SUPPORTING INFORMATION

A Pipeline to Enhance Ligand Virtual Screening: Integrating Molecular Dynamics and FLAP

Francesca Spyrakis [1]*, Paolo Benedetti [2], Sergio Decherchi [3,4], Walter Rocchia [3], Andrea Cavalli [5,6], Stefano Alcaro [7], Francesco Ortuso [7], Massimo Baroni [8] and Gabriele Cruciani [2]*

[1] Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 183, 41125, Modena, Italy; [2] Department of Chemistry, Biology and Biotechnology, University of Perugia, Via Elce di Sotto 8, 06123, Perugia, Italy; [3] CONCEPT Lab, Italian Institute of Technology, via Morego 30, 16163, Genova, Italy; [4] BiKi Technologies s.r.l., via XX Settembre 33, 16121, Genova, Italy; [5] CompuNet, Italian Institute of Technology, via Morego 30, 16163, Genova, Italy; [6] Department of Pharmacy and Biotechnology, University of Bologna, via Belmeloro 6, 40126, Bologna, Italy; [7] Department of Health Science, University Magna Graecia, Campus of Catanzaro, Campus "S. Venuta", Viale Europa, 88100, Catanzaro, Italy; [8] Molecular Discovery Limited, 215 Marsh Road, Pinner Middlesex-London HA5-5NE, United Kingdom.

*** Corresponding authors**

Gabriele Cruciani

Department of Chemistry, Biology and Biotechnology, University of Perugia, Via Elce di Sotto 8, 06123, Perugia, Italy.
phone: 0039 075 5855629
e-mail: gabri@chemiome.chm.unipg.it

Francesca Spyrakis

Department of Life Sciences, University of Modena and Reggio Emilia, Via Campi 183, 41125, Modena, Italy.
phone: 0039 059 2058604
e-mail: francesca.spyrakis@unimore.it

Benchmarking FLAP

Different benchmarking for comparing the performances of the FLAP software with respect to other virtual screening tools have already been performed, always reporting promising results.[1,2] In particular, in 2007, FLAP VS performance was tested and compared with those of DOCK, GOLD, and Glide docking methods. Very satisfactory results were obtained using FLAP in only a fraction of the time required by the other methods (0.5 h for FLAP, against the 6 h for Glide and GOLD and 10 h for DOCK).[1] The analyses were performed on three different targets, i.e. factor Xa, thymidine kinase, and estrogen receptor α. Ligands were obtained from the MDL Drug Data Report. Then, in 2010, FLAP performance was assessed using a subset of the DUD dataset containing 13 different targets each with more than 15 different chemotype classes. Excellent chemotype enrichments were obtained by both ligand-based and receptor-based approaches, of approximately 17-fold over random on average at a false positive rate of 1%.[2]
None of the aforementioned studies included PNP as possible target. We thus decided to compare FLAP performance in SRC VS with respect to other two well-known virtual screening software tools, i.e. Autodock Vina and Glide.

The results are reported in Table S1 and in Figure S1. For Glide only the AUC values given by the SP protocol, giving the best performances, were reported. Overall FLAP demonstrated to provide better results in all the different cases, both in terms of global AUC and of early enrichment. Moreover, the time necessary for running the VS on FLAP is about a few minutes for the considered database, while goes from hours to days for Glide and AutoDock Vina, respectively, according to the number of used CPUs.

AutoDock Vina. AutoDock Vina version 1.1.2 was used.[3] AutoDock Vina is a new generation docking software from the Molecular Graphics Laboratory at the Scripps Research Institute (La Jolla, CA). It achieves significant improvements in the average accuracy of the binding mode predictions, while being two orders of magnitude faster than AutoDock 4.
Input files were prepared with AutoDock Tools. The size of the binding site was defined by a box having 15 Å side lengths for X, Y and Z dimensions. The centroid of the binding site was set on the cocrystallized DATMe-Immucillin H ligand in the 3k8o structure and the exploration space was defined in a box having 10 Å side lengths. All the other structures were aligned to 3k8o. Since we were interested in the ranking, that is, the estimated free energy of binding, and not in the binding poses, one binding mode was generated. The energy range was set to 3 kcal/mol. Ligands were ranked according to the estimated free energy of binding in an increasing scale. The screening of the entire database was performed against all the X-ray structures and the ten medoids.

