Article

Assessment of the Fragment Docking Program SEED

Kenneth Goossens, Berthold Wroblowski, Cassiano Langini, Herman van Vlijmen, Amedeo Caflisch, and Hans De Winter*

Cite This: https://dx.doi.org/10.1021/acs.jcim.0c00556

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: The fragment docking program solvation energy for exhaustive docking (SEED) is evaluated on 15 different protein targets, with a focus on enrichment and the hit rate. It is shown that SEED allows for consistent computational enrichment of fragment libraries, independent of the effective hit rate. Depending on the actual target protein, true positive rates ranging up to 27% are observed at a cutoff value corresponding to the experimental hit rate. The impact of variations in docking protocols and energy filters is discussed in detail. Remaining issues, limitations, and use cases of SEED are also discussed. Our results show that fragment library selection or enhancement for a particular target is likely to benefit from docking with SEED, suggesting that SEED is a useful resource for fragment screening campaigns. A workflow is presented for the use of the program in virtual screening, including filtering and postprocessing to optimize hit rates.
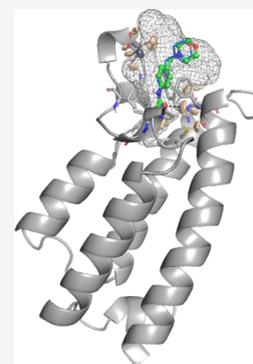
INTRODUCTION

The vast druglike chemical space has been estimated to contain over 10^{60} molecules.[1] Needless to say, no currently available techniques come close to being able to screen a molecular library of this size. For this reason, fragment-based drug design (FBDD) has become a standard lead discovery approach in the pharmaceutical industry since its introduction in the 1990s.[2] The main advantage of fragment-based approaches is that the fragment space can be sampled more efficiently than the druglike space.[3,4] The relatively low level of complexity of fragments also facilitates optimization by medicinal chemists, as well as increases efficiency and the success rate of screening campaigns. In the past, the importance of ligand efficiency has also been discussed.[5−7] Because fragment hits are so small, they are expected to bind very efficiently through strong contacts and provide a more suitable starting point for lead optimization than large ligands for which key interacting groups are not so easily distinguished from less important functional groups. However, due to their limited number of functional groups, absolute binding affinities of fragments are low compared to those of druglike molecules and binding is often promiscuous. Thus, identifying fragment hits remains challenging with any of the techniques currently used.

Currently, FBDD is a drug discovery approach that mainly uses biophysical techniques for measuring protein-ligand binding affinities.[8] In the past two decades, many efforts have been made to develop FBDD methods into valuable tools in both the academia and the industry. Biophysical techniques that are currently used for detection of fragment hits include nuclear magnetic resonance (NMR) spectroscopy, surface plasmon resonance (SPR), isothermal titration calorimetry (ITC), and X-ray crystallography.[8−10] Each of these methods

has its own advantages and disadvantages, so they are often combined in fragment screening campaigns.[11] Virtual screening (VS), in which compounds are identified by computational methods including docking and shape-based similarity screening,[12] has become a highly effective approach.[13−15] Fragment-based VS, in comparison, has been used less frequently. As mentioned before, the small size and relatively simple structure of fragments make it difficult to distinguish binders from nonbinders. Minor differences in affinity are a major obstacle for VS, where factors like scoring function inaccuracies, energy approximations, and limited conformational sampling also come into play. Furthermore, while many comparative studies have been published on docking of drug-sized small molecules, few studies describe the efficacy of fragment-based VS.[16−19] Standardized, well-defined data sets of fragment screening experiments that contain hits and nonhits do not exist at the time of writing, so it can be difficult to perform any meaningful evaluation for fragment-based docking tools.

As mentioned above, one of the advantages of FBDD is the ability to explore chemical space more efficiently. The relevant fragment space is, however, still orders of magnitude larger than the number of compounds that can be realistically validated by biophysical techniques.[20] As such, in silico screening of these fragment libraries provides an efficient and

Table 1. Properties of the 15 Data Sets Included in This Study

target name	number of fragments[a]	number of hits[b]	origin[c]	experimental screening method[d]	PDB id: structure 1/structure 2
BACE	805	20	ZoBio	SPR	5HDZ/2QP8
BLAC	734	6	ZoBio	SPR	4KZ4/4KZ5
BTK	2467	64	Janssen	ThermoFluor	in-house only
EPHA4	743	6	ZoBio	% inhibition assay	2XYU/2Y6O
FALZ	395	8	SGC	X-ray	FALZA-x0177/FALZA-x0309[e]
HPK1	1271	22	Janssen	displacement assay	in-house only
HSP90	724	69	ZoBio	TINS	3K98/2YEC
NUDT5	694	36	SGC	X-ray	5QJ4/5QJA
NUDT7	792	31	SGC	X-ray	5QGP/5QGY
OGA	751	61	Janssen	NMR	5UN9/in-house
PARP14	645	18	SGC	X-ray	5QHV/5QI3
PCAF	534	7	ZoBio	SPR	5FE8/5FE9
PIM1	587	19	ZoBio	X-ray	1XWS/3R04
NhDMX	736	10	ZoBio	NMR	in-house only
KHK	380	41	Janssen	X-ray	in-house only

[a]The number of fragments is the size of the fragment library after filtering out molecules with >2 rotatable bonds except for the FALZ, NUDT5, NUDT7, and PARP14 targets, for which the unfiltered libraries were screened. In Table S1, the size of the unfiltered libraries is given. [b]Number of hits in the filtered libraries. [c]SGC, structural genomics consortium; Janssen, Janssen internal data set; ZoBio, ZoBio internal data set. [d]X-ray, X-ray crystallography; SPR, surface plasmon resonance; NMR, nuclear magnetic resonance; TINS, target immobilized NMR screening. [e]The crystal structures are available at https://www.thesgc.org/ligand-bounds/falza.

low-cost alternative method to these existing biophysical techniques, and as such it can be a valuable technique to supplement fragment-based experimental techniques.

Solvation energy for exhaustive docking (SEED) is an open-source program for docking mainly rigid small molecules by an exhaustive search on discretized three-dimensional space. The development of SEED is hosted at https://gitlab.com/CaflischLab/SEED. The method was first published in 1999.[21] The force-field-based energy function used in SEED consists of four terms: (1) electrostatic and (2) van der Waals (vdW) interactions between the protein and the fragment, and electrostatic desolvation penalty of (3) the protein and (4) the fragment. An efficient numerical approach to the generalized Born equation was developed for calculating the electrostatic contribution to the binding free energy.[22] As a consequence, the docking of a fragment takes about 1−10 s depending on the definition of the binding site. Further details on the SEED docking protocol and energy function can be found in the original papers and in the documentation at https://caflischlab-seed.readthedocs.io/en/latest/index.html.[21,22] SEED has been used successfully in several fragment screening campaigns, with hit rates of up to 40%.[23−27] A review on SEED campaigns up to 2017 has been published recently.[28]

Aside from SEED, very few tools exist that are dedicated to fragment docking. LUDI, a rule-based docking program, is one of the only other specialized tools.[29] Generally, traditional docking programs like Flexx,[30] GOLD,[31] DOCK,[32] Glide,[33] Autodock,[34] and AutoDock VINA[35] are also suitable for fragment docking, but evaluation of performance is largely absent.[36,37] In this work, a retrospective docking study is carried out for analyzing the performance of SEED on 15 different protein targets for which experimental fragment screening data containing hits and nonhits is available. In addition, a comparison is presented with Glide, a well-established docking tool. The libraries range in size from nearly 400 to 3500 fragments, and the fraction of experimental hits ranges from 0.8 to 10.8%. The potential for enrichment is evaluated, and the impact of changes in protocols and filters on the performance of SEED is investigated. The optimal protocols for fragment-based screening in SEED are discussed, and a general workflow is suggested.

METHODS

Data Sets. Table 1 shows the 15 data sets used for the assessment of SEED and the comparison with Glide. The table lists the target names, number of fragments in the libraries, the hit rates, experimental techniques of the in vitro screening campaigns, and the PDB id of the proteins used for the VS. In addition, the total number of fragments before filtering and some further comments are provided in Table S1 as Supporting Information. X-ray-determined binding poses for the hits in the SGC data sets are publicly available and are provided in the Associated Content section.

We used the same fragment libraries for experimental and virtual screening. Minor differences in the virtual library are possible but are negligible and are related to fragments that failed the preprocessing (parametrization, missing structural information, etc.) for the docking software. The experimental library sizes and hit rates were left unaltered to allow for a fair comparison between experimental and virtual screening campaigns.

