Figures

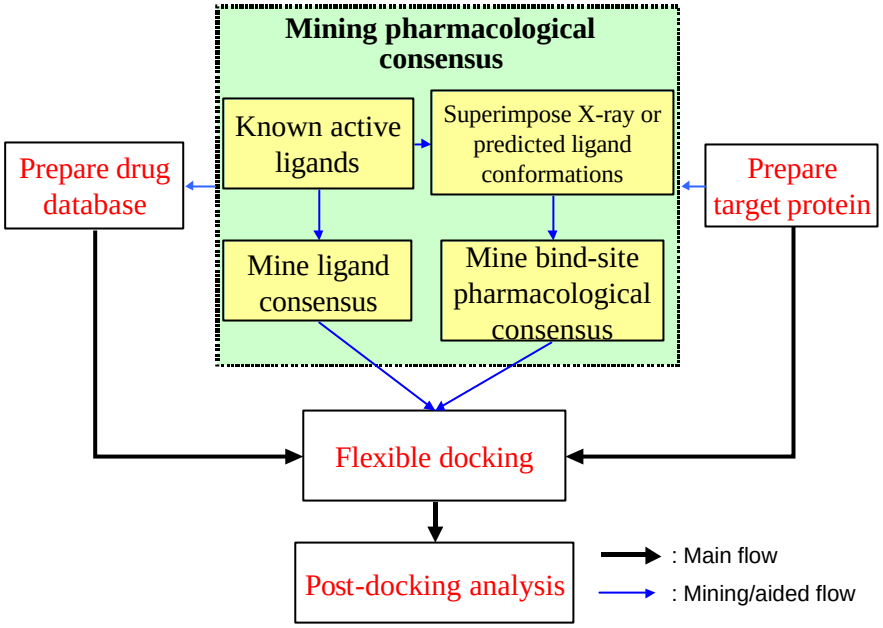


Figure 1. The main steps of GEMDOCK for virtual database screening, including the target protein and compound database preparation, flexible docking, and post-docking analysis. GEMDOCK mines a pharmacological consensus from the target protein and known active ligands when available.

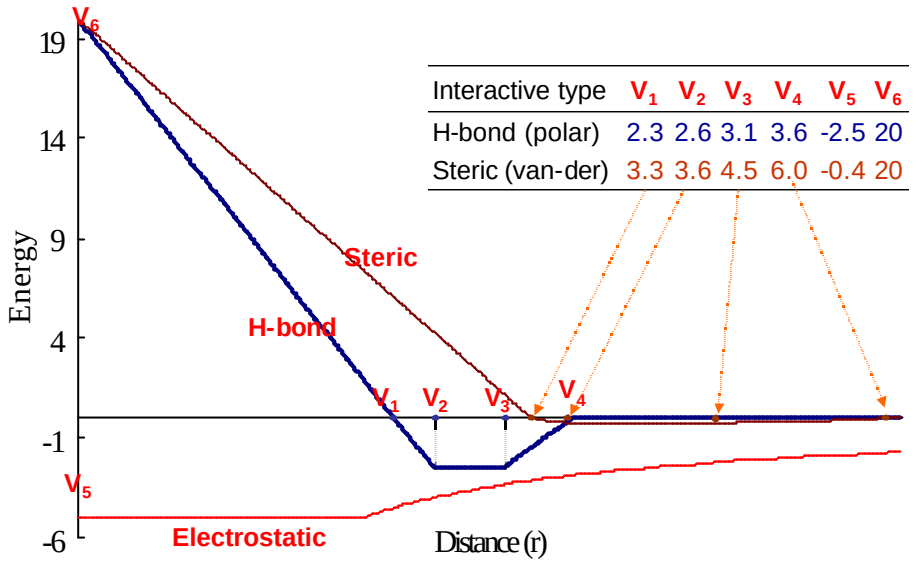


Figure 2. The linear energy function of pair-wise atoms for steric interactions (light line), hydrogen bonds (bold line), and electrostatic potential in GEMDOCK.