Glide. All target receptor models were preliminary submitted to the Protein Preparation Wizard module as implemented in the Schrödinger 2014-1 suite.[4,5] Such a procedure included optimised position hydrogen atoms and OPLS-2005 force field.[6,7] The resulting models were aligned using the CE algorithm.[8-11] The docking binding site was defined, for each target model, by means of a 27,000 $Å^3$ regular box centred onto the phosphate group located into the chain A active site. Glide[12-15] energy contour grids were computed at standard accuracy level. All target models were finally considered for docking simulation with respect to the compounds dataset. The

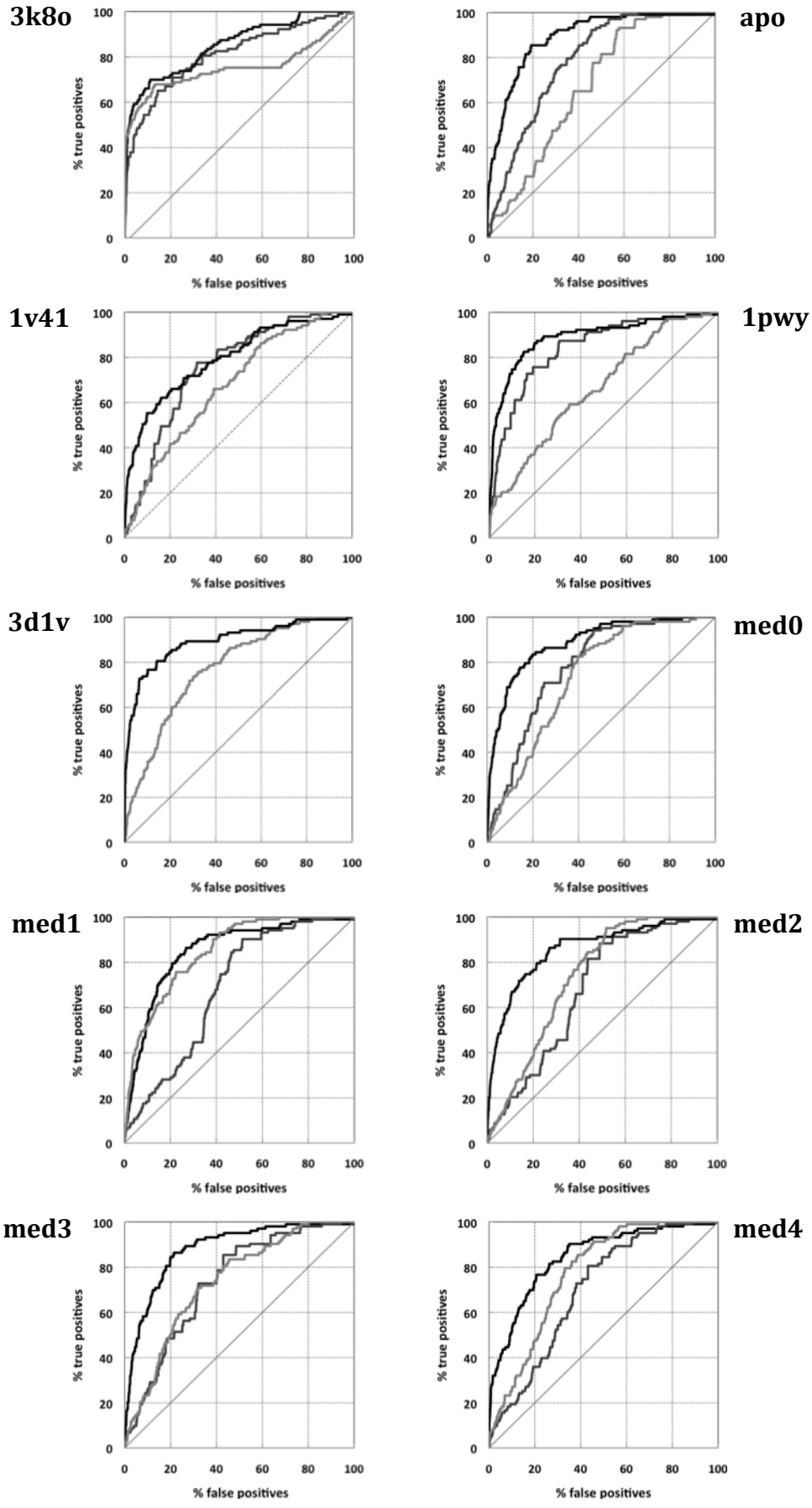
ligands conformation variability was taken into account by means of the Glide flexible algorithm. The High Throughput Virtual Screening (HTVS), the Standard Precision (SP) and the Extra Precision (XP) scoring functions were considered for the ligands targets affinity ranking. The screening of the entire database was performed against all the X-ray structures and the ten medoids.

