Freedom Space - Design of Ultra-Large Synthetically Accessible Chemical Spaces Using Machine Learning-based Reagent Filtering

Anna Kapeliukha,[1] Mariana Perebyinis,[1] Serhii Hlotov,[1] Mykola Protopopov,[1] Olga Tarkhanova,[1] Yurii Moroz[2]

[1]Chemspace LLC, 85 Winston Churchill Street, Kyiv 02094, Ukraine
[2]Enamine Ltd, 78 Winston Churchill Street, Kyiv 02094, Ukraine

CHEMSPACE
Delivering Discovery Solutions®

Introduction

Ultra-large chemical spaces are transforming drug discovery by increasing the chances of identifying novel drug candidates. Efficient use of these spaces requires that compounds be synthesizable, a challenge addressed through methods like synthesizability scores and rule-based filtering. However, improvements are still needed. One of the leading examples is Enamine REAL Space (48B compounds), which has a confirmed 80% synthesis success rate and is created from well-validated reactions and building blocks.
In this work, we present Freedom Space 3.0, a 5 billion compound space that is REAL-like and complementary to Enamine REAL and has a confirmed synthesis success rate of 75%+. Using this space as an example, we showcase a novel workflow for creating synthetically accessible chemical spaces from proprietary building block collections using machine learning-based reagent filtering.

Freedom Space 3.0
5B compounds

- 10 chemical transformations
- 75+ synthesis success rate
- Complementary to Enamine REAL

General workflow for the ML-assisted space generation

Use any kind of proprietary/ commercial building block collection. Building blocks from the Enamine in-stock collection can be added for space size/diversity.

Utilize ML models trained on reaction success data from years of Enamine REAL Space operation to filter the reagents.

Filtering the space by physicochemical properties according to the needs of your projects.

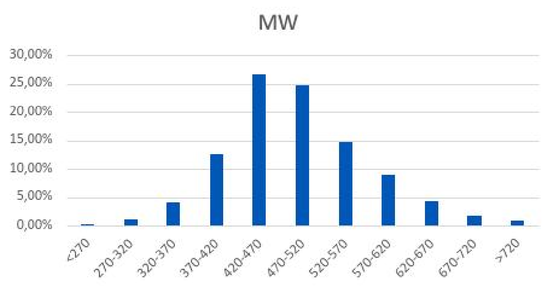
Building block collection analysis → **Reactions selection** → **Reagents filtering using ML models** → **Synthon creation and enumeration** → **Property filtering** → **Custom Synthetically Accessible Chemical Space**

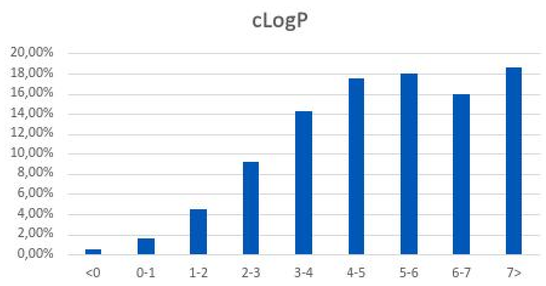
Selecting the building blocks for specific reaction types using in-house designed SMARTS patterns.

Creating the space in two formats: synthons and full enumerated molecules.

Freedom Space Overview

Physicochemical properties





Reaction diversity

7 ML-supported (6 two-component and 1 three-component) and 2 human-filtered reactions were used in the space creation.

Reaction Schemes

Theoretical synthesisability evaluation

To evaluate the synthesizability of the molecules in the space we utilized commonly used synthesizability scores: RAScore[1], SAscore[2] and performed retrosynthetic analysis using AiZynthFinder[3].

