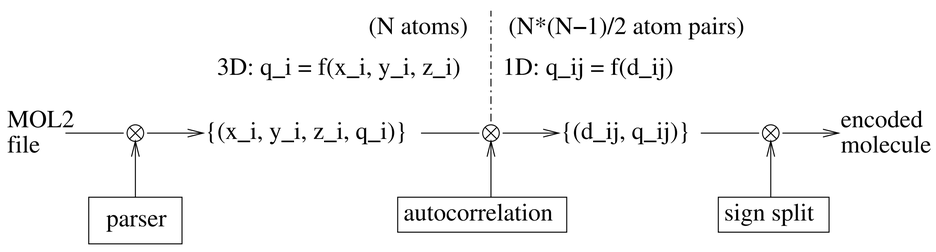
A rotation-translation invariant mol. desc. and its use in ligand-based virtual screening

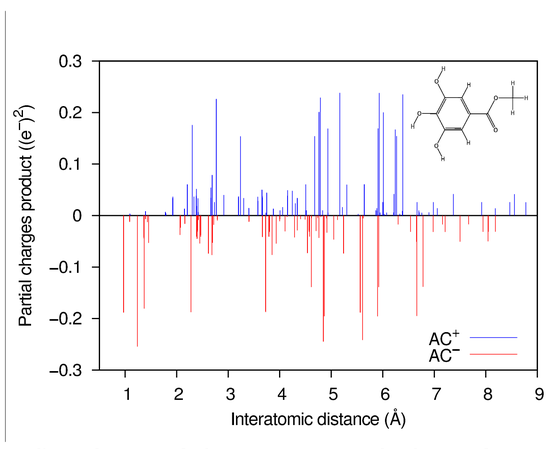
Francois Berenger and Kam Y. J. Zhang

Structural Bioinformatics Team, CLST, RIKEN, Yokohama, Japan

METHOD

Overview of ACPC's encoding pipeline, along with its output and the scoring pipeline[1,4]. ACPC is written in OCaml[7].





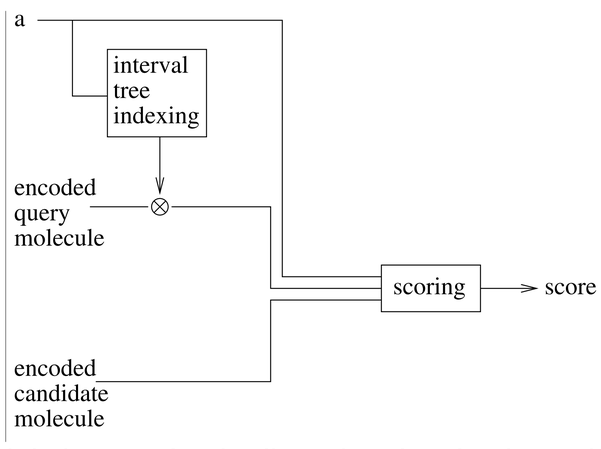


Figure 1 : Left: encoding of a molecule. Middle: positive (in blue) and negative (in red) values of the autocorrelation of partial charges for the ligand molecule shown in 2D. Right: scoring of two encoded molecules (one query molecule and a candidate molecule from a database). $\frac{1}{a}$ is the half-base length of a triangular kernel function used to transform the set of Kronecker deltas into a continuous function.

TRAINING AND TEST DATASETS

DUDE[5] contains 102 targets, 25660 ligands and 1060858 decoys. Preparation protocol: i) Open Babel[6] 2.3.9 unique InChI filtering ii) OMEGA[3] 2.4.6 lowest energy conformer iii) partial charges assigned with MMFF94x of MOE 2013.08 iv) molecules renamed to follow the scheme [active]InChIKey[TheInChIKey]. Training set; "DUDE-training" = 51 targets chosen randomly and only the 1st ligand of each target used as a query. Test set; "DUDE-test" = all 51 remaining targets with all their ligands.