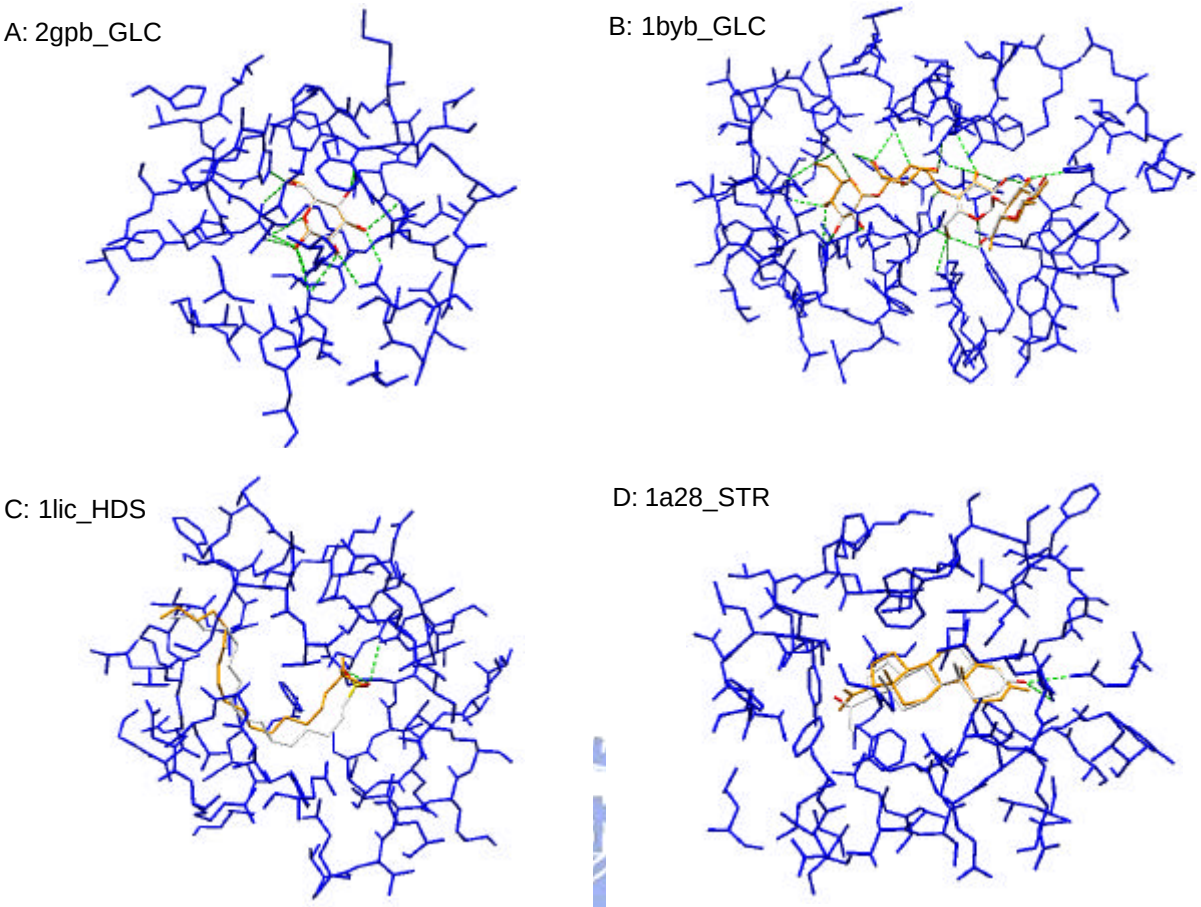


Figure 3. GEMDOCK results for four typical acceptable complexes (i.e., the RMSD value < 2.0 Å). The RMSD values of these complexes were less than 1.0 Å Most of the docking ligand groups (orange) were identical to the crystal ligand structures (CPK). The green dotted lines represent hydrogen bonds and cavities show in blue color.

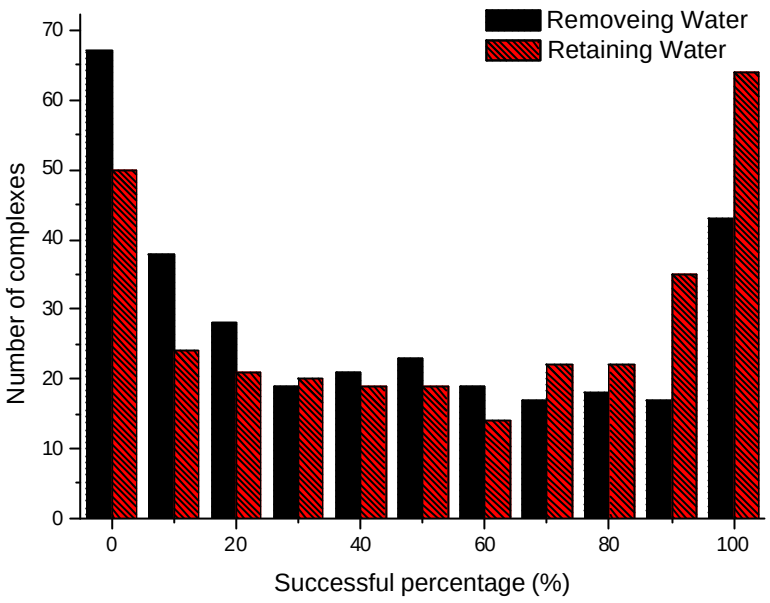


Figure 4. The successful percentages of GEMDOCK for retaining (red bar) and removing (black bar) structure water molecules in CCDC/Astex test set.