Table S1. Total AUC calculated for the SRC performed on the X-ray structures and on the medoids by FLAP, AutoDock Vina and Glide SP protocol. No value is reported for the SRC-3d1v.

VS	FLAP AUC	AutoDock Vina AUC	Glide SP AUC
SRC-3k8o	0.85	0.69	0.75
SRC-apo	0.89	0.67	0.79
SRC-1v41	0.79	0.76	0.67
SRC-1pwy	0.89	0.84	0.66
SRC-3d1v	0.90	*	0.67
SRC-med0	0.89	0.78	0.73
SRC-med1	0.85	0.68	0.85
SRC-med2	0.86	0.68	0.74
SRC-med3	0.88	0.74	0.74
SRC-med4	0.85	0.70	0.77
SRC-med5	0.89	0.70	0.65
SRC-med6	0.85	0.70	0.78
SRC-med7	0.92	0.77	0.76
SRC-med8	0.90	0.67	0.82
SRC-med9	0.86	0.82	0.83

*No value is reported because of the impossibility to prepare the input with AutoDock Tools.

Figure S1.


3k8o
apo
1v41
1pwy
3d1v
med0
med1
med2
med3
med4
% true positives
% false positives
0
20
40
60
80
100

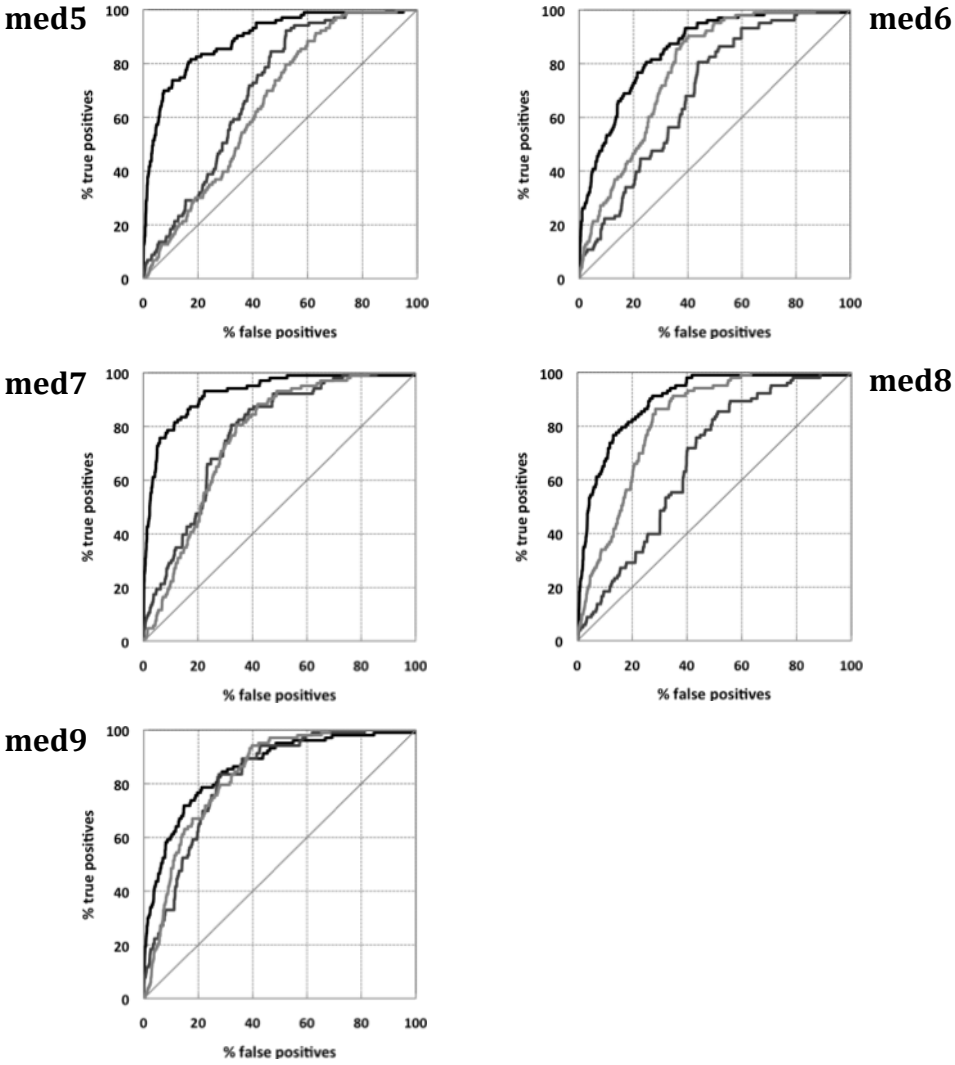


Figure S1. Comparison of the ROC curves for the single-structure-VS performed on the PNP X-ray structures and on the medoids extracted from the MD trajectory by FLAP (black line), AutoDock Vina (dark grey line) and Glide (light grey line).

Figure S2.

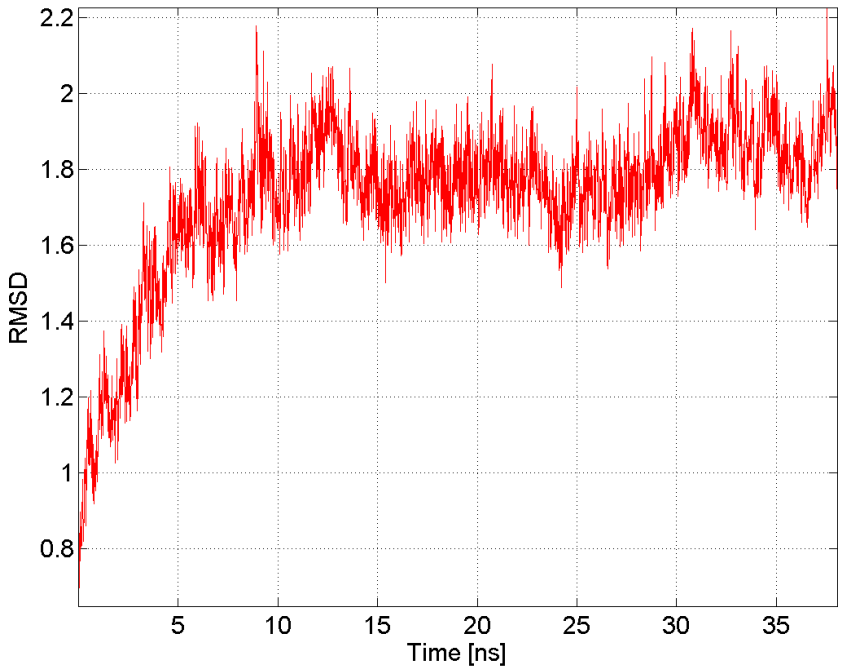


Figure S2. Backbone RMSD calculated for the PNP MD trajectory with respect to the first production frame, showing the structural stability along time. The value of the RMSD is around the crystal resolution.

Figure S3.

