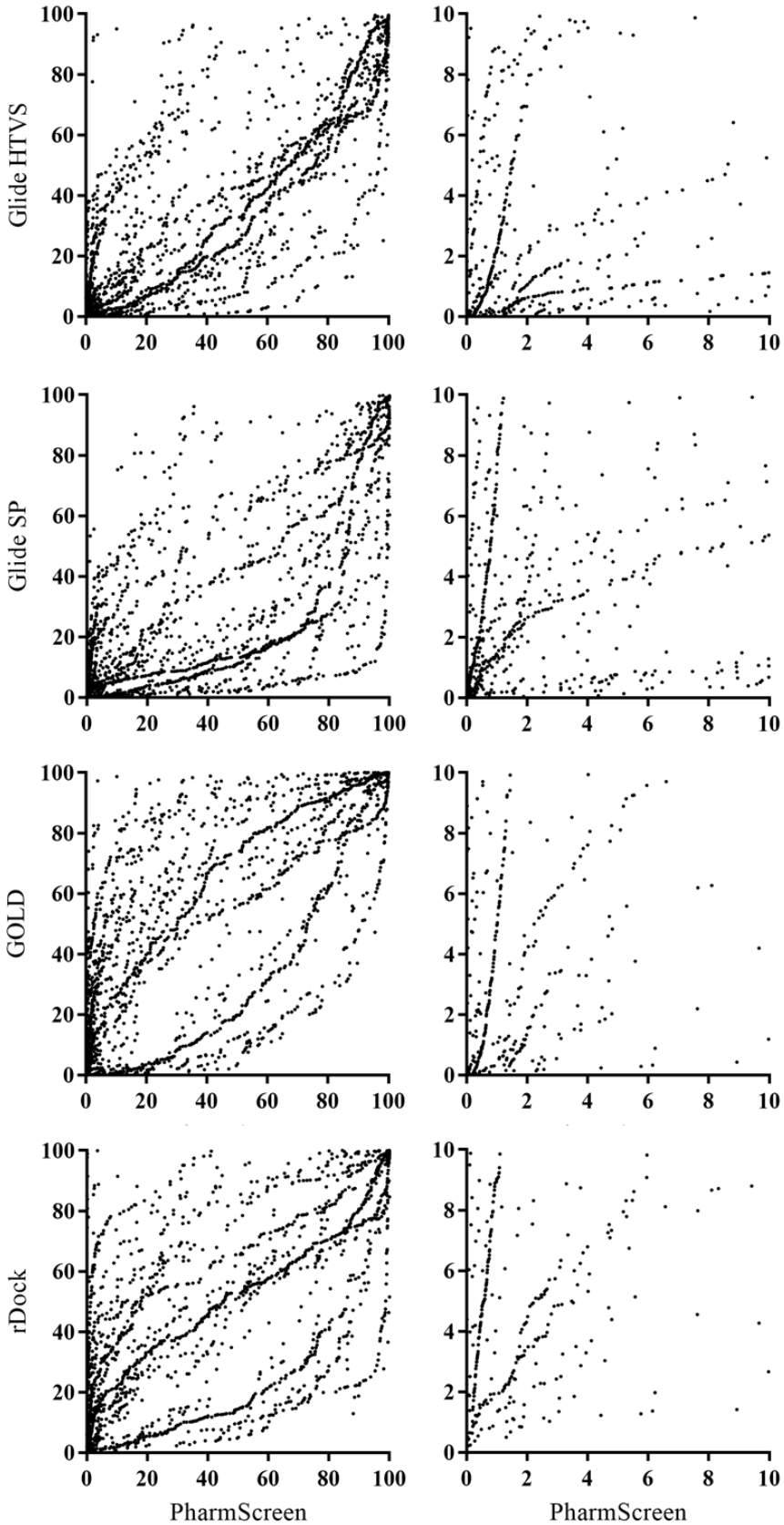

Figure 1. Distribution of the active compounds for the 40 targets included in the BS1 dataset as ranked by PharmScreen and by the four docking methods (Glide HTVS, Glide SP, rDock and GOLD). The position of the actives recovered in the LB and SB screenings has been normalized relative to the total

number of compounds in every set (from 1 to 100%). (Left) Representation obtained for the whole set of actives. (Right) Detailed view of the distribution of actives in the first 10%.

Assessment of the combination strategies derived using global similarity. The effect of combining LB and SB methods for the two datasets can be examined from the ROCe and AUC values reported in Tables 2-5. The performance of the combined strategies is different for the two datasets. For the BS1 dataset, the PR approach enhances the performance of Glide HTVS and Glide SP, reaching ROCe 0.5% values of 37.5, which compare with values of 27.3 and 28.3 for Glide HTVS and Glide SP respectively, whereas there is a slight improvement in the AUC (values of 0.75 and 0.77, which compare with 0.70 and 0.74 for Glide HTVS and Glide SP). In contrast, while PR also improves the VS outcome for the BS2 dataset in the case of Glide HTVS (ROCe 0.5% = 15.0, AUC = 0.72), there is in fact a worsening of the performance obtained with Glide SP (ROCe 0.5% of 35), as the ROCe 0.5% is reduced to 23.8, although the AUC increases from 0.69 to 0.79.

The results obtained for the two hybrid approaches, RR and CR, generally exhibit a large resemblance. For the BS1 dataset the RR and CR strategies lead to a slight improvement in recovering actives in the initial stages of the VS as compared to Glide HTVS and Glide SP, this trait being more relevant wit the CR strategy (ROCe 0.5% of 43.0). However, the largest impact is observed for the BS2 dataset, as the final CR ranking outperforms the behavior of Glide SP, leading to ROCe values ranging from 43.8 to 8.6. Nevertheless, the effect on the AUC is less significant.

Table 2. ROCe and AUC for the three combined strategies (PR, RR, CR) using PharmScreen and Glide HTVS. For each row, the best performing configuration is highlighted in bold.

Metric	**PR**	**RR** GS [a]	**CR** GS	**RR** LS	**CR** LS
BS1 dataset					
ROCE 0.5	37.5	30.4	39.7	31.7	**43.0**
ROCE 1	24.9	18.3	25.0	20.8	**26.1**
ROCE 2	15.8	11.3	14.4	13.2	**16.1**
ROCE 5	8.2	5.9	7.4	7.0	**8.3**
AUC	**0.75**	0.67	0.71	0.72	**0.75**
BS2 dataset					
ROCE 0.5	15.0	**17.5**	16.3	**17.5**	12.5
ROCE 1	8.8	**11.9**	10.0	10.6	8.1
ROCE 2	6.6	**7.8**	6.6	5.9	6.6
ROCE 5	**5.7**	5.5	3.4	5.0	4.0
AUC	**0.72**	0.56	0.59	0.56	0.58

[a] GS: Global (GS) and local (LS) similarity measurement performed with Tanimoto and Tversky coefficients.

Table 3. ROCe and AUC for the three combined strategies (PR, RR, CR) using PharmScreen and Glide SP. For each row, the best performing configuration is highlighted in bold.

Metric	**PR**	**RR** GS	**CR** GS	**RR** LS	**CR** LS
BS1 dataset					
ROCE 0.5	37.5	33.1	**43.0**	31.6	42.9
ROCE 1	24.1	22.1	27.4	22.6	**27.6**
ROCE 2	15.4	13.1	16.9	14.3	**17.7**
ROCE 5	8.2	6.7	8.5	8.0	**9.4**
AUC	0.77	0.71	0.76	0.76	**0.80**
BS2 dataset					
ROCE 0.5	23.8	35.0	**43.8**	30.0	37.5
ROCE 1	19.4	20.0	**28.8**	15.0	26.3
ROCE 2	12.5	12.5	**16.9**	9.7	15.3
ROCE 5	7.5	6.8	**8.6**	5.5	8.4
AUC	0.79	0.69	**0.74**	0.69	0.73

The consistency in the behavior of both rDock and GOLD for the two datasets is kept in their combined methods. Thus, the combination with PharmScreen within the PR framework leads to a significant improvement in the recovery of actives, as the ROCe 0.5% increases from 4.1 to 23.3 (BS1) and 5.0 to 13.8 (BS2) for rDock, and from around 17.0 to 28.0 (BS1) and 18.8 (BS2) for GOLD. The better performance obtained for the hybrid combination with GOLD reflects the larger recovery of actives provided by this docking method compared to rDock (see Tables 4 and 5).

With regard to the two hybrid approaches, the RR combination shows better performance than the CR approach, reaching ROCE 0.5% values close to 30.0 and 34.0 for the poses generated by rDock and GOLD, respectively. In these cases, therefore, the correction of the score provided by these docking methods with the similarity measurement between the best pose and the ligand template leads to a sizable improvement in the overall VS performance. This trait is also reflected in the AUC values, which increase from around 0.60 for the two docking methods to around 0.72 in the combined strategies.

The enhanced performance of the RR rescoring compared to the PR is remarkable from a computational point of view, because this latter strategy requires carrying out both LB and SB screenings of the compounds. In contrast, the RR (and CR) approach rely on a single (SB)

screening, as the LB similarity is performed only for the best docked pose of each compound, which leads to a marginal increase in the expensiveness of the VS.

Table 4. ROCe and AUC for the three combined strategies (PR, RR, CR) using PharmScreen and rDock. For each row, the best performing configuration is highlighted in bold.