Preparation of Protein Structures and Libraries of Fragments. It is generally known that small deviations in active-site root-mean-square deviation (RMSD) can have a strong impact on VS results.[38] Therefore, two different protein structures were prepared for each data set (with the exception of NhDMX) to account for the impact of small conformational changes in the binding site on the performance of the VS protocols. Reported results are subsequently the results of the best-performing conformation for any given protocol. In case multiple holo-crystal structures were available, selection of the two structures was based on a visual inspection of the conformational differences with preference to those structures having large differences. The selected binding pockets for the targets of the data sets were defined based on available X-ray crystal structure data. The binding site chosen was the site where the largest number of ligands or fragments was located in holo-protein crystal structures. Hydrogens, missing atoms,

and side chains were modeled using the Biologics Modeling package in the Schrödinger Suite 2019.[39,40] Explicit crystal water molecules were kept if they showed more than two interactions with protein residues and were consistently found in multiple crystal structures. Bond orders were assigned, and hydrogens were added. Next, the orientations of amide (Asn and Gln), hydroxyl (Ser, Thr, and Tyr), and thiol (Cys) groups and the tautomeric state of His residues were optimized using the exhaustive sampling option. Protonation states of the titratable side chains were calculated at pH 7.4 using the Schrödinger Biologics Modeling package. Finally, hydrogen positions were minimized while heavy atoms were constrained using the Protein Preparation Wizard in Maestro.[41]

Fragment libraries were imported into MOE as SMILES and were converted to two-dimensional structures.[42] Because SEED uses rigid conformations of the ligand for docking, molecules with more than two rotatable bonds were excluded except for the FALZ, NUDT5, NUDT7, and PARP14 data sets. For these four sets, filtering out fragments with more than two rotatable bonds would lead to libraries containing less than 300 fragments. The libraries were then imported into Maestro in the SDF format, and three-dimensional (3D) conformations were generated for all occurring protomers at pH 7 ± 1 using the LigPrep module, preserving the tautomeric form defined in the library. Then, up to 64 conformations were generated for docking with SEED for each fragment using ConfGen.[43]

Data Set Analysis. The conformational differences in the binding site between both structures of the same target were measured using MOE.[42] The protein sequences were first aligned and superimposed, and the RMSD was measured over all backbone atoms for the full protein and for all heavy atoms of residues within 4.5 Å of either ligand (Table S2). The number of rotatable bonds of the fragments was calculated using the Janssen in-house Third Dimension Explorer (3DX) tool.[44] Fragment charges were calculated using MOE.

Docking Methods. *Docking by SEED.* The evaluation of the binding energy in the program SEED consists of a force-field-based energy function with a continuum dielectric approximation of the protein–fragment electrostatic interactions and desolvation penalties by the generalized Born model.[21,22,45] The partial charges and vdW parameters for the atoms in the protein and in the fragments were taken from the CHARMM36 all-atom force field and the CHARMM general force field (CGenFF), respectively.[46−48] Importantly, CHARMM36 and CGenFF employ a consistent paradigm for the determination of partial charges and vdW parameters. The molecular surface was assigned as a dielectric discontinuity surface, and the dielectric constants were assigned values of 2.0 and 78.5 for the volume occupied by the solute and the solvent, respectively.

The different docking and scoring options that were evaluated are summarized in Table 2. SEED offers three methods for the actual placement of the fragments: (1) polar docking in which fragments are positioned and oriented such that at least one favorable hydrogen bond with a residue in the binding site is formed, (2) apolar docking in which fragments are placed in the hydrophobic regions of the binding site, and (3) a combination of (1) and (2). The docking itself was carried out using up to 64 conformers per fragment precalculated by ConfGen. Two screening protocols were thoroughly evaluated: (1) docking by combining polar and apolar placement of multiple conformers with the total energy (**Total**) as a scoring function (docking protocol **DP1**) and (2)

Table 2. Energy Terms in SEED, Ranking Methods, and Derived Docking Protocols[a]

description	abbreviation in text
Energy Terms	
total energy of the interaction between the fragment and protein	**Total**
electrostatic interaction between the fragment and protein, within the implicit solvent	**ElinW**
ElinW minus the free energy of solvation of the fragment	**Delec**
Van der Waals interaction between the fragment and protein	**vdW**
Ranking Methods	
ligand efficiency (**Total**/number of non-hydrogen atoms of the fragment)	**LE**
electrostatic efficiency (**ElinW**/number of non-hydrogen atoms of the fragment)	**ElectrE**
Van der Waals efficiency (**vdW**/number of non-hydrogen atoms of the fragment)	**vdWE**
Docking Protocols	
docking of multiple conformers of the fragment and scoring by **Total**	**DP1**
polar docking of multiple conformers of the fragment and scoring by **Delec**	**DP2**

[a]Rationales for the selection of these particular terms, methods, and protocols are given throughout the text.

polar docking alone of multiple conformers with **Delec** (see Table 2 for the definition) as a scoring function (**DP2**). For the **DP1** protocol, the influence of vdW efficiency filters (**vdWE**) of −1.1 or −1.6 kcal/mol per heavy atom were investigated. Robustness of the **DP1** protocol was also tested by varying the multiplicative factors of the vdW energy and fragment desolvation penalty contributions, solvation grid spacing, active-site definition, number of generated conformations per fragment, and the use of additional filters. As a result of an exploratory variation of parameters, the most generally applicable protocol included a 1.2 times rescaling of the vdW contribution, a 0.5 times rescaling of the fragment desolvation penalty, an electrostatic efficiency filter of −0.1 kcal/mol, a vdW efficiency filter of −1.1 kcal/mol, or no vdW efficiency filter for superficial binding sites (in this study, only the NhDMX binding site did not warrant a vdW efficiency filter).

Docking by Glide. In this study, the Glide docking tool was used to provide a baseline performance for fragment screening as it is one of the only tools that has previously been validated for fragment screening.[18,19] The settings recommended by Schrödinger for fragment screening were used for Glide docking. The docking grid was created in such a way that it resembled the SEED sampling space as good as possible. Docking in Glide was done using the Standard Precision algorithm.[33] The expanded sampling scheme was used, the number of poses per ligand for the initial phase of docking was increased to 50 000, the scoring window for keeping the initial poses was increased to 500 kcal/mol, and the best 1000 poses of every fragment were kept for energy minimization. The protein was kept rigid during docking.