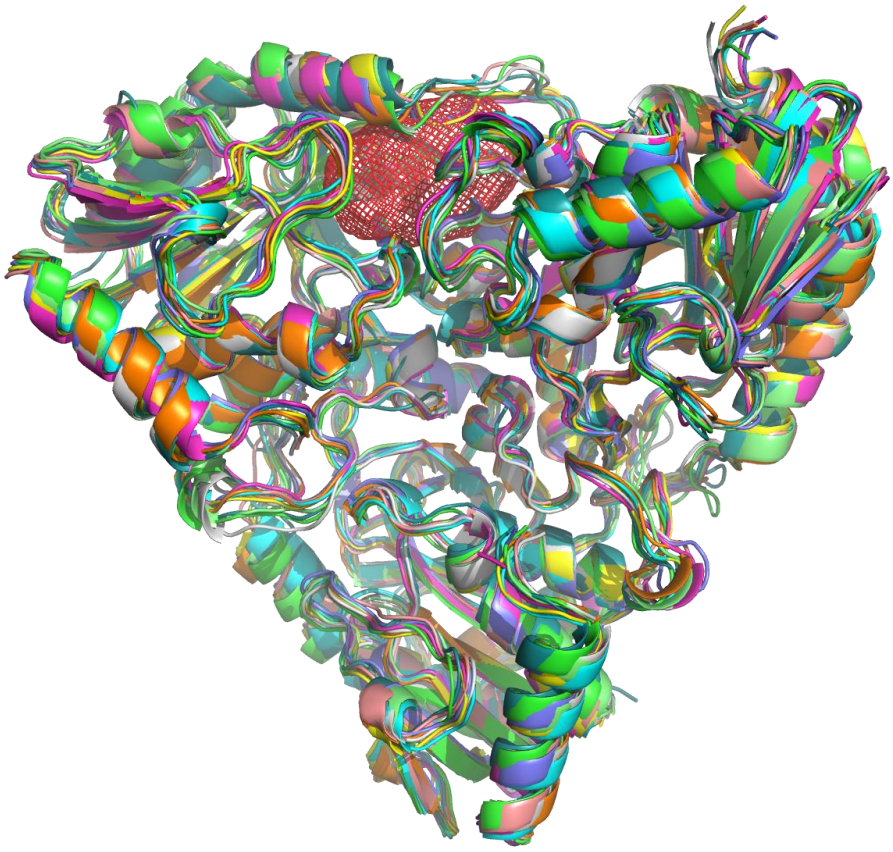


Figure S3. Superimposed clusters centers for the PNP trajectory. The binding site is indicated by the red mesh contour.

Figure S4.

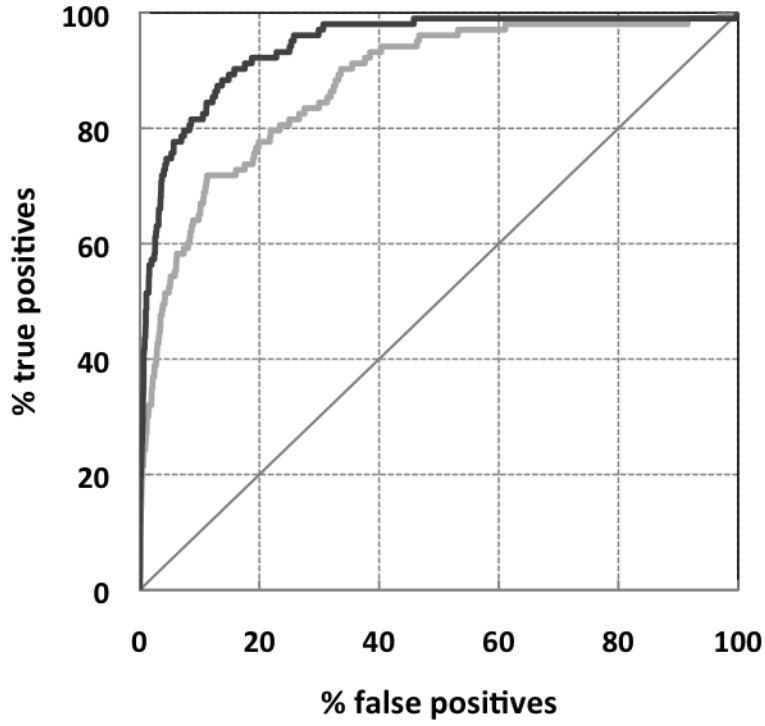


Figure S4. Comparison of the ROC curves for the LDA-SRC-3k8o (light grey line) and for the LDA-MRC-3k8o VS (dark grey line).

