

SimBioSys®

Technical Note

## Case Study: Virtual Screening

### Reference

- [1] Stahl, M.; Rarey, M. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.* **2001**, *44*, 1035-1042.
- [2] Schultz-Gasch, T.; Stahl, M. Binding site characteristics in structure-based virtual screening: evaluation of current docking tools. *J. Mol. Model.* **2003**, *9*(1), 47-57.

### Study Overview

Virtual screening is a particularly demanding task for molecular docking methods. In this scenario, many compounds are docked into the binding pocket of a receptor. From the, optionally many, generated poses, only the top scored one is used to represent each compound, and the scores are used to rank order the compounds based upon their predicted binding affinity. In the original study [1] seven pharmaceutically relevant targets were chosen that differ significantly in their binding sites. The objective of the study was to evaluate the performance of four different scoring functions in virtual screening using poses that were generated by a single docking software, FlexX. Each scoring function was used to pick the best pose for each compound and to rank order the docked molecules, and the enrichment factor was calculated. None of the originally tested scoring functions was shown to be robust enough to be considered generically successful, although a combination of FlexX and PLP, called ScreenScore, has proven to outperform the other scoring functions. In reference [2] the performance of additional two docking packages, Glide and FRED, was evaluated with the same test set and using various scoring functions. We have tested eHiTS against the same targets, and with the same sets of actives and decoys and have shown that eHiTS' enrichment results outperform the other tested docking programs and scoring functions in all cases.

### Methods

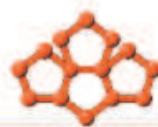
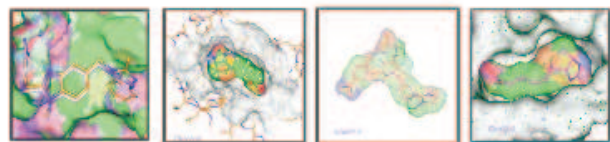
For six out of the seven original targets publicly available structures are available. We have screened compounds for cyclooxygenase-2 (Cox2), estrogen receptor, P38 MAP kinase, thrombin, gelatinase-A, and neuraminidase. Both active molecules, and the decoy set were provided by the authors of the original study. eHiTS 2009 was used both to generate the poses and to score them.

eHiTS does not require any preparation of the target or the docked ligands, making the software very easy to use compared to other commercial applications. The command line used to initiate the screening of each target by eHiTS was:

```
ehits.sh -complex PDB-CODE.pdb -ligand LIG_DB.sdf -accuracy 1 -toprank 1 -allowflat
```

where PDB-CODE.pdb is the target's structure as downloaded from the PDB, and LIG\_DB.sdf is the sdf file containing all the actives and decoys. Accuracy 1 should in general be the level of choice in a virtual screening run where large libraries of compounds are docked, and computational efficiency is a requirement. The **-toprank** flag dictates how many poses are sent to be output, and **-allowflat** indicates that planar 3D structures should be allowed.

All the active molecules docked successfully in their respective receptor's binding pocket. For each target few



SimBioSys®

## Technical Note

molecules from the decoy set failed to dock, but in any case the failure rate of between 0.4% and 2.7% was much lower than the 6% rate reported in the paper. Compounds which failed to dock were appended to the list in arbitrary order.

### Data

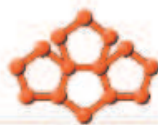
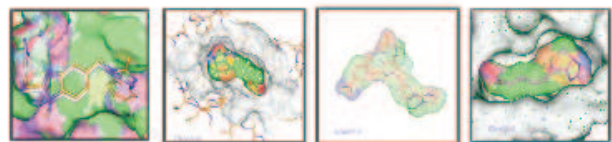
The authors provided a library of 7397 decoy molecules taken from the World Drug Index. Those compounds were selected after filtering the database for structures of relevant molecular weight, and represent a diverse set in terms of Daylight fingerprints.

The PDB codes and the number of actives supplied by the authors is given in the table below.