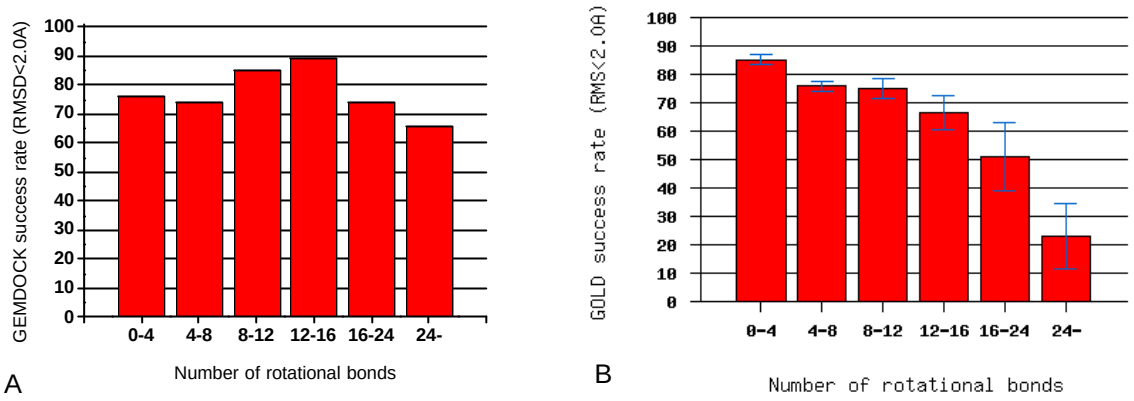


Figure 5. Comparison of successful rates between GEMDCOK and GOLD in different rotational bond levels. (A) Successful rates of GEMDOCK, (B) Successful rates of GOLD. This comparison shows GEMDOCK has stable performance (77% in average) in different rotational bond levels. Both GEMDCOK and GOLD use default settings respectively.

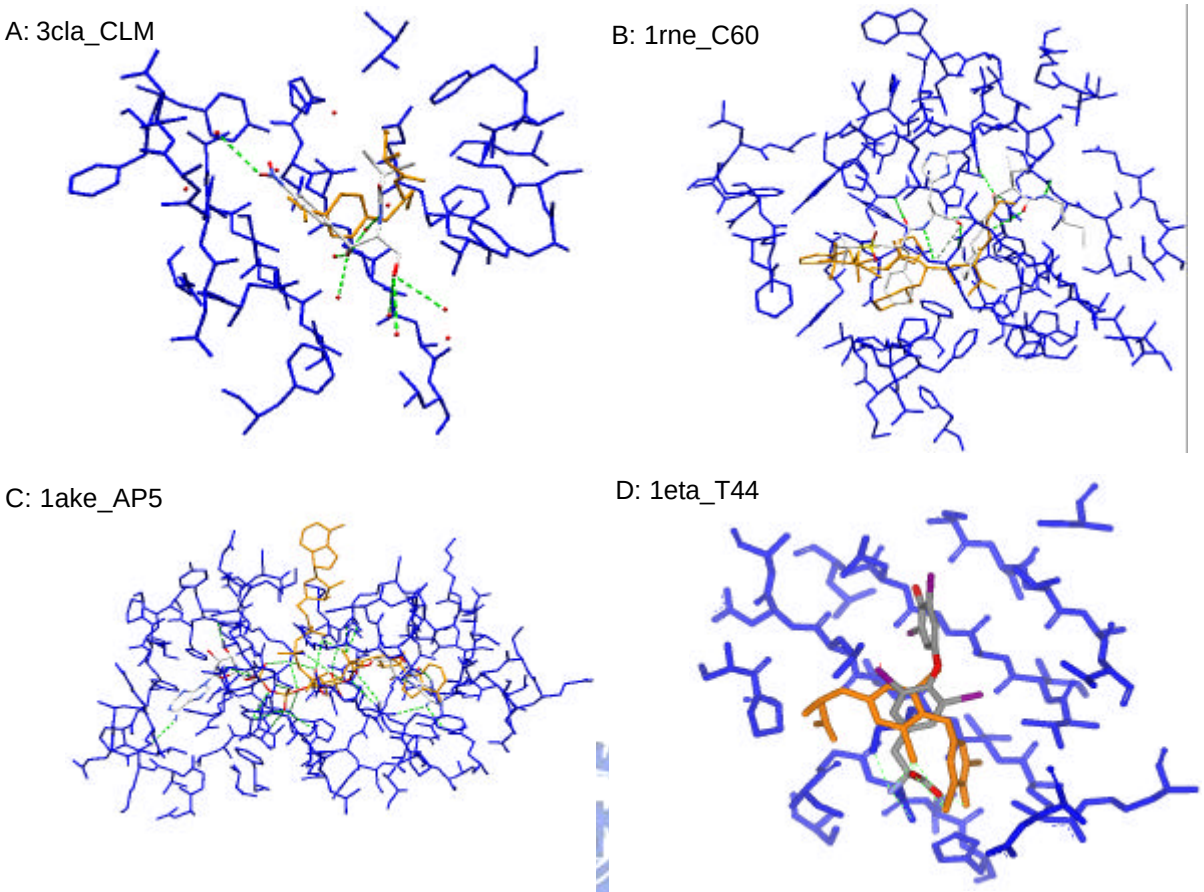
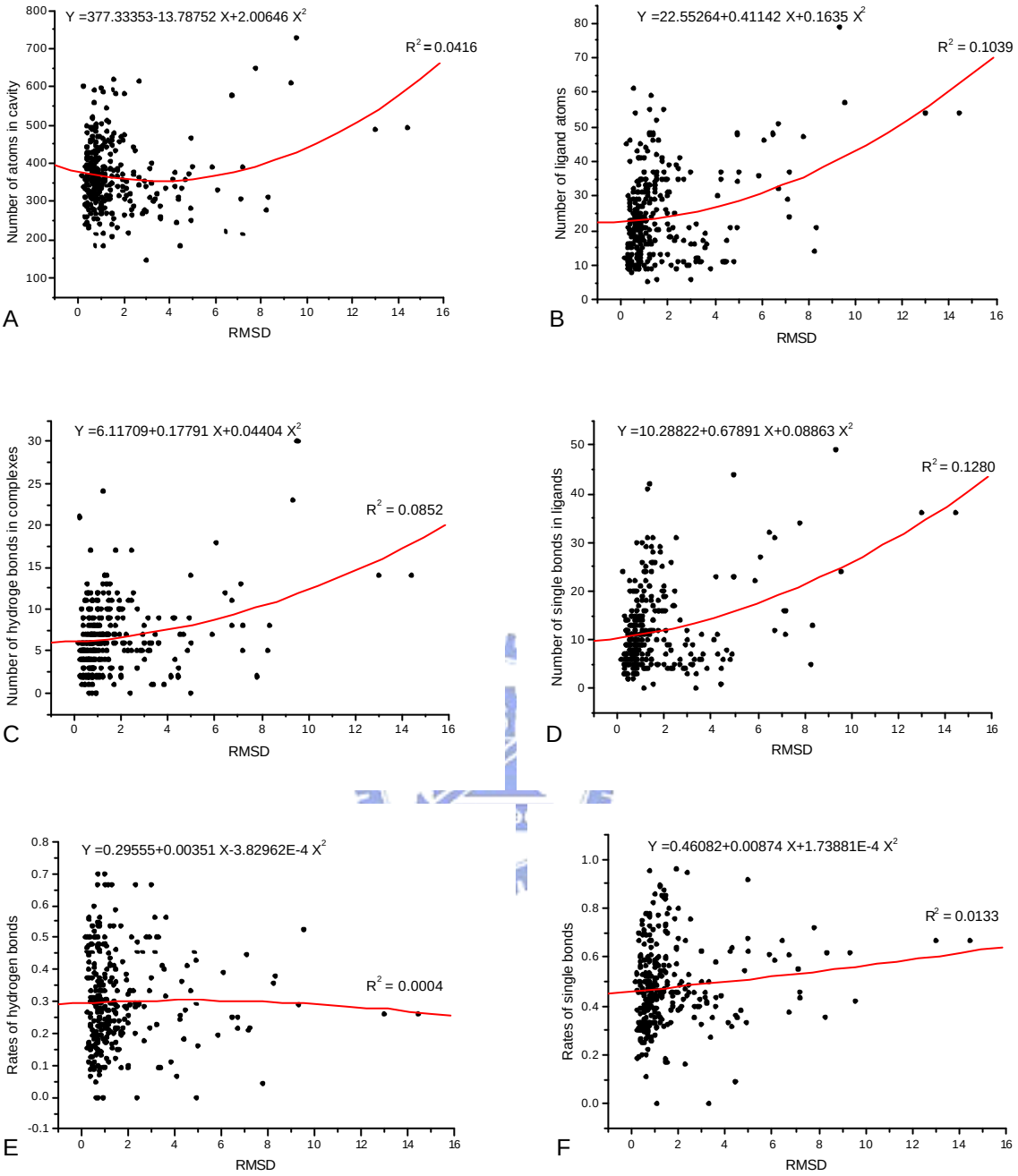