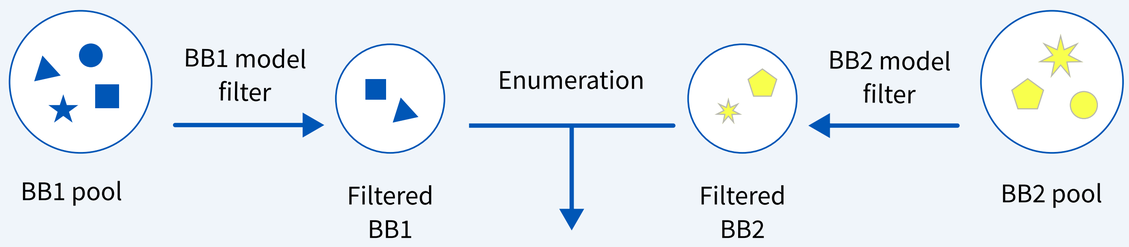
Practical synthesisability evaluation

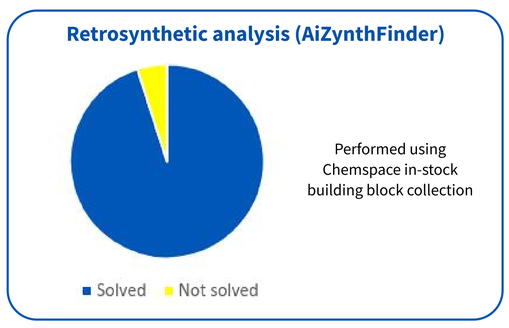
Freedom Space 3.0 has been commercially available for over a year and has been used in a number of commercial and academic projects. To showcase the synthesisability of the molecules in the space, we are showing some of the timelines of the projects it was used in.

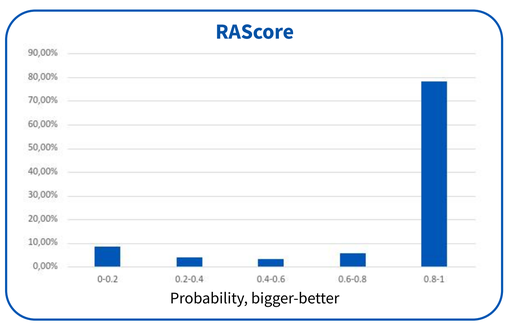
www.chem-space.com
a.kapeliukha@chem-space.com

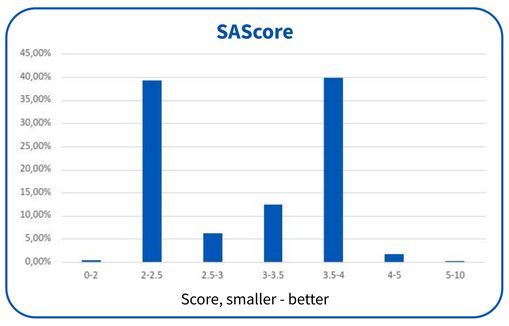
ML-based reagent filtering

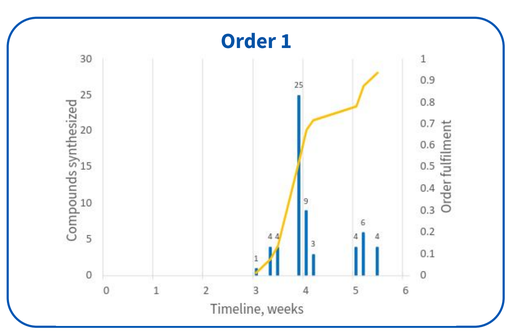
The reagents were filtered using machine-learning-based filters, trained on historical reaction success data from the operation of Enamine REAL. For each reagent, a neural network model was trained using a combined synthon-building block representation.











References:
1.Thakkar, A., Chadimová, V., Bjerrum, E. J., Engkvist, O., and Reymond, J.-L., 2021, Retrosynthetic accessibility score (RAscore) – rapid machine learned synthesizability classification from AI driven retrosynthetic planning: Chemical Science, v. 12, p. 3339–3349.
2.Ertl, P., and Schuffenhauer, A., 2009, Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions: Journal of Cheminformatics, v. 1.
3.Genheden, S., Thakkar, A., Chadimová, V., Reymond, J.-L., Engkvist, O., and Bjerrum, E., 2020, AiZynthFinder: a fast, robust and flexible open-source software for retrosynthetic planning: Journal of Cheminformatics, v. 12.