Target	PDB code	number of actives
COX-2	1CX2	128
estrogen receptor	1ERR	55
P38 MAP kinase	1OVE	25
Thrombin	1DWD	67
Gelatinase-A	1MMP	43
Neuraminidase	1IVF	17

### Results

Figure 1 shows the enrichment of inhibitors achieved for each of the seven targets. The enrichment is presented as an accumulated percentage of inhibitors contained in the top X% of the ranked database. The percentage of the database is given in a logarithmic scale in correspondence with figure 1 of the paper, to expand the [1%:10%] interval which is more significant to drug discovery scenarios



SimBioSys®

## Technical Note

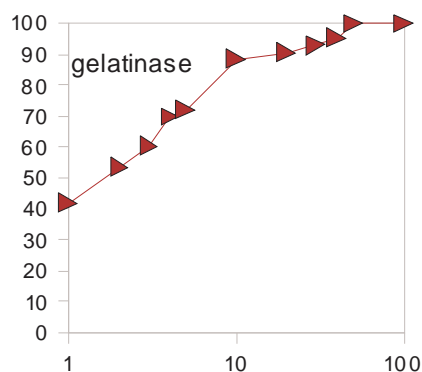
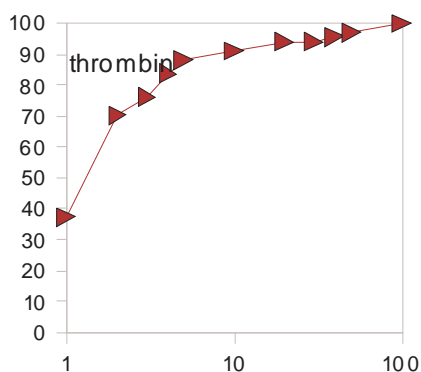
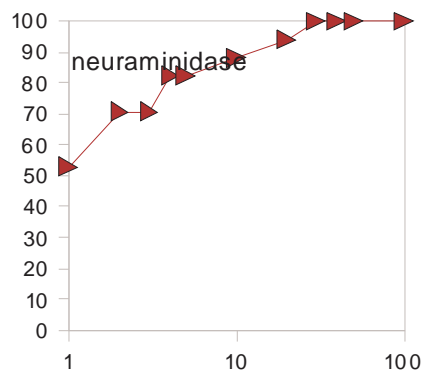
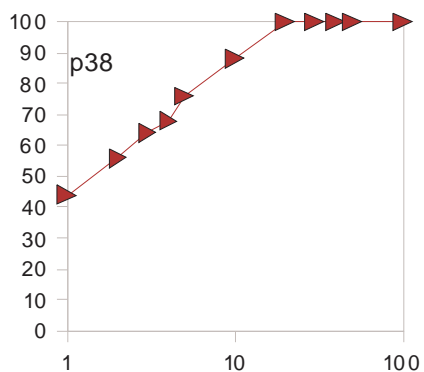
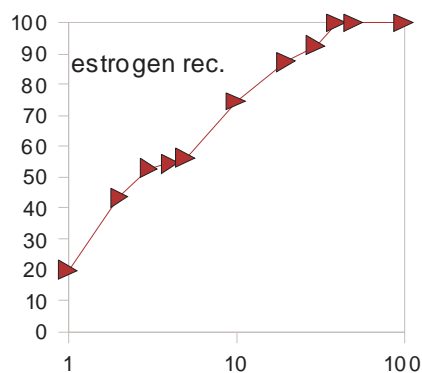
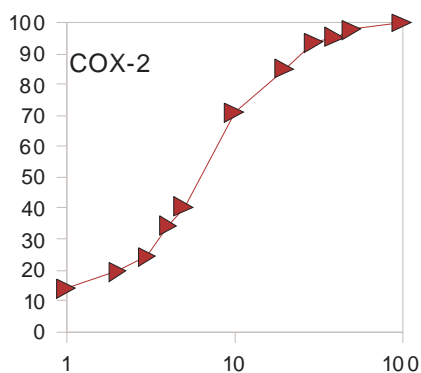
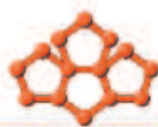
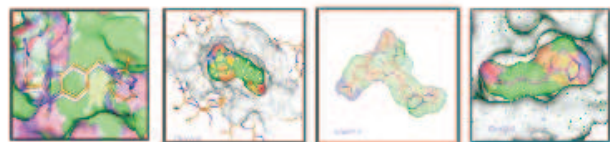


Figure 1. Enrichment of inhibitors for the six targets: percentage of active compounds as a function of the percentage of the ranked database.



