

# Estimation of synthetic accessibility score of drug-like molecules

Peter Ertl

Novartis Institutes for BioMedical Research, Basel, Switzerland

peter.ertl@novartis.com  
<http://peter-ertl.com>

Ranking, or prioritization of molecules is needed in many areas of modern drug discovery process. Molecules are normally prioritized according to criteria such as drug-like properties, natural-product likeness, predicted biological activity or freedom to operate with respect to intellectual property. Since sooner or later the selected structures will be resynthesized or derivatized, prioritization also by their synthetic accessibility, should be included early on.

The method for estimation of synthetic accessibility score (SAscore) which we have developed at Novartis and is presented here is based on a combination of fragment contributions and a complexity penalty. Fragment contributions have been calculated based on the analysis of fragments in one million representative molecules from PubChem and therefore one can say that they capture historical synthetic knowledge stored in this database. The molecular complexity score takes into account the presence of non-standard structural features, such as large rings, non-standard ring fusions, stereocomplexity and molecule size. The method has been validated by comparing calculated SAscores with ease of synthesis as estimated by experienced medicinal chemists for a set of 40 molecules. The agreement between calculated and manually estimated synthetic accessibility is very good with  $r^2 = 0.89$ .

Various possible applications of the new synthetic accessibility score in drug discovery processes will be also discussed, including purchasing samples for screening, selecting hits from high-throughput screening for follow-up, or ranking molecules generated by various *de novo* design approaches.

See also:

P. Ertl and A. Schuffenhauer, Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions, *Journal of Cheminformatics*, 1:8 2009 (open access)  
<http://www.jcheminf.com/content/1/1/8>