What Structures are Claimed in Patents?

How to apply this knowledge when creating novel drugs

Medicinal chemistry transformations from patent literature

Designing and optimizing novel drugs require both creativity and knowledge. Using the Matched Molecular Pairs method is one way of supporting this process. Commonly, MMP is used to connect structural changes of drug molecules to corresponding changes in assay readouts (Figure 1).

The MMP method was used to extract all synthetically available transformations described in the patent database SureCHEMBL. Accordingly, it is possible to get an overview of how often medicinal chemists have used certain transformations, irrespective of their optimization parameters (Table 1).

SureChEMBL is a publicly available large-scale resource containing compounds extracted from the full text, images and attachments of patent documents. The data are extracted from the patent literature according to an automated text and image-mining pipeline, updated on a daily basis.

www.surechembl.org

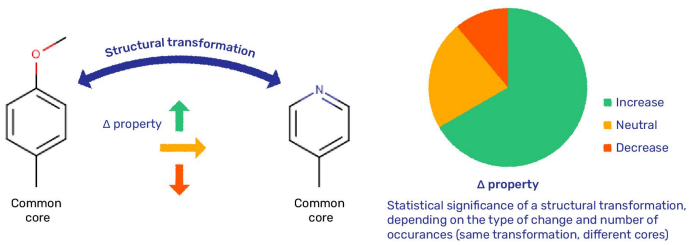

Figure 1. Description of the MMP method [1]. A structural transformation of an R-group is compared to other compounds with the same scaffold. The number of such transformations ("occurrences") can be connected to a statistically significant increase (or decrease) of a selected parameter of interest for optimization. In this case, only the number of occurring transformations in SureCHEMBL were counted, without a connection to any data..

Table 1. Data behind extracted transformations

Numbers	Data points used to create MMP set
61	Years of deposited patent applications
600 MB	of text
1.35 M	patent applications
20 M	exemplified compounds
1.4 M	Unique transformations
1-20	Transformation size (#atoms)
> 1000	Transformations with >1000 examples
>50.000	Transformations with > 50 examples

