NRGRank: Coarse-grained structurally-informed ultra-massive virtual screening

Thomas Descoteaux[1] and Rafael Najmanovich[1*]

[1]Departement of Pharmacology and Physiology, Faculty of Medicine, Université de Montréal

* To whom correspondence should be addressed: rafael.najmanovich@umontreal.ca

ABSTRACT

NRGRank is a coarse-grained structurally-informed software capable of performing ultra-massive virtual screening with accuracy comparable to docking-based methodologies but with 100 fold speed increase. NRGRank is based on a coarse-grained evaluation of pairwise atom-type pseudo-energy interactions that implicitly accounts for compound and side-chain flexibility as well as limited backbone movements. We compare NRGRank to docking-based virtual screening software Glide, Autodock Vina and DOCK 3.7 on the DUD-E virtual screening benchmark using enrichment factors at 1% (EF1). We observe broad variations of EF1 values across targets, structural models and methods. For apo form or AlphaFold2 models, out of a subset of 37 targets from DUD-E, NRGRank has better EF1 values than Glide for 12 and 13 targets respectively. Even in holo form, where the accuracy of docking software increases, NRGRank has better EF1 values than 13, 10 and 5 targets out of 37 compared to AutoDock Vina, DOCK 3.7 and Glide respectively. Comparing the rank of true binders in Glide and NRGRank shows that true binders ranked in the top 1% are complementary between methods irrespective of the target form (AlphaFold, apo or holo). That is, utilizing NRGRank detects binders that are missed by Glide (and presumably other methods), whereas those found by Glide are missed by NRGRank. Furthermore, we observe that most hits found by NRGRank within the top 50 predictions remain once the top 1% of predictions when re-scored with Glide, but the hit rate within the top 50 predictions increase. NRGRank can evaluate one molecule in 0.3 s on average, enabling a modern laptop with 8 cores to screen 1,000,000 molecules in 24 hours, up to two orders of magnitude faster than the reported speed of DOCK 3.7, AutoDock Vina running on GPUs and Glide. NRGRank occupies a unique niche among tools for virtual screening being insensitive to structural inaccuracies but with comparable accuracy as state-of-the-art docking methods and fast as AI-based methods but without the dangers of overfitting as it is based on 780 pseudo-energy parameters. Combined with the fact that NRGRank does not require extensive or expensive computational resources or expert pre-processing of targets, it is unique in making high-performance ultra-massive virtual screening accessible to all.

INTRODUCTION

Among the different approaches in drug design, sifting through vast diverse compound sets to identify potential candidate compounds offers the possibility of exploring novel areas of chemical space. While cutting-edge robotic screening systems can evaluate an impressive 200,000 compounds per day [1], their high costs make them largely inaccessible and impractical to explore the several billion potential compounds that are even today accessible through synthesis-on-demand with a small number of synthetic steps and moderate fragment libraries. The landscape of molecule libraries is continuously expanding, with only the subset of lead-like compounds (MW ≤ 460, -4 ≤ SlogP ≤ 4.2, HBA ≤ 9, HBD ≤ 5, Rings ≤ 4, RotBonds ≤ 10) from the Enamine Real library (under 10% of the entire library) boasting a staggering 7.88 billion compounds, approximately 80% of which have a high likelihood of successful synthesis [2]. Conducting high throughput experimental screening with a small fraction of these molecules presents significant challenges in terms of time and expense, not to mention the considerable resources required for synthesis and transportation.

Virtual high-throughput screening (vHTS) offers a promising alternative, particularly for exploring novel libraries in the early stages of drug design. Such methodologies may take many forms. One common strategy involves ranking the molecules in a library through a calculated score and prioritizing those with the most favorable scores for further investigation. Research indicates that enlarging the library size utilized in virtual screens can increase hit rate [3,4]. As libraries grow, algorithms become more robust and computational power more widespread, ultra-massive vHTS (uHTS) is beginning to impact drug development, often identifying weak binding compounds [5] and sometimes strong binders [6].

In 2023, the Target Central Resource Database (TCRD) contained approximately 17,500 proteins in the human proteome for which there are no known binders and limited knowledge of protein function [7]. Recent advances in protein structure prediction brought about by AlphaFold 2 [8] have produced reliable structural models of all these proteins, enabling the potential use of uHTS for identifying modulators of protein function to significantly accelerate research.

A recent study [9] compared the effect of using different types of target structures on enrichment factors at 1% (EF1) for the Glide docking method [10] on 37 targets from the Directory of Useful Decoys – Enhanced (DUD-E) benchmark dataset [11]. The authors show that utilizing an AlphaFold model or apo (unbound) structure halved the average EF1 compared to that obtained with a holo protein structure. While extra refinement steps may be taken to enhance the accuracy of virtual screening on apo or model structures, such refinement involves the utilisation of holo (bound) forms of the target. This result highlights the sensitivity of Glide, considered by some to be the industry standard in docking simulations, to the quality of the structure of the target. Therefore, while Glide outperforms all other methods using holo proteins, its performance is the same as all other methods, including NRGRank as shown below, for apo proteins or AlphaFold models. It is unclear how Autodock Vina, DOCK 3.7 perform when using apo protein structures or AlphaFold models.

The accessibility of docking methods is often affected by additional requirements. A recent literature review [12] reveals that out of the 12 most widely used virtual screening software packages, only 4 are accessible for academic use free of charge: AutoDock Vina [13,14], DOCK 3.7 [15], FRED [16] and LibDock [17]. Some open-source software require closed-source third-party applications, for which academic licenses are mandatory during target or ligand preparation. For instance, DOCK

3.7 relies on several third-party software components for optimal performance, such as OpenEye Omega, OEChem, JChembase, Corina, and Chimera, all of which entail academic or paid licenses (source: https://wiki.docking.org/index.php/Install_DOCK_3.7). These constraints further add to the entry barrier for researchers seeking to utilize virtual screening.

Recently, numerous methodologies leveraging artificial intelligence have emerged [18]. However, these approaches come with notable drawbacks. Primarily, they tend to struggle in scenarios outside their training data. Despite exhibiting promising performance, particularly when using data they were trained on [19], they often perform poorly when confronted with unfamiliar situations [20]. It is still unclear how well such methodologies perform considering the several types of bias that may be present. From the trivial bias of not having a dataset split containing an evaluation set and optimizing parameters to improve test-set performance (Goodhart's law), the still trivial analog bias (different targets in the dataset being evolutionarily related or binders of the same target sharing common scaffolds) to the more subtle decoy bias described next. The Directory of Useful Decoys-Extended (DUD-E) [11] is widely used as a benchmark for virtual screening with physics-based docking methods. However, The DUD-E dataset embodies decoy bias by virtue of the property matching [21], making its use specifically for artificial-intelligence methods problematic. By selecting decoys that have matched properties but are topologically different, a classifier that learns a large set of possible topological variations maintaining physical properties, may effectively learn the complement of that set (true binders), i.e., the possible topologies that maintain the physical properties but are not decoys. By creating alternative decoy sets, Chen *et al.* demonstrated that CNN-based models achieve high accuracy due to analog or decoy biases rather than learning underlying rules that govern ligand-protein interactions [21]. In this work, we compare NRGRank, a physics-based method, to other physics-based methods on the DUD-E benchmark dataset. Given the current absence of a widely used benchmark dataset devoid of decoy bias, we do not compare NRGRank to any AI-based methods.

The surge in popularity of GPU (Graphics Processing Unit)-based software, exemplified by tools like Uni-Dock [22] and Vina-GPU 2.0 [23], is primarily driven by their remarkable speed enhancements compared to traditional CPU-based applications. For instance, Uni-Dock can dock 38.2 million molecules within 12 hours, leveraging the power of 100 NVIDIA V100 GPUs. However, while the speed gains are undeniably impressive, it is essential to acknowledge the limitations of GPU-based docking software. One prominent drawback is their reliance on technologies like CUDA, as seen in the case of the most recent version of Uni-Dock (v1.1). CUDA is a programming model exclusive to NVIDIA GPUs, meaning that the software cannot be utilized by users lacking compatible graphics cards. This accessibility hurdle is compounded by the recent "AI boom" which has spurred an unprecedented demand for GPUs, consequently driving up their prices significantly and creating an economic barrier to their utilization in vHTS.

Here we present NRGRank, a python-based coarse-grained structure-informed software capable screening a million compounds per day in a typical laptop CPU, requiring no manipulation of the target structure with complementary results and comparable accuracy in terms of EF1 values to those of existing docking methods, particularly for apo protein structures and AlphaFold models, removing barriers for ultra-massive virtual screening and opening the doors to exploring both chemical and protein space broadly and unbiasedly.

RESULTS AND DISCUSSION

PERFORMANCE ON APO PROTEIN STRUCTURES AND ALPHAFOLD MODELS

Virtual screening software generally overlook the dynamic nature of side chains in favour of speed. This is especially important when using targets with apo protein structures and AlphaFold models. Apo proteins (without a bound ligand) can exhibit significant conformational flexibility. In their apo form, the binding pocket may not be well-defined or may undergo conformational changes upon ligand binding. In 90% of binding-sites at least one side-chain undergoes a rotamer change upon ligand binding and in approximately a third of cases the rotamer change is critical for ligand binding [24]. AlphaFold provides highly accurate predicted protein structures, but these predictions generally account for the full range of conformational flexibility that might be present in a native protein. Predicted models may have uncertainties in side-chain positions or the modelled conformation reflect a local minimum incompatible with ligand binding, which traditional docking methods do not handle well in virtual screening scenarios. To address this limitation, NRGRank adopts a coarse-grained methodology that implicitly accounts for side chain and ligand flexibility. NRGRank calculates binding energy on a per index-cube basis rather than a fixed distance. That is to say, the contribution to the overall interaction energy of a ligand atom within an index-cube interacting with a protein atom in an adjacent index-cube is the same irrespective of the specific locations of each atom within their respective index cubes. With the current value for the index-cube length (6.56 Å) we obtain the same accuracy in AF2, apo and holo structures (see below), suggesting that most ligand conformational changes, side chain rotamer changes or minor backbone movements are likely maintaining the relevant atoms within the same index-cube, reducing sensitivity to the effect of dynamics in ligand binding while at the same simplifying and speeding up the algorithm. We experimented with a wide range of index-cube length values (Supplementary Figure S1), as well as number of rotamers per ligand (Supplementary Figure S2) or rotations per axis (Supplementary Figure S3). on the DUD-E dataset and observe that EF1 results to be insensitive to variations of these parameters around the adopted default values: a single conformer per ligand, 9 rotations per axis (729 in total) and an index-cube length of 6.56 Å as discussed in the methods section.

We utilize the DUD-E dataset to compare the performance of NRGRank against that of Glide for apo form protein structures or AlphaFold2 models. The comparison is restricted to Glide as we could not find published data for the same apo targets or AlphaFold2 models for DOCK 3.7 or AutoDock Vina. These 37 targets selected by the authors of the manuscript evaluating Glide [9] are cases for which apo structures with respect to the known binders could be identified [25] . The results show considerable variability in EF1 for both software programs when analyzing apo and AF2 structures for all targets. Specifically, the average EF1 value for NRGRank with apo structures is 5.4 ± 5.7, while for Glide with apo structures the EF1 average is 10.4 ± 8.1. For AF2 structures, NRGRank yields an average EF1 value of 5.4 ± 6.0, while Glide has a value of 11.0 ± 10.9. A paired Student's t-test between NRGRank and Glide yields p-values of 0.003 for apo structures and 0.004 for AF2 structures. The EF1 values for individual apo and AF2 targets is shown in Figure 1. Despite the average EF1 differences, NRGRank have higher individual EF1 values than Glide for 14 targets (representing just over a third of this DUD-E subset) for apo form and AF2 models (Figures 1B and 1C). The individual values of EF1 for each target in the different forms for each software can be found in Supplementary Table 1.

Despite the differences in performance between NRGRank and Glide for apo or AF2 models, each software identifies distinct subsets of true binders. Figure 2 show the rank of individual true binders across all 37 DUD-E targets for which we can compare our results with those of Glide for AF2, apo and holo targets. There is a high number of ligands with rank above 1% for both NRGRank and Glide (quadrant 1 in all panels) representing true positive binders missed by both software.