Metric	**PR**	**RR** GS	**CR** GS	**RR** LS	**CR** LS
BS1 dataset					
ROCE 0.5	23.3	31.5	21.8	**39.1**	25.6
ROCE 1	17.0	19.2	15.6	**24.6**	18.0
ROCE 2	11.1	12.2	10.8	**15.6**	13.1
ROCE 5	6.3	6.4	6.3	**7.8**	7.7
AUC	0.70	0.68	0.72	0.73	**0.76**
BS2 dataset					
ROCE 0.5	13.8	**27.5**	22.5	25.0	11.3
ROCE 1	10.6	**18.8**	16.3	17.5	15.0
ROCE 2	6.9	**11.6**	11.3	**11.6**	10.9
ROCE 5	4.5	**6.4**	5.9	6.0	5.8
AUC	**0.75**	0.70	0.73	0.72	0.73

Table 5. ROCe and AUC for the three combined strategies (PR, RR, CR) using PharmScreen and GOLD. For each row, the best performing configuration is highlighted in bold.

Metric	**PR**	**RR** GS	**CR** GS	**RR** LS	**CR** LS
BS1 dataset					
ROCE 0.5	28.0	34.2	29.2	**45.9**	35.3
ROCE 1	19.9	20.6	19.2	**27.8**	23.4
ROCE 2	12.3	12.4	11.9	**16.1**	14.8
ROCE 5	6.9	6.4	6.4	**8.3**	7.6
AUC	0.66	0.68	0.66	**0.73**	0.72
BS2 dataset					
ROCE 0.5	18.8	**31.3**	25.0	25.0	22.5
ROCE 1	15.6	16.3	21.3	**17.5**	16.9
ROCE 2	9.4	**11.9**	12.8	10.3	12.5
ROCE 5	5.5	**6.8**	6.8	6.1	5.9
AUC	0.77	**0.72**	**0.72**	**0.72**	0.71

The analysis of the results obtained for the individual sets reveals the occurrence of significant improvements due to the use of combined strategies, which are partially masked in the analysis of the average results presented in Tables 2-5. This is illustrated by the behavior observed for the set of dihydrofolate reductase (DHFR) ligands in the BS2 dataset. A narrow and deep pocket defines a single binding mode in this target (Figure 2). Moreover, the chemical structure of the template molecule and 24 hits share a pyrido[2-3]pyrimidine ring, which can form HBs with three residues (Glu30A, Ile7A, and Val115A) at the bottom of the pocket,[71] thus favoring the definition of a unique binding mode. Most of the hits docked with Glide, rDock and GOLD exhibit a high overlap with the template compound (Figure 2; see also the Supporting Information Figure S2). These conditions are suited for the application of the hybrid RR protocol, as the LB-guided reranking of the docked poses leads to an effective

enhancement in the VS performance. At this point, let us note that the BS1 dataset also contains a subset of compounds targeting the DHFR enzyme, and the RR approach performs better than the rest of methods, as noted in ROCe 0.5% values that are increased from 25 (Glide SP), 15 (rDock) and 35 (GOLD) to 75 for the RR method (see Supporting Information Tables S4-S7).

Conversely, trypsin (in BS1 dataset) presents an open and superficial pocket, where a significant part of the crystallized ligand is exposed to the solvent (Figure 3).[73] In this case, the ligands exhibit a higher diversity in their binding mode, and hence the RR method, which relies on the similarity of the docked pose to the reference template, performs worse. Nevertheless, the challenges posed by the existence of multiple binding modes are alleviated by the CR strategy.

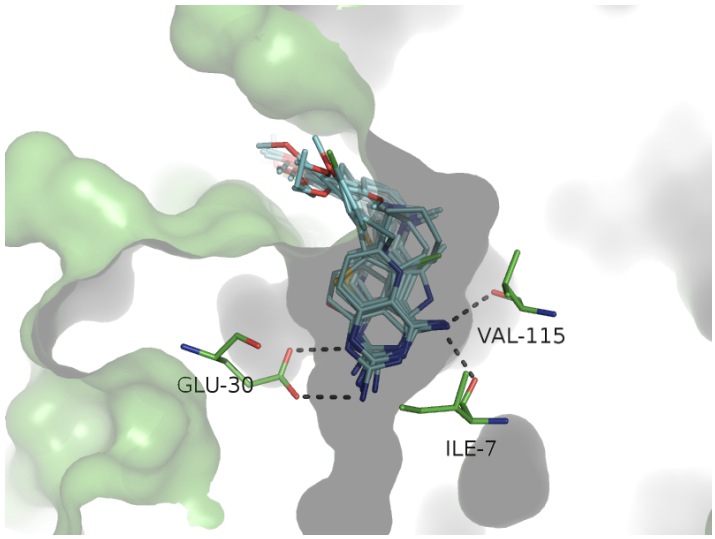

Figure 2. Binding mode of DHFR (PDB code: 1KMV) ligands. The surface of the pocket is shown in green, and the co-crystallized template molecule is shown as green sticks. The hit molecules containing a pyrido[2-3]pyrimidine ring docked with Glide are shown as cyan sticks. Dotted lines represent the HBs formed with Glu30, Ile7, and Val115A.

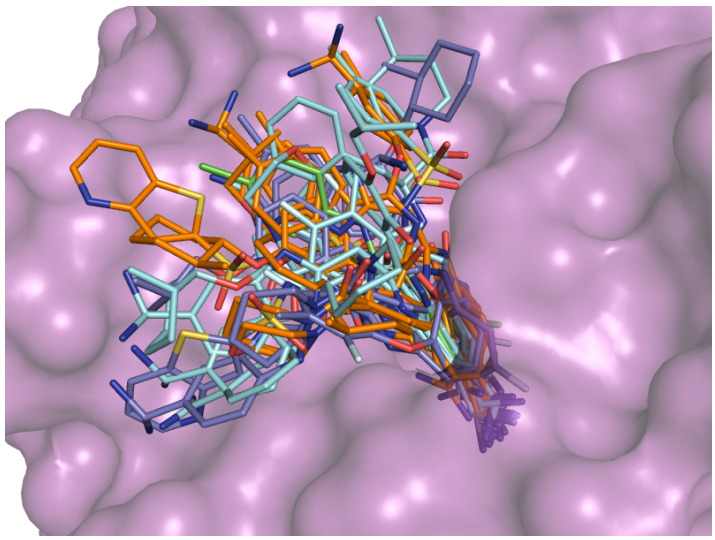

Figure 3. Binding mode of beta-trypsin, Trypsin_BS1, (PDB code: 1BJU, purple), co-crystallized reference molecule (green), docked molecules using Glide (cyan), rDock (orange), and GOLD (blue).

In light of the trends discussed above for DHFR and trypsin, we have analyzed the influence of the physicochemical features of the binding pocket on the distribution of the ROCe (5%) values determined for the RR strategy. In particular, we have examined both the geometrical characteristics and the balance between hydrophobic and hydrophilic residues, as several studies have highlighted the relevance of these features in determining the druggability of the binding pocket,[74-76] and are used to score the suitability of protein cavities to accommodate drug-like compounds in several predictors, such as fPocket[77] and DoGSiteScorer.[78] For our purposes here, these descriptors were determined for the whole set of targets included in the datasets using fPocket (https://fpocket.sourceforge.net). The results clearly indicate a worsening of the ROCe values with the increase in the size of the binding cavity, as estimated from either the solvent-exposed surface (data not shown) or the total volume of the pocket (Figure 4). This effect can be attributed to the larger number of possible arrangements that can be adopted by the ligand in pockets characterized by a large curvature. On the other hand, the hydrophilic/hydrophobic character of the cavity was quantified by using the hydrophobicity score implemented in fPocket, which estimates the mean hydrophobicity for all residues implicated in the binding site using the residue-based hydrophobicity scale reported by

Monera et al.[79] The results suggests a tendency to provide larger ROCe values for either polar or apolar cavities, whereas lower values are obtained for pockets with an intermediate balance between hydrophilicity and hydrophobicity. These traits, which are common for the different docking methods, must be taken with caution, because the simplified description afforded by the normalized hydrophobicity of the pocket may be less relevant that the 3D distribution of polar and apolar groups in the binding cavity.