Figure 6. GEMDCOK results for four factors for unacceptable examples (i.e., RMSD value > 2.0 Å). (A) The critical structure water molecules are removed from binding site (3cla). (B) The ligand structure is highly flexible and large (1rne). (C) Significant clashes between protein and ligand atoms (1ake). (D) Specific protein-ligand interactions do not take into consideration in our scoring function (for example, the interaction I..O in 1eta). The docking ligand conformations are orange and the crystual ligand conformations are CPK. Hydrogen bonds are shown in green dotted lines and cavities are shown in blue.


Y =377.33353-13.78752 X+2.00646 X2
R2 = 0.0416
Number of atoms in cavity
RMSD
A
Y =22.55264+0.41142 X+0.1635 X2
R2 = 0.1039
Number of ligand atoms
RMSD
B
Y =6.11709+0.17791 X+0.04404 X2
R2 = 0.0852
Number of hydroge bonds in complexes
RMSD
C
Y =10.28822+0.67891 X+0.08863 X2
R2 = 0.1280
Number of single bonds in ligands
RMSD
D
Y =0.29555+0.00351 X-3.82962E-4 X2
R2 = 0.0004
Rates of hydrogen bonds
RMSD
E
Y =0.46082+0.00874 X+1.73881E-4 X2
R2 = 0.0133
Rates of single bonds
RMSD
F

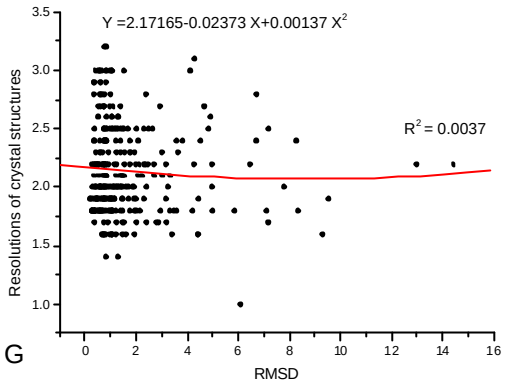


Figure 7. Affected factors analysis of GEMDOCK performance. (A) Cavity sizes, (B) Ligand sizes, (C) Number of hydrogen bonds in native binding, (D) Single bonds in flexible ligand conformations, (E) Ratio of hydrogen bonds, (F) Ratio of single bonds, (G) Resolutions of crystal structures.