Quality Metrics. To test the utility of the SEED tool, we focused mostly on evaluation of the screening power, a measure for the ability of a scoring function to prioritize true binders by ranking. Visual inspection of hit fragments was done to inspect for anomalies but was not used as a performance criterion. To investigate the screening power, the enrichment factor (EF), the true positive rate (TPR), the area under the
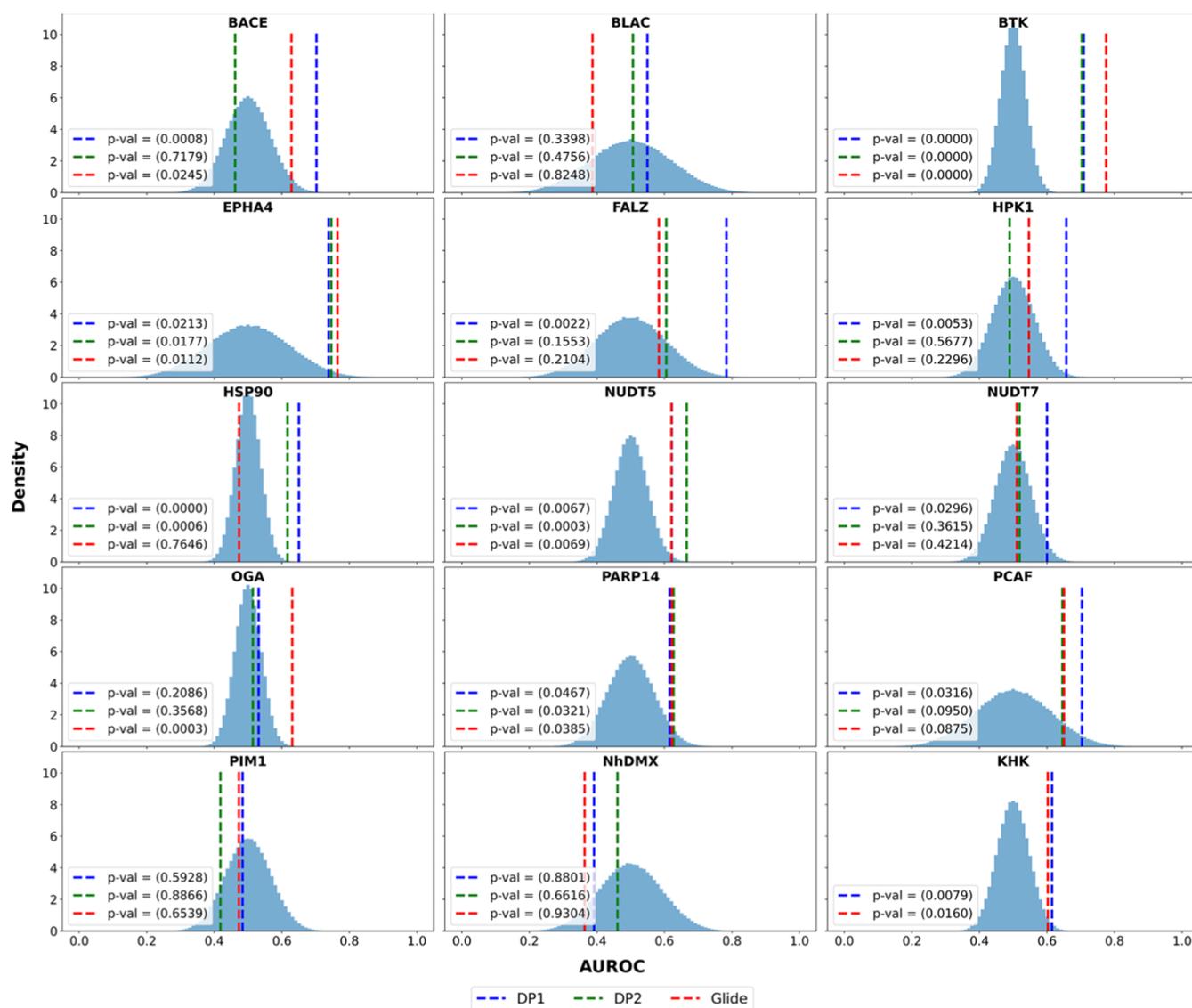
Figure 1. For each target, the AUROC metrics for DP1 (blue dashed lines), DP2 (green dashed lines), and Glide (red dashed lines) are compared to the distribution of AUROCs for random ranking of the actives (light-blue histograms). The random null distributions were generated with 1 000 000 samples by drawing the ranks of the actives uniformly without replacement. The reported p-values are calculated from the empirical cumulative distribution function of the AUROC null distribution and are used to test whether the screening protocols perform better than random on the overall library. The indicators for the DP1 and Glide protocols overlap for NUDT5. The active site of KHK is very hydrophobic. Using the DP2 protocol for this site results in almost no successfully docked fragments, so the DP2 protocol was omitted from the results.

receiver operating characteristic (AUROC/AUROC$_x$), and the power metric (PM)[49] were used as metrics.

The AUROC is defined as the area under the receiving operating characteristic (ROC) curve, which is empirically constructed by plotting the true positive rate (TPR; see below) as a function of the false positive rate (FPR) for all possible thresholds. The AUROC can be interpreted as the probability that the docking protocol we are using will rank a randomly chosen positive control before a randomly chosen negative one,[50,51] and it is used to quantify the modelwide screening performance. To test statistically whether the rankings derived from our protocols perform better than random on the specific data sets, we analyzed, for each library we simulated, the null model distribution of AUROCs under the assumption that the ranks of the actives are uniformly distributed.[52,53] For each case, we used 1 000 000 samples, which were generated by randomly drawing the ranks of the actives without replacement.

The EF$_x$ at x % of the ranked fragment library is calculated as

$$\text{EF}_x = \frac{\#\text{HITS}_{\text{sel}}}{N_{\text{sel}}} \times \frac{N_{\text{tot}}}{\#\text{HITS}_{\text{tot}}} \tag{1}$$

with EF$_x$ being the EF for the situation in which the top-x % of the ranked library is defined as a hit, $\#\text{HITS}_{\text{sel}}$ is the number of true hits in the top-x % of the ranked library, $\#\text{HITS}_{\text{tot}}$ is the number of true hits in the entire library, N_{tot} is the total library size, and N_{sel} is the size in compounds of the first x % of the ranked library.

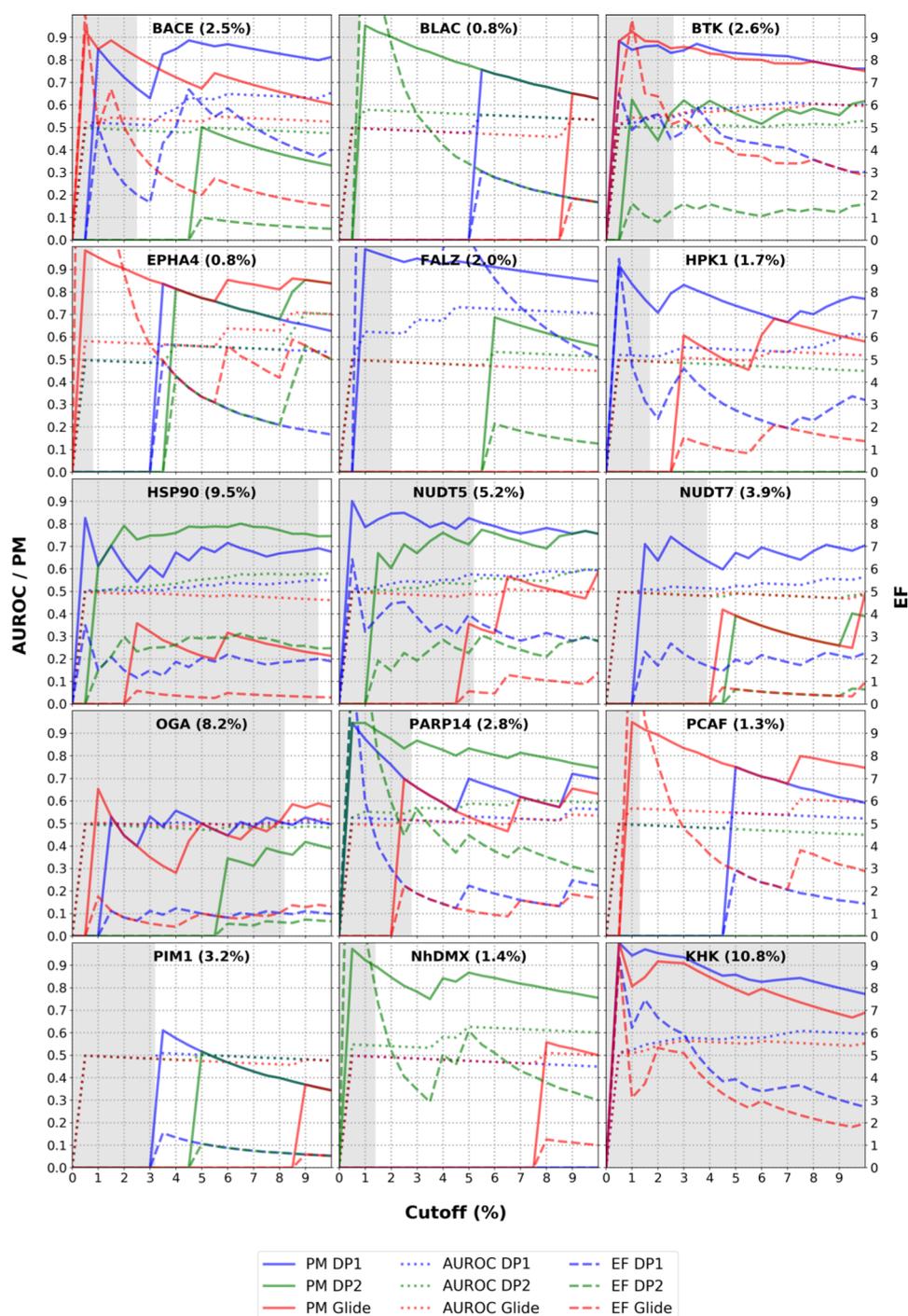The sensitivity metric, also called the TPR or recall, is calculated as

Figure 2. Comparison of the $AUROC_x$ and PM (left y-axes) and EF (right y-axes) metrics as a function of the scoring cutoff that is used to separate predicted actives from inactives (x-axes), calculated from the **DP1**, **DP2**, or **Glide** docking on each of the 15 targets. The reported values are generated by taking the scoring averages of the best-performing conformation when two targets were available for each protocol. Gray-shaded areas indicate the experimental fraction of hits for each target set, which is also reported in parentheses. The figure was generated by ranging the scoring cutoff between 0 and 10% in increments of 0.5%. For reasons of clarity, the maximum plotted value of the EF metric has been limited to 10. The active site of KHK is very hydrophobic. Using the DP2 protocol for this site results in almost no successfully docked fragments, so the DP2 protocol was omitted from the results.

$$TPR_x = \frac{\#HITS_{sel}}{\#HITS_{tot}} \qquad (2)$$

with TPR_x being the TPR calculated for the situation in which the top-x % of the ranked library is defined as a hit.

