Benchmarking Commercial Conformer Ensemble Generators

Nils-Ole Friedrich[1], Agnes Meyder[1], Christina de Bruyn Kops[1], Kai Sommer[1], Florian Flachsenberg[1], Matthias Rarey[1], Johannes Kirchmair[1]*

[1]Center for Bioinformatics, University of Hamburg, Bundesstraße 43, Hamburg 20146, Germany

* J. Kirchmair. E-mail: kirchmair@zbh.uni-hamburg.de. Tel.: +49 (0)40 42838 7303

Recently we published a cheminformatics pipeline that allows the fully automated extraction of high-quality structures of protein-bound ligands from the PDB. [1] The criteria applied for the selection include the fit of the molecular structures to the electron density, the physicochemical and structural properties of the ligands, and many more. With this approach we compiled the Platinum Dataset, which with its 4626 structures is the most comprehensive collection of such structural data available.

In the initial publication, the Platinum Dataset was used for benchmarking seven freely available algorithms for conformer ensemble generation. In continuation of this work, we present an extended performance analysis of algorithms on a further refined version of the Platinum Dataset. The benchmarking study now also includes the most popular commercial algorithms for conformer ensemble generation: ConfGen [2], ConfGenX (Schrödinger), Cxcalc (ChemAxon), iCon (Inte:Ligand), OMEGA (OpenEye) [3] and three algorithms of the Molecular Operating Environment (MOE, Chemical Computing Group) [4].

We found the commercial tools to be much more similar in performance than the free tools and have more reliable success rates overall. Few commercial algorithms performed better than the best performing free tool RDKit. [5] ConfGenX and OMEGA reached nearly the same high accuracy, but OMEGA was faster and achieved a favorable balance of accuracy, ensemble size and runtime. We also analyzed the influence of different force fields on the performance of RDKit and ConfGenX.

The Platinum Dataset is available at http://www.zbh.uni-hamburg.de/platinum_dataset.

[1] Friedrich, N.; Meyder, A.; de Bruyn Kops, C.; Sommer, K.; Flachsenberg, F.; Rarey, M.; Kirchmair, J. High-Quality Dataset Of Protein-Bound Ligand Conformations And Its Application To Benchmarking Conformer Ensemble Generators, *J. Chem. Inf. Model.,* **2017**.
[2] Watts, K.S.; Dalal, P.; Murphy, R.B.; Sherman, W.; Friesner, R.A.; Shelley, J.C., ConfGen: A Conformational Search Method for Efficient Generation of Bioactive Conformers, *J.Chem. Inf. Model.,* **2010**, 50, 534-546
[3] Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T., Conformer Generation with OMEGA: Algorithm and Validation Using High Quality Structures from the Protein Databank and the Cambridge Structural Database, *J. Chem. Inf. Model.,* **2010**, 50, 572-584.
[4] Molecular Operating Environment (MOE), version 2013.08; Chemical Computing Group Inc., Montreal, QC, **2016**.
[5] Tosco, P.; Stiefl, N.; Landrum, G. Bringing the MMFF Force Field to the RDKit: Implementation and Validation. *J. Cheminf.,* **2014**, 6, 37.