Screening Commercial Compounds

When Creating Novel Drugs

Can leads or even drugs of the future be found in commercial space?

The size of the drug-like chemical space has been estimated be around 10^{63} molecules. Somewhere in this huge space, we should expect to find the drugs of the future. Aiming to make this space more accessible to scientists, several chemistry vendors have launched commercial databases of drug-like compounds, available off-the-shelf or with a predicted high synthetic availability. Many pharma companies are also launching proprietary databases [1] (*Figure 1*).

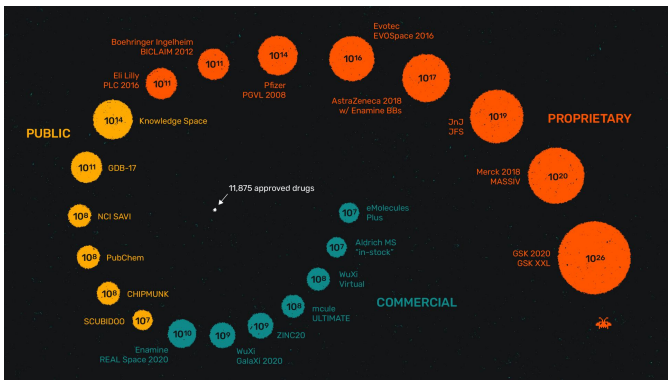

Figure 1. The chemical space of available small molecules is continuously increasing.

How to search commercial space

When making a virtual screen in chemical space, the scientists usually starts with a template, being a reference compound or a hit from a screen. One of the fastest way to screen large databases is via fingerprint methods [2].

When to search commercial space

In order to investigate what kind of molecules can be found in commercial space, we have run fingerprint-based similarity searches for leads, drugs and their analogues in 10 different small molecule drug discovery projects [3]. The results show that the highest similarity scores and the highest number of analogues are found for less decorated templates (*Figure 2*). This applies to both lead-like and drug-like compounds. Examples of templates and hits are shown in *Figure 3*.

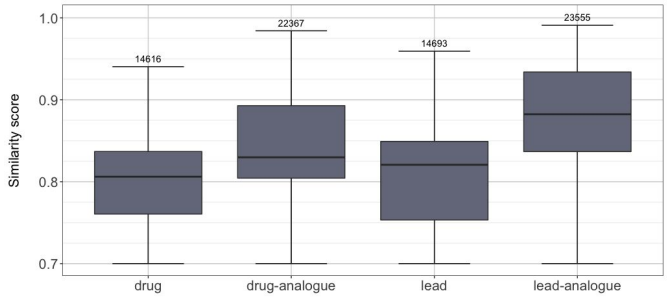

Figure 2. Overview of mean similarity score and number of hits in Enamine Space with a Tanimoto similarity score > 0.7 for leads, drugs and their respective analogues from 10 different drug discovery projects [3]. Visualization uses Five-number Summary