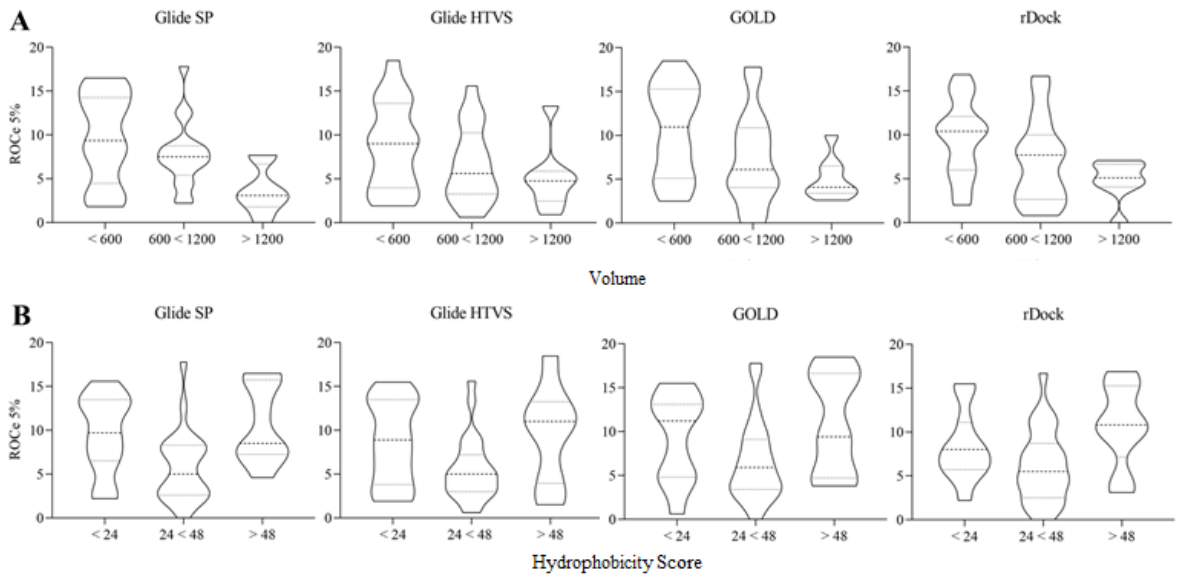

Figure 4. Distribution of ROCe (5%) values according to the volume and hydrophobicity of the binding pockets in the set of targets. (Top) Volume (Å^3) of the pockets (the maximum value is close to 1.900 Å^3; the number of targets in the three categories is 18, 17 and 9). (Bottom) Hydrophobicity score (the maximum value is 72; the number of targets in each category is 11, 23 and 9). ADA set was discarded, the values reported was not properly computed.

Performance of hybrid approaches with measurements of local similarity. Although CR appears to be suitable to correct the limitations of both docking and similarity measurements,

there are cases where this strategy leads to a negligible improvement in the reranking of compounds. This is exemplified by catechol O-methyltransferase (COMT in BS1 dataset), which also has a large solvent-exposed pocket that enables multiple binding modes (Figure 4, left).[80] In this case, the limited impact of the CR technique can be ascribed to not only to the limited accuracy of docking methods in predicting the structural arrangement of the ligand, but also to the different size of the ligands, which tends to penalize the molecular overlay with the template when a global similarity measurement is used. As an example, let us note the comparison of five active compounds (Figure 5, right) docked in a small portion of the binding site, but with a molecular size notably smaller than the reference template. None of them was reranked in the top 5% due to the penalized measurement of the global similarity made with the Tanimoto index. Therefore, the performance of the hybrid rankings, although promising for enhancing the ligand enrichment in VS, may be limited by the bias introduced in measurements of global similarity between molecules with disparate sizes. This limitation may be alleviated by the use of local similarity, which may be estimated with the Tversky coefficient.

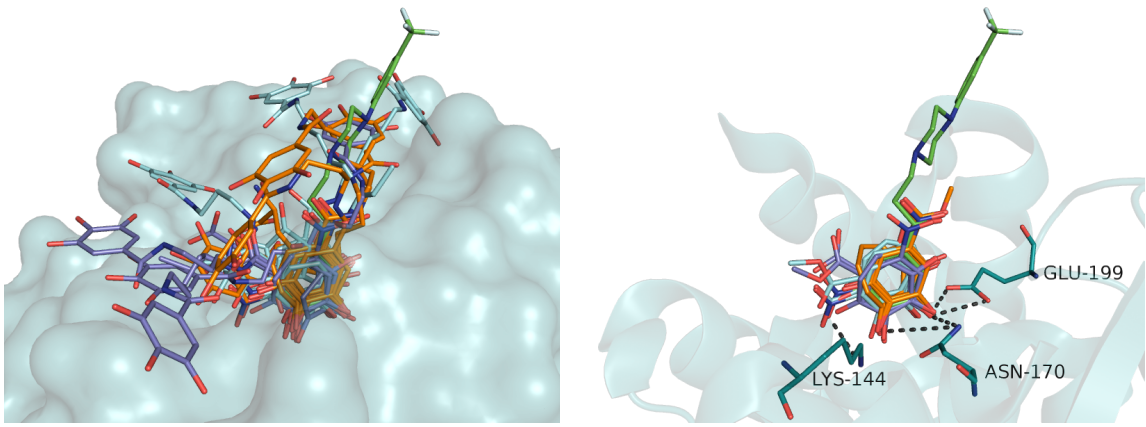

Figure 5. (Left) Binding mode of catechol O-methylltransferase (COMT in BS1 dataset; PDB code: 1H1D, shown as blue surface) with all ligands of DHFR_BS1 docked using Glide, rDock, and GOLD. (Right) Overlay of the reference molecule (co-crystallized structure BIA) and five catechol-containing ligands with a notorious difference in size relative to the reference compound. The reference molecule is shown as green sticks, and hits docked using Glide, rDock, and GOLD are shown as cyan, orange

and blue sticks, respectively. Hydrogen bonds between the reference compound and Lys144, Asn170 and Glu199 are shown as dashed lines.

Tables 2-5 show the ROCe and AUC values obtained upon application of local similarity measurements (the results for all individual sets are reported in the Supporting Information Tables S4-S7). Both RR and CR hybrid techniques outperform the ranking obtained from pure LB and SB (Glide, GOLD) VS tools in both BS1 and BS2 datasets, whereas the RR approach seems to be the best protocol for recovering actives with rDock. Indeed, the RR strategy leads to a notable increase in both ROCe for most BS1 sets. On the other hand, the AUC values obtained when RR and CR are used with Glide remain mostly unaffected, while larger improvements are obtained upon combination with GOLD and rDock.

Inspection of Tables 2-5 suggest that local similarity yields better results for the BS1 dataset, whereas measurements of global similarity are better for the BS2 data set. This behavior can be related to the different sensitivity of the hits included in BS1 and BS2 to measurements of structural resemblance to the reference ligand (Figure 6), which was estimated from the comparison of the molecular connectivity encoded in Morgan fingerprints (radius = 2) using both global (Tanimoto) and local (Tversky) similarity metrics calculated by RDKit.[58] For the BS1 dataset, 75% of the hits have a global similarity lower than 0.4, which increases up to 0.6 when local similarity is used. In contrast, the similarity of the hits in BS2 increases up to 0.4 upon replacement of the global similarity by the local one.

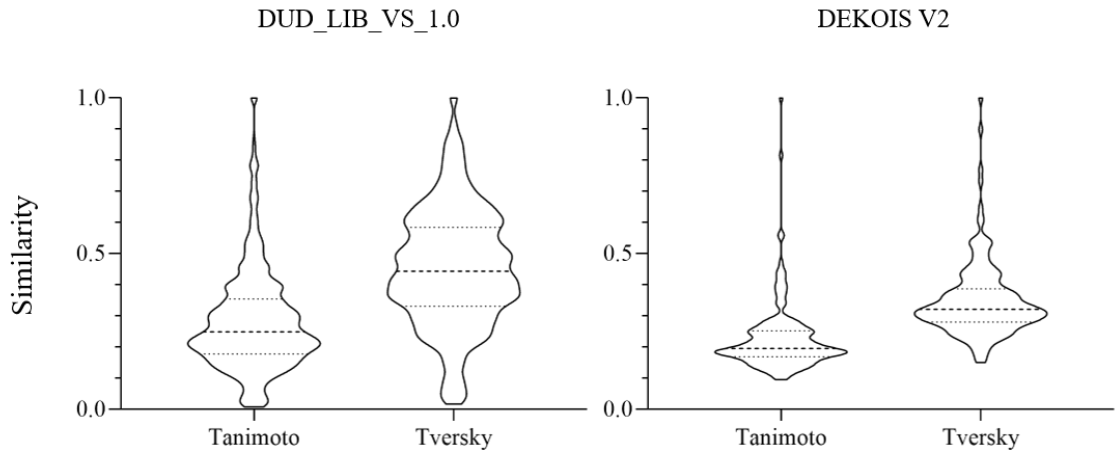

Figure 6. Distribution of hits in the BS1 (DUD_LIB_VS_1.0) and BS2 (DEKOIS V2) datasets according to the structural resemblance to the reference ligand.

These findings may raise the question whether the actives early recovered in the VS are those with larger resemblance according to their molecular connectivity with the template. To this end, we have examined the structural resemblance of the actives recovered in the ROCe 5% from the RR in conjunction with the four docking methods. The structural resemblance has been determined using Morgan fingerprints (radius = 2) calculated by RDKit[58] and measured with both global and local similarity metrics. As noted before, the usage of local similarities tends to increase the number of recovered actives, especially for the BS1 dataset. However, the number of compounds with high similarity (> 0.7) increases roughly twofold (from around 60 to 135 compounds) when the local similarity replaces the global one, this factor being increased to a factor ranging from 3 (Glide HTVS) to 4 (Glide SP) when the similarity threshold is reduced to 0.5.