The $AUROC_x$ metric is a cutoff-dependent extension of the AUROC and was derived by calculating the area under the

ROC as defined by plotting the single TPR_x as a function of the single false positive rate (FPR_x), calculated for a given activity cutoff rate x %. The $AUROC_x$ is expected to oscillate locally around the value of 0.5 for uniform random ranking and to decrease linearly, in regions where no actives are ranked.

Finally, the PM_x metric was calculated according to[49]

$$PM_x = \frac{TPR_x}{TPR_x + FPR_x} \qquad (3)$$

The EF is a metric that is useful in evaluating the enrichment within a specified fraction of the scored library, relative to the experimental hit rate. However, this metric is strongly dependent on the total number of fragments and number of actives in the library. Notably, EF does not discriminate between hit fragments below the threshold value and does not hold into account any fragments that are not within the threshold, making it very dependent on the chosen threshold. The TPR metric expresses the percentage of fragments retrieved within a specified fraction of the scored library. It has most of the drawbacks of the EF but does not take into account the ratio of hits to nonbinders, which reduces the variance on this metric. The AUROC is another commonly used metric with a statistical background, unlike the EF and the TPR. It is a measure of the docking performance over the full library. The main disadvantage of the AUROC is that there is no bias toward early retrieval of fragments, which is not representative of the aim of a virtual screening experiment. Finally, the power metric is a statistically solid alternative for the TPR. It combines the TPR with the false positive rate so that it is no longer as sensitive to the threshold.

In the context of virtual screening on prospective targets, we are primarily interested in the early enrichment as resource constraints usually allow for experimental validation of only a (small) fraction of the compound library. For this reason, we have also considered metrics that focus on the early enrichment region of the virtual screens along with the AUROC, which is commonly used to evaluate docking performance.

■ RESULTS

The SEED performance was mainly judged based on screening power, which is the ability of a scoring function to distinguish true binders from inactive compounds.[54] While docking power (i.e., the correct prediction of the binding mode observed experimentally, e.g., in crystal structures) is also an important metric to be considered in fragment screening, the amount of binding pose information in the curated data sets is limited. However, previous reports have provided evidence that SEED is able to accurately predict binding poses of fragments.[23−26,55,56]

Performance Compared to Random Ranking. Figure 1 compares the rankings from the different docking protocols (**DP1**, **DP2**, and **Glide**) with the histograms of the AUROCs for simulated uniformly distributed ranks of the actives. When two conformations of the same target were used for docking, we kept the scores for the best-performing conformation with the protocol. The distributions are all centered around 0.5, which is the value commonly used as the random baseline for binary classifiers, but their standard deviations depend on the specific composition of the library in terms of the total number of compounds and actives.[52] In our cases (as in virtual screening in general), the number of hits is much smaller than the number of inactives and we observe a lower standard deviation for larger fractions of actives in libraries of the same size (compare, for example, EPHA4 and OGA) and for larger libraries containing a similar fraction of hits (compare BTK and PARP14). To test whether AUROCs above 0.5 are statistically better than random, we calculated p-values from the empirical cumulative distribution function of the null

model. The p-values represent the probability that random ranking could achieve a performance equal to or better than the respective screening protocol.

For most of the targets, there is at least one screening protocol that achieves better-than-random performance. Specifically, **DP1**, **DP2**, and **Glide** rankings are better than the null model in 11, 5, and 7 cases, respectively (at a significance level $\alpha = 0.05$). All of the protocols are better than uniform ranking for BTK, EPHA4, NUDT5, PARP14, and KHK, whereas they all fail on BLAC, PIM1, and NhDMX. Figure S1 also shows the AUROC values for each individual protein conformation and for a "mixed approach" in which the highest score for any given fragment is used to calculate the AUROC, regardless of the protein conformation that originated the pose. The AUROC is an overall measure of performance, which considers the position of all of the actives and is strongly deteriorated by actives that are docked in the wrong pose, as they usually rank poorly; this problem can particularly affect SEED when docking very flexible molecules. However, in the context of virtual screening, rankings with low AUROC can still be valuable if they manage to rank some actives in the top fraction of compounds.

Comparing SEED with Glide. In Figure 2, the performance in terms of early enrichment of the two docking protocols of SEED (**DP1** and **DP2**) is compared to **Glide**. For each of the 15 targets, the EF, PM, and $AUROC_x$ metrics of the three different screens were calculated at cutoff values (the x variable in eq 1) ranging between 0 and 10% in increments of 0.5%. Analysis of these data shows that there is no screening method that performs consistently better than the others, but some trends can be deduced.

Focusing on Figure 2, it is clear that for some of the targets no convincing virtual screening results are obtained. This is particularly true for the three targets NUDT7, OGA, and PIM1 for which the maximal EFs calculated using a cutoff of 5% on the ranked hitlist are all below or close to 2, the $AUROC_x$ values are around 0.5 (implying a random model), and the PM metric is between 0.5 and 0.7. Virtual screening results for the other targets are better. We observe that SEED, using a combined polar and apolar docking approach and with the total energy as the scoring function (**DP1**), performs in 8 out of the 15 targets (BACE, BTK, FALZ, HPK1, HSP90, NUDT5, PARP14, KHK) better than or equal to the **DP2** and **Glide** protocols. The other tested docking protocol of SEED (**DP2**), in which only a polar docking was used in combination with **Delec** as the scoring function (Table 2), performs well in 4 out of the 15 targets (BLAC, HSP90, PARP14, NhDMX). Finally, good results are found for **Glide** in the case of five out of the 15 targets (BACE, BTK, EPHA4, PCAF, KHK).

In Table 3, a similar comparison between the three docking protocols is given. In this case, the sensitivity metric is used to compare the protocols and to analyze how well each of the methods is capable of identifying hits within the cutoff limit imposed by the actual fraction of hits contained in the data set (fHits). In the case of an ideal virtual screening model, the expected outcome for the TPR_{fHits} metric in Table 3 would be 100%, meaning that all of the available hits in the data set would be retrieved by the method using a cutoff equal to this fraction. Comparison of Table 3 with Figure 2 reveals some interesting conclusions. There is a good agreement overall of the results from Figure 2 with those from Table 3. Protocols and data sets that perform well in terms of the $AUROC_x$ and

Table 3. Sensitivity Analysis of the Different Docking Protocols Applied to Each of the 15 Targets[a]

target name	fHits: fraction of hits (%)[b]	TPR$_{fHits}$ (%)[c]		
		DP1	DP2	Glide
BACE	2.5	**5.0**	0.0	**10.0**
BLAC	0.8	0.0	**16.7**	0.0
BTK	2.6	**11.1**	3.2	**12.7**
EPHA4	0.8	0.0	0.0	**16.7**
FALZ	2.0	**25.0**	0.0	0.0
HPK1	1.7	**4.6**	0.0	0.0
HSP90	9.5	**18.8**	**24.6**	2.9
NUDT5	5.2	**19.4**	**16.7**	2.8
NUDT7	3.9	6.4	0.0	0.0
OGA	8.2	8.2	4.9	9.8
PARP14	2.8	**5.6**	**16.7**	**5.6**
PCAF	1.3	0.0	0.0	**14.3**
PIM1	3.2	0.0	0.0	0.0
NhDMX	1.4	0.0	**10.0**	0.0
KHK	10.8	**26.8**	[d]	**22.0**

[a]The cutoff value that was used to calculate the sensitivity (TPR) was set equal to the fraction of hits in each data set (fHits). Bold TPR values indicate the sets that performed well according to Figure 2. [b]The fraction of hit fragments in the library. [c]The true positive rate (TPR) calculated at a cutoff value corresponding to the fHits value of each target. In an ideal screening model, this TPR$_{fHits}$ value should be equal to 100%. [d]The active site of KHK is very hydrophobic. Using the **DP2** protocol for this site results in almost no successfully docked fragments, so the **DP2** protocol was omitted from the results.

EF metrics also perform accordingly in terms of the sensitivity metric. This is especially the case for the majority of the 15 targets (BACE, BLAC, BTK, EPHA4, FALZ, HPK1, HSP90, NUDT5, PARP14, PCAF, NhDMX, and KHK). For these cases, a good correlation between the TPR$_{fHits}$ and EF is observed, indicating a high enrichment at early cutoff fractions for some of the screening models. For the three targets (NUDT7, OGA, and PIM1) for which none of the three methods are performing well when evaluated by the calculated EF, PM, and AUROC$_x$ values in Figure 2, a similar conclusion can be drawn from the TPR$_{fHits}$ metric in Table 3.