SimBioSys®

Technical Note

Figure 2 is a reproduction of figure 3 in the original study, with the eHiTS results included. It demonstrates the percentage of inhibitors recovered at the top 2%, 5% and 10% of the screened database. eHiTS clearly outperforms the reported enrichments for all targets, including that achieved with the optimized ScreenScore.

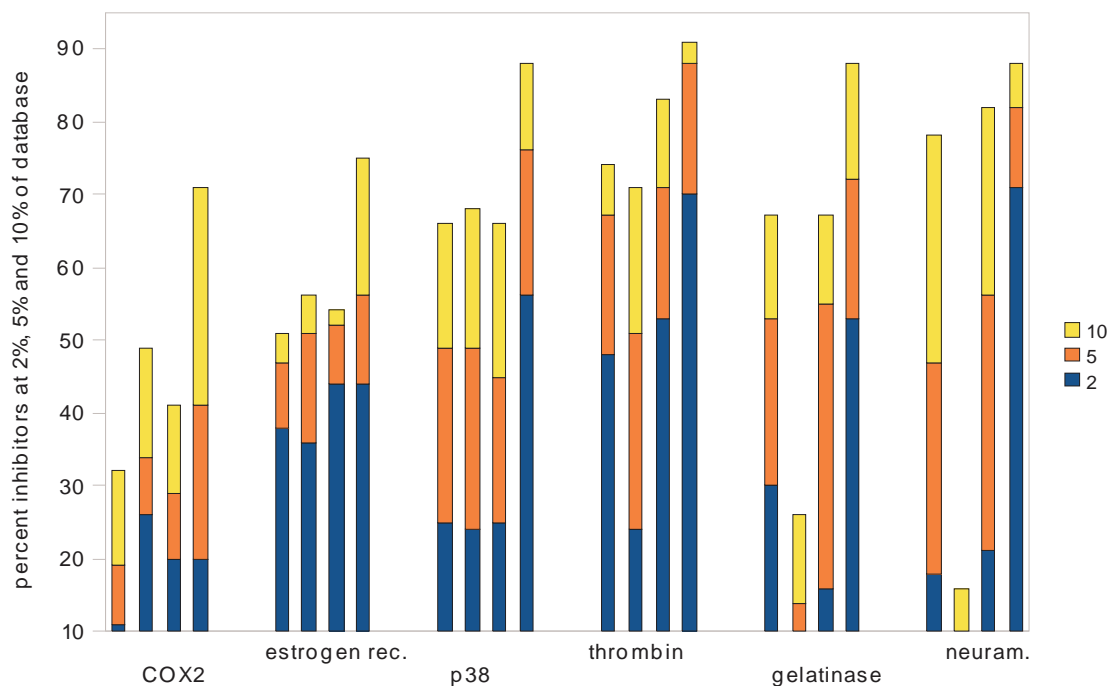


Figure 2. Comparison of enrichments for the six targets. eHiTS 2009 (right most in each target) compared to FlexX results achieved with FlexX, PLP and ScreenScore scoring functions.

In the following figure, eHiTS enrichments are compared to the best enrichments achieved with each of the tested docking programs, as shown in Figure 2 of reference [2]. Note that none of the other tested software applications achieves its best performance consistently with the same scoring function. This comparison is therefore a particularly strong evidence for the power of eHiTS docking and scoring capabilities. eHiTS clearly outperforms the other tools in 4 cases, whereas it is in second places for the other two cases.