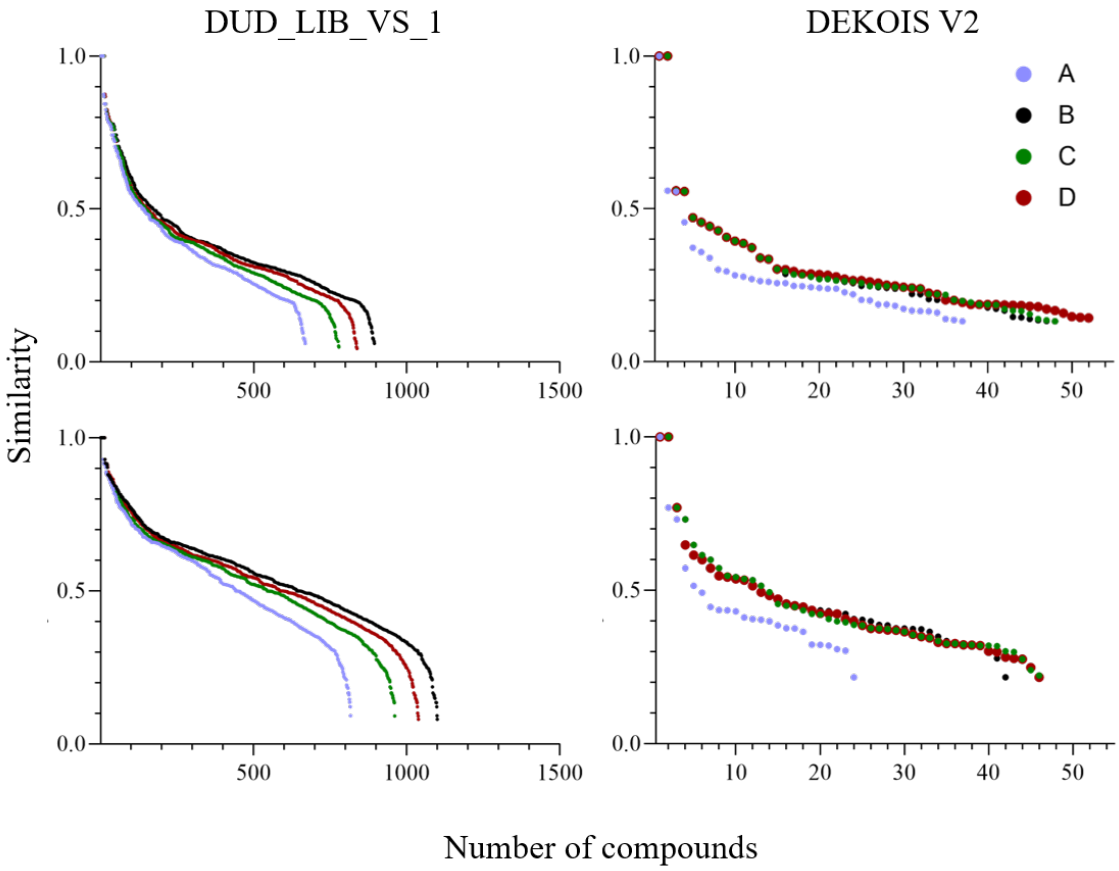

Figure 7. Distribution of actives in the first 5% of the ROCe curve obtained with the RR strategy according to the structural (Morgan fingerprint using RDKit, radius = 2) similarity to the reference ligand. Results determined by using (top) global and (bottom) local similarities for the BS1 (DUD_LIB_VS_1.0) and BS2 (DEKOIS V2) data sets. Results obtained for the combined RR strategy with Glide HTVS, Glide SP, GOLD and rDock are shown as blue, black, red and green dots, respectively.

The behavior of the combined LB+SB strategies is not uniform for all the sets, a trait already recognized in previous studies.[24] Besides the physicochemical features of the binding pocket and the docking method used in the SB screening, the enrichment may be influenced by the similarity metrics adopted in the comparison with the reference molecule. As an example, let us note that four out of the five hits discussed above (Figure 5, right) are rescored within 1% of the ranking using local similarity measurements for the three docking tools. In addition,

Figure 8 shows the hits found in the ROCe 0.5% using local similarity for tyrosine kinase c-src (SRC) and peroxisome proliferator-activated receptor gamma (PPARγ).[81,82] These systems were chosen due to the low number of actives recovered at this ROCe level when a global similarity is used: no actives were recovered for both molecular systems using Glide and rDock coupled to RR, whereas 1 and 3 actives were identified using GOLD for PPARγ and SRC, respectively. In contrast, local similarity measurements lead to the recovery of 3/17, 2/6, and 2/16 actives for PPARγ/SRC at the ROCe 0.5% level using Glide, rDock, and GOLD, respectively.

With regard to the BS2 targets, the use of local similarity leads to slightly lower averaged values compared to the global similarity ones. This primarily stems from the results obtained for HIV-1 protease (HIVPR_BS2) for the hybrid methods derived from GOLD and Glide, and for dihydrofolate reductase (DHFR_BS2) in the case of rDock, whereas the results remain unaltered for the rest of targets (see the Supporting Information Tables S4–S7). In particular, the decoys found at ROCe 0.5% with local similarity are ranked below position 90 when global similarity is used for HIVPR_BS2. Similarly, three out of the six first decoys found for DHFR_BS2 at the ROCe 0.5% are located below position 41 with the Tanimoto coefficient. This behavior arises from the comparison of compounds with dissimilar sizes, as shown in Figure 9, which displays the first decoy recovered for the RR approach in HIV1PR_BS2[83] using GOLD and Glide, and DHFR_BS2[84] with rDock.

A

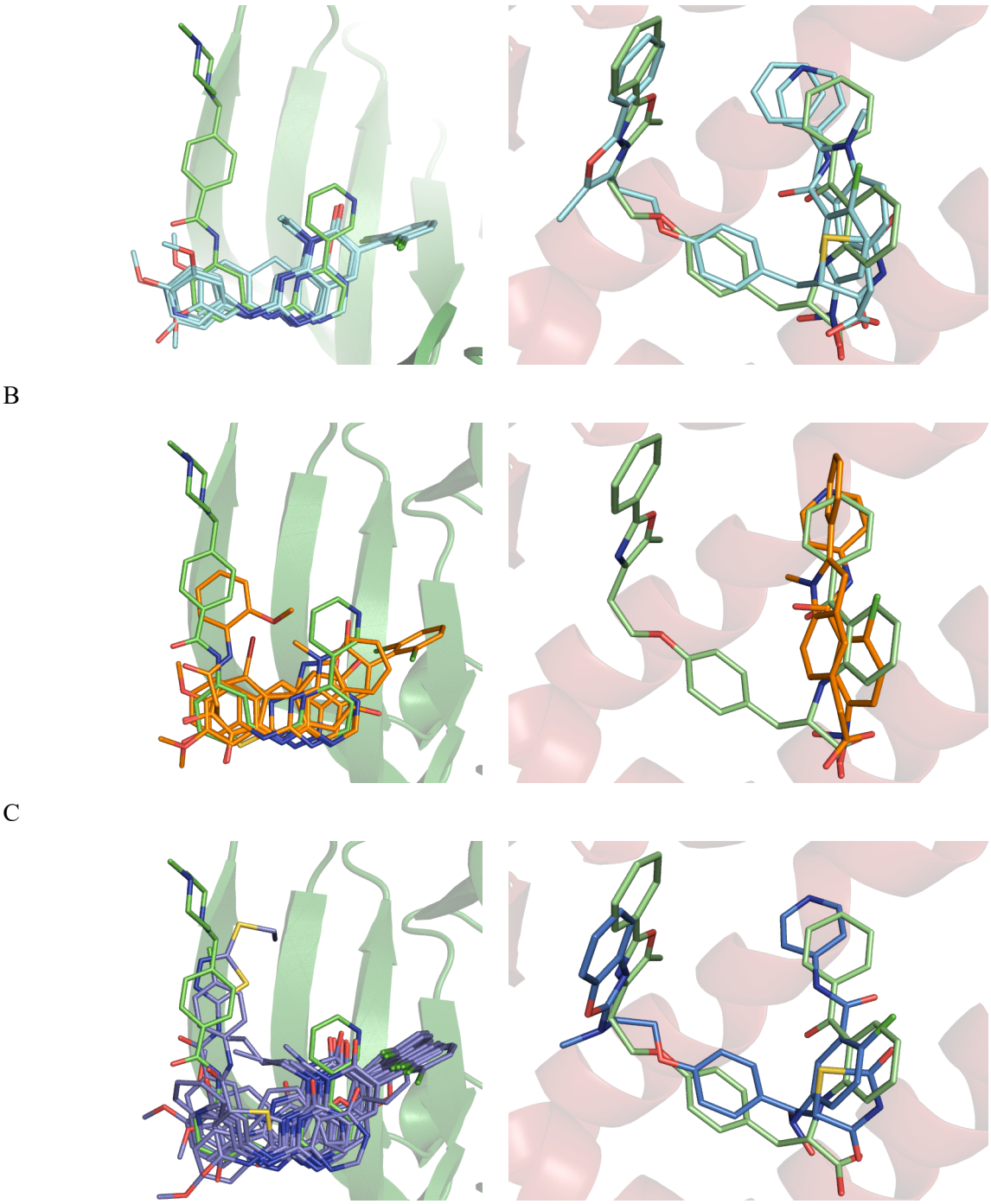

Figure 8. (Left) Binding mode of ligands to tyrosine kinase c-src (SRC in BS1 dataset; PDB code: 1Y57, green) at the ROCe 0.5% level. (Right) Binding mode of peroxisome proliferator-activated receptor gamma (PPARγ in BS1 data set; PDB code: 1FM9, green) at the ROCe 0.5% level. In the two cases the co-crystallized reference molecule is shown as green sticks, and hits docked using (A) Glide, (B) rDock, and (C) GOLD are shown as cyan, orange and blue sticks, respectively.