However, from a practical point of view, this is not the desired screening power one expects from a typical virtual screening, especially when one wants to evaluate larger virtual data sets. Similar conclusions can be drawn for NUDT7, OGA, and PIM1, and in particular, the latter target seems to be the most problematic for all of the three evaluated virtual screening methods. One can therefore conclude that neither SEED nor Glide is performing well in the case of these three targets. If we combine the results of the AUROC measurements in Figure 1 and the TPR$_{fHits}$ values of Table 3, we can isolate the data sets for which the different screening protocols provide a consistent performance across the full library in combination with a significant early enrichment. As a baseline for good performance, we can select data sets for which the AUROC is statistically better than a random distribution and for which the TPR$_{fHits}$ is more than twice the experimental hit rate. For **DP1**, this is the case for seven targets (BACE, BTK, FALZ, HPK1, NUDT5, PARP14, KHK). For **DP2**, this is only the case for three targets (HSP90, NUDT5, PARP14), and for **Glide**, this is the case for five targets (BACE, BTK, EPHA4, PARP14, KHK). By combining the performance metrics, we can say that

the **DP1** protocol is more reliable than either the **DP2** or **Glide** protocol.

Fragment Flexibility. The performance of SEED when docking more "flexible" fragments was also tested and compared to **Glide**. All fragments having more than two rotatable bonds were defined as flexible fragments. The performance of **Glide** was compared to the best-performing protocol of SEED (**DP1**), using four different data sets for which a sufficient number of flexible fragments were available (BTK, KHK, OGA, and HPK1). Instead of filtering each of the libraries for fragments with less than three rotatable bonds, the full libraries were used for screening. The number of fragments in each data set can be found in Table S1.

The results are summarized in Figure 3 and may be compared to the corresponding data in Figure 2, the latter being generated for fragments with two or less rotatable bonds. While in the cases of BTK, HPK1, and KHK as targets, **DP1** outperformed **Glide** as a screening tool (Figure 2), this is no longer the case when more flexible fragments are used. It is interesting to note that in the cases of BTK and HPK1, SEED performance is roughly the same with and without the addition of the flexible molecules, whereas **Glide** improves with the addition of the flexible set. On the contrary, for KHK, both SEED and GLIDE worsen with respect to the case without the flexible ligand. In fact, both protocols seem—from a qualitative point of view as shown in Figure 3—to perform almost equally well. This illustrates one of the limitations of SEED, namely, that both the protein and fragment are kept rigid to allow for fast screening, but this design choice comes at the cost of performing less well in the case when flexible molecules need to be screened.

◼ DISCUSSION

Data Sets. There is a low public availability of fragment docking data sets. Available data sets are often suboptimal for benchmarking due to a combination of low experimental hit rates and small library size or, conversely, very high experimental hit rates and large library size. Therefore, some data sets in this study are not as useful for benchmarking compared to the sets that are often used in the assessment of docking tools for druglike molecules. The lack of targets with available fragment X-ray data is still an issue. For example, several targets that were used for screening by X-ray crystallography had multiple fragment-binding pockets. As we cannot make a distinction between experimental hits that bind in the pocket selected for VS and fragment hits that bind elsewhere with most experimental screening procedures, the enrichments reported here for both SEED and Glide are likely significantly lower than the actual enrichments resulting from the screening studies. In other cases, inconsistencies could be seen in the hits found by different in vitro screening assays, which is a known issue.[10] In some data sets, a significant portion of hits is likely false positive, as a result of either aspecific binding to the protein surface or detergents or binding in different protein domains. Initially, 24 different targets were considered with data sets provided by Janssen, SGC, and ZoBio. After considering the factors discussed here to the best extent, only 15 data sets remained for which the experimental data were robust enough to be included in the study.

The fragment libraries in this study are also relatively small, resulting in a relatively high variance on the reported results. The libraries that were used range from 394 fragments to 3521
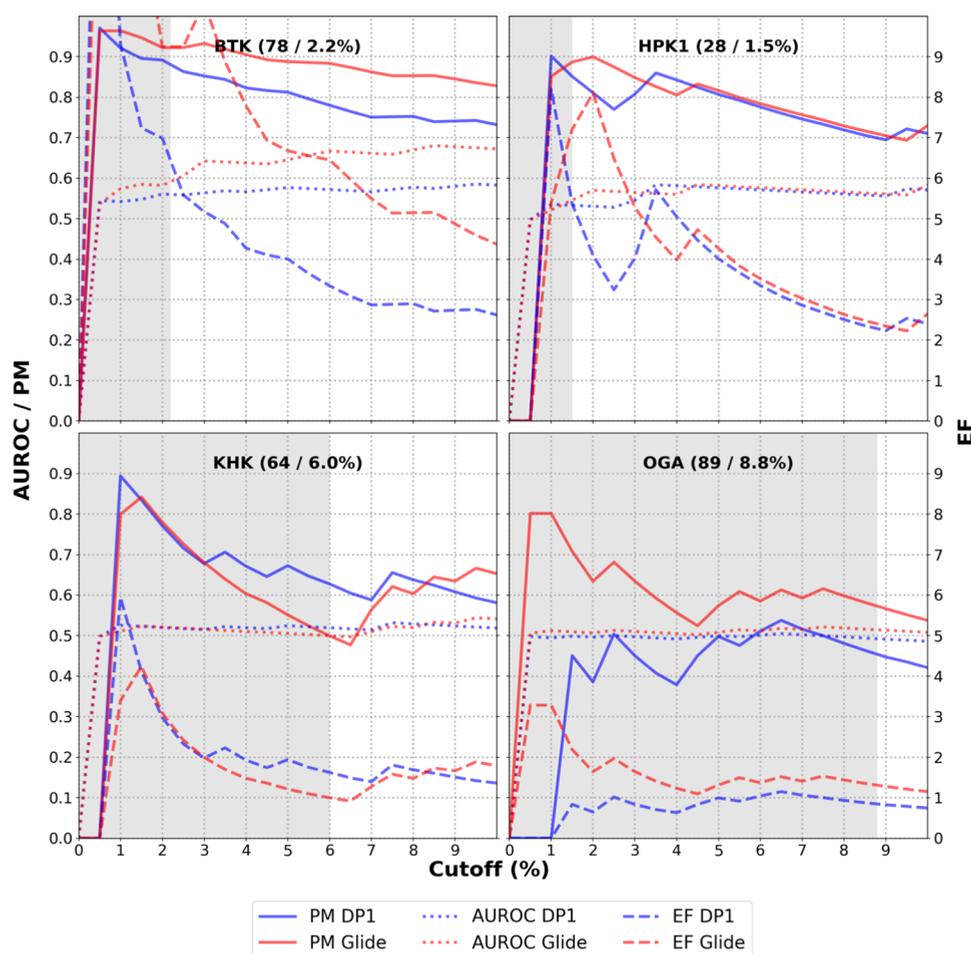
Figure 3. Comparison of the AUROC$_x$ and PM metrics (left *y*-axes) and EF (right *y*-axes) metrics as a function of the virtual screening cutoff (*x*-axes), calculated from the **DP1** and **Glide** virtual screens on four targets in combination with fragments containing more than two flexible torsions. The figure was generated using an identical protocol as for Figure 2. Panel labels describe the targets with both the actual number of hit fragments and experimental hit rates in parentheses.

fragments, while fragment libraries can often contain on the order of $10^4 - 10^5$ fragments. The number of experimental hits among the data sets varies between 0.8 and 10.8%. In data sets with high hit rates, it is expected to obtain lower EFs. These factors can result in noise on the results while also making comparative metrics ambiguous. Despite these shortcomings, it was still possible to design a workflow that provides enrichment for nearly all tested data sets.

