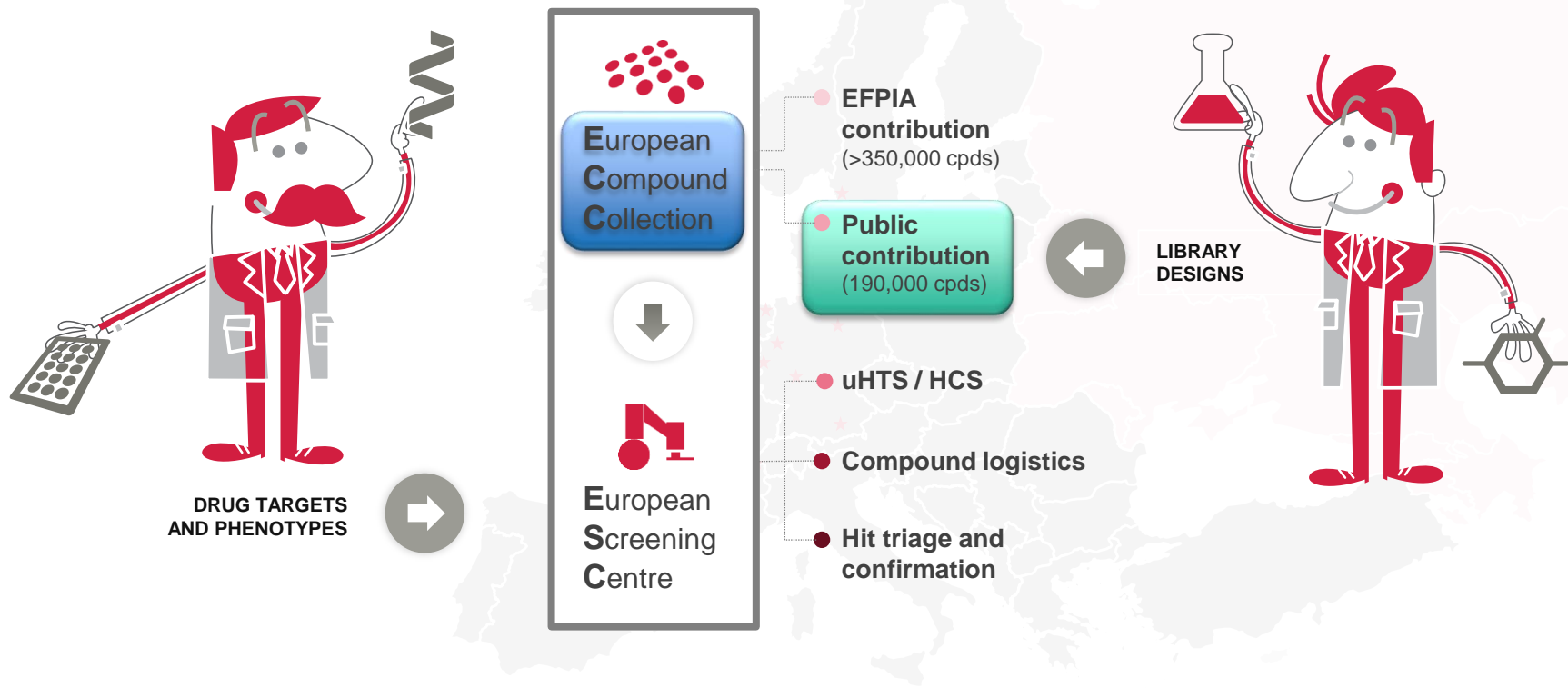


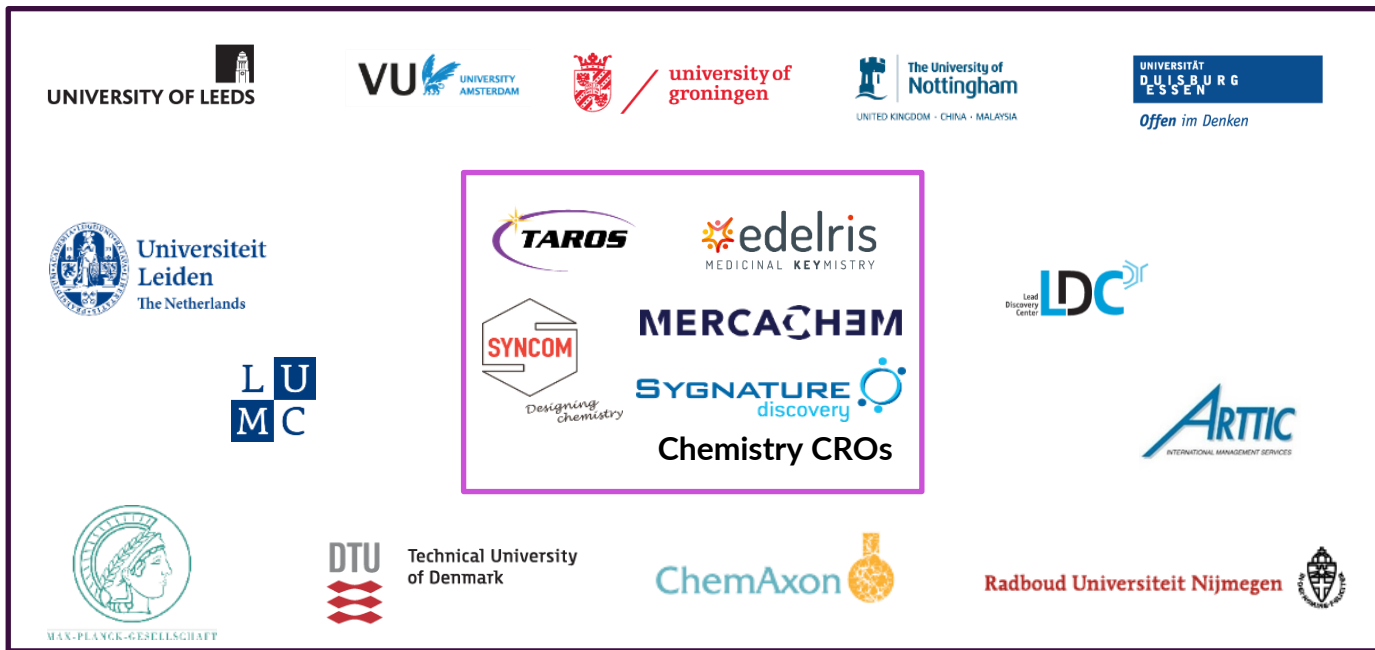
# The ELF compound collection : a novel and diverse drug-like screening collection

*This