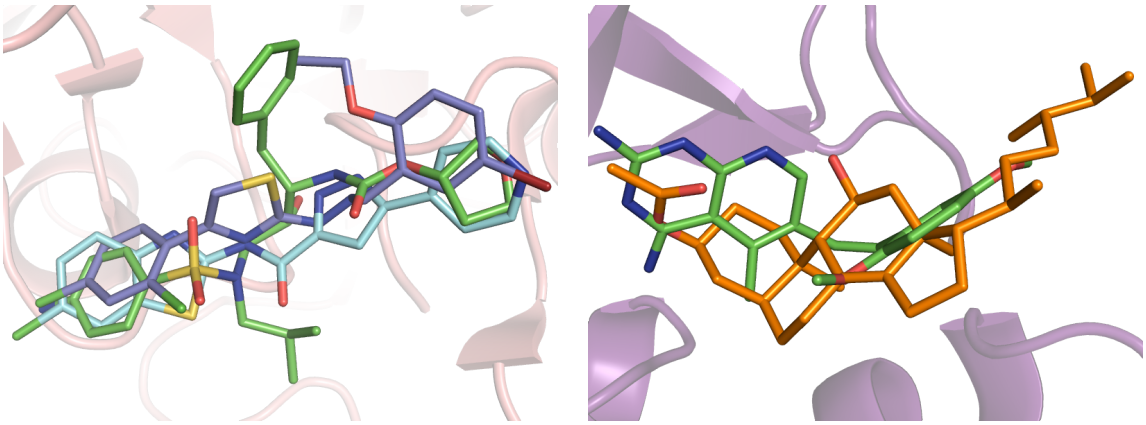

Figure 9. (Left) Binding mode of ligands to HIV-1 protease (HIVPR in BS2 dataset; PDB code: 3NU3, light red) at the ROCe 0.5% level. The co-crystallized reference molecule is shown as green sticks, and decoys docked with Glide and GOLD as cyan and blue sticks, respectively. (Right) Binding mode of ligands to dihydrofolate reductase (DHFR in BS2 dataset; PDB code: 1NHZ, purple) at the ROCe 0.5% level. The co-crystallized reference molecule is shown as green sticks, and the decoy docked using rDock as orange sticks.

As a final remark, the robustness of the hybrid RR and CR strategies used in conjunction with local similarity measurements is supported by the heatmaps provided in Figure 10, which displays the hierarchical position of the LB and SB methods as well as of the different LB+SB approaches for each of the 44 targets according to ROCe 1% and AUC (see Supporting Information Figures S3 and S4 for heatmaps to ROCe 0.5%, 2% and 5%). Figure 10 supports the better performance of the CR when used in combination with Glide_SP, as it recovers the largest number of hits according to the two metrics not only relative to PharmScreen and Glide_SP, but also in comparison with PR and RR. Nevertheless, this latter approach is the most robust strategy when used in combination with rDock and especially for GOLD, which can be attributed to a better performance of this latter method in predicting the pose of the docked ligands.

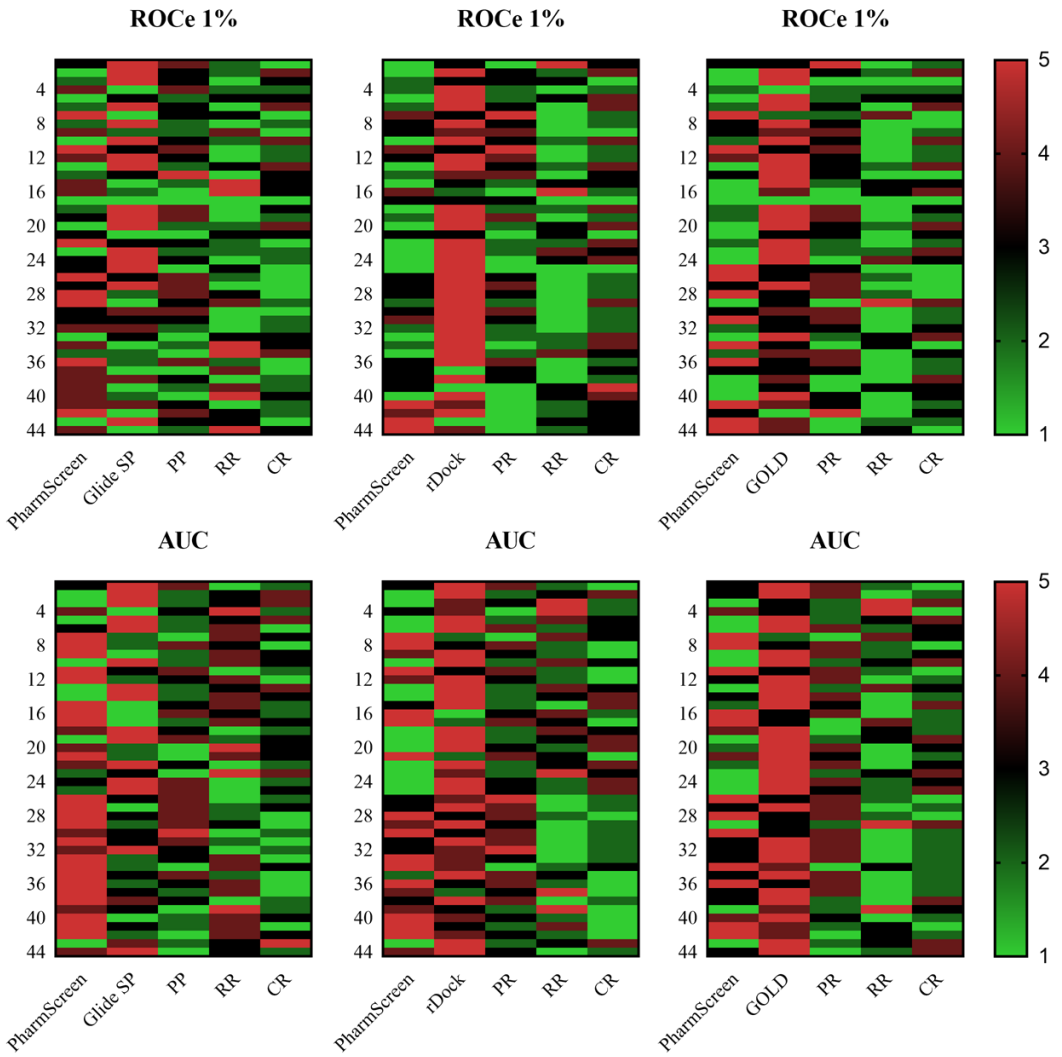

Figure 10. Heatmap of the hierarchical position of the LB, SB, and combined (PR, RR, and CR in conjunction with local similarity measurements) techniques for each of the 44 targets according to the ROCe 1 % and AUC. Results obtained relative to Glide, rDock and GOLD are provided in the left, middle and right columns. The color scale is indicative of the hierarchical position (from first to fifth) of a given technique according to the number of identified actives (best position in green, and worst position in red).

Analysis of chemotype diversity. An interesting outcome of VS is the feasibility of finding molecules with novel chemical scaffolds that expand the chemical diversity of active compounds. To evaluate the chemical diversity of the selected hits, a weighting scheme based on the ROC metric was applied.[49] As this analysis requires a clustered set, only the targets

included in the BS1 dataset were considered. Table 6 shows the average awROCe at different percentages for PharmScreen, Glide SP, rDock, GOLD and the two best combined methods, RR and CR coupled to local similarity, for each SB tool. The combined methods achieve the best overall performance in chemotype enrichment, in line with the ROCe values. In the case of the combined methods derived from Glide SP, CR improves the performance of RR, which is nevertheless the best combined method when used in conjunction with rDock and GOLD. In fact, the RR derived from GOLD poses is the best option among all others in the initial percentages of awROCe (results for individual sets are provided in the Supporting Information Tables S8-S10).

Table 6. awROCe metrics derived for PharmScreen, Glide SP, rDock, GOLD and the hybrid RR and CR methods coupled to each docking program. The best configuration is highlighted in bold for each column.

	awROCe 0.5%	**awROCe 1%**	**awROCe 2%**	**awROCe 5%**
PharmScreen	25.8	17.2	10.3	5.9
Glide SP				
Docking	29.0	19.6	12.7	6.7
RR	37.0	24.9	15.1	7.9
CR	41.5	26.3	**16.9**	**8.9**
rDock				
Docking	2.9	4.0	4.3	3.4
RR	35.1	22.5	13.9	6.8
CR	24.3	17.9	12.6	7.1
GOLD				
Docking	13.8	10.7	8.3	4.1
RR	**45.5**	**26.6**	15.0	7.5
CR	31.4	22.7	14.5	7.4

Figure 11 shows the binding site of thrombin as an example of the dependence of the hybrid RR on the docking pose. For this set, an active ligand is recovered at the awROCe 0.5% level

using GOLD poses, and none when docked poses generated by Glide and rDock are considered. The active recovered on the top of the RR obtained combining PharmScreen local similarity measurement with GOLD exhibits a high overlap with the reference molecule, while enabling the formation of the hydrogen bond between the ligand with Asp189 and Ser214 (Figure 11).[85] In contrast, rDock and Glide yield poses that have lower overlap with the template, leading to lower similarity scores.