However, there are large numbers of molecules that have a rank below 1% for one software but not for the other (quadrants 2 and 4 in all panels). Only a very small fraction of true positives are detected by both software within the top 1% (quadrant 3 in all panels). While intuitively users of virtual screening docking software would be tempted to use multiple software and focus on the intersection of top results as independent validation (quadrant 3), our results indicate that with this strategy most true binders for any target would be discarded. Therefore, despite any differences in EF1 values, utilization of NRGRank may detect true binders that are missed by other software.

PERFORMANCE ON HOLO PROTEIN STRUCTURES

The evaluations of virtual screening software performance most likely to represent real usage conditions for virtual screening in our view are those using apo forms or modeled structures as in the previous section. However, docking software in virtual screening is often evaluated using holo protein structures [26], in part because side-chain flexibility although important is generally not considered in virtual screening. We utilize the holo form structures provided with the DUD-E dataset to compare NRGRank to widely used docking software for virtual screening such as DOCK 3.7 and AutoDock Vina. As we could find data for Glide only for the subset of 37 targets selected by the authors analysing Glide [9]. We restrict the analysis to the 37 targets. NRGRank achieves an average EF1 of 6.9 ± 8.0 for holo form targets while Glide obtains an average EF1 of 22.4 ± 13.2. The same situation observed for Apo and AF2 targets is true here again. Namely, when comparing the ranks of individual binders, it is still true that true binders detected by NRGRank within the top 1% are largely missed by Glide and those found by Glide among the top 1% predictions are missed by NRGRank. Therefore, despite a vastly lower EF1 for holo protein structures, it is still true, even in the extreme comparison between Glide and NRGRank using holo protein structures, that NRGRank is complementary to Glide (Figure 2C).

For the same subset of DUD-E, the average EF1 values for AutoDock Vina and DOCK 3.7 are respectively 10.0 ± 8.6 and 14.2 ± 10.8 . Despite EF1 values for Holo structures being smaller for NRGRank, the variance for all software is so large that EF1 values do not represent large statistical differences (Figure 3A). Comparing the EF1 of individual targets between NRGRank and other software for holo target structures (Figure 3B and 3C) shows that NRGRank has higher or equal EF1 values for 13 targets relative to AutoDock Vina and 10 relative to DOCK 3.7. Despite the nearly fourfold difference in average EF1 values with respect to Glide, for 5 targets, NRGRank still has better EF1 values than Glide (Figure 3D).

RE-SCORING NRGRANK RESULTS WITH GLIDE

NRGRank is a method for ranking small-molecules in virtual screening, unlike docking methods, it is not designed to predict the correct ligand pose and generate a ligand-protein complex structure. In situations where a docked pose in virtual screening is required, we envision the use of a docking software (such as AutoDock Vina, DOCK 3.7, Glide or FlexAID among others) to be used to re-score the output of NRGRank. Additionally, as docking methods are more accurate, they can also be used to re-score a fraction of top results obtained by NRGRank. In order to simulate such a situation with Glide, we re-scored the top 1% predictions of DUD-E compounds (including both true binders and decoys) for a given target in the rank order defined by their Glide scores. The relative position in the ranked list for each target of the true binders represents the effect of docking such molecules with Glide. In practical terms we decided to utilize only targets with at least 50 true binders among the subset of 37 targets for which we have glide ranks and calculate how many of true binders are among the top 50 predicted molecules before and after re-scoring with Glide.

This experiments reflects a real scenario in which one would use NRGRank to rank a large dataset, re-score a small fraction of top ranked molecules and test experimentally 50 molecules for the presence of binders. The value 50 for the number of molecules to be tested experimentally was chosen here to balance having a number that would be reasonable to test experimentally and the number of targets in the subset of DUD-E for which Glide scores are available. This experiment was performed using AF2 models, as well as apo and holo protein forms for 34 targets (marked with an asterisk in Supplementary Table 1). For AF2 models, a total of 329 binders were among the 1% NRGRank predictions with 117 among the top 50 predictions for all 34 targets (quadrants 2 and 3 in Figure 4A). Of these 117, 111 remain among the top 50 predictions even after re-scoring (quadrant 3 in Figure 4A), the remaining 6 are lost with re-scoring (quadrant 2 in Figure 4A) and an additional 69 binders that had ranks above 50 with NRGRank are added to the top 50 list after re-soring (quadrant 4 in Figure 4A). Re-scoring increases the fraction of binders in the top 50 from 45% to 64% (Figure 4B shows the hit rate distribution). For apo targets (Figures 4C and 4D), we have 145 in quadrant 1 (missed with or without re-scoring), 147 in quadrant 3 (in the top 50 before and after re-scoring), 11 in quadrant 2 (lost upon re-scoring) and 84 in quadrant 4 (added to the top 50 upon re-scoring). The fraction of binders in the top 50 increases from 40% to 59%. For holo targets (Figures 4E and 4F), we observe 135, 11, 217 and 122 in quadrants 1, 2, 3 and 4 respectively with a fraction of binders in the top 50 increasing from 47% to 90% upon re-scoring. While re-scoring is beneficial, it is by no means essential in order to find binder even among the top 50 predictions. In X out of the 32 targets, we find at least one binder among the top 50 predictions. The hit rate distributions show that for all three types of targets the re-scoring shifts the distributions right, increasing the hit rate for the 32 targets (Figure 4B, D & F).

EVALUATION OF NON-PROPERTY MATCHED DECOYS

By virtue of the high level of similarity between known binders and decoys (assumed to be non-binders), the DUD-E dataset defines a lower bound on the accuracy of any biophysics-based vHTS method (not so for artificial intelligence systems as per the discussion in the introduction). To simulate a more common scenario we conducted an experiment to assess the capability of NRGRank in distinguishing true binders from non-property matched molecules. The Enamine real diversity subset (ERDS) compound library at 1% comprising was used for this experiment. At the time of download this library contains 67.5 million compounds (and is available for download from our GitHub repository for reproducibility purposes), currently the equivalent dataset available from Enamine contains over 90 million compounds. The true binders for each target in DUD-E were selected from ChEMBL as molecules matching any of the following criteria: IC50, EC50, K_i or K_d values ≤ 1 μM. This is the same criteria used to select binders by the original DUD-E authors. We plot in Figure 5 for each target the average score for ERDS against that for true binders. For all targets, the average score for known binders is considerably better than for random molecules. This result shows that NRGRank can discriminate binders from random molecules. For 79 out of 102 DUD-E targets, the top ranked molecule is a true binder but considering the overlap between the distributions (Supplementary Figure S4), for all targets there are large numbers of ERDS molecules with ranks better than the rank of known binders. For 23 targets, ERDS molecules have better scores than known binders. For example, for the β_1-adrenergic G-protein coupled receptor (DUD-E ID adrb1) two ERDS molecules rank better than the best binder.

The PDB structure assigned to adrb1 by DUD-E is 2VT4. The receptor is the turkey adrb1 bound to the high affinity antagonist cyanopindolol (pdb ligand code P32). The affinity of this molecule for the guinea pig adrb1 ranges from 27 to 40 pM [27], equivalent to -14.30 to -14.18 kcal/mol. We

utilized our Surfaces [28] software to predict the ΔG for P32 and obtained a ΔG of -10.01 kcal/mol. We then proceeded to dock the top-ranking ERDS compound (Enamine ID s_22____22609224____10640652 or S_652 for short) with our software FlexAID [29] followed by analysis with Surfaces of the top ranked pose. The predicted ΔG for S_652 is -14.75 kcal/mol, suggesting that this molecule has more opportunities for favourable interactions with the target, in part due to the larger size. Many of the relevant interactions observed between P32 and adrb1 in 2VT4 [30] are equally present in our model for S_652 interacting with adrb1 (Figure 6). We re-scored molecule S_652 on all 102 targets with FlexAID and find a Pearson's correlation coefficient between the NRGRank scores and FlexAID scores of 0.62 (Supplementary Figure S5). Our experience with FlexAID is that a score below -200 AU is generally indicative of successful binding. In the case of S_652, the FlexAID score is -258.03 AU. The median FlexAID score for S_652 on all 102 DUD-E targets is -87.73. S_652 is among the top 100 ranked molecules for 40 targets. This may hint at this molecule being an artifact of the method. However, only 12 targets have FlexAID scores below -200, including among them adrb2 (also known to also tightly bind cyanopindolol). Of these, there are only 3 DUD-E targets with better FlexAID scores than adrb1 (glcm, cp2c9 and ppara with scores of -269.63 AU, -263.27 AU and -259.13 AU respectively). Whereas the NRGRank ranks point to S_652 as a potential artifact, the FlexAID scores do not.

The fact that S_652 is rank among the top 100 molecules for 40 DUD-E targets suggests that the scoring function in NRGRank captures signals beyond what makes the binding-sites of these diverse targets different. One possibility is a signal from backbone atoms. However, a version of NRGRank only considering interactions with protein side-chain atoms had lower performance (average EF1) on the DUD-E than the version presented here.

Considering the discussion above, we believe it is beneficial to subject top-ranking compounds to additional scrutiny including docking against a number of unrelated targets to determine, as is the case with S_652 on adrb1, if the compound ranks highly on multiple unrelated targets.

ANALYSIS OF TOP SCORING COMPOUNDS ON CLOSELY RELATED TARGETS

The capability of binding a particular target without binding similar cross-reactivity targets is a desired objective in drug design. As NRGRank coarse-graining 'bundles' multiple atoms in index-cubes and sums over 27 index-cubes, it could render the overall contribution to scoring of small binding site differences insignificant and make it impossible to differentiate similar binding-sites in terms of candidate binding compounds. In order to confirm that it is possible to find unique top scoring molecules for closely related proteins, we screened the ERDS compound library of 67.5M molecules on three closely related kinases.

The proto-oncogene serine/threonine kinase PIM1 is considered as a potential target in multiple types of cancer including prostate and triple negative breast cancer [31–37]. Two other kinases PIM2 and PIM3 belong to the same family. There are six differences between residues lining the PIM1, PIM2 and PIM3 binding-sites (Figure 7). However, the side-chains of four among these are facing away from the cavity. Of the remaining two positions with cavity facing reside side-chains with differences among the three PIM kinases, There are two differences between PIM1 and PIM2: $VAL126^{PIM1}$-$ALA122^{PIM2}$ and $GLU124^{PIM1}$-$LEU120^{PIM2}$, while between PIM1 and PIM3 only one $VAL126^{PIM1}$-$ALA129^{PIM3}$ and between PIM2 and PIM3 also only one difference: $LEU120^{PIM2}$-$GLU127^{PIM3}$ (Figure 7).

Only four molecules within the top 100 ranked molecules for PIM1, are also among the top 100 ranked molecules for either PIM2 or PIM3 with a significant number ranked 10,000 or higher in

either (Figure 8). Similarly, there are several top-ranking molecules for PIM2 (Supplementary Figure S6) and PIM3 (Supplementary Figure S7) that rank more poorly on the other members of the family. Furthermore, the top 100 ranked molecules for PIM2 have better ranks on PIM3 than for PIM1 (most molecules lie under the diagonal in Supplementary Figure S6) in agreement with the smaller number of mutations between 2 and 3, particularly considering the residues with side-chains facing the cavity. The same is true for top ranking PIM3 molecules, they rank better in PIM2 than in PIM1 (Supplementary Figure S7), but in this case, the molecules have better ranks for PIM1 (the points are shifted left) in agreement with the fact that there are less mutations between PIM3 and PIM1 versus PIM2. Additionally,

SPEED AND PERFORMANCE

NRGRank needs 0.36 +/- 0.23s seconds on average to screen one molecule when testing the whole DUD-E dataset. This speed comes as a result of the pre-processing, the simplicity of the scoring function that unlike the original FlexAID scoring function [38] does not necessitate the calculation of surface areas in contact and lastly, due to the fact that for a given ligand atom, all interacting protein atoms within the 26 cubes of 6.56 Å plus the cube containing the ligand atom contribute equally to the docking score. This makes it possible to account implicitly for different ligand conformations. We tested the effect of varying the number of ligand conformers on the average EF1 obtained on the DUD-E diverse set containing 8 proteins, one for each class, and found no advantage in scoring more than one conformer per ligand (Supplementary Figure S2). Likewise, we also varied the number of rotations per axis and found no improvement in EF1 when probing more than 9 rotations per axis (Supplementary Figure S3).