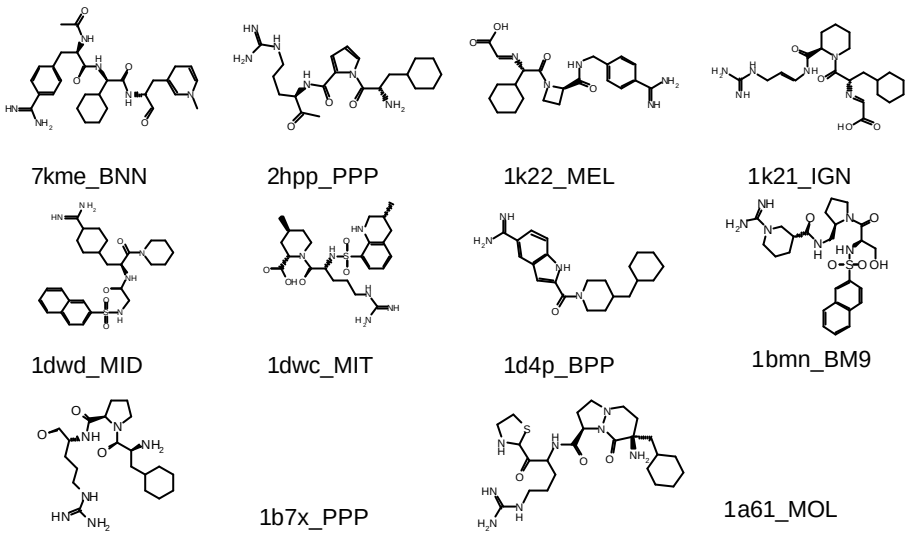


Figure 8. Ten known human a-thrombin ligands are docked against the thrombin complex (PDB code 1dwd) for evaluating docking accuracy and screening performance. Each ligand was denoted in four characters followed three characters. First four letters are PDB code and later three letters denote the ligand name.

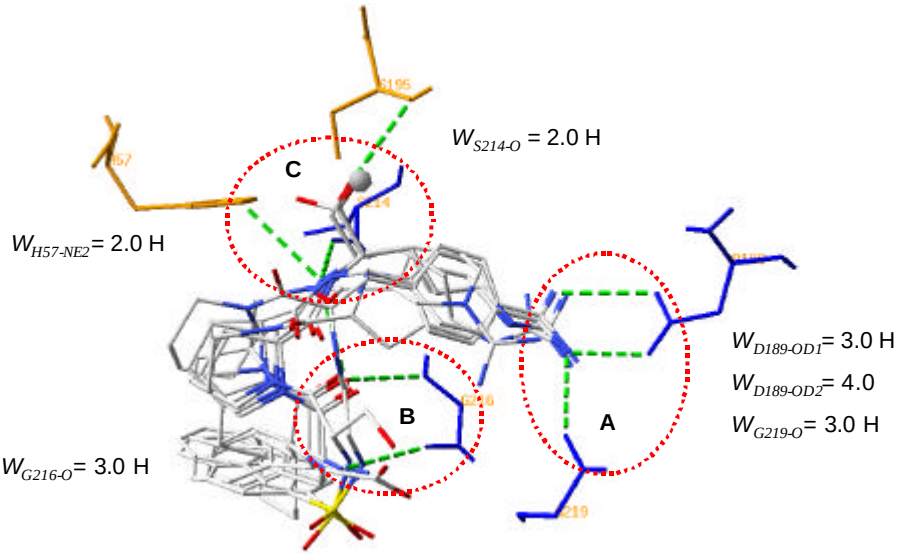


Figure 9. The binding-site pharmacological consensuses are identified by superimposing ten known human α-thrombin ligands against the reference protein 1dwd Six pharmacological preferences and interacitons are identified and circled as A (acetyl-p-amidinophenyl group), B (amino and ketone group) and C (amino group). The pharmacological weights are denoted in $W_{position}$ and the interaction types are labeled in H, V and none behind weight values. The dotted lines indicate hydrogen bonds.