Performance: SEED versus Glide. From the results of this study, it can be concluded that SEED generally outperforms Glide for screening of small fragments. The **DP1** SEED protocol performs statistically better than random for 11 out of the 15 studied targets when considering the AUROC, while this is only the case for 7 out of 15 targets with the **Glide** docking protocol. If we also consider the early enrichment based on the TPR metric in addition to the AUROC values, the **DP1** SEED protocol performs well for seven targets, while this is only the case for five targets with the **Glide** protocol. However, it goes without saying that both tools have their advantages and disadvantages. The Glide user interface is very user-friendly and allows for comparatively fast and easy protein and fragment preparation, which can be done completely in Maestro. SEED, in contrast, is a command line tool that relies on other programs for protein and fragment preparation. Also, Glide is not limited to rigid fragment

docking. It has a fast algorithm that samples conformational space without relying on multiple input conformations, making it more suitable for larger, more flexible fragments. Finally, Glide also has the option of postdocking energy minimization of the best poses and can account for induced fit to some extent by allowing some binding site residues to rotate. Postdocking energy minimization is also possible in SEED, but it is limited to rigid-body translation and rotation of the fragment only. On the other hand, SEED is an open-source tool. It is more customizable than Glide to suit the needs of the user. Settings regarding sampling can be fine-tuned to find a balance between efficiency and accuracy, and even the scoring function can be customized to adjust to the needs of the user. The SEED algorithm explores the space defined by the user exhaustively. This should, in practice, result in better reproducibility, exploration of all possible poses, and an accurate score being assigned to the top pose, assuming that an adequate diversity of conformations is provided in the case of flexible molecules.

It is important to emphasize that the results reported in this study are based on the ability of both tools to rank fragments by raw scoring functions. In practice, various other considerations can be made, including analysis of incorrectly buried hydrogen bond donors and acceptors, visual inspection of binding poses, knowledge-based filtering of fragments, and
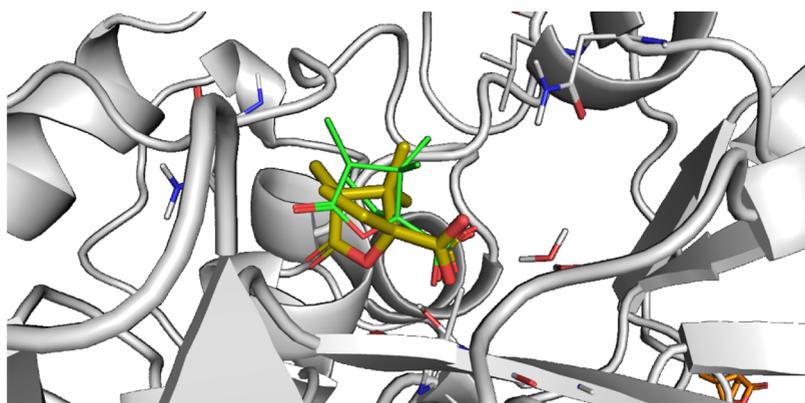
Figure 4. Using the **ElinW** ranking to correctly predict binding poses of charged fragments. The crystal structure pose of a fragment binding to the BLAC active site (PDB ID: 4KZ7) is shown in gold sticks. The SEED top pose using the **ElinW** ranking is shown in green lines. While the carboxylate groups overlap very well in the binding poses, the bulky group of the fragment is not positioned correctly (RMSD = 3.3 Å).

rescoring with other scoring functions.[17,57−61] Using these methods for further postprocessing, one can drastically improve the hit rate in screening campaigns, as can be seen in successful campaigns.[20,62,63]

Defining the Binding Site. The selection of binding site residues for fragment docking in SEED is defined by the user in the input file. It is recommended to use SEED on targets for which a holo-protein crystal structure is available. As with regular small-molecule docking, using a holo-protein crystal structure will help consider conformational changes in the binding pocket upon ligand binding as small rotamer changes in active-site residues can already have a sizeable impact on docking results. They can either prevent fragments from being docked due to clashes if the site is too small or result in top fragment poses with unlikely contacts if the site is too big. Therefore, it is also suggested that there is a general idea of an expected binding motif if one expects good results from fragment docking. Bromodomains and protein kinases have well-defined interactions with their natural ligand acetylated lysine and ATP cosubstrate, respectively, which facilitates the preparation of a SEED docking run and contributes to the success of SEED campaigns on these targets.[23−27,56]

Unlike many other docking tools, the binding site for SEED is not defined by a "docking box" in which nonexhaustive sampling is performed but is based on vectors generated according to user-picked residues to allow for exhaustive sampling by rotations around these vectors. With information about the binding motif, it is possible to limit the number of chosen residues, which improves performance in both binding pose prediction and enrichment. It is, however, also possible to use SEED without information about the binding site. In such a case, tools exist that can help identify promising binding sites. For example, in the 29 structures from 15 different targets considered in this study, the SiteFinder module of MOE was able to recognize the screened binding site as the "best" binding site for 24 of the structures and as the second-best binding site for four of the structures. Tools like SiteFinder only give an estimate of the region of interest though, as the binding site is usually larger than the region(s) in which fragments will bind with high efficiency. Other tools probing the binding site more precisely to find fragment-binding hotspots also exist.[64] The biggest issue that remains is the possibility of conformational change upon ligand binding. A possible solution to this is to sample binding site flexibility through molecular dynamics simulations.[63] In this study, we

also attempted to find a way to correlate results with certain protocols to characteristics of the protein binding site. Unfortunately, we were unable to find any clear correlation. What we did find is that the standard **DP1** SEED protocol that we recommend is outperformed by a protocol that utilizes the "ElinW" energy score for ranking some of the data sets. This is discussed in the next section.

Unresolved Issue: SEED and Charged Fragments. A current known issue of the SEED scoring function, and an issue with force-field-based methods in general, is that it is difficult to find a unifying scoring criterion that works for fragments with different formal charges. Models based on fixed partial charges cannot correctly capture polarization, making it very difficult to distinguish between charged and uncharged molecules (we use the term uncharged fragment for molecules devoid of formal charges). As mentioned in Table 2, the "**Delec**" energy is the difference between protein/fragment electrostatic energy and free energy of hydration. For charged fragments, **Delec** is always very unfavorable (large and positive) because it is not possible to compensate their very favorable free energy of hydration by electrostatic interactions with the protein. This issue carries over to a ranking based on the total energy score.

The SEED docking parameters used in this study result in charged fragments not being accurately ranked for binding sites with (several) charged side chains. While this is not a big concern if additional filters are included in a workflow such as discussed below, it might have an impact on the enrichment in cases where no additional postprocessing is applied. BACE and BLAC have binding sites with negatively and positively charged side chains, respectively, and many of the hits are fragments bearing a formal charge opposite to the ones in the binding site. It has been previously established that regular SEED scoring protocols are not appropriate for ranking of charged fragments,[56] and tests on the BACE and BLAC data sets have shown that a modified version of the **DP1** protocol using the **ElinW** ranking method instead of the "**Total**" ranking method (see Table 2) can significantly improve enrichment when charged fragments are expected to have contacts with charged residues. In fact, different rankings can be used separately to retrieve pools of compounds more enriched in polar, charged, and nonpolar fragments. Considering the BLAC library, none of the hits containing carboxylic acids passed the applied filters or ranked within the 10% cutoff on the ranked hitlist in our initial screenings. When using an
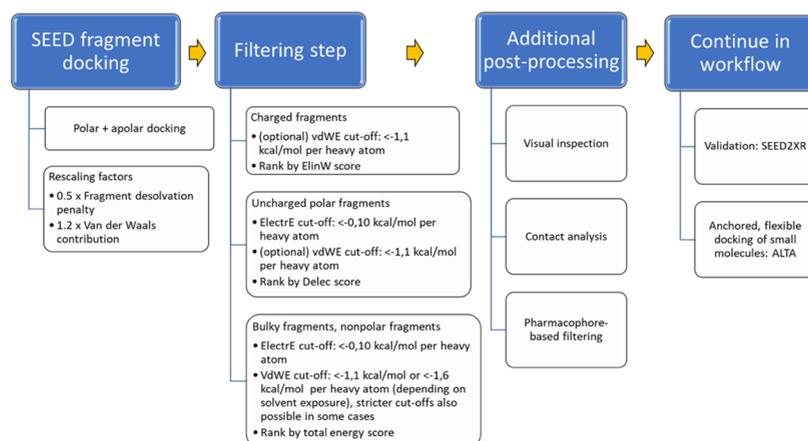
Figure 5. Suggested SEED fragment screening workflow. ElectrE, electrostatic efficiency; vdWE, vdW efficiency. SEED2XR, ALTA; see the reference for a detailed explanation of both workflows.[28]

alternate protocol by sorting the ranked compounds by the **ElinW** ranking method, several of these hits were retrieved. Of the two hits for which crystal structures are available (PDB ID: 4KZ7, 4KZ9), one was found within the top 10% of fragments with this protocol, with a pose similar to the crystal structure pose (see Figure 4). The other fragment was ranked 691 out of 734 compounds but was found to bind to a different region after inspecting the crystal structure.

