SpaceLight – Fingerprint Screening

BioSolveIT
expect actives!

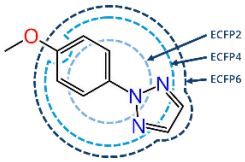
SpaceLight is a Chemical Space exploration tool which screens vast combinatorial compound spaces for close analogs of a query compound based on molecular fingerprints.

How does SpaceLight work?

SpaceLight was developed in collaboration with the University of Hamburg (ZBH) and utilizes topological fingerprint similarity such as the as the well-known Extended-Connectivity Fingerprints (ECFPs) and Connected Subgraph Fingerprint (CSFP).

What are molecular fingerprints?

Molecular fingerprints are essential cheminformatics tools to encode, describe and compare structural features of compounds. Along the different fingerprint subtypes which are explained below, users of SpaceLight can select several variants (also known as "iterations") of those to fine-tune their results.



Variants describe the order of neighbor shells that encode the features of atoms in a molecule. The variant number indicates the shell size of the generated substructures used for the similarity search. Larger substructures cover more features and atom arrangements of the molecule, yet smaller differences receive a penalty in their similarity score which can affect the ranking of the results. Therefore, it can sometimes be required to fine-tune the screening by applying different variants of the fingerprint subtype.
For instance, the variants of ECFPs describe the effective diameter of the largest feature of iterations performed. An example given above shows which area of the structure corresponds to its ECFP variant.

What fingerprint subtypes can be searched with SpaceLight?

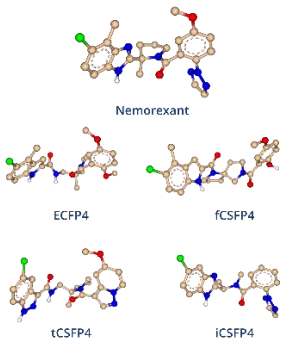
Alongside the well-known ECFP4 fingerprint, new "CSFP" methods are available to retrieve compounds of interest from vast Chemical Spaces based on the challenges and scenarios related to individual drug discovery projects.
ECFP: The extended-connectivity fingerprint. Its variant ECFP4 belongs to the best-performing fingerprints in small molecule virtual screening and target prediction benchmarks. Considers only circular features for similarity
Further, SpaceLight also supports a variety of Connected Substructure (or Subgraph) Fingerprints (CSFPs) and their variants:

- **fCSFP:** A fingerprint for fine-grained similarity measurement.
- **tCSFP:** A topological fingerprint method which dis-covers analogs with similar arrangement of function-alities.
- **iCSFP:** An independent fingerprint method using the maximum-common substructure as the similarity descriptor. Literally billions of molecules can be searched in 5 minutes

		EFCP	fCSFP	iCSFP	tCSFP
chemical substructures	circular	•			
	all		•	•	•
	element	•	•	•	•
	connectivity	•	•		•
	connectivity in substructure		•	•	
	valence	•	•	•	
atom properties	valence in substructure		•	•	
	aromaticity				•
	formal charge	•			
	weight	•			
	ring membership	•	•		

Source for accessible compounds

The molecules retrieved from dedicated Chemical Spaces with SpaceLight are accessible by synthesis. The molecules are not fantasy molecules that require complex synthesis, but instead they can be ordered from the compound makers — or be synthesized in one or two steps in-house. For this realized workflow, we collaborate with Enamine, WuXi LabNetwork, OTAVA, Chemspace and eMolecules in creating actionable make-on-demand Chemical Spaces.



Use the compounds for machine learning approaches or structure-based calculations (e.g., docking studies) to narrow down the best candidates for follow up.

Literature

Bellmann, L.; Penner, P.; Rarey, M. Connected Subgraph Fingerprints: Representing Molecules Using Exhaustive Subgraph Enumeration. *J. Chem. Inf. Model.* **2019**, *59* (11), 4625–4635. https://doi.org/10.1021/acs.jcim.9b00571.

Bellmann, L.; Penner, P.; Rarey, M. Topological Similarity Search in Large Combinatorial Fragment Spaces. *J. Chem. Inf. Model.* **2021**, *61* (1), 238–251. https://doi.org/10.1021/acs.jcim.0c00850.