Figure S5.

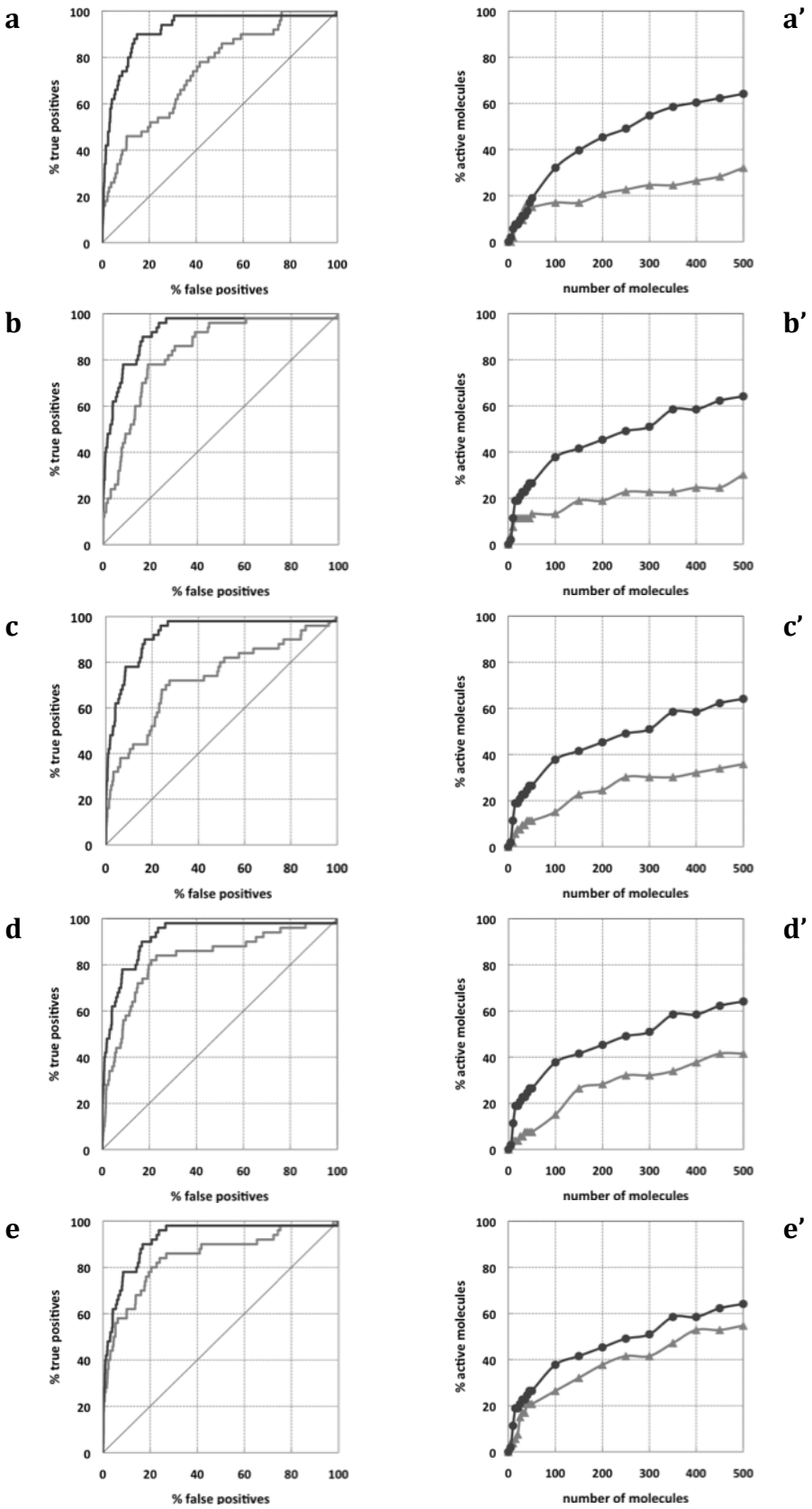

a
a'
b
b'
c
c'
d
d'
e
e'
% true positives
% false positives
% active molecules
number of molecules
0
20
40
60
80
100
0
100
200
300
400
500