DATA AND CODE AVAILABILITY

NRGRank can be downloaded from our GitHub repository at github.com/NRGlab/NRGRank, which includes installation and the necessary example files for running it. The ERDS library of Enamine compounds used in this work is available at our Zenodo data depository. We chose to provide a copy of this compound library as it is no longer available from Enamine, having been superseded by a 96.7M library (that should contain the 67.5M subset we use). The Zenodo data depository also contains the ranks for all binders and decoys for each of the DUD-E targets, to permit others to perform detailed comparisons as we were able to do only because the team behind Glide chose to make the data public. We think this should be the norm and not the exception. The Zenodo data depository also contains the binding-site definitions used for each DUD-E target. This depository can be found at http://dx.doi.org/10.5281/zenodo.14841476. NRGRank can also be used as part of the NRGSuite-Qt [39] to screen pre-loaded compound databases including FDA approved drugs, all ligands present in the PDB and all tetrapeptides. Additionally, the NRGSuite-Qt offers the possibility for users to load their own ligand datasets in SMILE format. The NRGSuite-Qt PyMOL plugin also includes GetCleft [29] for the detection and refinement of cavities/binding-sites, FlexAID for re-scoring and pose determination, NRGTEN [40] for the generation of target conformational ensembles and IsoMIF [41] for the detection of binding-site similarities, particularly useful when screening ligands present in the PDB [42] .

CONCLUSIONS

NRGRank is structurally-informed ultra-massive virtual screening software based on the exhaustive quantification of pairwise atomic interaction energies using 40 atom types. NRGRank shows comparable performance in virtual screening experiments to docking software Glide, AutoDock Vina and DOCK 3.7, particularly for targets in apo form and AlphaFold2 models while

being between one and three orders of magnitude faster than these software. A re-scoring experiment with Glide proves that re-scoring finds additional binders but even without re-scoring NRGRank finds different and substantial subset of binders to Glide, the re-scoring process effectively enrich the top results with those binders originally missed by NRGRank without losing the ones found uniquely by NRGRank. We conjecture that the same is true for any pair of virtual screening software. That is, that each has unique biases that lead to the detection of different sets of binders. Therefore, instead of seeking to use different virtual screening software to validate results, users are more likely to benefit from the union of top results from various software rather than the intersection of top results. We show that NRGRank can discriminating known binders in large libraries and that re-scoring can help judge if a molecule represents a screening artifact or a candidate for further analysis. Lastly, we show that NRGRank can detect different top ranking molecules even for very closely related targets. The simplicity of NRGRank, which does not require expert pre-processing or target refinement, combined with its minimal computational resource requirements, equivalent accuracy for apo form or model targets, complementary results to docking-based screening software, and high speed, makes it a valuable addition to the arsenal of computational tools for ultra-massive virtual screening, lowering the barriers to its adoption.

ACKNOWLEDGEMENTS

We thank the Digital Research Alliance of Canada for providing computational resources. RJN is a member of the Quebec Network for Research on Protein Function, Engineering, and Applications (PROTEO).

METHODS

Three inputs are necessary to run NRGRank: Target structure, binding-site definition and ligand structures. The Target structure and binding-site definition are used in a pre-processing step to create a lookup grid of pre-calculated pairwise atom type interactions. This pre-processed grid and the ligand structures are then used by NRGRank to probe exhaustively each ligand rotation and translation within the binding-site. The target and ligands are required to be in Mol2 format with SYBYL atom types assigned.

TARGET AND LIGAND PREPARATION

NRGRank utilizes a scoring function based on SYBYL atom types. As such, it is necessary to map both protein and ligand atoms into SYBYL atom types. The Mol2 structure file format automatically assigns SYBYL atom types to both protein and ligand atoms. The first step in preparing the target is converting the original protein structure and ligand files to the Mol2 format. There are many ways to do this, for example using PyMOL [43] or Open Babel [44]. RDKit [45] was used to generate one conformer for each ligand from SMILES. We used the ETKDGv3 conformer generator in EmbedMultipleConfs from the rdkit.Chem.AllChem module.

BINDING-SITE DEFINITION

Binding-sites were defined as previously [38,46] using the GetCleft algorithm, which is a C programming language implementation of the Surfnet algorithm [47]. Briefly, the algorithm defines a binding-site as the largest set of intersecting spheres that can be placed at the midpoint between any pair of atoms without overlapping with the van der Walls radii of any protein atom and with sphere radii within lower (r_l) and upper (r_u) values. Careful choice of the lower and upper values permits to fine-tune the shape of the binding-site volume without "spilling over" across cavities that ought to be considered unconnected by visual inspection. In cases where adjustments were

necessary, we fine-tuned the sphere radii away from default values. Targets in the DUD-E dataset are provided in the holo form. In these cases, the GetCleft cavity containing the bound ligand was selected for the experiments using r_l and r_u 1.75 Å and 3.75 Å respectively. For experiments with apo and AF2 structures, we utilized the structures provided by the authors [9] and the equivalent GetCleft cavity. That is, the cavity in the apo or AF2 structure that overlaps with the ligand-bound cavity in the holo form when the two protein structures are superimposed. It is worth mentioning that we do not constraint the definition of the binding-site volume in any way with the bound ligand in the holo form, a practice often employed in the benchmarking of docking algorithms and that greatly biases the analysis by severely restricting the volume of search space. As NRGRank performs an exhaustive search, the additional search space at most could impact execution time. In the case of PIM1, PIM2 and PIM3, we selected the binding pocket with the largest volume, that coincides with the ATP binding pocket.

TARGET PRE-PROCESSING

A pre-processing step is required to create a lookup grid of pseudo-energies for each atom type based on the distribution of protein atoms in the structure. The Mol2 protein structure file is loaded into the software and atoms are indexed into a grid of index-cubes. As NRGRank is somewhat insensitive to the choice of index-cube width (Supplementary Figure S1), we chose a width with a physical explanation, namely equal to two times the maximum van der Waals atom radius of all Sybyl atom types (1.88 Å) augmented by the radius of water (1.44 Å), or 6.56 Å. As such, an index cube may contain two atoms in contact, as defined by Surfaces [28] and FlexAID [38]. Subsequently, for each of these cubes, the energy matrix developed for FlexAID is used to calculate the pseudo-energy of each possible Sybyl atom type with all protein atoms within the central and adjacent cubes, for a total of 27 cubes. The binding energy for one Sybyl atom type is given by:

$$\epsilon_{x'} = \sum_{i=1}^{n} \epsilon_{x'i'} \tag{1}$$

where $\epsilon_{x'}$ is the total energy for an atom x of Sybyl atom type x' located in a cube and all atoms located within 1 cube excluding other ligand atoms. The pairwise interaction energy between atom x and atom i of Sybyl atom type i' is given by $\epsilon_{x'i'}$. Therefore, a ligand atom of a particular atom-type will contribute an amount $\varepsilon_{x'}$ irrespective of where the protein atoms that contributes to that interaction is located within a 7622.11 $Å^3$ volume (equal to the volume of 27 cubes with 6.56 Å side). As a result, NRGRank is insensitive to ligand conformers, side-chain rotamers and minor backbone movements, any movements that maintain the atoms within the same interaction volume contribute equally.

The binding site volume provided as input is utilized to generate ligand anchor points as a grid spaced at intervals of 1.5 Å. Each anchor point represents a position at which each ligand will be tested.

The center of geometry of each ligand is placed on each anchor point in the binding site. For each anchor point, the ligand is rotated 9 times per axis to sample 729 different orientations. The binding energy of a pose is the cumulative sum of the binding energies associated with each ligand atom, as calculated during the pre-processing phase. The final score for a compound corresponds to the lowest pseudo-energy obtained in the exhaustive search. Given the simplified nature of the scoring,

the pose associated to that score is not saved, only the scores and relative ranking of all molecules tested.

EVALUATION OF PERFORMANCE ON APO PROTEIN STRUCTURES AND ALPHAFOLD MODELS

We utilized the apo, holo, and AlphaFold 2 structures on a subset of 37 DUD-E targets as provided by the authors for comparisons with Glide [9]. The structures provided in maegz format were converted to mol2 files utilizing Schrödinger Maestro (Release 2024-2; Schrödinger, LLC, New York, NY, 2024). The methodology outlined in the binding site definition section was used for holo structures. To mimic the binding site preparation for Glide, the holo binding site definition was used for apo and AlphaFold 2 structures.

When calculating ligand percentile rank from Glide, the best scoring result for each molecule was retained, and duplicate or worse scores were discarded.

EVALUATION OF PERFORMANCE ON HOLO PROTEIN STRUCTURE

For each of the 37 targets used in the Glide paper, the crystal structure was retrieved from the DUD-E database in the pdb format and converted to the Mol2 format usingPyMOL. Next, binding sites were defined around crystallized ligands using GetCleft with parameters as described above or adjusted when the ligand did not fit within the binding site. Target structures and ligand Mol2 files from the DUD-E website were then preprocessed using NRGRank. Subsequently, NRGRank screening was ran on the Narval cluster managed by Calcul Québec (calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca). This required approximately 30 minutes for the 102 targets in DUD-E. Finally, EF1 was calculated as described by Jain *et al.* [48].

EVALUATION OF NON-PROPERTY MATCHED DECOYS

All 102 targets were obtained from the DUD-E dataset. Known binders were obtained from ChEMBL [49,50] and respect the following criteria: EC50, K_i, K_d values ≤ 1 μM. The decoys are a set of diverse molecules comprising 67.5 million compounds from the Enamine REAL diversity subset. This resulted in the evaluation of over 6.7billion molecules. The ERDS 67.5M compounds dataset and the list of ChEMBL known binders for DUD-E targets are available online at the Zenodo data depository.

REFERENCES

1. Michael, S. *et al.* A Robotic Platform for Quantitative High-Throughput Screening. *ASSAY Drug Dev. Technol.* **6**, 637–657 (2008).

2. Grygorenko, O. O. *et al.* Generating Multibillion Chemical Space of Readily Accessible Screening Compounds. *iScience* **23**, 101681 (2020).

3. Lyu, J. *et al.* Ultra-large library docking for discovering new chemotypes. *Nature* **566**, 224–229 (2019).

4. Lyu, J., Irwin, J. J. & Shoichet, B. K. Modeling the expansion of virtual screening libraries. *Nat Chem Biol* **19**, 712–718 (2023).

5. Singh, I. *et al.* Structure-Based Discovery of Inhibitors of the SARS-CoV-2 Nsp14 N7-Methyltransferase. *J. Med. Chem.* **66**, 7785–7803 (2023).

6. Gahbauer, S. *et al.* Docking for EP4R antagonists active against inflammatory pain. *Nat. Commun.* **14**, 8067 (2023).

7. Kelleher, K. J. *et al.* Pharos 2023: an integrated resource for the understudied human proteome. *Nucleic Acids Res.* **51**, D1405–D1416 (2022).

8. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

9. Zhang, Y. *et al.* Benchmarking Refined and Unrefined AlphaFold2 Structures for Hit Discovery. *J. Chem. Inf. Model.* **63**, 1656–1667 (2023).

10. Friesner, R. A. *et al.* Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal Of Medicinal Chemistry* **47**, 1739–1749 (2004).

11. Mysinger, M. M., Carchia, M., Irwin, J. J. & Shoichet, B. K. Directory of useful decoys, enhanced (DUD-E): better ligands and decoys for better benchmarking. *Journal Of Medicinal Chemistry* **55**, 6582–6594 (2012).

12. Sabe, V. T. *et al.* Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *Eur. J. Med. Chem.* **224**, 113705 (2021).

13. 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.* **31**, 455–461 (2010).

14. Eberhardt, J., Santos-Martins, D., Tillack, A. F. & Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **61**, 3891–3898 (2021).

15. Balius, T. E., Tan, Y. S. & Chakrabarti, M. DOCK 6: Incorporating hierarchical traversal through precomputed ligand conformations to enable large-scale docking. *J. Comput. Chem.* (2023) doi:10.1002/jcc.27218.

16. McGann, M. FRED Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **51**, 578–596 (2011).

17. Diller, D. J. & Merz, K. M. High throughput docking for library design and library prioritization. *Proteins: Struct., Funct., Bioinform.* **43**, 113–124 (2001).

18. Murugan, N. A., Priya, G. R., Sastry, G. N. & Markidis, S. Artificial intelligence in virtual screening: Models versus experiments. *Drug Discov. Today* **27**, 1913–1923 (2022).

19. Zheng, L., Fan, J. & Mu, Y. OnionNet: a Multiple-Layer Intermolecular-Contact-Based Convolutional Neural Network for Protein–Ligand Binding Affinity Prediction. *ACS Omega* **4**, 15956–15965 (2019).

20. Buttenschoen, M., Morris, G. M. & Deane, C. M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.* **15**, 3130–3139 (2023).