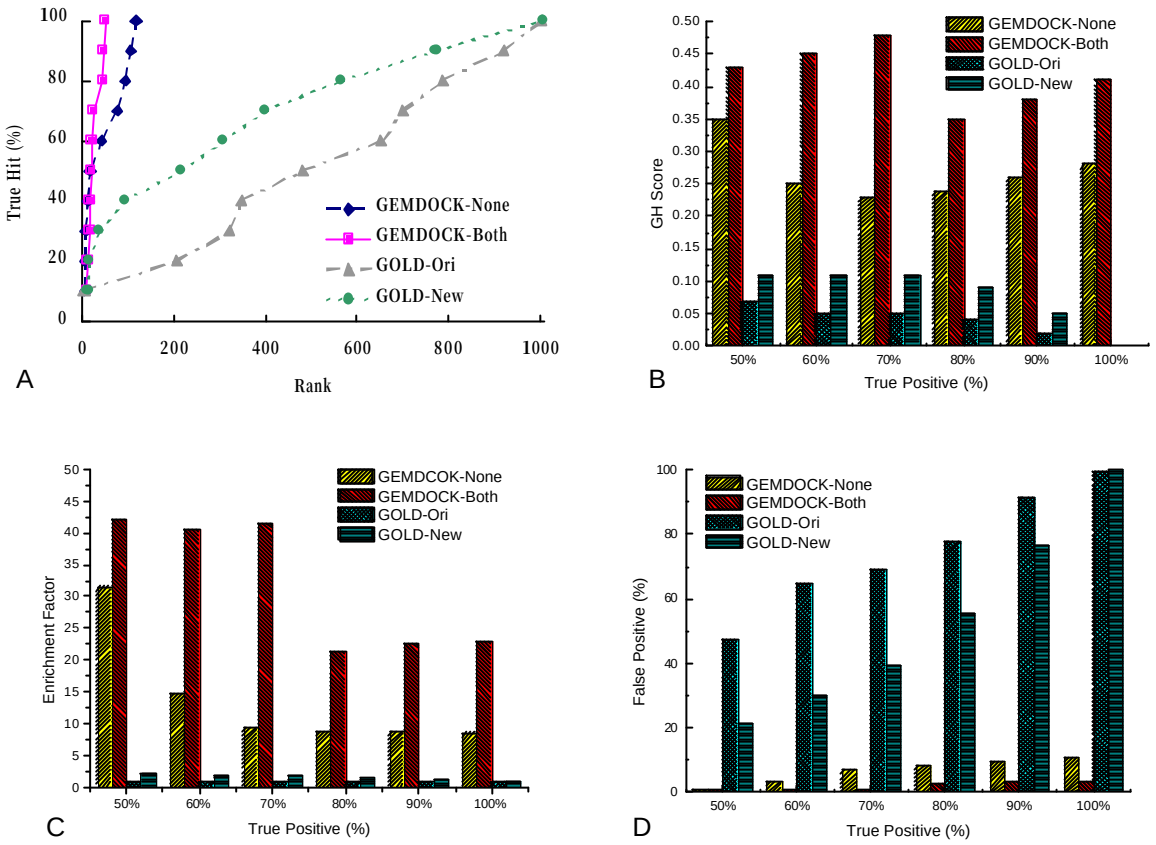


Figure 10. GEMDOCK screening accuracies of thrombin are assessed by (A) True hit, (B) GH scores, (C) Enrichment factor and (D) False positive rate against true positive rates ranging from 50% to 100%. The performances of GEMDOCK are superior when using both ligand preferences and receptor pharmacological interaction preferences.