PARAMETER SETUP

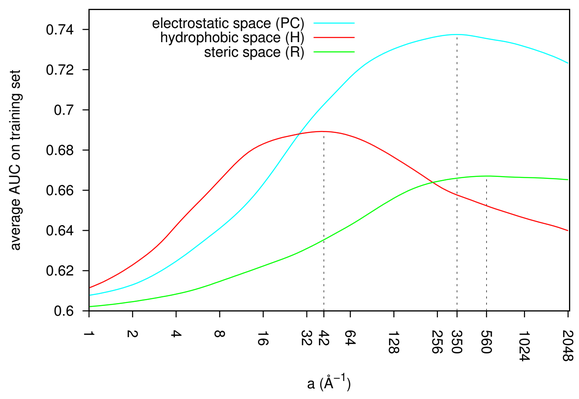


Figure 2 : Using the training set, an optimal value for the ***a*** parameter was found for each feature space (dotted lines).

On the training set, scores from each feature space were also centered then scaled. Then, a combination of weights was found to improve the average AUC and ER at 1% reached by ACPC. The weights are 0.8 for the electrostatic space, 0.2 for the steric space and 0.5 for the hydrophobic space.

VIRTUAL SCREENING PERFORMANCE

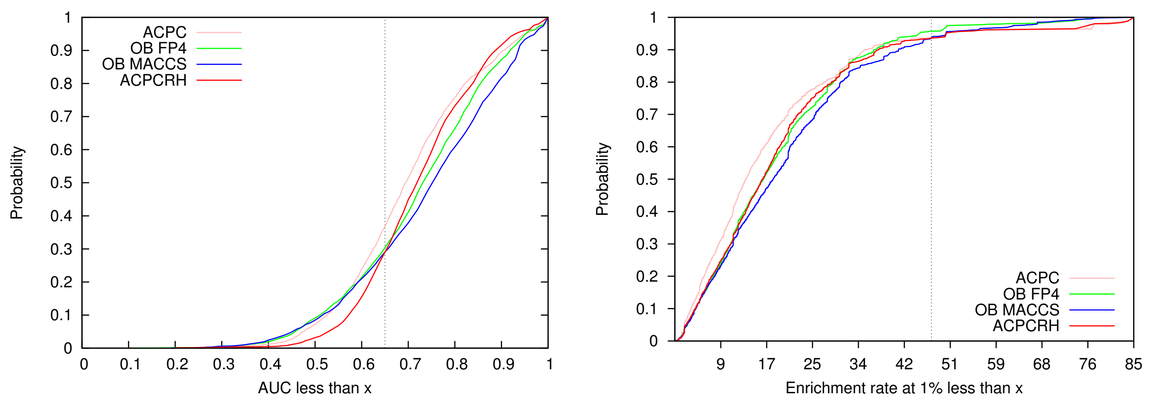


Figure 3 : Left: AUCs' cumulative distribution functions for ACPC in the electrostatic feature space, Open Babel's (OB) FP4 and MACCS implementations and ACPC using three feature spaces (ACPCRH) on DUDE-test. Right: enrichment rates at 1% in the same experiment.

CONSENSUS QUERIES

Five query molecules per target were chosen randomly on DUDE-training. Then, the min rank was used to rank-order molecules when using up to five query molecules for consensus in the electrostatic feature space.

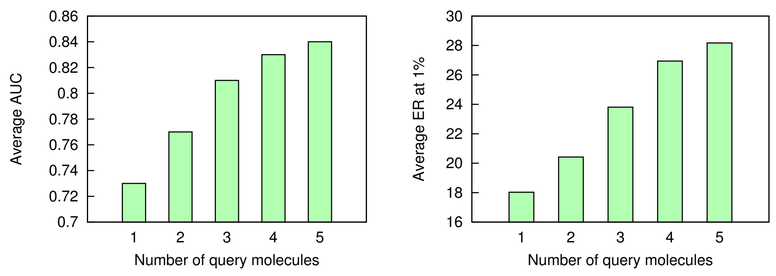


Figure 4 : Left: average AUC as a function of the number of query molecules used during consensus. Right: average enrichment rate at 1% in the same experiment.