21. Chen, L. *et al.* Hidden bias in the DUD-E dataset leads to misleading performance of deep learning in structure-based virtual screening. *PLoS ONE* **14**, e0220113 (2019).

22. Yu, Y. *et al.* Uni-Dock: GPU-Accelerated Docking Enables Ultralarge Virtual Screening. *J. Chem. Theory Comput.* **19**, 3336–3345 (2023).

23. Ding, J. *et al.* Vina-GPU 2.0: Further Accelerating AutoDock Vina and Its Derivatives with Graphics Processing Units. *J. Chem. Inf. Model.* **63**, 1982–1998 (2023).

24. Gaudreault, F., Chartier, M. & Najmanovich, R. J. Side-chain rotamer changes upon ligand binding: common, crucial, correlate with entropy and rearrange hydrogen bonding. *Bioinformatics* **28**, i423–i430 (2012).

25. Guterres, H., Park, S.-J., Jiang, W. & Im, W. Ligand-Binding-Site Refinement to Generate Reliable Holo Protein Structure Conformations from Apo Structures. *J. Chem. Inf. Model.* **61**, 535–546 (2021).

26. Xu, M., Shen, C., Yang, J., Wang, Q. & Huang, N. Systematic Investigation of Docking Failures in Large-Scale Structure-Based Virtual Screening. *ACS Omega* **7**, 39417–39428 (2022).

27. Engel, G., Hoyer, D., Berthold, R. & Wagner, H. (±)[125Iodo]cyanopindolol, a new ligand for β-adrenoceptors: Identification and quantitation of subclasses of β-adrenoceptors in guinea pig. *Naunyn-Schmiedeberg's Arch. Pharmacol.* **317**, 277–285 (1981).

28. Teruel, N., Borges, V. M. & Najmanovich, R. Surfaces: a software to quantify and visualize interactions within and between proteins and ligands. *Bioinformatics* **39**, btad608 (2023).

29. Gaudreault, F. & Najmanovich, R. J. FlexAID: Revisiting Docking on Non-Native-Complex Structures. *J. Chem. Inf. Model.* **55**, 1323–1336 (2015).

30. Warne, T. *et al.* Structure of a β1-adrenergic G-protein-coupled receptor. *Nature* **454**, 486–491 (2008).

31. Speers, C. *et al.* Identification of novel kinase targets for the treatment of estrogen receptor-negative breast cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **15**, 6327–6340 (2009).

32. Bianchini, G. *et al.* Prognostic and therapeutic implications of distinct kinase expression patterns in different subtypes of breast cancer. *Cancer research* **70**, 8852–8862 (2010).

33. Mahata, S. *et al.* PIM1/STAT3 axis: a potential co-targeted therapeutic approach in triple-negative breast cancer. *Med Oncol* **39**, 74 (2022).

34. Yuan, Y. *et al.* PIM1 promotes hepatic conversion by suppressing reprogramming-induced ferroptosis and cell cycle arrest. *Nat Commun* **13**, 5237 (2022).

35. Zhao, W., Qiu, R., Li, P. & Yang, J. PIM1: a promising target in patients with triple-negative breast cancer. *Med Oncol* **34**, 142 (2017).

36. Brasó-Maristany, F. *et al.* PIM1 kinase regulates cell death, tumor growth and chemotherapy response in triple-negative breast cancer. *Nat Med* **22**, 1303–1313 (2016).

37. Merkel, A. L., Meggers, E. & Ocker, M. PIM1 kinase as a target for cancer therapy. *Expert Opin Inv Drug* **21**, 425–436 (2012).

38. Gaudreault, F. & Najmanovich, R. J. FlexAID: revisiting docking on non-native-complex structures. *Journal of chemical information and modeling* **55**, 1323–1336 (2015).

39. Galdino, G. T., DesCôteaux, T., Teruel, N. & Najmanovich, R. NRGSuite-Qt: A PyMOL plugin for high-throughput virtual screening, molecular docking, normal-mode analysis, the study of molecular interactions and the detection of binding-site similarities. *bioRxiv* 2025.01.23.634566 (2025) doi:10.1101/2025.01.23.634566.

40. Mailhot, O. & Najmanovich, R. The NRGTEN Python package: an extensible toolkit for coarse-grained normal mode analysis of proteins, nucleic acids, small molecules and their complexes. *Bioinformatics* **37**, 3369–3371 (2021).

41. Chartier, M. & Najmanovich, R. Detection of Binding Site Molecular Interaction Field Similarities. *J. Chem. Inf. Model.* **55**, 1600–1615 (2015).

42. Rodríguez-Salazar, C. A. *et al.* Ebola virus VP35 interacts non-covalently with ubiquitin chains to promote viral replication. *PLOS Biol.* **22**, e3002544 (2024).

43. DeLano, W. *The PyMOL Molecular Graphics System. Schrödinger, LLC. (Www.Pymol.Org)*.

44. O'Boyle, N. M. *et al.* Open Babel: An open chemical toolbox. *J. Cheminformatics* **3**, 33 (2011).

45. Riniker, S. & Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **55**, 2562–2574 (2015).

46. Gaudreault, F., Morency, L.-P. & Najmanovich, R. J. NRGsuite: a PyMOL plugin to perform docking simulations in real time using FlexAID. *Bioinformatics* **31**, 3856–3858 (2015).

47. Laskowski, R. A. SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. *J Mol Graphics* **13**, 323–330 (1995).

48. Jain, A. N. Bias, reporting, and sharing: computational evaluations of docking methods. *Journal of Computer-Aided Molecular Design* **22**, 201–212 (2008).

49. Davies, M. *et al.* ChEMBL web services: streamlining access to drug discovery data and utilities. *Nucl. Acids. Res.* **43**, W612–W620 (2015).

50. Zdrazil, B. *et al.* The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res.* **52**, D1180–D1192 (2023).

Tables and Figures

Tables

Table 1. Average EF1 values for different software on the subset of 37 DUD-E targets by target structure

	Target type[a]		
	AF2	**Apo**	**Holo**
NRGRank	5.4 ± 6.0	5.4 ± 5.7	6.9 ± 8.0
AutoDock Vina	NA	NA	10.0 ± 8.6
DOCK 3.7	NA	NA	14.2 ± 10.8
Glide	11.0 ± 10.9	10.4 ± 8.1	22.4 ± 13.2

Figures

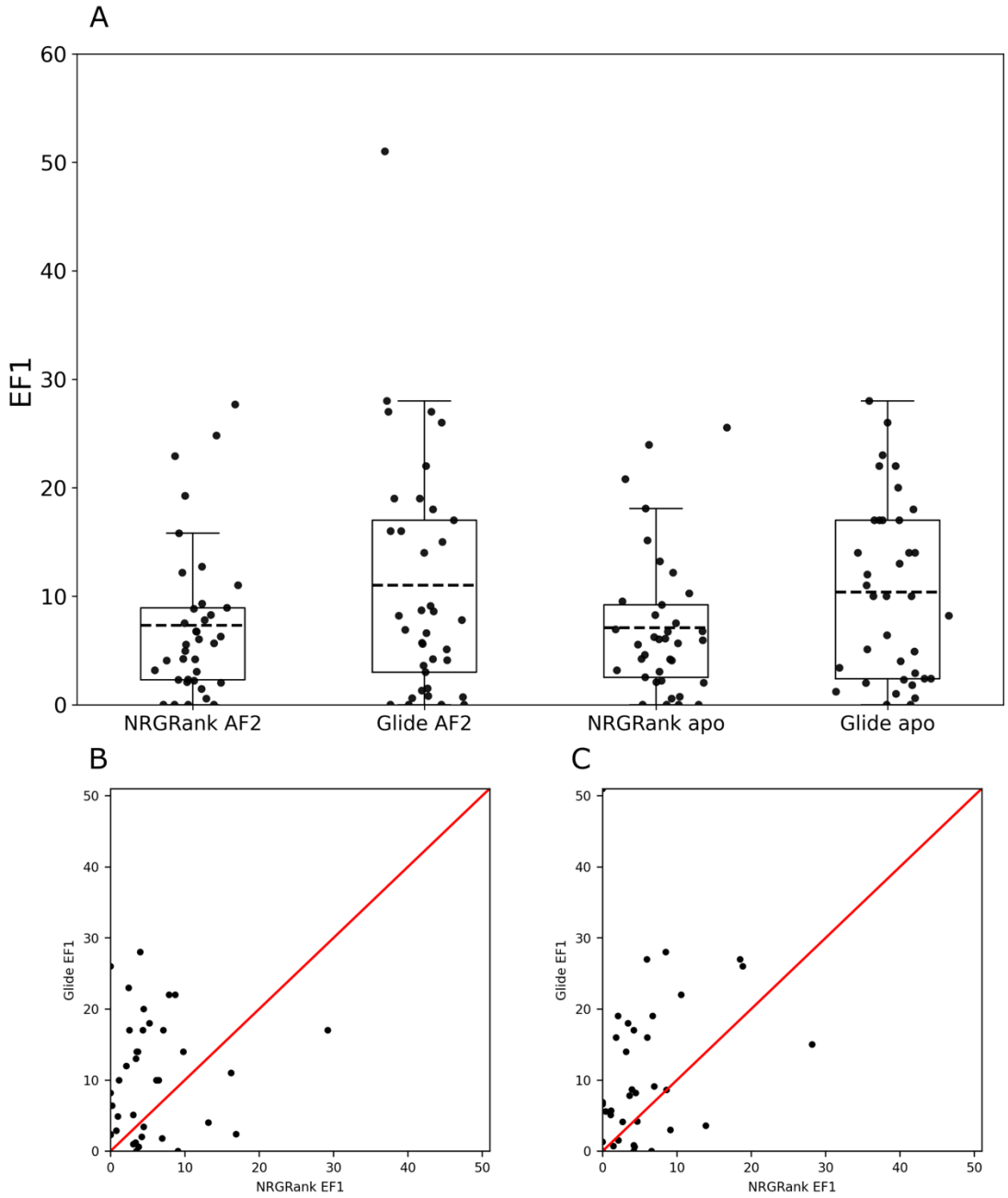


Figure 1. Enrichment factors for non-holo targets. EF1 for 37 DUD-E for AlphaFold2 (AF2) models and apo form (apo) targets for Glide and NRGRank (A) and comparison of EF1 for individual targets showing that for 12 target in apo form (B) and 13 targets for AF2 models (C) out of 37 , NRGRank has better EF1 values.