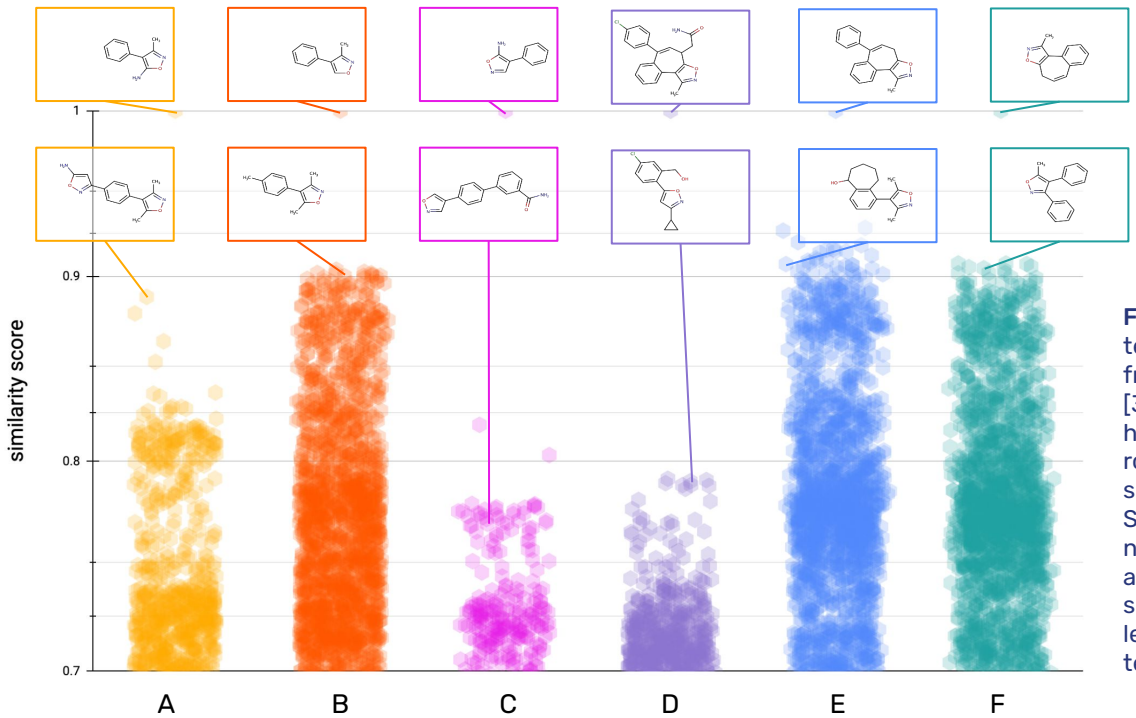

Figure 3. Examples of templates used (A-F) from the BRD4 project [3] (top row) to find hits (examples in 2nd row) with a similarity score > 0.7 in Enamine Space. Both the number of hits found and the similarity scores are higher for less decorated templates.

Table 6. Catalog of available databases in Design Hub

Purchasable compounds

- eMolecules
- MolPort
- Enamine REAL
- Enamine Store
- Mcule Ultimate

IP and scientific literature

- Reaxys
- SureChEMBL
- DrugBank
- ChEMBL
- PubChem

Submit your additional requests of other commercial or literature databases to ml-support@chemaxon.com

You can also plug-in your own internal databases, virtual spaces etc.

# hits	Mean time to retrieve hits (s)	
	similarity	substructure
200	3.8	3.4
2000	3.9	3.5
20.000	4.3	4.3

Table 7. Mean time for retrieving hits from Enamine Space using 50 templates

Extra step to sort substructure searches

Substructure searches are complementary to fingerprint searches, in case of looking for analogues to a certain core structure. Depending on the project-related question, close or diverse analogues are of the highest interest. Figure 4 shows examples of the substructure hits for a certain lead template [3], sorted by similarity to the query. The scientist can select which type of substructure hits (more or less similar) are desired to follow up on.

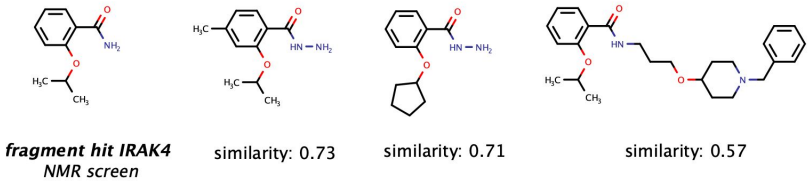

Figure 4. Lead structure from the IRAK4 project [3] and examples of substructure hits in Enamine Space, sorted by similarity to the query

Similarity score in the eyes of a chemist

Depending on what type of compounds are stored in the database, the hits with the highest similarity scores are not always favored by a medicinal chemist. For instance, many chemists would likely consider the hits in the RIPK1 project to be more similar to the template than in the SMN2 project, despite the similarity scores being lower for the RIPK1 analogues (*Figure 5*). Thus, additional filtering steps are often required to extract the best hits.

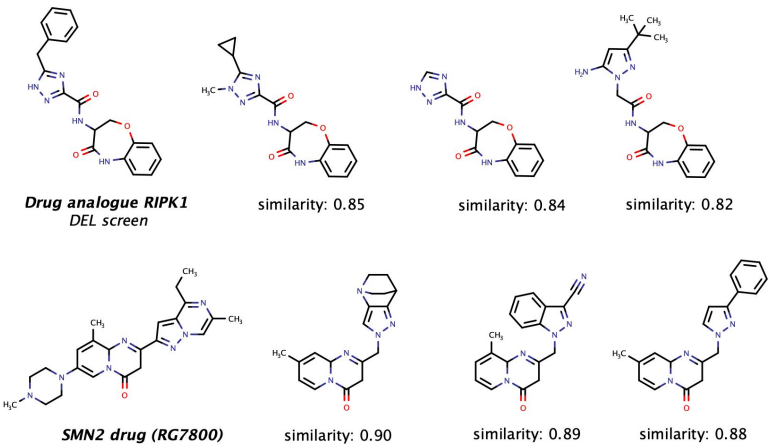

Figure 5. Examples of templates and the hits from Enamine Space:
a. Drug-analogue from the RIPK project [3]
b. Drug from the SMN2 project [3]

Plugins to "Your Own Space" and commercial resources

Design Hub offers plugins to several commercial and literature databases, as well as the possibility to generate and store "Your Own Space", available to search via MadFast fingerprint similarity as well as substructure search (*Table 6*). Searching in these databases returns hits within a few seconds (*Table 7*).

1. Warr W. Report on an NIH Workshop on Ultralarge Chemistry Databases. ChemRxiv. Cambridge: Cambridge Open Engage; 2021;
2. https://chemaxon.com/products/madfast
3. Brown et al, Where do recent small molecule clinical development candidates come from? J. Med. Chem. 2018, 61, 21, 9442–9468