

de novo generated combinatorial library design

Supplementary material

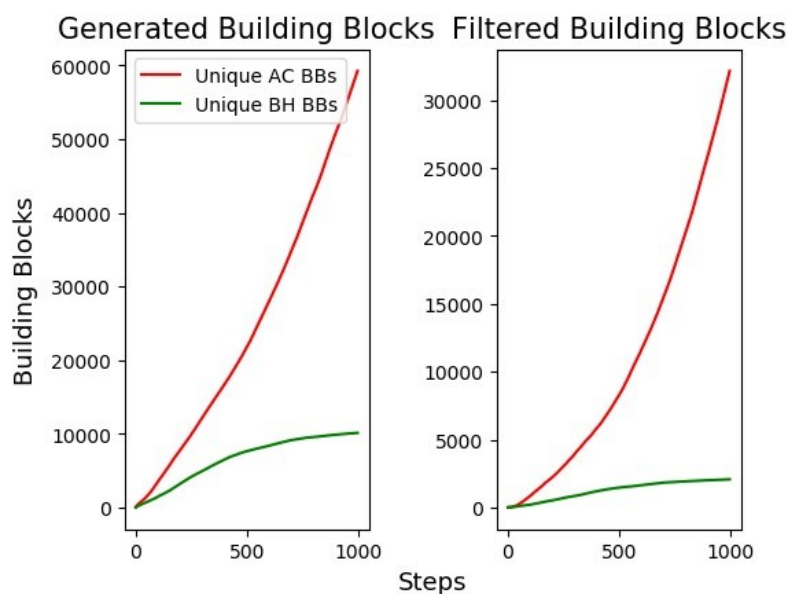


Figure S.1 A plot of the generated number of unique building blocks (BBs) using LibINVENT for 1000 epochs using a learning rate of 128 and a learning rate of 5×10^{-6} . Left: All generated building blocks. Right: Building blocks scoring higher than 0.8 according to the trained QSAR model and further filtered to match the desired reactions of Amide Coupling (AC) and Buchwald-Hartwig (BH) according to SMARTS¹ substructure match.

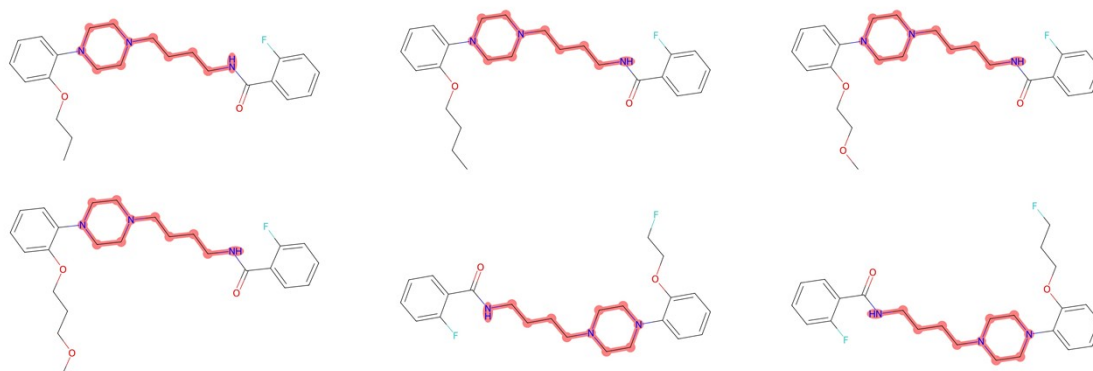


Figure S.2 Example of generated building blocks attached to the scaffold selected to have a low diversity using a greedy sampling algorithm. The computed logdeterminant for this selection was -8.26. The used scaffold is highlighted for visual clarity.

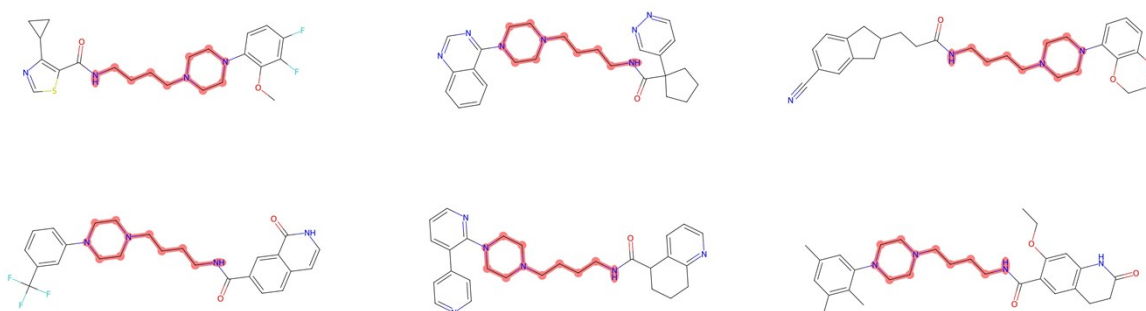


Figure S.3 Example of generated building blocks attached to the scaffold selected to have a high diversity using a greedy sampling algorithm. The computed logdeterminant for this selection was -0.53. The used scaffold is highlighted for visual clarity.

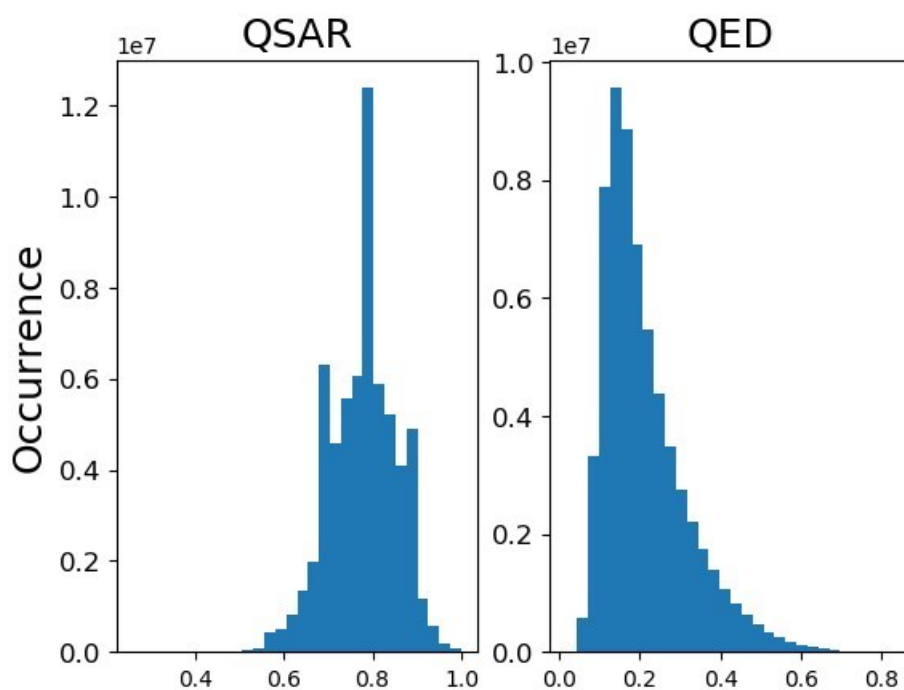


Figure S.4 The distribution of QSAR (left) and QED (right) values, for all products made from the generated building blocks up to 4 reaction steps, i.e., the full space optimized over by the sampling algorithm. Median QSAR=0.78, Median QED= 0.18.

1. Daylight, SMARTS - A Language for Molecular Patterns, <https://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>, (accessed 2023-02-28, 2023).