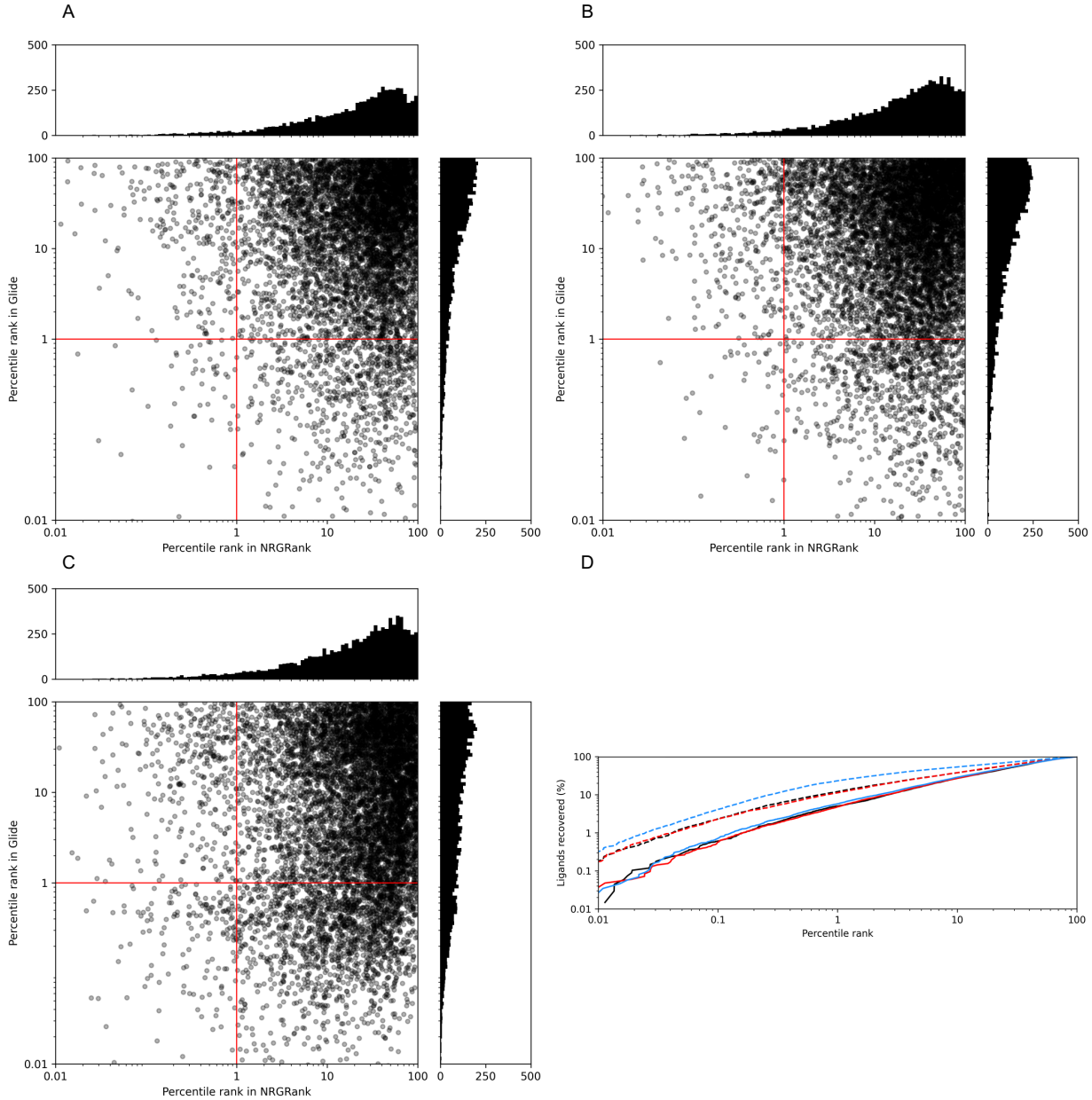


Figure 2. Comparison of rank of true binders in NRGRank and Glide for 37 AF2 (A), apo (B) and holo (C) targets. The number of true binders missed by either software (quadrant 1) are 5475, 6856 and 6164 respectively for AF2, apo and holo targets respectively. The number of binders identified by NRGRank but missed by Glide (quadrant 2) are 289, 352 and 337 respectively for AF2, apo and holo. The number missed by NRGRank but identified by Glide (quadrant 4) are 723, 886 and 1087 respectivelyfor AF2, apo and holo targets. Those found by both software (quadrant 3) are 50, 42 and 161 respectively for AF2, apo and holo targets.

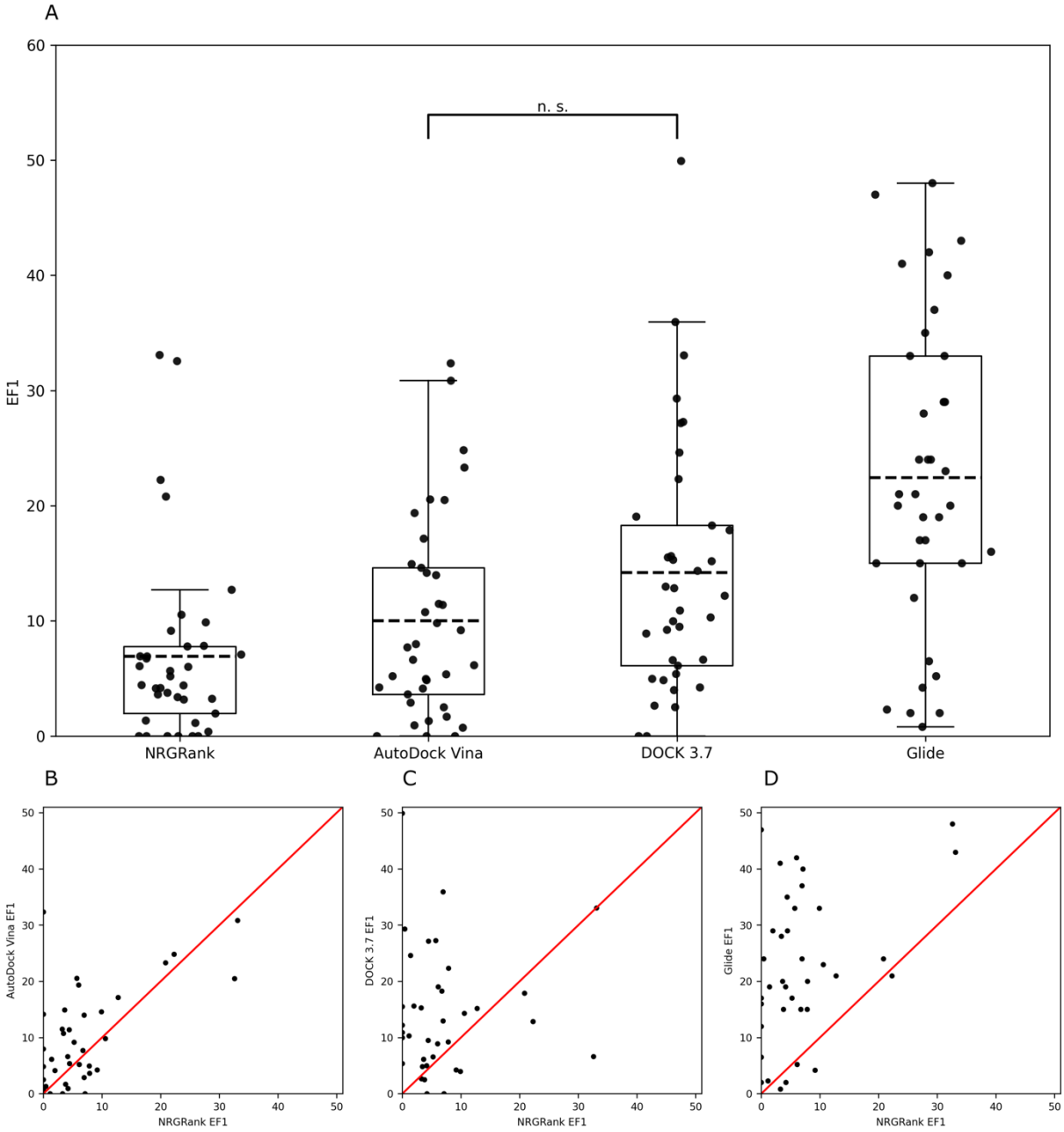


Figure 3. Comparison of the enrichment factor of NRGRank for 37 DUD-E targets in holo form for NRGRank, . p-Values for pairwise student t-test of comparisons: NRGRank vs Vina 2.15×10^{-2}, Vina vs Dock 7.45×10^{-2}, NRGRank vs Dock 2.36×10^{-3}, NRGRank vs Glide 1.32×10^{-8}, Glide vs Dock 5.08×10^{-3}. NRGRank has better IF1 values for 13, 10 and 5 targets in holo form compared to AutoDock Vina, DOCK 3.7 and Glide respectively.

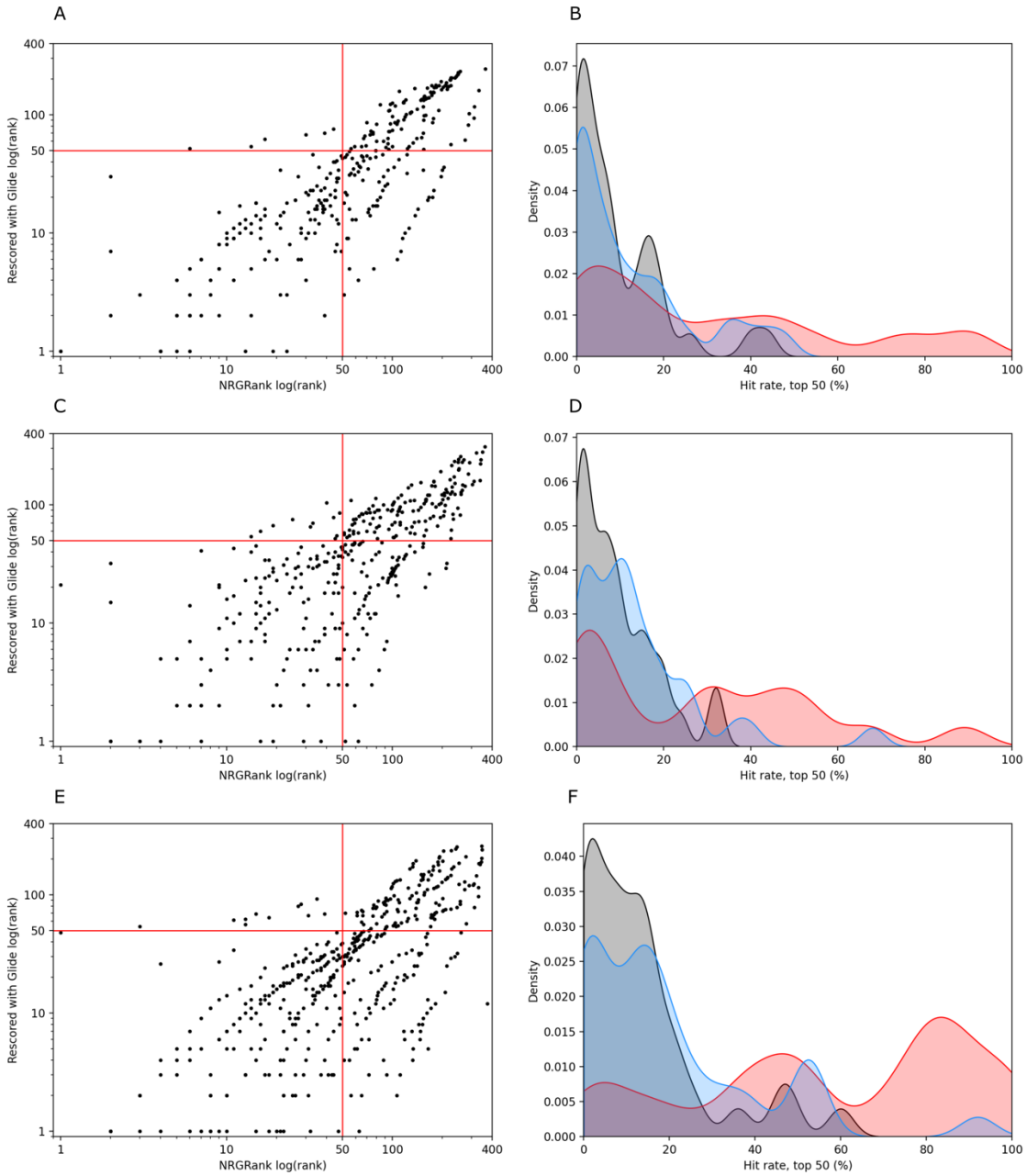


Figure 4. Effect of re-scoring NRGRank hits with Glide. For each of 32 targets in DUD-E with at least 50 true binders, we re-scored the top 1% ranked molecules (both binders and decoys) by re-ordering the molecules in the relative order they appear in the published Glide results and determined how many molecules remain, are added or removed from the top 50. Quadrant 1 in panels A, C & E represent binders with rank above 50 before or after re-scoring. Quadrant 3 represent molecules that were in the top 50 predictions and remain after re-scoring. Quadrant 2 represents binders that that were removed from the top 50 upon re-scoring. Lastly, in quadrant 4, are binders that were added to the top 50 upon re-scoring. Panels B, D & F show the hit rate distributions.