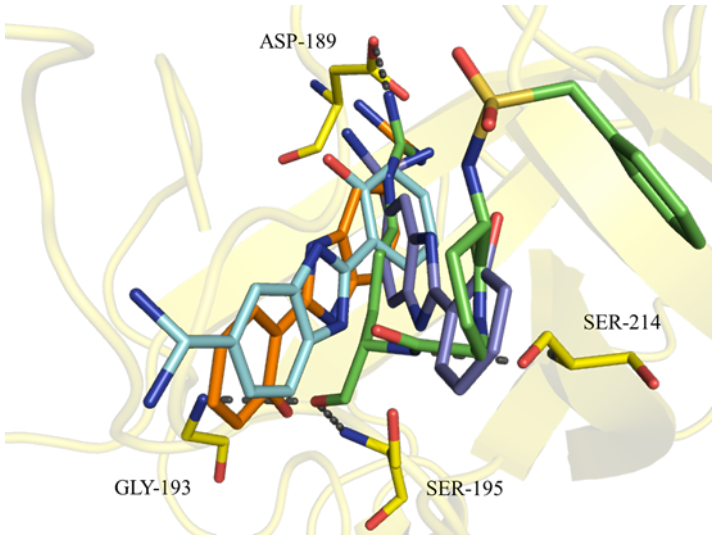

Figure 11. Binding mode of thrombin (PDB code 1BA8, yellow), co-crystallized reference molecule (green sticks), and the docked hit from Glide (cyan sticks), rDock (orange sticks), and GOLD (blue stick). Hydrogen bond interactions between the reference compound and Asp189, Ser214, Ser195, and the backbone of Gly193 are shown as dashed lines.

The heatmap of robustness for the awROCe 1% is shown in Figure 12 (individual values at all percentages are shown in the Supporting Information Tables S8-S10, and heatmaps at awROCe 0.5%, 2% and 5% in Figures S5 and S6). For all the percentages considered in this study, a maximum of two sets are classified in third position using RR in conjunction with the poses generated by rDock and GOLD. Likewise, only a single set is classified in third position when the CR approach is used with the poses taken from Glide SP. Overall, the improvement

of chemotype enrichment (awROCe) gives support to the synergy of combining LB and SB methods according to the RR and CR schemes.

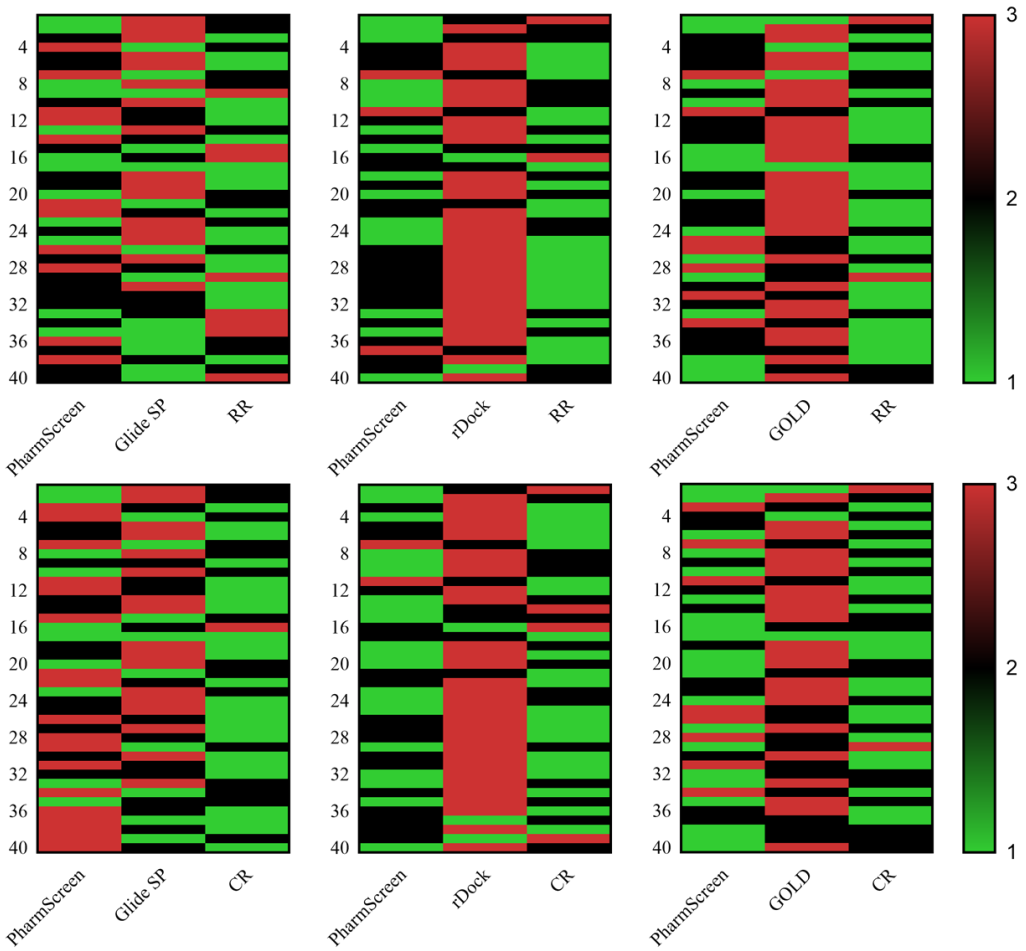

Figure 12. Heatmap of the hierarchical position of the LB, SB, and hybrid (top) RR and (bottom) CR, both determined in conjunction with local similarity measurements, according to the awROCe 1% level. The method ranked first is shown in green, the second in black and the third in red.

CONCLUSIONS

Since (de)solvation is fundamental for the binding of ligands to their pockets in macromolecular targets, it can be expected that active ligands bound to the same pocket share hydrophobic/hydrophilic features that are complementary to the residues that shape the binding pocket, thus providing clues about the molecular determinants that define the binding

mode. Inspired by the concept of 'consensus scoring',[86] hydrophobic similarity can be used to complement the information provided by SB methods in order to enhance the feasibility of discriminating between active and inactive compounds, leading to the three combined strategies examined here in conjunction with state-of-the-art molecular docking (Glide, rDock, and GOLD) programs.

The results obtained for the 44 sets of targets support the synergy of the hybrid LB+SB approaches, as the combined ranking consistently shows better performance than using only either LB or SB methods. Among the proposed protocols, CR and RR using partial similarity yield the best average performance in recovering actives in the datasets. The results also highlight the influence due to the specific docking formalism in driving the overall performance of two combined methods. In particular, the best performance is obtained by combining PharmScreen's hydrophobic similarity in conjunction with Glide within the CR approach, and with rDock and GOLD within the RR framework. Nevertheless, since the results are largely sensitive to the specific physicochemical features of the target, it seems desirable to confirm the suitability of these combined strategies in future blind challenges (http://drugdesigndata.org).

Let us note that the use of these strategies only implies a reduced increment in the computational expensiveness, making them promising to refine the results of VS studies. Finally, an essential feature of the combined methods introduced herein is that 3D similarity calculations are independent of the generation of docking poses. Hence, any existing ranking might also be re-evaluated based on 3D similarity calculations relative to suitable templates taken from available experimental data.

SUPPORTING INFORMATION

The Supporting Information is available free of charge at

https: //pubs.acs.org/doi/

Configuration files for docking tools, list of targets included in datasets BS1 and BS2, list of changes in protein residues, molecules discarded in docking calculations with SP and HTVS modes of Glide, ROCe and awROCe results for individual sets, graphical representation of the binding mode of DHFR with docked hits using rDock and Gold, and heatmaps of the hierarchical position of RR and CR regarding pure VS methods for ROCe and awROCe 0.5%, 2% and 5%.

Datasets available at the GitHub repository https://github.com/Pharmacelera/Molecular-Similarity-and-Docking.

AUTHOR INFORMATION

Corresponding Author

* E-mail: fjluque@ub.edu.

ORCID

Javier Vázquez: 0000-0003-4400-6378

Alessadro Deplano: 0000-0002-8451-5831

Enric Herrero: 0000-0001-7837-3593

F. Javier Luque: 0000-0002-8049-3567

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Dr. I. Giangreco, Dr. F. Stanzione and Dr. X. Barril for valuable suggestions and critical reading of the manuscript. We are grateful to the Cambridge Crystallographic Data Cenre (CCDC) for providing access to GOLD. We thank the Ministerio de Economia y Competitividad (MINECO; SAF2017-88107-R, DI-14-06634; AEI/FEDER UE) and the

Generalitat de Catalunya (2017SGR1746, 2015-DI-052) for financial support. The Consorci de Serveis Universitaris de Catalunya (CSUC) is acknowledged for computational resources (Molecular Recognition project). J.V. is fellowship from the Generalitat de Catalunya.

REFERENCES

(1) Sperandio, O.; Miteva, M.; Villoutreix, B. Combining Ligand- and Structure-Based Methods in Drug Design Projects. *Curr. Comput. Aided-Drug Des.* **2008**, *4*, 250–258.

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

(3) Lavecchia, A.; Di Giovanni, C. Virtual Screening Strategies in Drug Discovery: A Critical Review. *Curr. Med. Chem.* **2013**, *20*, 2839–2860.