analysis was performed by Dr. Chimed Jansen (Symeres) and Dr. Hugues Lemoine (Edelris)*

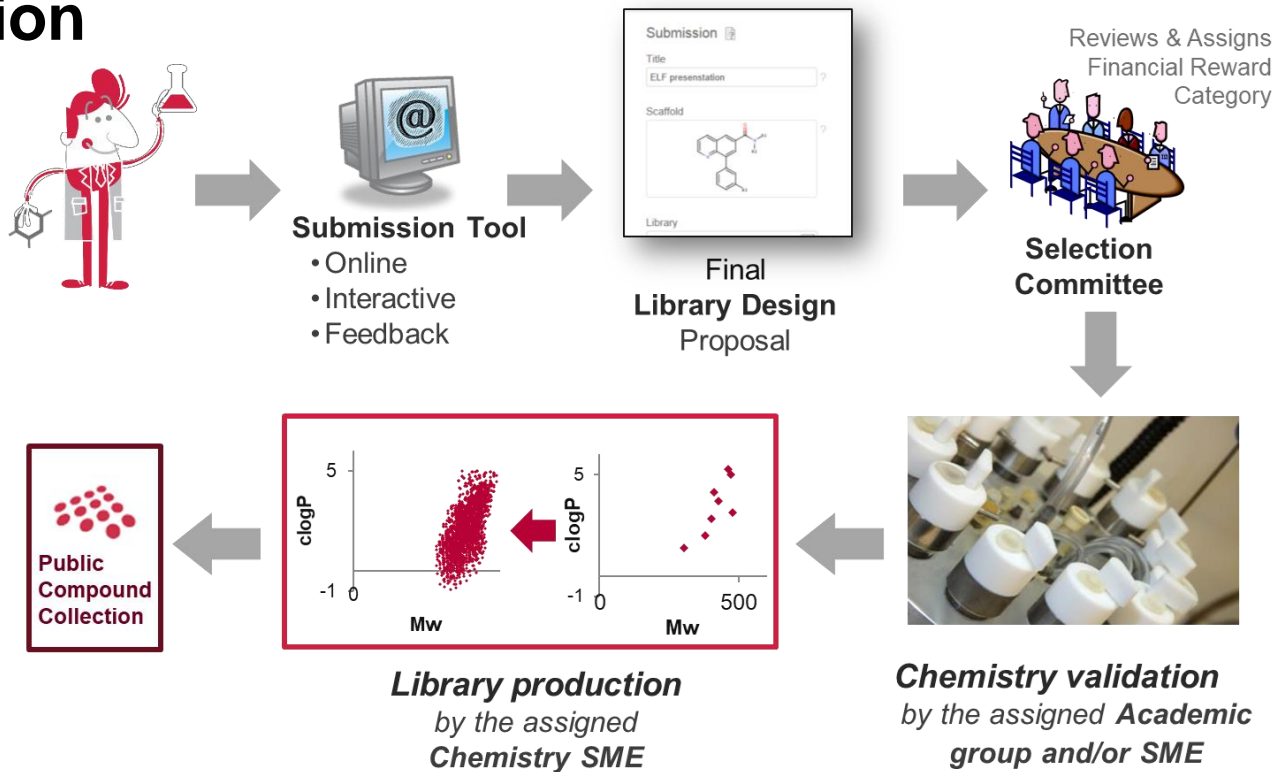
# A unique library is available for screening