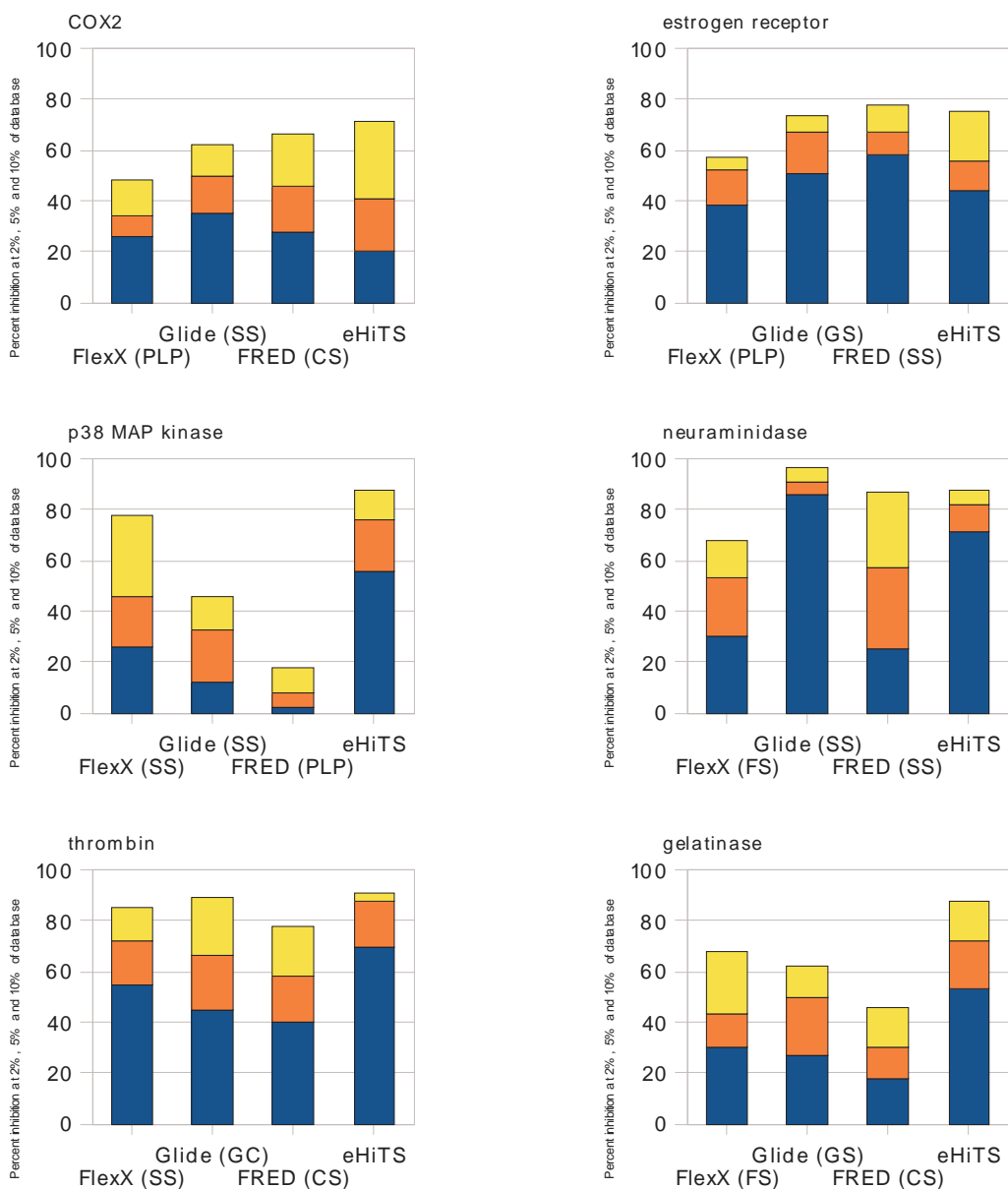
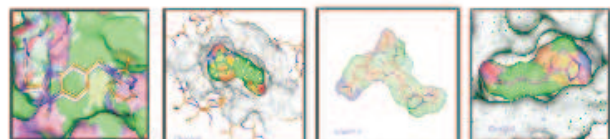


Figure 3. Enrichment of inhibitors for the six targets. Comparison of eHiTS 9.0 to the best enrichment achieved for FlexX, Glide and FRED in reference [2].

eHiTS 2009 was therefore shown to be an outstanding tool for virtual screening, performing very well for a diverse set of targets with an exceptional ease of use.