CoLibri – Building Spaces

BioSolveIT
expect actives!

CoLibri is a toolkit for the creation of individual, combinatorial Chemical Spaces. Using defined chemical reactions, it creates large molecule data sets which can be screened with Chemical Space exploration tools.

How does CoLibri work?

Synthetic access is the key to success in practically all pharmaceutical research projects. What can be made – easily enough – decides in the end what is actually made. The collective know-how of chemical reactions in a company is therefore the most valuable intellectual property pool to draw from. Why not capture this know-how and make it accessible throughout the organization?
CoLibri can do exactly that. Starting from what is known in the literature, it lets you transform your collective know-how into a Chemical Space, constituting huge numbers of virtual but synthetically accessible compounds.
CoLibri comprises three commandline tools to create individual, combinatorial Chemical Spaces: ReactionSynthesizer, FragspaceMerger and SpaceLightDBCreator.

ReactionSynthesizer

ReactionSynthesizer is a command line tool that takes one or more reaction definitions as an input and generates an individual fragment space for every one of them. In addition, the reaction synthesizer can enumerate product molecules for a reaction.

FragspaceMerger

Another command line tool implemented in CoLibri is the fragment space merger (= FragspaceMerger). It takes the output of the ReactionSynthesizer (multiple individual fragment spaces) and merges them to a single concise fragment space.

SpaceLightDBCreator

The molecular fingerprint screening method SpaceLight requires a distinct Chemical Space format. SpaceLightDBCreator transforms your input reactions into a molecular fingerprint readable format that supports the well-known Extended-Connectivity Fingerprints (ECFPs) and Connected Subgraph Fingerprint (CSFP) and their variants.

Advantages

- Generate new intellectual property
- Expand your literature compound toolkit
- Find purchasable analogs
- Save time and money

De novo design & synthetic access

De novo design was first invented to think outside the box, to generate new ideas, and to come up with entirely novel solutions. However, its major drawback so far was the lack of synthetic access, hence medicinal chemists often objected to these designs. CoLibri overcomes this issue by generating compound collections based on defined chemistry rules, resulting in synthetically-accessible new chemical entities (NCEs).

Can I create my own Chemical Space?

Many companies have a unique collection of in-house chemistries and building blocks, and therefore want to set up their own space. This will also create the highest level of IP-value. **Pfizer** did that and built PGVL. They turned 358 combinatorial libraries based in validated reaction protocols into a virtual Chemical Space, resulting in 3 trillion chemically accessible compounds.

Find hidden Gems

Scientists at **AstraZeneca** did a systematic study about the orthogonality of different similarity search methods. Specifically, it was determined across a number of data sets for a variety of targets, which method uniquely found active molecules, namely active molecules that the other methods missed. Later, this study was repeated using FTrees confirming that FTrees performed best according to this benchmark.

No active site – where do I start?

If you are working on a GPCR project (or any other project lacking a protein structure) and cannot follow the traditional structure-based design approach, mining a virtual Chemical Space will be a useful approach to generate inspiring new chemistries as starting points. **Boehringer-Ingelheim** used this approach to propel a GPR119 project. They started out with four structurally different literature compounds as queries to search their in-house virtual, BI-CLAIM to generate 10,000 FTrees hits. These were post-processed by a 3D shape filter based on one of the more rigid query compounds. Specific cores of the molecules were selected by the chemistry team as "activity anchors" and focused libraries around these designed and synthesized. This approach led to two new GPR119-agonist hit series, one of which subsequently resulted in a new lead class.

Discover distant neighbors

Using the FTrees similarity search method, **Gedeon Richter** found novel Histamine H_4 receptor and Serotonin transporter ligands. This study also nicely confirms that FTrees is well able to jump chemical classes and find remotely similar molecules – active against the same target – with distinctly different scaffolds.

Literature

Korn, M.; Ehrt, C.; Ruggiu, F.; Gastreich, M.; Rarey, M. Navigating Large Chemical Spaces in Early-Phase Drug Discovery. *Curr. Opin. Struct. Biol.* **2023**, *80*, 102578.
https://doi.org/10.1016/j.sbi.2023.102578.

Warr, W. A.; Nicklaus, M. C.; Nicolaou, C. A.; Rarey, M. Exploration of Ultralarge Compound Collections for Drug Discovery. *J. Chem. Inf. Model.* **2022**, *62* (9), 2021–2034.
https://doi.org/10.1021/acs.jcim.2c00224.

Warr, W. Report on an NIH Workshop on Ultralarge Chemistry Databases. *ChemRxiv* **2021**, 0–44.
https://doi.org/10.26434/chemrxiv.14554803.v1

BioSolveIT © 2026