# Public Compound Consortium (2013-2018)



# Design and synthesis of the Public Compound Collection



# PCC: Physical chemical properties

Distribution of physchem properties  
of 4 datasets :

PCC (200k)

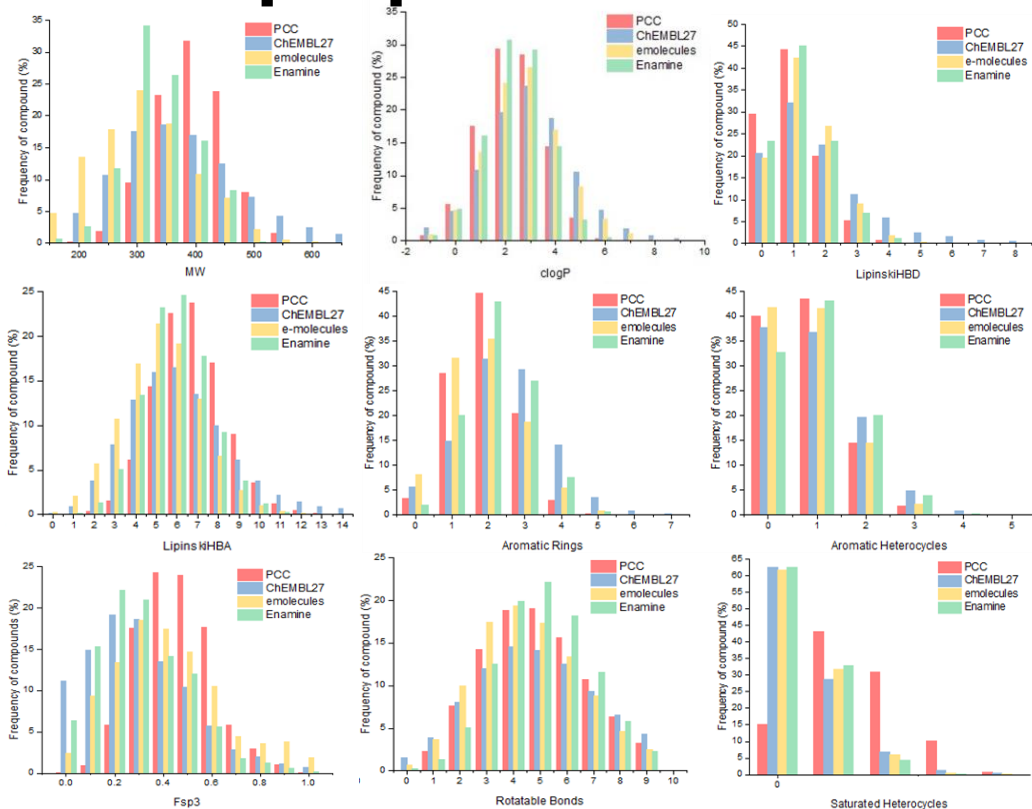
Enamine (1.5M)

E-Molecules (16M)

ChEMBL (1.5M)