(4) Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S. Role of Computer-Aided Drug Design in Modern Drug Discovery. *Arch. Pharm. Res.* **2015**, *38*, 1686–1701.

(5) Klopmand, G. Concepts and Applications of Molecular Similarity, by Mark A. Johnson and Gerald M. Maggiora, Eds., John Wiley & Sons, New York, 1990, 393 P. In *Journal of Computational Chemistry*; John Wiley & Sons, Ltd, 1992; Vol. 13, pp 539–540.

(6) Maldonado, A. G.; Doucet, J. P.; Petitjean, M.; Fan, B.-T. Molecular Similarity and Diversity in Chemoinformatics: From Theory to Applications. *Mol. Divers.* **2006**, *10*, 39–79.

(7) Villoutreix, B. O.; Renault, N.; Lagorce, D.; Sperandio, O.; Montes, M.; Miteva, M. A. Free Resources to Assist Structure-Based Virtual Ligand Screening Experiments. *Curr. Protein Pept. Sci.* **2007**, *8*, 381–411.

(8) Banegas-Luna, A.-J.; Cerón-Carrasco, J. P.; Pérez-Sánchez, H. A Review of Ligand-Based Virtual Screening Web Tools and Screening Algorithms in Large Molecular Databases in the Age of Big Data. *Future Med. Chem.* **2018**, *10*, 2641–2658.

(9) Hein, M.; Zilian, D.; Sotriffer, C. A. Docking Compared to 3D-Pharmacophores: The Scoring Function Challenge. *Drug Discov. Today Technol.* **2010**, *7*, e229–e236.

(10) Barril, X.; Morley, S. D. Unveiling the Full Potential of Flexible Receptor Docking Using Multiple Crystallographic Structures. *J. Med. Chem*. **2005**, *48*, 4432–4443.

(11) Cavasotto, C. N.; Orry, A. J. W. Ligand Docking and Structure-Based Virtual Screening in Drug Discovery. *Curr. Top. Med. Chem*. **2007**, *7*, 1006–1014.

(12) Cappel, D.; Wahlstrom, R.; Brenk, R.; Sotiffre, C. A. Probing the Dynamic nature of Water Molecules and Their Influences on Ligand Binding in a Model Binding Site. *J. Chem. Inf. Model*. **2011**, *51*, 2581–2594.

(13) Chen, Y.-C. Beware of Docking. *Trends Pharmacol. Sci*. **2015**, *36*, 78–95.

(14) Sotriffer, C. Docking of Covalent Ligands: Challenges and Approches. *Mol. Inf*. **2018**, *37*, 1800062.

(15) Palacio-Rodríguez, K.; Lans, I.; Cavasotto, C. N.; Cossio, P. Exponential Consensus Ranking Improves the Outcome in Docking and Receptor Ensemble Docking. *Sci. Rep*. **2019**, *9*, 5142.

(16) Wilson, G. L.; Lill, M. A. Integrating Structure-Based and Ligand-Based Approaches for Computational Drug Design. *Future Med. Chem.* **2011**, *3*, 735–750.

(17) Borges, N. M.; Rodrigues Sartori, G.; Ribeiro, J. F. R.; Rocha, J. R.; Martins, J. B. L.; Montanari, C. A.; Gargano, R. Similarity Search Combined with Docking and Molecular Dynamics for Novel HAChE Inhibitor Scaffolds. *J. Mol. Model.* **2018**, 24–41.

(18) Anighoro, A.; Bajorath, J. A Hybrid Virtual Screening Protocol Based on Binding Mode Similarity. In *Rational Drug Design*; Humana Press, New York, NY, 2018; pp 165–175.

(19) da Silva Figueiredo Celestino Gomes, P.; Da Silva, F.; Bret, G.; Rognan, D. Ranking Docking Poses by Graph Matching of Protein–Ligand Interactions: Lessons Learned from the D3R Grand Challenge 2. *J. Comput. Aided. Mol. Des.* **2018**, *32*, 75–87.

(20) Banoglu, E.; Çalışkan, B.; Luderer, S.; Eren, G.; Özkan, Y.; Altenhofen, W.; Weinigel, C.; Barz, D.; Gerstmeier, J.; Pergola, C.; et al. Identification of Novel Benzimidazole Derivatives as Inhibitors of Leukotriene Biosynthesis by Virtual Screening Targeting 5-Lipoxygenase-Activating Protein (FLAP). *Bioorg. Med. Chem.* **2012**, *20*, 3728–3741.

(21) Smith, J. R.; Evans, K. J.; Wright, A.; Willows, R. D.; Jamie, J. F.; Griffith, R. Novel Indoleamine 2,3-Dioxygenase-1 Inhibitors from a Multistep in Silico Screen. *Bioorg. Med. Chem.* **2012**, *20*, 1354–1363.

(22) Drwal, M. N.; Agama, K.; Wakelin, L. P. G.; Pommier, Y.; Griffith, R. Exploring DNA Topoisomerase I Ligand Space in Search of Novel Anticancer Agents. *PLoS One* **2011**, *6*, e25150.

(23) Swann, S. L.; Brown, S. P.; Muchmore, S. W.; Patel, H.; Merta, P.; Locklear, J.; Hajduk, P. J. A Unified, Probabilistic Framework for Structure- and Ligand-Based Virtual Screening. *J. Med. Chem.* **2011**, *54*, 1223–1232.

(24) Svensson, F.; Karlén, A.; Sköld, C. Virtual Screening Data Fusion Using Both Structure- and Ligand-Based Methods. *J. Chem. Inf. Model.* **2012**, *52*, 225–232.

(25) Tan, L.; Geppert, H.; Sisay, M. T.; Gütschow, M.; Bajorath, J. Integrating Structure- and Ligand-Based Virtual Screening: Comparison of Individual, Parallel, and Fused Molecular Docking and Similarity Search Calculations on Multiple Targets. *ChemMedChem* **2008**, *3*, 1566–1571.

(26) Larsson, M.; Fraccalvieri, D.; Andersson, C. D.; Bonati, L.; Linusson, A.; Andersson, P. L. Identification of Potential Aryl Hydrocarbon Receptor Ligands by Virtual Screening of Industrial Chemicals. *Environ. Sci. Pollut. Res.* **2018**, *25*, 2436–2449.

(27) Meslamani, J.; Li, J.; Sutter, J.; Stevens, A.; Bertrand, H.-O.; Rognan, D. Protein–Ligand-Based Pharmacophores: Generation and Utility Assessment in Computational Ligand Profiling. *J. Chem. Inf. Model.* **2012**, *52*, 943–955.

(28) Anighoro, A.; Bajorath, J. Three-Dimensional Similarity in Molecular Docking: Prioritizing Ligand Poses on the Basis of Experimental Binding Modes. *J. Chem. Inf.*

Model. **2016**, *56*, 580–587.

(29) Anighoro, A.; Bajorath, J. Binding Mode Similarity Measures for Ranking of Docking Poses: A Case Study on the Adenosine A2Areceptor. *J. Comput. Aided. Mol. Des.* **2016**, *30*, 447–456.

(30) Anighoro, A.; Bajorath, J. Compound Ranking Based on Fuzzy Three-Dimensional Similarity Improves the Performance of Docking into Homology Models of G-Protein-Coupled Receptors. *ACS Omega* **2017**, *2*, 2583–2592.

(31) Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of Protein - Ligand Interactions. Docking and Scoring: Successes and Gaps. *J. Med. Chem.* **2006**, *49*, 5851−5855.

(32) Tirado-Rives, J.; Jorgensen, W. L. Contribution of Conformer Focusing to the Uncertainty in Predicting Free Eneregies for Protein- Ligand Bindning. *J. Med. Chem.* **2006**, *49*, 5880−5884.

(33) Luque, F. J.; Barril, X.; Orozco, M. Fractional Description of Free Energies of Solvation. *J. Comput. Aided. Mol. Des.* **1999**, *13*, 139–152.

(34) Muñoz, J.; Barril, X.; Hernández, B.; Orozco, M.; Javier Luque, F. Hydrophobic Similarity between Molecules: A MST-Based Hydrophobic Similarity Index. *J. Comput. Chem.* **2002**, *23*, 554–563.

(35) Muñoz-Muriedas, J.; Perspicace, S.; Bech, N.; Guccione, S.; Orozco, M.; Luque, F. J. Hydrophobic Molecular Similarity from MST Fractional Contributions to the Octanol/Water Partition Coefficient. *J. Comput. Aided. Mol. Des.* **2005**, *19*, 401–419.

(36) Ginex, T.; Muñoz-Muriedas, J.; Herrero, E.; Gibert, E.; Cozzini, P.; Luque, F. J. Development and Validation of Hydrophobic Molecular Fields from the Quantum Mechanical IEF/PCM-MST Solvation Models in 3D-QSAR. *J. Comput. Chem*. **2016**, *37*, 1147–1162.

(37) Ginex, T.; Vazquez, J.; Gilbert, E.; Herrero, E.; Luque, F. J. Lipophilicity in Drug Design: An Overview of Lipophilicity Descriptors in 3D-QSAR Studies. *Future Med. Chem.* **2019**, 11, 1177–1193.