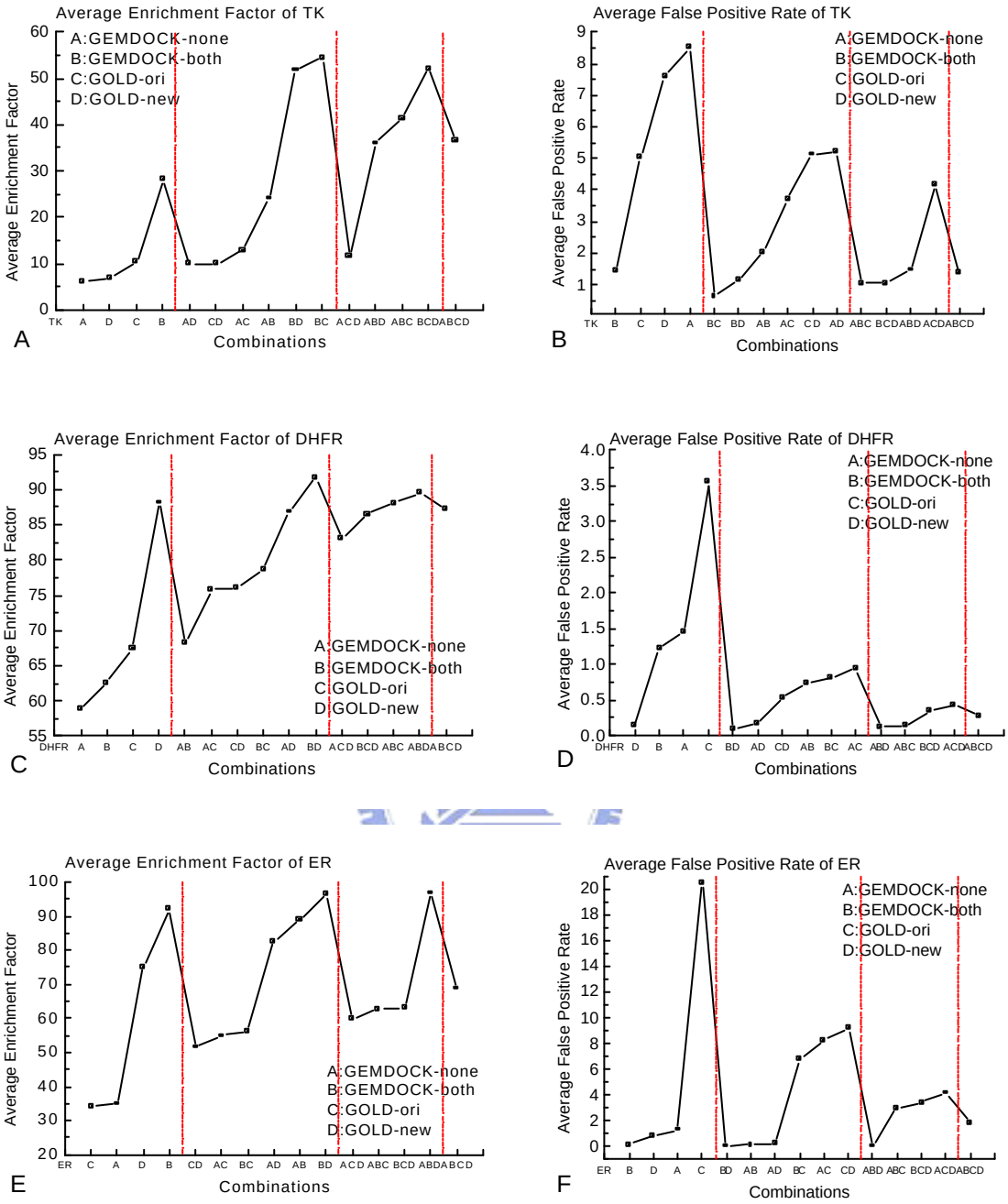


Figure 11. Average enrichment factor and average false positive rate of three virtual screening targets. (A) Average enrichment factor and (B) Average false positive rate of TK. (C) Average enrichment factor and (D) Average false positive rate of DHFR. (E) Average enrichment factor and (F) Average false positive rate of ER. The label in x-axis is the combination of rankings. The label of methods corresponds to table 10. (None: GEMDOCK without pharmacological preferences, both: GEMDOCK with pharmacological preferences, Ori: GOLD with GoldScore, and new: GOLD with recombinant GoldScore)

