SYBA: Bayesian estimation of synthetic accessibility of organic compounds

CURRENT STATUS: UNDER REVISION

Journal of Cheminformatics BMC

Milan Voršilák
Ustav molekularni genetiky Akademie Ved Ceske Republiky
ORCiD: https://orcid.org/0000-0002-8923-1627

Michal Kolář
Ustav molekularni genetiky Akademie Ved Ceske Republiky
ORCiD: https://orcid.org/0000-0002-4593-1525

Ivan Čmelo
University of Chemistry and Technology Prague
ORCiD: https://orcid.org/0000-0001-7787-8653

Daniel Svozil
University of Chemistry and Technology
daniel.svozil@gmail.com*Corresponding Author*
ORCiD: https://orcid.org/0000-0003-2577-5163

DOI:
10.21203/rs.2.22597/v1

SUBJECT AREAS
Analytical Biochemistry

KEYWORDS
synthetic accessibility, Bayesian analysis

Abstract

SYBA (SYnthetic Bayesian Accessibility) is a fragment based method for the rapid classification of organic compounds as easy- (ES) or hard-to-synthesize (HS). SYBA is based on the Bayesian analysis of the frequency of molecular fragments in the database of ES and HS molecules. It was trained on ES molecules available in the ZINC15 database and on HS molecules generated by the Nonpher methodology. SYBA was compared with a random forest, that was utilized as a baseline method, as well as with other two methods for synthetic accessibility assessment: SAScore and SCScore. When used with their suggested thresholds, SYBA improves over random forest classification, albeit marginally, and outperforms SAScore and SCScore. However, with thresholds optimized by the analysis of ROC curves, SAScore improves considerably and yields similar results as SYBA. Because SYBA is based merely on fragment contributions, it can be used for the analysis of the contribution of individual molecular parts to compound synthetic accessibility. Though SYBA was developed to quickly assess compound synthetic accessibility, its underlying Bayesian framework is a general approach that can be applied to any binary classification problem. Therefore, SYBA can be easily re-trained to classify compounds by other physico-chemical or biological properties. SYBA is publicly available at https://github.com/lich-uct/syba under the GNU General Public License.

Full-text

Due to technical limitations, full-text HTML conversion of this manuscript could not be completed. However, the manuscript can be downloaded and accessed as a PDF.

Additional File 1

Due to technical limitations, Additional File 1 is not available in this version.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

Additional file 1.xlsx
Additional file 4.xlsx
Additional file 3.xlsx
Additional file 2.docx