While the **ElinW** ranking method yields better enrichment for BACE and BLAC, significant early enrichment for these targets is still obtained with the **DP1** and **DP2** protocols, respectively (see Figure 2). However, the pool of best-ranked fragments using these protocols is vastly different compared to using the **ElinW** ranking method. More specifically, we observed that a bigger fraction of charged, polar, and apolar fragments can be retrieved using the **ElinW** ranking method, the **DP2** protocol, and the **DP1** protocol, respectively. With respect to these observations, we encourage the separate ranking of charged and uncharged compounds. This approach has been taken previously in SEED screening experiments as well.[27,56]

Docking Flexible Fragments with SEED. The performance of SEED when not using a filter on the number of rotatable bonds in a fragment was assessed. The results of this comparison showed that SEED performs worse when more flexible fragments are included in the library while also performing slightly worse than Glide in this case. This is possibly due not to approximations in the scoring function itself but to the pose generation. The computational efficiency of screening flexible fragments is also something to keep in mind when using SEED. While Glide does not slow down significantly when fragment size increases, SEED is several times slower when docking flexible molecules, and this is because of two reasons. First, flexible molecules are generally larger than rigid fragments and have more functional groups. As SEED uses an exhaustive sampling algorithm, all functional groups will be positioned and sampled along the user-defined vectors, causing a significant increase in sampling per molecule. Second, Glide handles sampling of fragments with rotatable bonds relatively efficiently. In SEED, on the other hand, multiple conformations of flexible molecules have to be docked individually. The number of input conformations needed for a sufficient amount of sampling increases exponentially with the number of rotatable bonds. However, lower enrichments are

also obtained in Glide when using unfiltered libraries, compared to using a filter on the number of rotatable bonds. Because of this, virtual docking of large, flexible molecules is generally dissuaded.

Use Cases of SEED. As this study was a validation of the ability of SEED to outperform random fragment selection and provide early enrichment of hits, no extensive postprocessing was done to avoid any kind of bias. In reality, the SEED tool is a lot more flexible in usage than what is represented here. For the benchmarking, top percentages of ranked fragments were chosen to obtain representative values. However, there is currently no single protocol that is suited for all fragment types. As such, it is highly recommended to not use raw percentage cutoffs for fragment docking. Even though the study demonstrates a significant enrichment of libraries, the use of a customized approach with manual postprocessing will likely lead to hit rates that are higher than presented in this study. From the results of screening 15 data sets and exploration of various parameters, we propose a workflow to be used for virtual fragment screening with SEED (Figure 5).

Based on our results, we believe that virtual screening with SEED is a useful approach to prioritize fragments for experimental follow-up. Given the commercial availability of very large fragment-sized compound collections, custom libraries for specific targets are likely to have higher hit rates than a standard fragment library. Alternatively, a standard library can be augmented with fragments resulting from a virtual screen.

■ CONCLUSIONS

We have assessed the usefulness of the high-throughput docking program SEED for virtually screening libraries of fragments and conclude that the observed enrichments can be useful for selecting fragment libraries for experimental screening. While for flexible fragment docking Glide is faster than SEED, we found that SEED outperforms Glide in the majority of cases. Experimental fragment screening results can be inconsistent, with inaccuracies caused by aspecific protein binding and weak binding, which give rise to false positives and false negatives, respectively. Furthermore, fragments can bind to alternative binding sites on a protein, resulting in additional false positive hits. Therefore, the true EFs for virtual fragment screening are probably higher than reported. Furthermore, applying appropriate postprocessing and adjusting the used

protocol or workflow on a case-to-case basis also have a significant impact on virtual screening results.

Although SEED is able to provide significant fragment library enrichment, we observed that prediction of the correct binding pose is not reliable, especially in cases where fragments do not have a consistent binding motif in a binding site. Supplementing screening with X-ray crystallography is therefore essential before proceeding with fragment growing for hit or lead development.

We developed a best-performing SEED virtual fragment screening workflow, which combines three different protocols for charged, polar, and nonpolar fragments.

AVAILABILITY OF SOFTWARE

The most recent version of SEED is available at https://gitlab.com/CaflischLab/SEED. Recently, an efficient GUI has been developed that can be used for fragment library preparation and binding site definition for SEED screening. ACGUI is available from one of the authors (AC) as it is still under development and the backend for calculations is hosted on a relatively small compute cluster.

AVAILABILITY OF DATASETS

The confirmed hits for FALZ, NUDT5, NUDT7, and PARP14 and their respective X-ray-determined binding poses are available at https://www.thesgc.org/fragment-screening.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.0c00556.

> Full fragment libraries for FALZ, NUDT5, NUDT7, and PARP14 targets (Tables S1−S2, Figure S1) (PDF)

AUTHOR INFORMATION

Corresponding Author

Hans De Winter − *Department of Pharmaceutical Sciences, Laboratory of Medicinal Chemistry, University of Antwerp, 2610 Wilrijk, Belgium;* ⊙ orcid.org/0000-0002-4450-7677; Email: Hans.DeWinter@uantwerpen.be

Authors

Kenneth Goossens − *Department of Pharmaceutical Sciences, Laboratory of Medicinal Chemistry, University of Antwerp, 2610 Wilrijk, Belgium;* ⊙ orcid.org/0000-0003-4127-5724

Berthold Wroblowski − *Janssen Research and Development, 2340 Beerse, Belgium;* ⊙ orcid.org/0000-0002-9980-1646

Cassiano Langini − *Department of Biochemistry, University of Zurich, Zurich CH-8057, Switzerland;* ⊙ orcid.org/0000-0003-3637-7048

Herman van Vlijmen − *Janssen Research and Development, 2340 Beerse, Belgium;* ⊙ orcid.org/0000-0002-1915-3141

Amedeo Caflisch − *Department of Biochemistry, University of Zurich, Zurich CH-8057, Switzerland;* ⊙ orcid.org/0000-0002-2317-6792

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.0c00556

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

ABBREVIATIONS

3DX, Third Dimension Explorer; ALTA, Anchor-based Library TAiloring; AUROC, area under the receiver operating characteristic; CGenFF, CHARMM general force field; EF, enrichment factor; FBDD, fragment-based drug design; FPR, false positive rate; ITC, isothermal titration calorimetry; NMR, nuclear magnetic resonance; PM, power metric; RMSD, root-mean square deviation; ROC, receiver operating characteristic; SEED, solvation energy for exhaustive docking; SPR, surface plasmon resonance; SGC, structural genomics consortium; TPR, true positive rate; VdW, van der Waals; VS, virtual screening

REFERENCES

(1) Kirkpatrick, P.; Ellis, C. Chemical space. *Nature* **2004**, *432*, 823.

(2) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering high-affinity ligands for proteins: SAR by NMR. *Science* **1996**, *274*, 1531−1534.

(3) Hall, R. J.; Mortenson, P. N.; Murray, C. W. Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* **2014**, *116*, 82−91.

(4) Fink, T.; Bruggesser, H.; Reymond, J. L. Virtual exploration of the small-molecule chemical universe below 160 Daltons. *Angew. Chem., Int. Ed.* **2005**, *44*, 1504−1508.

(5) Kuntz, I. D.; Chen, K.; Sharp, K. A.; Kollman, P. A. The maximal affinity of ligands. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9997−10002.

(6) Kenny, P. W. The nature of ligand efficiency. *J. Cheminf.* **2019**, *11*, No. 8.

(7) Erlanson, D. A. Introduction to fragment-based drug discovery. *Top. Curr. Chem.* **2012**, *317*, 1−32.

(8) Doak, B. C.; Norton, R. S.; Scanlon, M. J. The ways and means of fragment-based drug design. *Pharmacol. Ther.* **2016**, *167*, 28−37.

(9) Linke, P.; Amaning, K.; Maschberger, M.; Vallee, F.; Steier, V.; Baaske, P.; Duhr, S.; Breitsprecher, D.; Rak, A. An Automated Microscale Thermophoresis Screening Approach for Fragment-Based Lead Discovery. *J. Biomol. Screening* **2016**, *21*, 414−421.

(10) Schiebel, J.; Radeva, N.; Koster, H.; Metz, A.; Krotzky, T.; Kuhnert, M.; Diederich, W. E.; Heine, A.; Neumann, L.; Atmanene, C.; Roecklin, D.; Vivat-Hannah, V.; Renaud, J. P.; Meinecke, R.; Schlinck, N.; Sitte, A.; Popp, F.; Zeeb, M.; Klebe, G. One Question, Multiple Answers: Biochemical and Biophysical Screening Methods Retrieve Deviating Fragment Hit Lists. *ChemMedChem* **2015**, *10*, 1511−1521.

(11) Chen, H.; Zhou, X.; Wang, A.; Zheng, Y.; Gao, Y.; Zhou, J. Evolutions in fragment-based drug design: the deconstruction-reconstruction approach. *Drug Discovery Today* **2015**, *20*, 105−113.