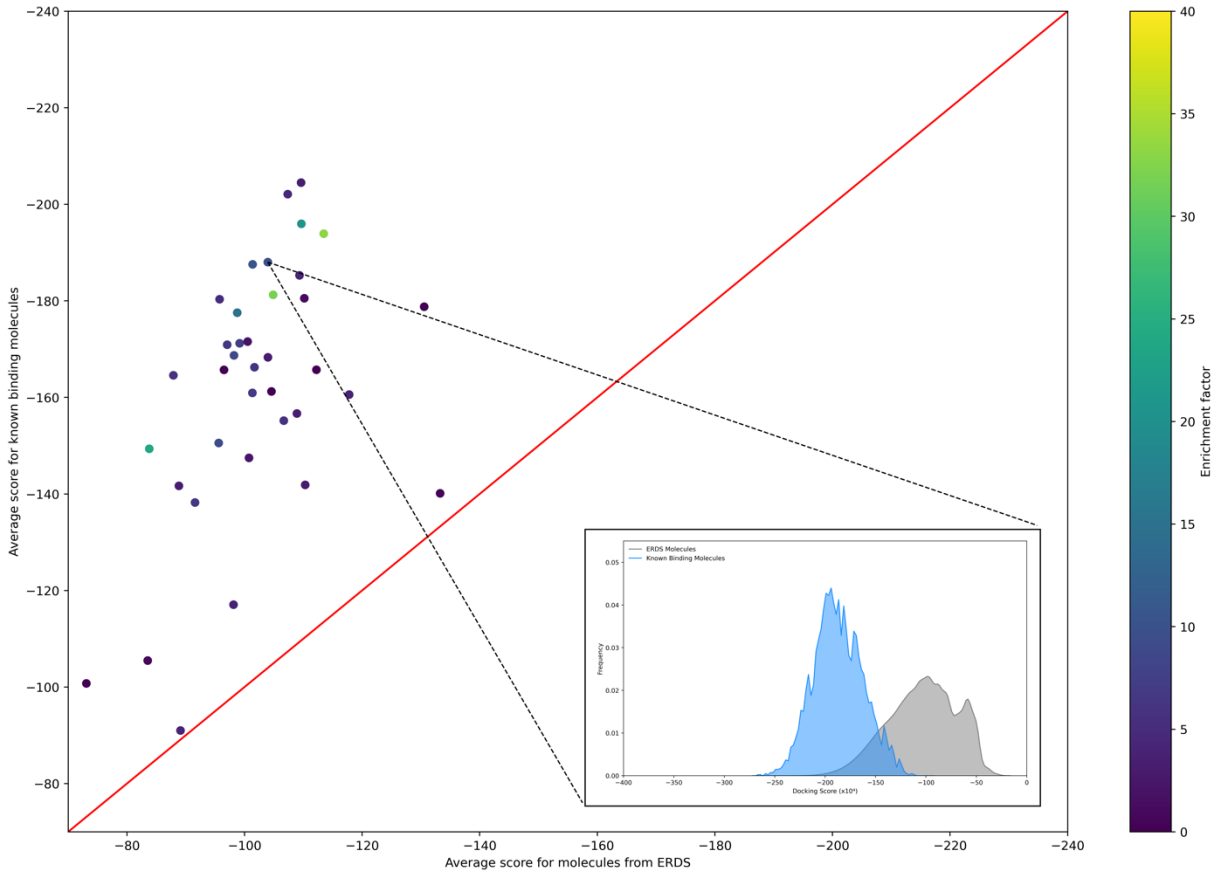


Figure 5: Ranking ERDS (Enamine REAL Diversity subset) on 37 targets of the DUD-E dataset. Average docking score for known binders and the average for the ERDS compounds for each target (the more negative the better the score is). The inset shows the BRAF protein kinase, with 4048 binders in DUD-E (blue distribution inset). The list of ERDS and known binder distribution averages, EF1 and number of known binders can be found in Supplementary Table 2.

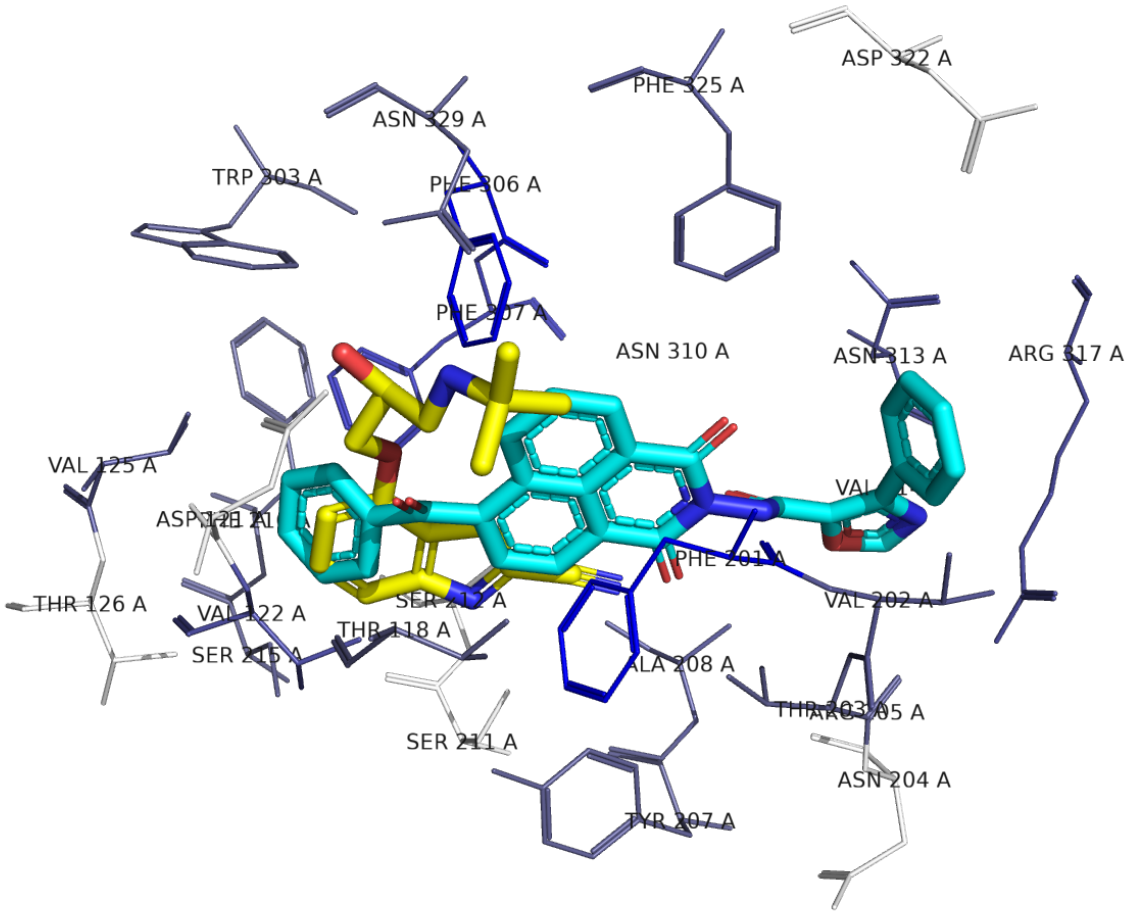


Figure 6: FlexAID docking pose of ERDS compound S_652 on adrb1. The ERDS compound s_22____22609224____10640652 (S_652 in cyan), found as the top-ranking ERDS compound by NRGRank, was docked with FlexAID. The top scoring pose shares several moieties with the compound P32 (cyanopindolol in yellow) in the PDB structure 2VT4 used as a target. Binding site side-chains are colored in a white to blue scale. The brighter the shade of blue, the more that residue contributes to the interaction with S_652.

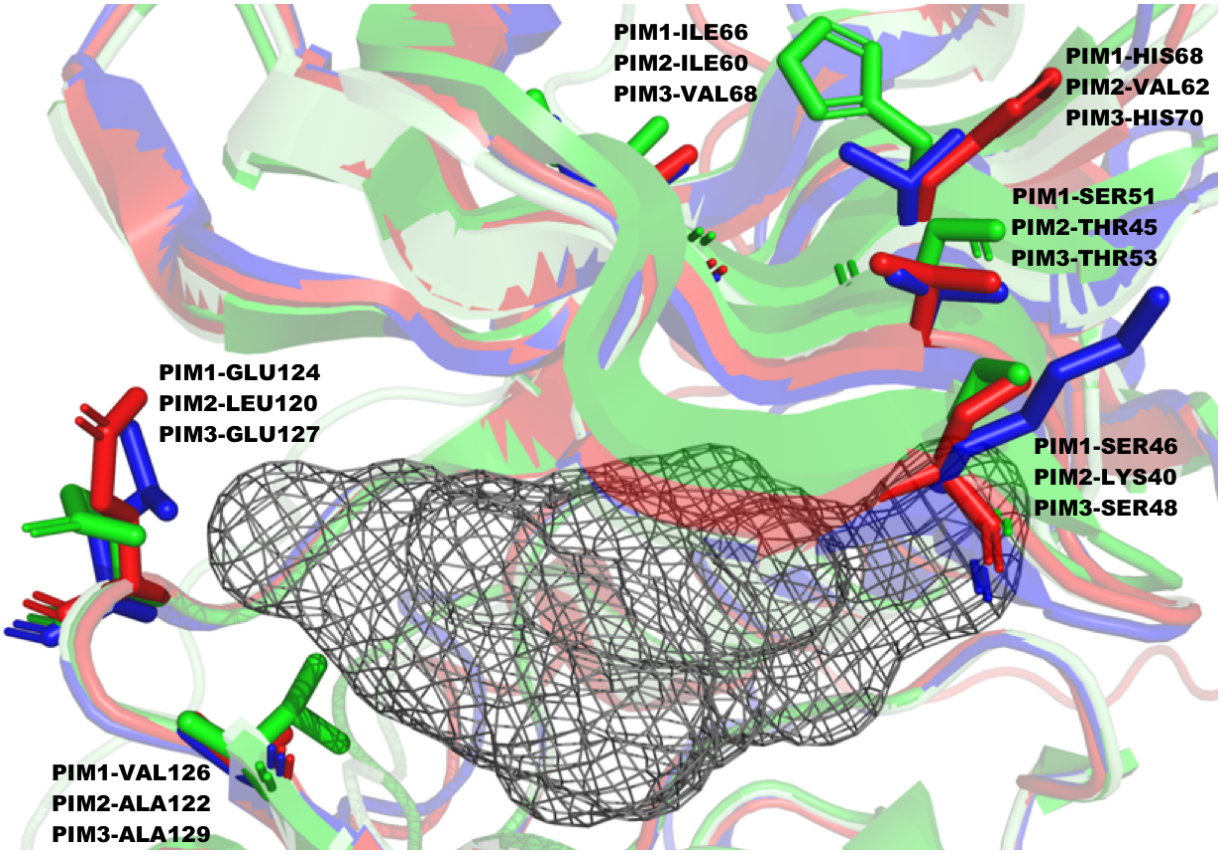


Figure 7: Binding-site differences between PIM1, PIM2 and PIM3. The C-terminal domain is depicted at the top. There are 6 positions with differences between PIM1 (green), PIM2 (blue) and PIM3 (red) on residues lining the ATP binding-site cavity (gray). The four mutated residues on the C-terminal domain have their side-chains facing away from the ATP binding-site. Only the two mutated positions on the hinge region between the N and C-terminal domains, have their side chains facing the cavity. Of these, only one position have a change in amino-acid type. The second position maintains a hydrophobic residue. The structures used for PIM1, 2 and 3 are 1XWS, AF-Q9P1W9-F1-model_v4 and AF-Q86V86-F1-v4 respectively.