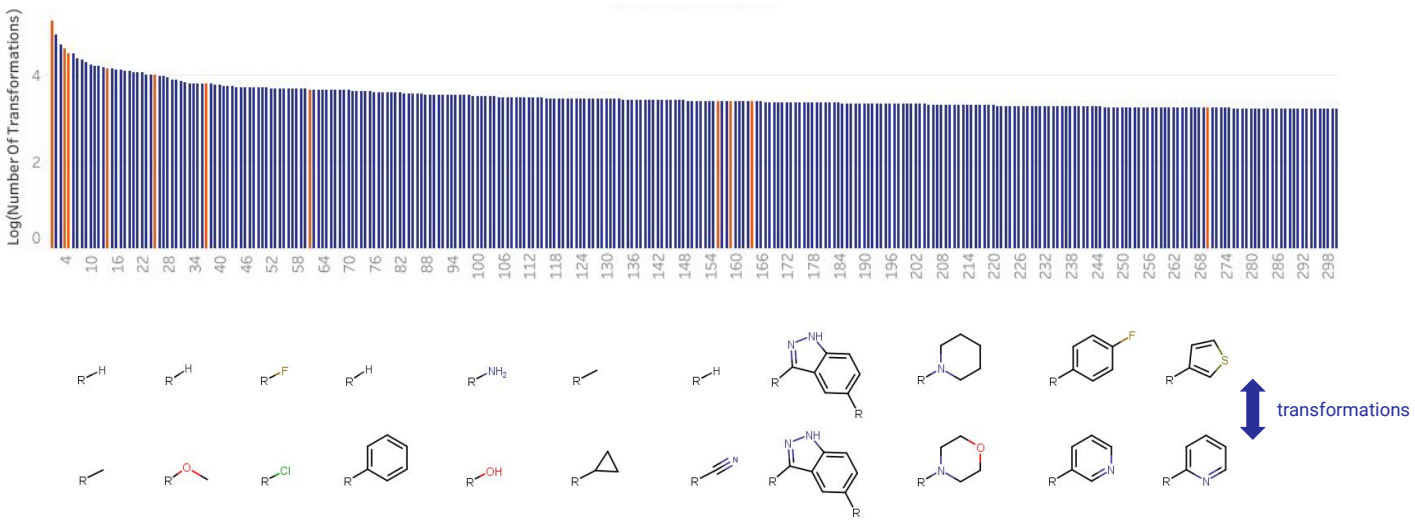

Figure 2. Number of occurrences and a few selected examples (orange bars) from the top 300 transformations (blue bars) in small molecule drug discovery projects, extracted from SureCHEMBL

What Structures are Claimed in Patents?

How to apply this knowledge when creating novel drugs

Use cases for patent literature MMP transformations

The MMP transformations from SureCHEMBL can be used in different ways to create analogues to a seed compound:

1. Based on the **most common** transformations [2]: automatic creation of compounds that are "expected" to be made in a project – making sure you don't forget any.

2. Based on the **least common** transformations: creation of analogues that are "unexpected" – compounds a medicinal chemist would not immediately think about, but could increase the chance of creating novel compounds

These analogues can then be filtered through any additional virtual screening cascade prior to selection for synthesis (Figure 3).

What is Design Hub?

Design Hub is an integrated DMTA application by ChemAxon for medicinal chemistry teams. Built on the best-in-class chemical drawing capabilities of Marvin JS, structure storage options of JChem Microservices, the application connects scientific rationale with compound tracking and computational resources needed for rational drug design.
This structured data then enables productivity boosts such as team Kanban boards, automatic status updates for compounds, or a universal query capability that combines chemical, text and metadata options.

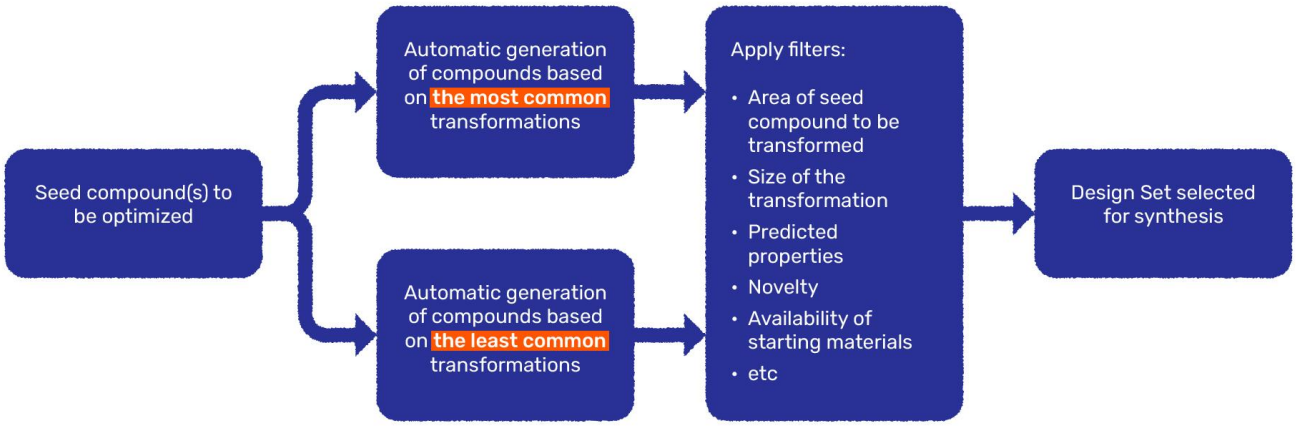

Figure 3. Example of workflows applying SureCHEMBL MMP transformations for creation of Design Sets

Download the 500 most common transformations from SureCHEMBL
Try it out yourself in Design Hub

[1] Hussain and Rea, Journal of Chemical Information and Modeling 2010 50 (3), 339-348