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

(13) McInnes, C. Virtual screening strategies in drug discovery. *Curr. Opin. Chem. Biol.* **2007**, *11*, 494−502.

(14) Hou, T.; Xu, X. Recent development and application of virtual screening in drug discovery: an overview. *Curr. Pharm. Des.* **2004**, *10*, 1011−1033.

(15) Sledz, P.; Caflisch, A. Protein structure-based drug design: from docking to molecular dynamics. *Curr. Opin. Struct. Biol.* **2018**, *48*, 93−102.

(16) Loving, K.; Salam, N. K.; Sherman, W. Energetic analysis of fragment docking and application to structure-based pharmacophore hypothesis generation. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 541−554.

(17) Marcou, G.; Rognan, D. Optimizing fragment and scaffold docking by use of molecular interaction fingerprints. *J. Chem. Inf. Model.* **2007**, *47*, 195−207.

(18) Sandor, M.; Kiss, R.; Keseru, G. M. Virtual fragment docking by Glide: a validation study on 190 protein-fragment complexes. *J. Chem. Inf. Model.* **2010**, *50*, 1165−1172.

(19) Kawatkar, S.; Wang, H.; Czerminski, R.; Joseph-McCarthy, D. Virtual fragment screening: an exploration of various docking and scoring protocols for fragments using Glide. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 527−539.

(20) Kolb, P.; Kipouros, C. B.; Huang, D.; Caflisch, A. Structure-based tailoring of compound libraries for high-throughput screening: discovery of novel EphB4 kinase inhibitors. *Proteins* **2008**, *73*, 11−18.

(21) Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive docking of molecular fragments with electrostatic solvation. *Proteins* **1999**, *37*, 88−105.

(22) Majeux, N.; Scarsi, M.; Caflisch, A. Efficient electrostatic solvation model for protein-fragment docking. *Proteins* **2001**, *42*, 256−268.

(23) Lolli, G.; Caflisch, A. High-Throughput Fragment Docking into the BAZ2B Bromodomain: Efficient in Silico Screening for X-Ray Crystallography. *ACS Chem. Biol.* **2016**, *11*, 800−807.

(24) Spiliotopoulos, D.; Wamhoff, E. C.; Lolli, G.; Rademacher, C.; Caflisch, A. Discovery of BAZ2A bromodomain ligands. *Eur. J. Med. Chem.* **2017**, *139*, 564−572.

(25) Spiliotopoulos, D.; Zhu, J.; Wamhoff, E. C.; Deerain, N.; Marchand, J. R.; Aretz, J.; Rademacher, C.; Caflisch, A. Virtual screen to NMR (VS2NMR): Discovery of fragment hits for the CBP bromodomain. *Bioorg. Med. Chem. Lett.* **2017**, *27*, 2472−2478.

(26) Zhu, J.; Caflisch, A. Twenty Crystal Structures of Bromodomain and PHD Finger Containing Protein 1 (BRPF1)/ Ligand Complexes Reveal Conserved Binding Motifs and Rare Interactions. *J. Med. Chem.* **2016**, *59*, 5555−5561.

(27) Marchand, J. R.; Lolli, G.; Caflisch, A. Derivatives of 3-Amino-2-methylpyridine as BAZ2B Bromodomain Ligands: In Silico Discovery and in Crystallo Validation. *J. Med. Chem.* **2016**, *59*, 9919−9927.

(28) Marchand, J. R.; Caflisch, A. In silico fragment-based drug design with SEED. *Eur. J. Med. Chem.* **2018**, *156*, 907−917.

(29) Bohm, H. J. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 593−606.

(30) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470−489.

(31) 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.

(32) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269−288.

(33) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739−1749.

(34) Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: Applications of autodock. *J. Mol. Recognit.* **1996**, *9*, 1−5.

(35) 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.

(36) Sheng, C.; Zhang, W. Fragment informatics and computational fragment-based drug design: an overview and update. *Med. Res. Rev.* **2013**, *33*, 554−598.

(37) Hoffer, L.; Renaud, J. P.; Horvath, D. Fragment-based drug design: computational & experimental state of the art. *Comb. Chem. High Throughput Screening* **2011**, *14*, 500−520.

(38) 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.

(39) Jacobson, M. P.; Friesner, R. A.; Xiang, Z.; Honig, B. On the Role of the Crystal Environment in Determining Protein Side-chain Conformations. *J. Mol. Biol.* **2002**, *320*, 597−608.

(40) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins* **2004**, *55*, 351−367.

(41) Sastry, G. M.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 221−234.

(42) ULC, CCG. Molecular Operating Environment (MOE) 2019.1, Montreal, QC, 2019.

(43) Watts, K. S.; Dalal, P.; Murphy, R. B.; Sherman, W.; Friesner, R. A.; Shelley, J. C. ConfGen: a conformational search method for efficient generation of bioactive conformers. *J. Chem. Inf. Model.* **2010**, *50*, 534−546.

(44) Cedeno, W.; Alex, S.; Jaeger, E. P.; Agrafiotis, D. K.; Lobanov, V. S. An Integrated Data Management Framework for Drug Discovery − From Data Capturing to Decision Support. *Curr. Top. Med. Chem.* **2012**, *12*, 1237−1242.

(45) Scarsi, M.; Apostolakis, J.; Caflisch, A. Continuum Electrostatic Energies of Macromolecules in Aqueous Solutions. *J. Phys. Chem. A* **1997**, *101*, 8098−8106.

(46) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(47) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., 3rd Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.

(48) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.

(49) Lopes, J. C. D.; Dos Santos, F. M.; Martins-Jose, A.; Augustyns, K.; De Winter, H. The power metric: a new statistically robust enrichment-type metric for virtual screening applications with early recovery capability. *J. Cheminf.* **2017**, *9*, No. 7.

(50) Hanley, J. A.; McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29−36.

(51) Bamber, D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J. Math. Psychol.* **1975**, *12*, 387−415.

(52) 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.

(53) 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*, No. 225.

(54) Su, M.; Yang, Q.; Du, Y.; Feng, G.; Liu, Z.; Li, Y.; Wang, R. Comparative Assessment of Scoring Functions: The CASF-2016 Update. *J. Chem. Inf. Model.* **2019**, *59*, 895−913.

(55) Lafleur, K.; Dong, J.; Huang, D.; Caflisch, A.; Nevado, C. Optimization of inhibitors of the tyrosine kinase EphB4. 2. Cellular

potency improvement and binding mode validation by X-ray crystallography. *J. Med. Chem.* **2013**, *56*, 84−96.

(56) Marchand, J. R.; Dalle Vedove, A.; Lolli, G.; Caflisch, A. Discovery of Inhibitors of Four Bromodomains by Fragment-Anchored Ligand Docking. *J. Chem. Inf. Model.* **2017**, *57*, 2584−2597.

(57) Verdonk, M. L.; Berdini, V.; Hartshorn, M. J.; Mooij, W. T.; Murray, C. W.; Taylor, R. D.; Watson, P. Virtual screening using protein-ligand docking: avoiding artificial enrichment. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 793−806.

(58) Jacquemard, C.; Drwal, M. N.; Desaphy, J.; Kellenberger, E. Binding mode information improves fragment docking. *J. Cheminf.* **2019**, *11*, No. 24.

(59) Budin, N.; Majeux, N.; Caflisch, A. Fragment-Based flexible ligand docking by evolutionary optimization. *Biol. Chem.* **2001**, *382*, 1365−1372.

(60) Giordanetto, F.; Jin, C.; Willmore, L.; Feher, M.; Shaw, D. E. Fragment Hits: What do They Look Like and How do They Bind? *J. Med. Chem.* **2019**, *62*, 3381−3394.

(61) Zhao, H.; Huang, D. Hydrogen bonding penalty upon ligand binding. *PLoS One* **2011**, *6*, No. e19923.

(62) Teotico, D. G.; Babaoglu, K.; Rocklin, G. J.; Ferreira, R. S.; Giannetti, A. M.; Shoichet, B. K. Docking for fragment inhibitors of AmpC beta-lactamase. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7455−7460.

(63) Ekonomiuk, D.; Su, X. C.; Ozawa, K.; Bodenreider, C.; Lim, S. P.; Otting, G.; Huang, D.; Caflisch, A. Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. *J. Med. Chem.* **2009**, *52*, 4860−4868.

(64) Rathi, P. C.; Ludlow, R. F.; Hall, R. J.; Murray, C. W.; Mortenson, P. N.; Verdonk, M. L. Predicting "Hot" and "Warm" Spots for Fragment Binding. *J. Med. Chem.* **2017**, *60*, 4036−4046.