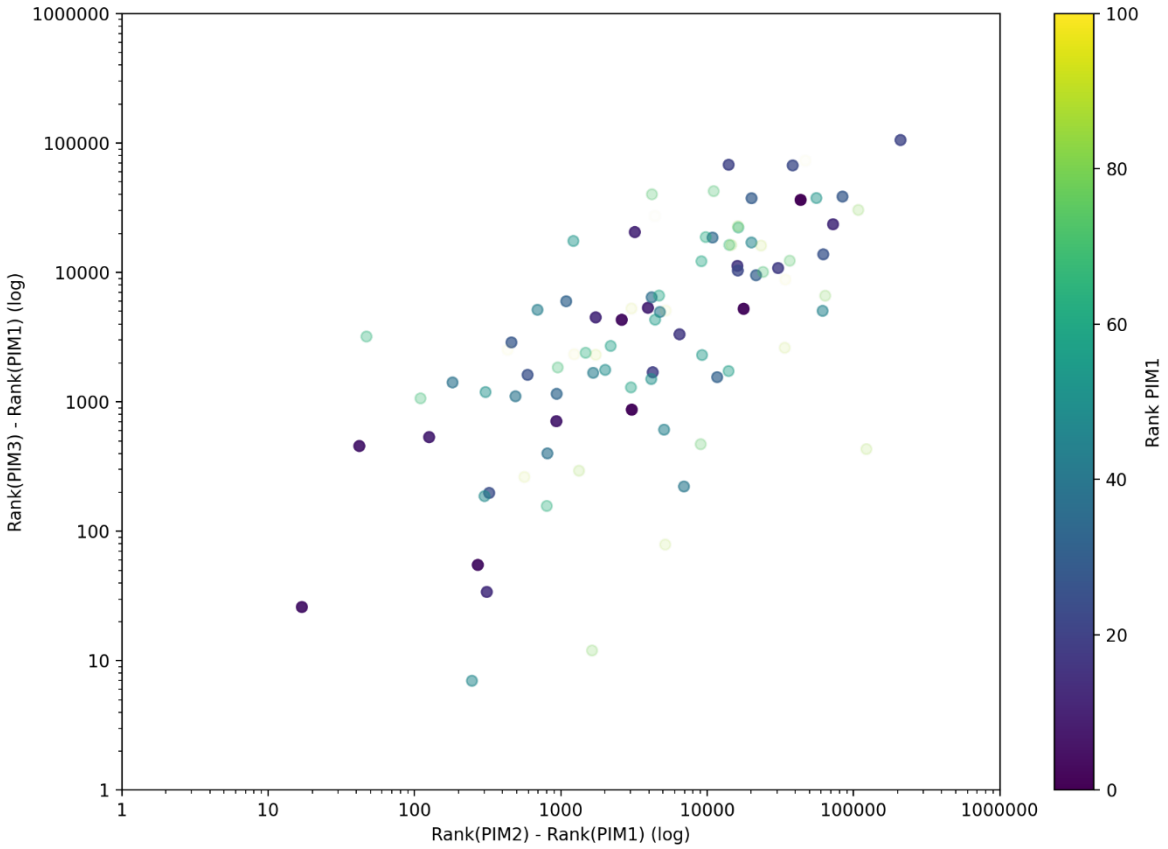
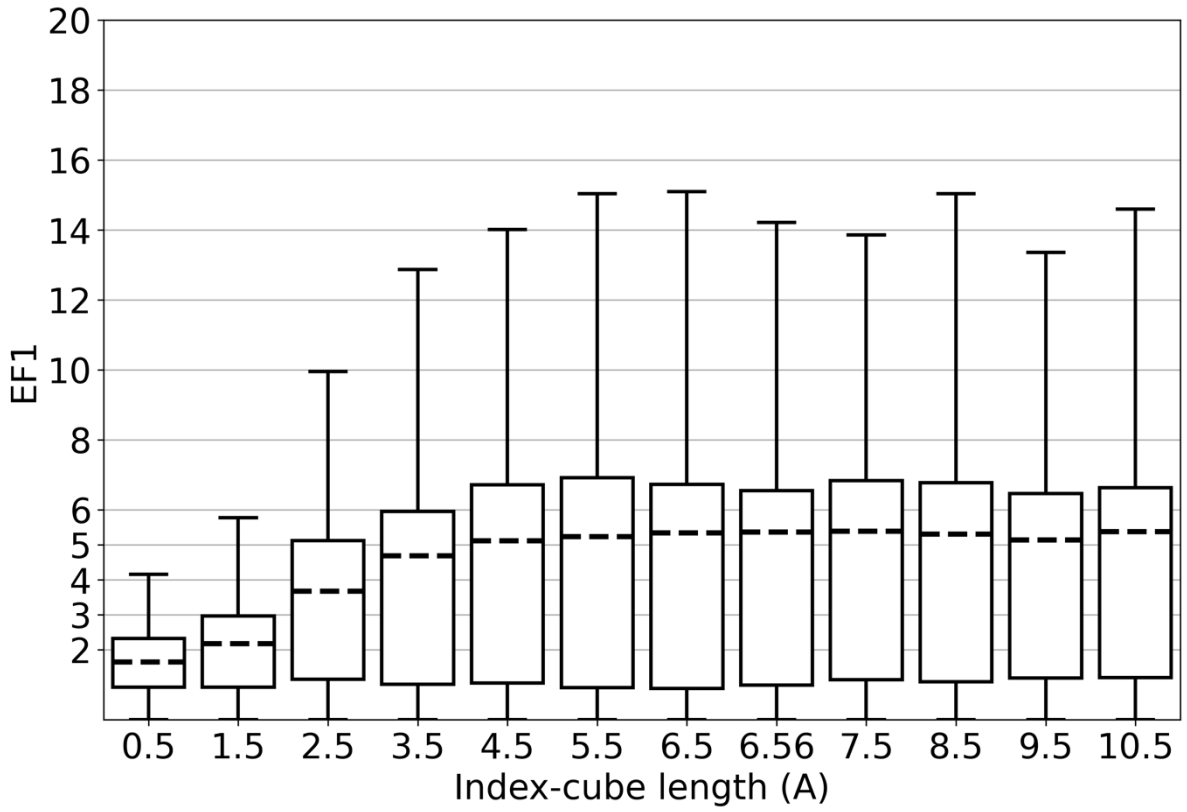
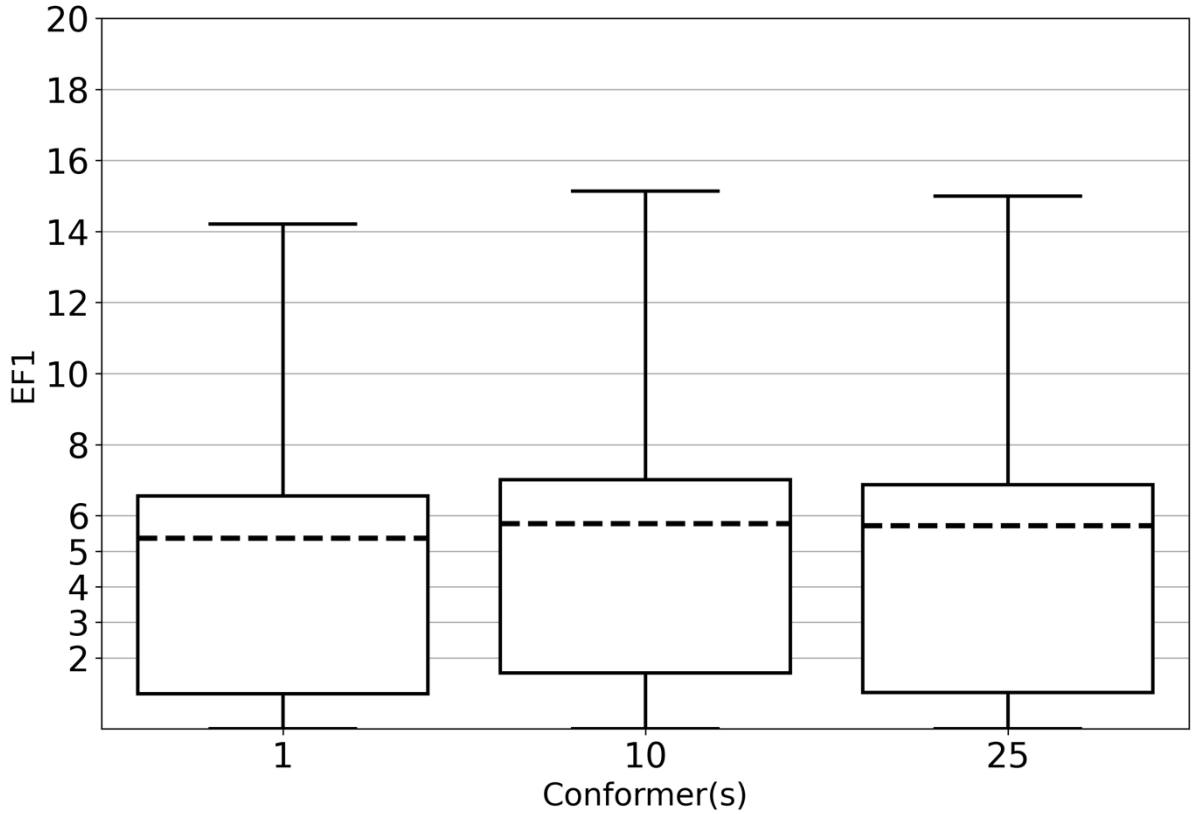


Figure 8: Rank comparison on PIM2 and PIM3 for the top 100 molecules ranked for PIM1. Among molecules in the top 100, a large number have poor ranks (10,000 and above) on PIM2 and PIM3.