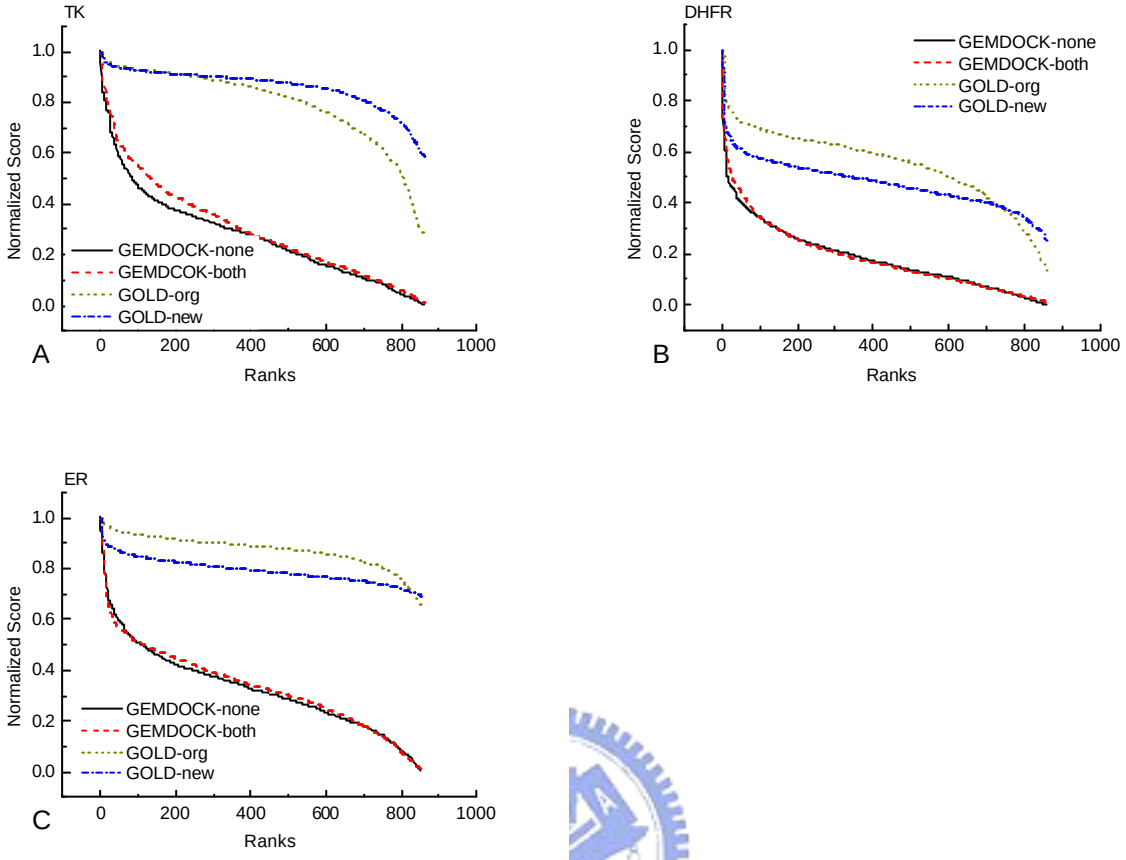


Figure 12. Ranks and score curves for three virtual screening targets. (A) TK, (B) DHFR, and (C) ER. Scoring of four methods has been normalized with equation 16. Scores of GOLD and GEMDOCK have highly divergences in performances. The label of methods corresponds to table 10. (None: GEMDOCK without pharmacological preferences, both: GEMDOCK with pharmacological preferences, Ori: GOLD with GoldScore, and new: GOLD with recombinant GoldScore)

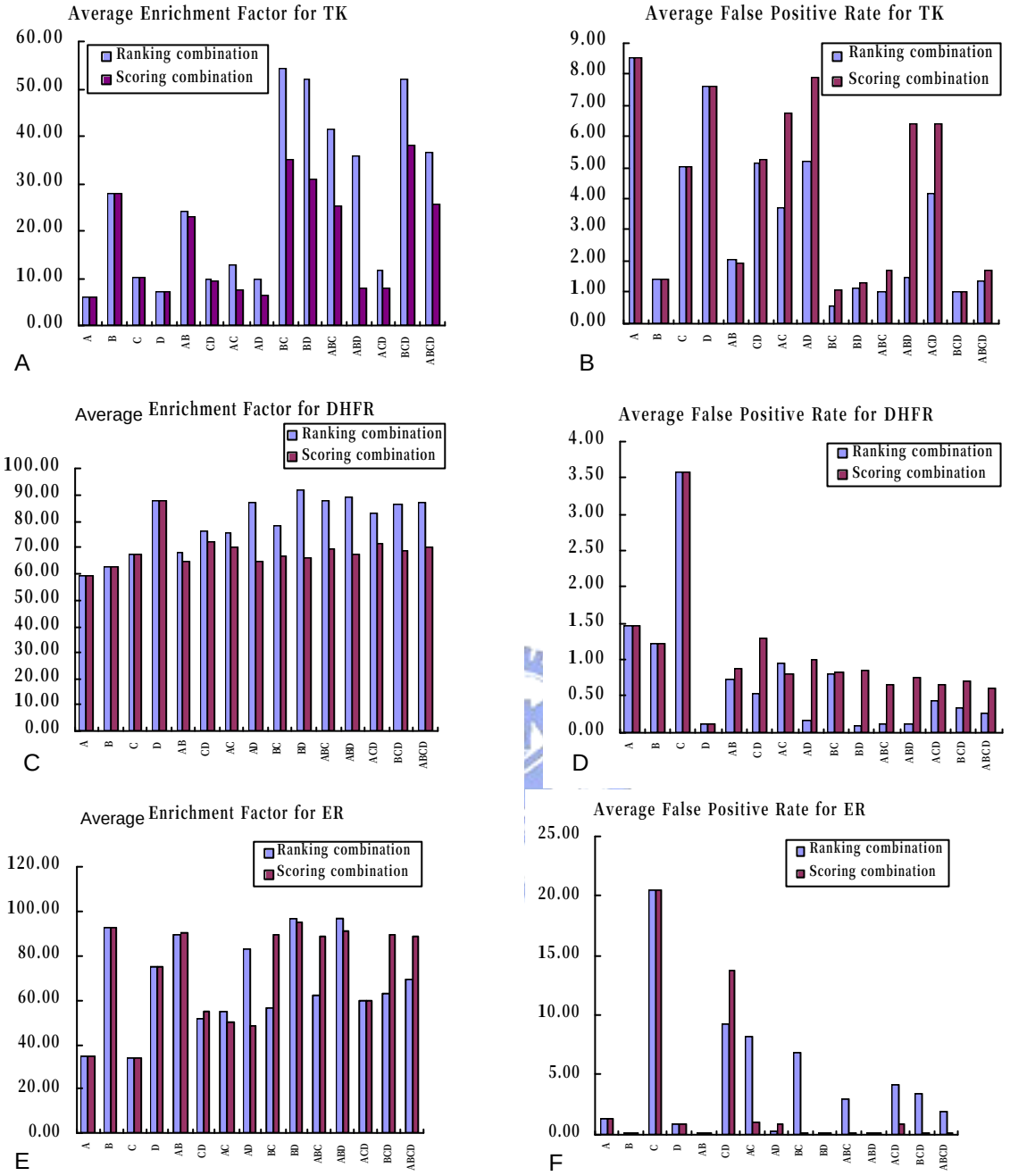


Figure 13. Average enrichment factor and false positive rate comparison for ranking and scoring combinations. (A) Average enrichment factor and (B) Average false positive rate of TK. (C) Average enrichment factor and (D) Average false positive rate of DHFR. (E) Average enrichment factor and (F) Average false positive rate of ER. The label in x-axis is the composition of methods.