QUERY MOLECULE ANNOTATION

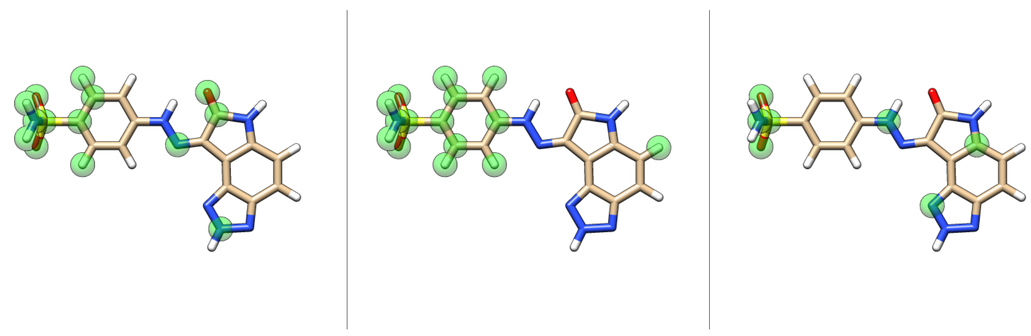


Figure 5 : In any feature space, the query molecule can be annotated based on atomic contribution to AUC (Left: electrostatic space, Middle: steric space, Right: hydrophobic space). Here, the 1st ligand of the cdk2 target has green balls highlighting atoms which if removed would decrease the AUC reached by the query molecule in the given feature space.

SPEED

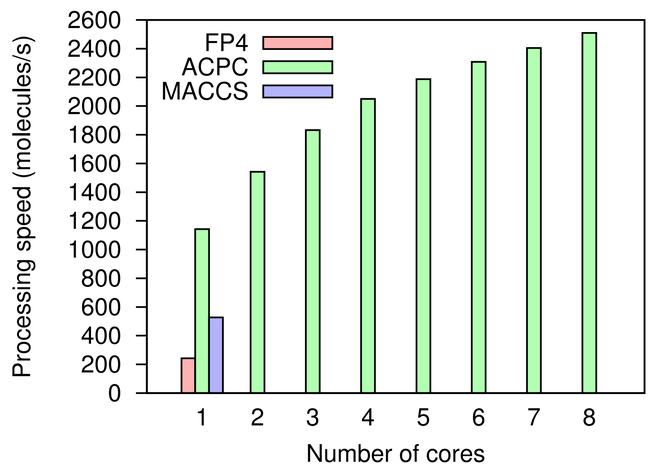


Figure 6 : Speed comparison between ACPC in the electrostatic feature space (Parmap[2] is used for parallelization) and Open Babel 2.3.9's MACCS and FP4 implementations. For each method, runtimes were averaged over three runs after a warmup run. The target was hivpr from DUDE (26450 ligands and decoys).

INSTALL ACPC

1 - Install and configure OPAM (cf. `http://opam.ocaml.org/`)
2 - Type: 'opam install acpc.1.2'

REFERENCES

[1] F Berenger, A Voet, XY Lee, and KYJ Zhang. A rotation-translation invariant molecular descriptor of partial charges and its use in ligand-based virtual screening. *Journal of Cheminformatics*, 6(1):23, 2014.

[2] M Danelutto and R Di Cosmo. A Minimal Disruption Skeleton Experiment: Seamless Map and Reduce Embedding in OCaml. *Procedia Computer Science*, 9(0):1837 – 1846, 2012. Proceedings of the International Conference on Computational Science, ICCS 2012.

[3] Paul C. D. Hawkins and Anthony Nicholls. Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures. *Journal of Chemical Information and Modeling*, 52(11):2919–2936, 2012.

[4] G Moreau and P Broto. The autocorrelation of a topological structure: a new molecular descriptor. *Nouveau journal de chimie*, 4(6):359 – 360, 1980.

[5] MM Mysinger, M Carchia, JJ Irwin, and BK Shoichet. Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *Journal of Medicinal Chemistry*, 55(14):6582–6594, 2012.

[6] N O'Boyle and M Banck and C James and C Morley and T Vandermeersch and G Hutchison. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1):33, 2011.

[7] X Leroy and D Doligez and A Frisch and J Garrigue and D Rémy and J Vouillon. *The OCaml system release 4.01*. INRIA, 2013.