Figure S5. Comparison of the ROC curves (left panels) and of the percentage of active molecules found in the first 500 ranked molecules (right panels) for the single-structure-VS and for the MRC-VS performed on the reduced dataset.
ROC curves (**a**) and active molecules % (**a'**) of the SRC-3k8o' VS (light grey line) and of the LDA-MRC-3k8o' VS (dark grey line). ROC curves (**b**) and active molecules % (**b'**) of the SRC-apo' VS (light grey line) and of the LDA-MRC-apo' VS (dark grey line). ROC curves (**c**) and active molecules % (**c'**) of the SRC-1v41' VS (light grey line) and of the LDA-MRC-1v41' VS (dark grey line). ROC curves (**d**) and active molecules % (**d'**) of the SRC-1pwy' VS (light grey line) and of the LDA-MRC-1pwy' VS (dark grey line). ROC curves (**e**) and active molecules % (**e'**) of the SRC-3d1v' VS (light grey line) and of the LDA-MRC-3d1v' VS (dark grey line). In the case of SRC-1v41', molecules were ranked according to the H*N1 FLAP score.

REFERENCES

1. Baroni, M.; Cruciani, G.; Sciabola, S.; Perruccio, F.; Mason, J.S. A Common Reference Framework for Analyzing/Comparing Proteins and Ligands. Fingerprints for Ligands and Proteins (Flap): Theory and Application. *J. Chem. Inf. Model.* **2007**, *47*, 279-294.
2. Cross, S.; Baroni, M.; Carosati, E.; Benedetti, P.; Clementi, S. Flap: Grid Molecular Interaction Fields in Virtual Screening. Validation Using the Dud Data Set. *J. Chem. Inf. Model.* **2010**, *50*, 1442-1450.
3. 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.
4. 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.
5. Schrödinger Suite 2014-1 Protein Preparation Wizard; Epik Version 2.7, Schrödinger, Llc, New York, Ny, 2013; Impact Version 6.2, Schrödinger, Llc, New York, Ny, 2014; Prime Version 3.5, Schrödinger, Llc, New York, Ny, 2014.
6. Jorgensen, W.; Maxwell, D.; Tirado-Rives, J. Development and Testing of the Opls All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225-11236.
7. Jorgensen, W.; Tirado-Rives, J. The Opls [Optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.
8. Jia, Y.; Dewey, T.G.; Shindyalov, I.N.; Bourne, P.E. A New Scoring Function and Associated Statistical Significance for Structure Alignment by Ce. *J. Comput. Biol.* **2004**, *11*, 787-799.
9. Pekurovsky, D.; Shindyalov, I.N.; Bourne, P.E. A Case Study of High-Throughput Biological Data Processing on Parallel Platforms. *Bioinformatics* **2004**, *20*, 1940-1947.
10. Shindyalov, I.N.; Bourne, P.E. Protein Structure Alignment by Incremental Combinatorial Extension (Ce) of the Optimal Path. *Protein Eng.* **1998**, *11*, 739-747.
11. Shindyalov, I.N.; Bourne, P.E. An Alternative View of Protein Fold Space. *Proteins* **2000**, *38*, 247-260.
12. *Glide, Version 6.2*, Schrödinger, LLC: New York, NY, 2014.
13. 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.
14. Friesner, R.A.; Murphy, R.B.; Repasky, M.P.; Frye, L.L.; Greenwood, J.R.; Halgren, T.A.; Sanschagrin, P.C.; Mainz, D.T. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J. Med. Chem.* **2006**, *49*, 6177-6196.
15. Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* **2004**, *47*, 1750-1759.