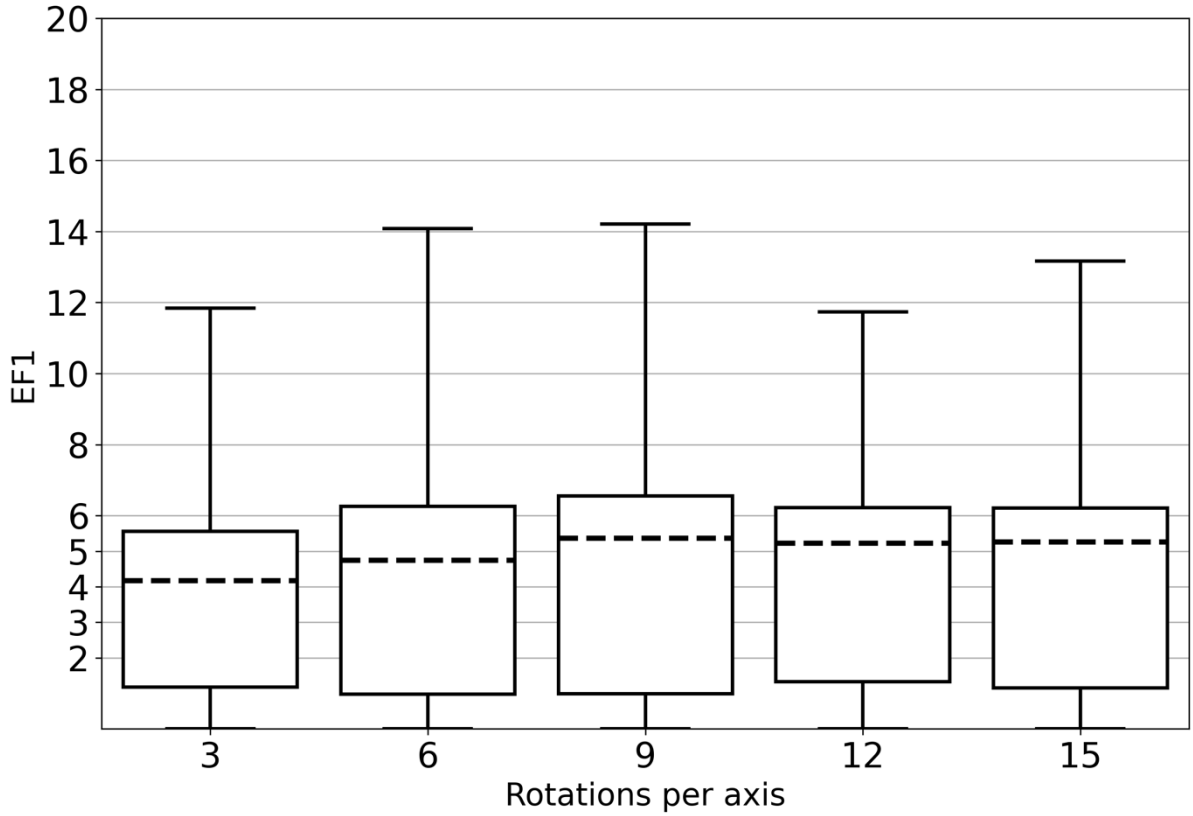
Supplementary Figures



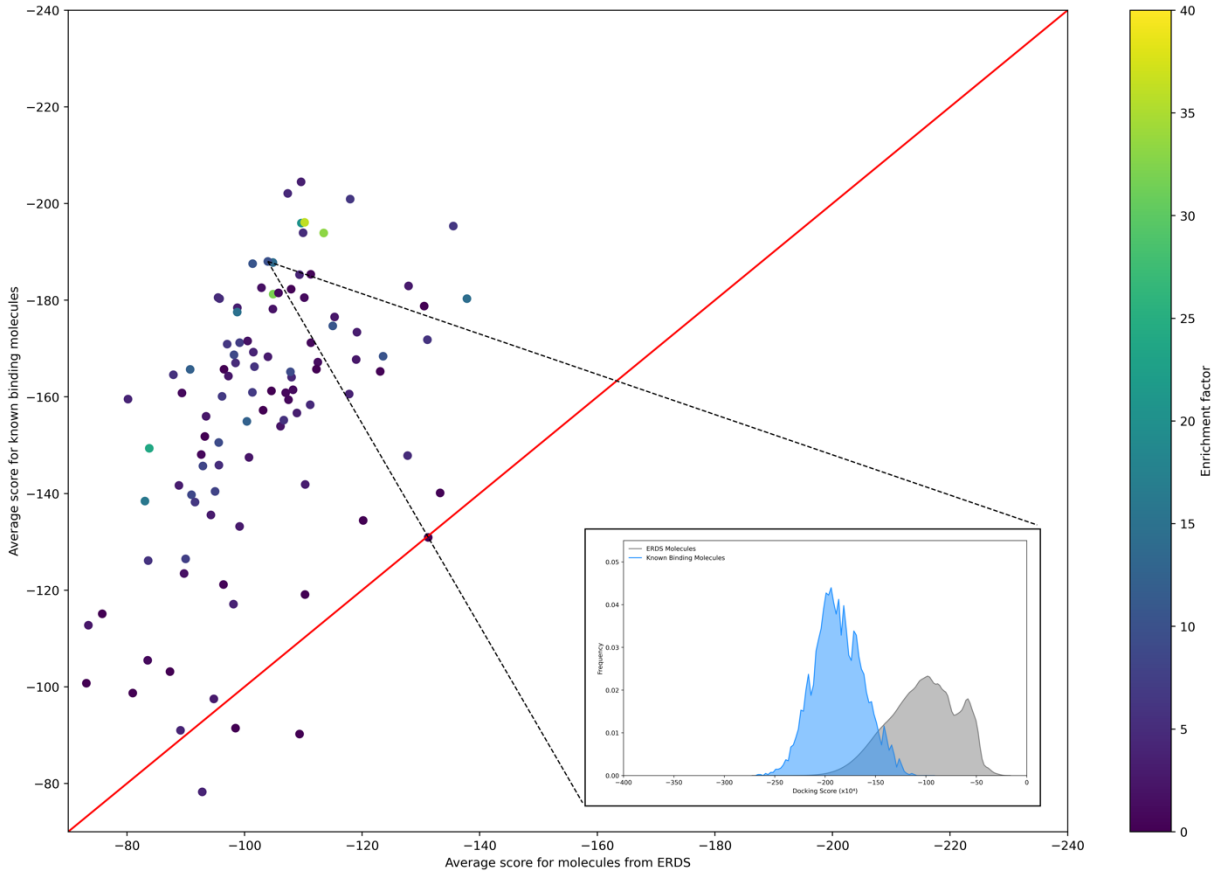
Supplementary Figure S1. Effect of index-cube length on EF1.



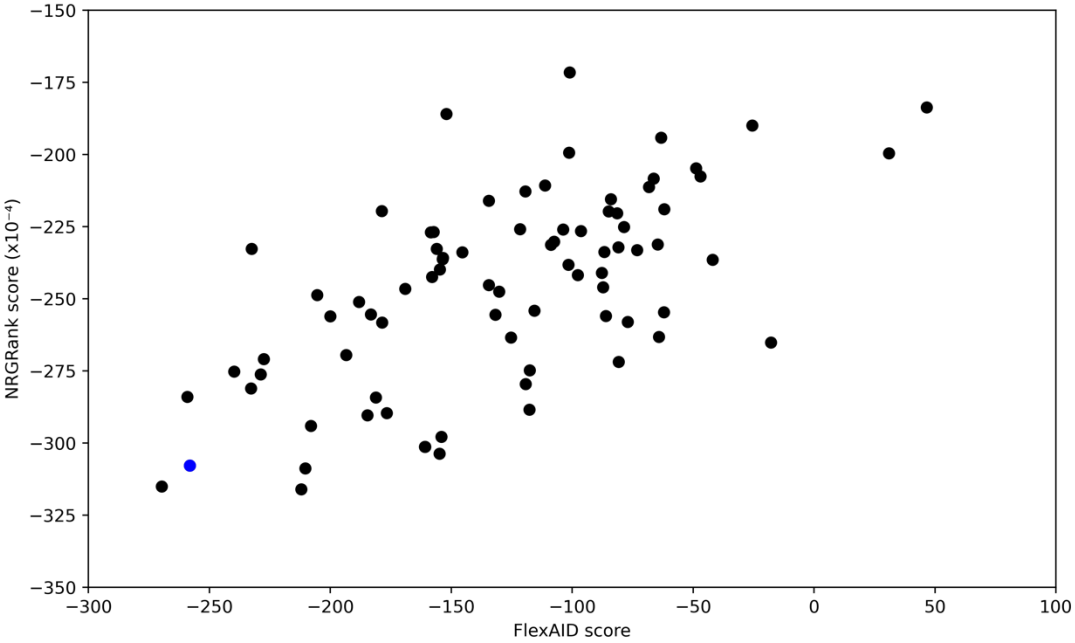
Supplementary Figure S2. Effect of number of conformers used on EF1.



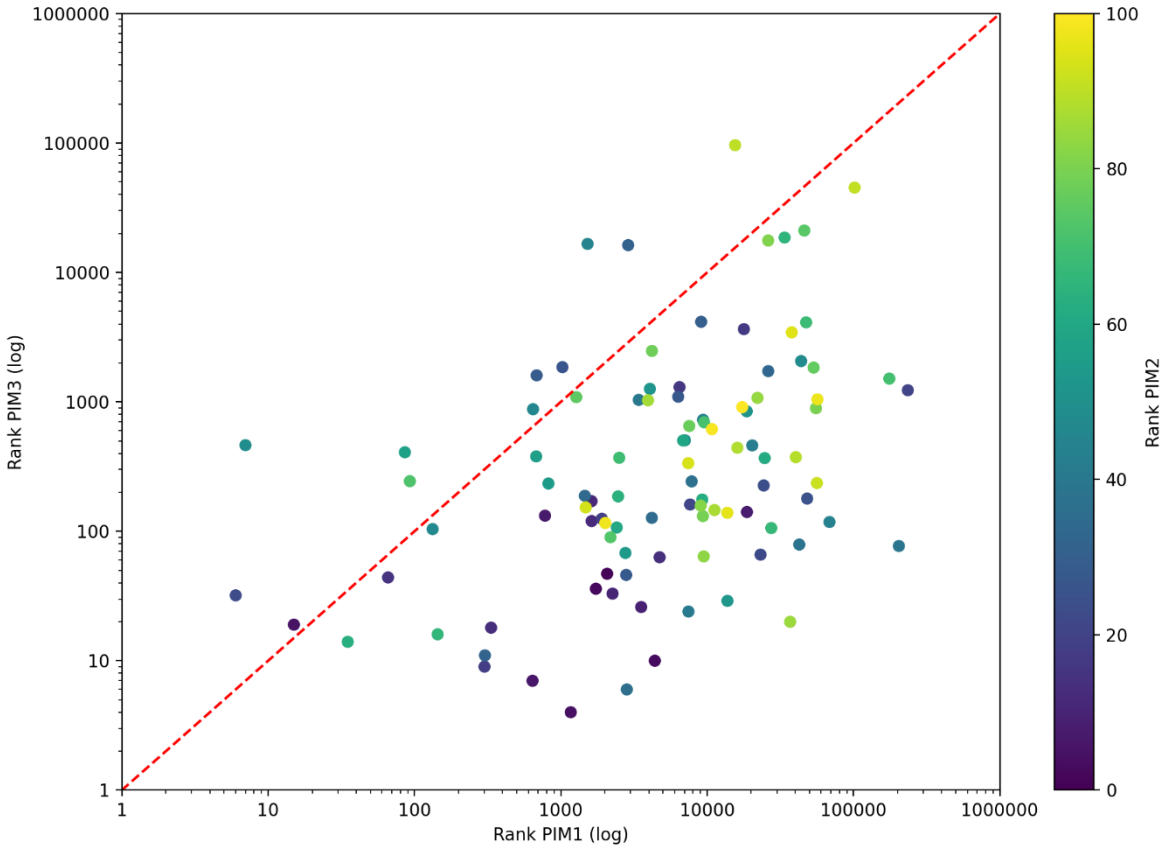
Supplementary Figure S3. Effect of number of rotations per axis on EF1.



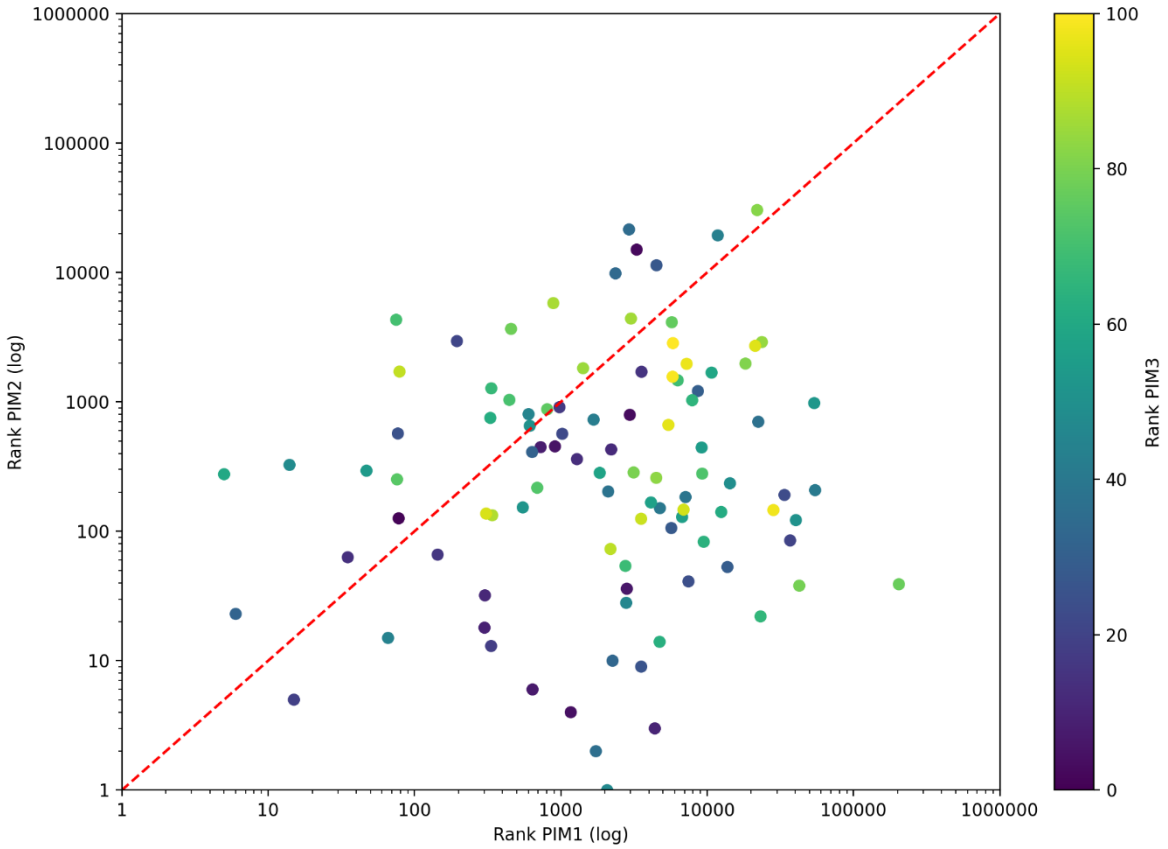
Supplementary Figure S4. Docking ERDS (Enamine REAL Diversity subset) on 102 targets of the DUD-E dataset. Average docking score for known binders and the average for the ERDS compounds for each target (the more negative the better the score is). The list of ERDS and known binder distribution averages, EF1 and number of known binders can be found in Supplementary Table 2.



Supplementary Figure S5. Comparison of the FlexAID and NRGRank scores for molecule S_652 on 101 DUD-E targets. Target adrb1 is shown in blue. The data shows a Pearson's correlation coefficient 0.62.



Supplementary Figure S6: Rank comparison on PIM1 and PIM3 for the top 100 molecules ranked for PIM2. Among molecules in the top 100, a number have poor ranks (10,000 and above) particularly for PIM1 but also for PIM3, in agreement with the fact that PIM2 has less mutations with respect to PIM3 than to PIM1.



Supplementary Figure S7: Rank comparison on PIM1 and PIM2 for the top 100 molecules ranked for PIM3. Among molecules in the top 100, a number have poor ranks (10,000 and above) particularly for PIM1 but also some for PIM2. While PIM3 has 4 mutations compared to PIM2 and only 3 compared to PIM1, it has only one mutation for residues whose side-chains face the binding-site cavity. The cavity side-chain facing mutation from PIM3 to PIM1 is an ALA to VAL, a more conservative mutation than that between PIM3 and PIM2 where the alteration is a GLU to LEU. This is reflected on the fact that ranks for the top PIM3 molecules on PIM1 is shifted to the left when compared to the ranks of the top 100 molecules for PIM2 on PIM1 (Supplementary Figure S5).