(38) Vázquez, J.; Deplano, A.; Herrero, A.; Ginex, T.; Gibert, E.; Rabal, O.; Oyarzabal, J.; Herrero, E.; Luque, F. J. Development and Validation of Molecular Overlays Derived from Three-Dimensional Hydrophobic Similarity with PharmScreen. *J. Chem. Inf. Model.* **2018**, *58*, 1596–1609.

(39) Luque, F. J.; Curutchet, C.; Muñoz-Muriedas, J.; Bidon-Chanal, A.; Soteras, I.; Morreale, A.; Gelpí, J. L.; Orozco, M. Continuum Solvation Models: Dissecting the Free Energy of Solvation. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3827–3836.

(40) Curutchet, C.; Orozco, M.; Luque, F. J. Solvation in Octanol: Parametrization of the Continuum MST Model. *J. Comput. Chem.* **2001**, *22*, 1180–1193.

(41) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. RDock: A Fast, Versatile

and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput. Biol.* **2014**, *10*, e1003571.

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

(43) Schrodinger. Glide. LLC: New York.

(44) 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.; et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(45) 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.

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

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

(48) 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.

(49) Jahn, A.; Hinselmann, G.; Fechner, N.; Zell, A. Optimal Assignment Methods for Ligand-Based Virtual Screening. *J. Cheminform.* **2009**, *1*, 14.

(50) Oprea, T. I.; Davis, A. M.; Teague, S. J.; Leeson, P. D. Is There a Difference between Leads and Drugs? A Historical Perspective. *J. Chem. Inf. Comput. Sci.* **2002**, *41*, 1308–1315.

(51) Barker, E. J.; Gardiner, E. J.; Gillet, V. J.; Kitts, P.; Morris, J. Further Development of Reduced Graphs for Identifying Bioactive Compounds. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 346–356.

(52) Bauer, M. R.; Ibrahim, T. M.; Vogel, S. M.; Boeckler, F. M. Evaluation and Optimization of Virtual Screening Workflows with DEKOIS 2.0 – A Public Library of Challenging Docking Benchmark Sets. *J. Chem. Inf. Model.* **2013**, *53*, 1447–1462.

(53) Anighoro, A.; Bajorath, J. Three-Dimensional Similarity in Molecular Docking: Prioritizing Ligand Poses on the Basis of Experimental Binding Modes. *J. Chem. Inf. Model.* **2016**, *56*, 580–587.

(54) Chen, X.; Liu, M.; Gilson, M. BindingDB: A Web-Accessible Molecular Recognition Database. *Comb. Chem. High Throughput Screen.* **2001**, *4*, 719–725.

(55) Forti, F.; Barril, X.; Luque, F. J.; Orozco, M. Extension of the MST Continuum Solvation Model to the RM1 Semiempirical Hamiltonian. *J. Comput. Chem.* **2008**, *29*, 578–587.

(56) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem.* **2006**, *27*, 1101–1111.

(57) MolVS 0.1.1: Molecule Validation and Standardization. https://pypi.org/project/MolVS/ (accessed Apr 11, 2018).

(58) Landrum, G. RDKit: Open-Source Cheminformatics. 2006.

(59) Luque, F. J.; Orozco, M. Version Locally Modified MOPAC 6.0. University of Barcelona 2008.

(60) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nuc. Acids Res*. **2000**, *28*, 235–242.

(61) RCSB Protein Data Bank http://www.pdb.org.

(62) Maestro, Schrödinger Release 2019-3. Schrödinger, LLC, New York, NY, 2019.

(63) 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.

(64) PharmScreen. Release 2019. Pharmacelera, Barcelona, 2019.

(65) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge Structural Database. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72*, 171–179.

(66) Tversky, A. Features of Similarity. *Psychol. Rev.* **1977**, *84*, 327–352.

(67) Maggiora, G. M.; Petke, J. D.; Mestres, J. A General Analysis of Field-Based Molecular Similarity Indices. *J. Math. Chem*. **2002**, *31*, 251–270.

(68) 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.

(69) Zhao, W.; Hevener, K. E.; White, S. W.; Lee, R. E.; Boyett, J. M. A Statistical Framework to Evaluate Virtual Screening. *BMC Bioinformatics* **2009**, *10*, 225.

(70) Sheridan, R. P. Alternative Global Goodness Metrics and Sensitivity Analysis: Heuristics to Check the Robustness of Conclusions from Studies Comparing Virtual Screening Methods. *J. Chem. Inf. Model.* **2008**, *48*, 426–433.

(71) Jain, A. N.; Nicholls, A. Recommendations for Evaluation of Computational Methods. *J. Comput. Aided. Mol. Des.* **2008**, *22*, 133–139.

(72) Klon, A. E.; Heroux, A.; Ross, L. J.; Pathak, V.; Johnson, C. A.; Piper, J. R.; Borhani, D. W. Atomic Structures of Human Dihydrofolate Reductase Complexed with NADPH and Two Lipophilic Antifolates at 1.09 and 1.05 Å resolution. *J. Mol. Biol*. **2002**, *320*, 677–693.

(73) Presnell, S. R.; Patil, G. S.; Mura, C.; Jude, K. M.; Conley, J. M.; Bertrand, J. A.; Kam,

C. M.; Powers, J. C.; Williams, A. D. Oxyanion-Mediated Inhibition of Serine Proteases. *Biochemistry* **1998**, *37*, 17068–17081.

(74) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-Based Maximal Affinity Model Predicts Small-Molecule Druggability. *Nat. Biotechnol.* **2007**, *25*, 71–75.

(75) Hajduk, P. J.; Huth, J. R.; Fesik, S. W. Druggability Indices for Protein Targets Derived from NMR-Based Screening Data. *J. Med. Chem.* **2005**, *48*, 2518–2525.

(76) Egner, U.; Hillig, R. C. A Structural Biology View of Target Drugability. Expert Opin. *Expert Opin. Drug Discovery* **2008**, *3*, 391–401.

(77) Schmidtke, P.; Barril, X. Understanding and Predicting Druggability. A High-Throughput Method for Detection of Drug Binding Sites. *J. Med. Chem.* **2010**, *53*, 5858–5867.

(78) Volkamer, A.; Kuhn, D.; Rippmann, F.; Rarey, M. DoGSiteScorer: A Web Server for Automatic Binding Site Prediction, Analysis and Druggability Assessment. *Bioinformatics* **2012**, *28*, 2074-2075.

(79) Monera, O. D.; Sereda, T. J.; Zhou, N. E.; Kay, C. M.; Hodges, R. S. Relationship of sidechain hydrophobicity and alpha-helical propensity on the stability of the single-stranded amphipathic alpha-helix. *J. Pept. Sci*. **1995**, *1*, 319–329.

(80) Bonifacio, M. J.; Archer, M.; Rodrigues, M. L.; Matias, P. M.; Learmonth, D. A.; Carrondo, M. A.; Soares-da-Silva, P. Kinetics and Crystal Structure of Catechol-O-Methyltransferase Complex with Co-Substrate and a Novel Inhibitor with Potential Therapeutic Application. *Mol. Pharmacol*. **2002**, *62*, 795–805.

(81) Cowan-Jacob, S. W.; Fendrich, G.; Manley, P. W.; Jahnke, W.; Fabbro, D.; Liebetanz, J.; Meyer, T. The Crystal Structure of a c-Src Complex in an Active Conformation Suggests Possible Steps in c-Src Activation. *Structure* **2005**, *13*, 861–871.

(82) Gampe, R. T., Jr.; Montana, V. G.; Lambert, M. H.; Miller, A. B.; Bledose, R. K.; Milburn, M. V.; Kliewer, S. A.; Wilson, T. M.; Xu, H. E. Asymmetry in the PPARγ/RXRα Crystal Structure Reveals the Molecular Basis of Heterodimerization Among Nuclear Receptors. *Mol. Cell* **2000**, *5*, 545–555.

(83) Shen, C. H.; Wang, Y. F.; Kovalevsky, A. Y.; Harrison, R. W.; Weber, I. T. Amprenavir Complexes with HIV-1 Protease and its Drug-Resistant Mutants Altering Hyrdrophobic Clusters. *Febs J.* **2010**, *277*, 3699–3714.

(84) Kauppi, B.;, Jakob, C.; Färnegårdh, M.; Yang, J.; Ahola, H.; Alarcon, M.; Calles, K.; Engström, O.; Harlan, J.; Muchmore, S.; Ramqvist, A. K.; Thorell, S.; Ohman L.; Greer J.; Gustafsson, J. A.; Carlstedt-Duke, J.; Carlquist, M. The Three-Dimensional Structures of Antagonistic and Agonistic Forms of the Glucocorticoid Receptor Ligand-binding Domain: RU-486 Induces a Transconformation that Leads to Active Antagonism. *J. Biol. Chem.* **2003**, *278*, 22748–22754.

(85) Krishnan, R.; Zhang, E.; Hakansson, K.; Arni, R. K.; Tulinsky, A.; Lim-Wilby, M. S.;

Levy, O. E.; Semple, J. E.; Brunck, T. K. Highly Selective Mechanism-Based Thrombin Inhibitors: Structures of Thrombin and Trypsin Inhibited with Rigid Peptidyl Aldehydes. *Biochemistry* **1998**, *37*, 12094-12103.

(86) Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus Scoring: A Method for Obtaining Improved Hit Rates from Docking Databases of Three-Dimensional Structures into Proteins. *J. Med. Chem.* **1999**, *42*, 5100–5109.