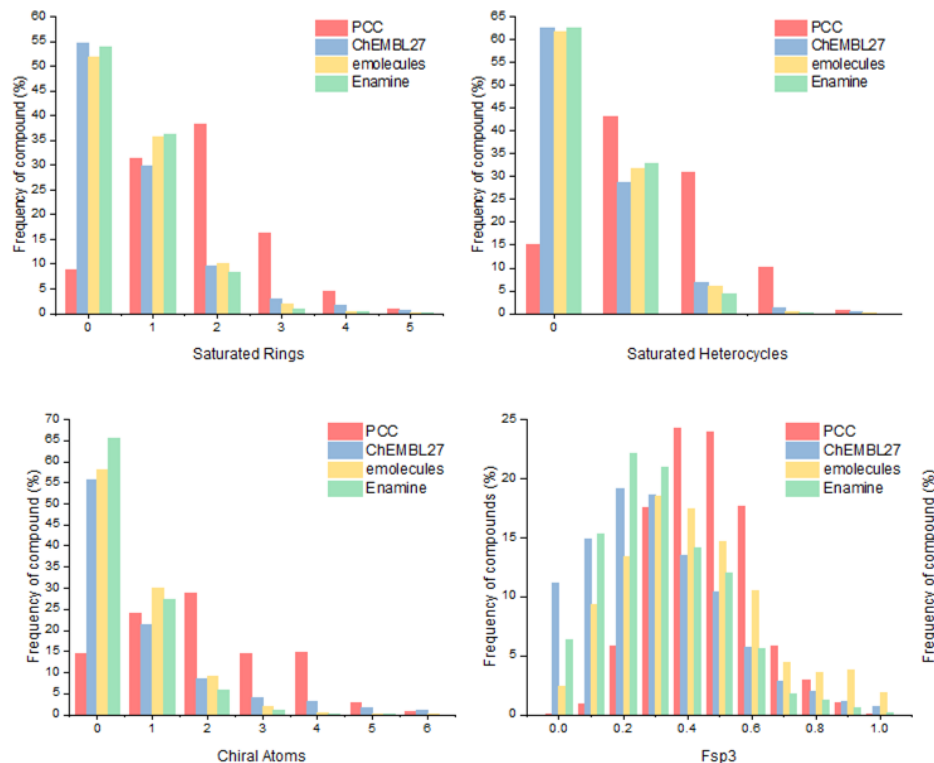
In short, the library can be  
summarized:

- Three dimensional/sp<sup>3</sup> rich
- Higher MW
- Low LogP



# PCC: A closer look at three dimensionality

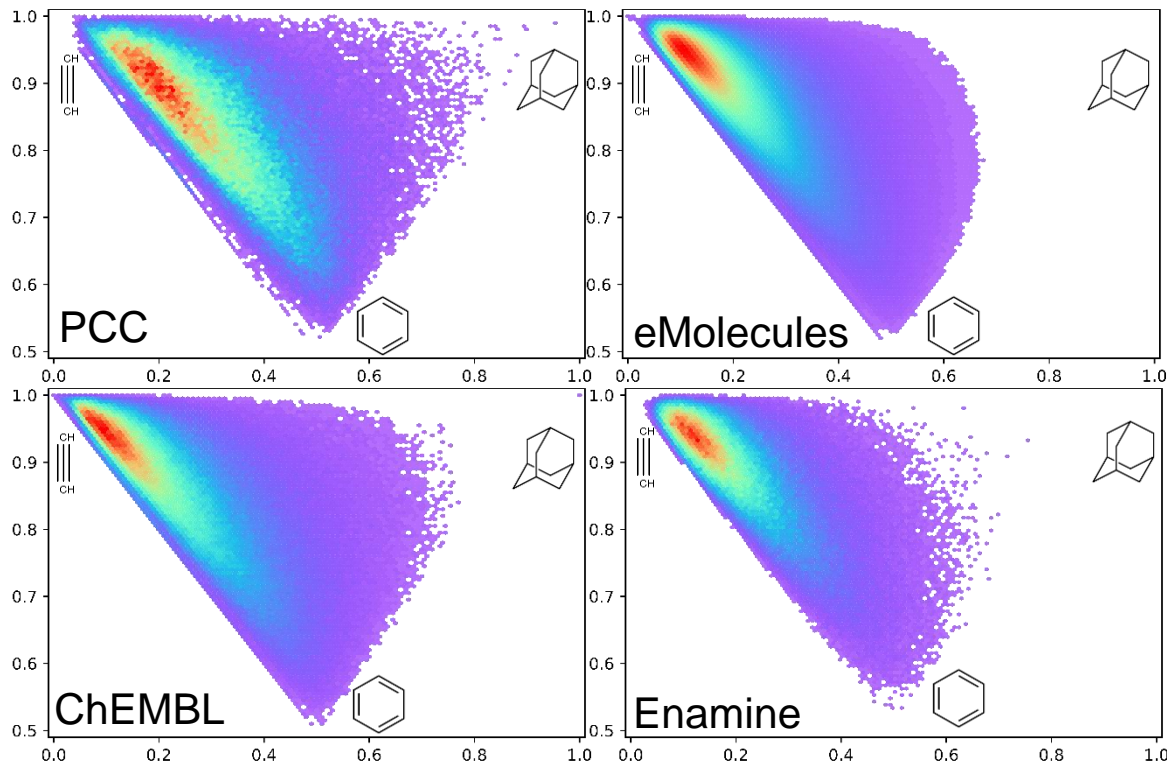
A tendency towards the synthesis of largely flat molecules within Drug Discovery is well documented despite planarity forming a potential hinderance to clinical success. Several key metrics can help show how the PCC breaks this trend by containing a high fraction of saturated carbons.



# PCC: PMI

Plotting the Principal Moment of Inertia can be used to gauge the distribution of three dimensionality in a library.

Here we see the PCC shows a greater spread from rod like towards both disc like and sphere like.

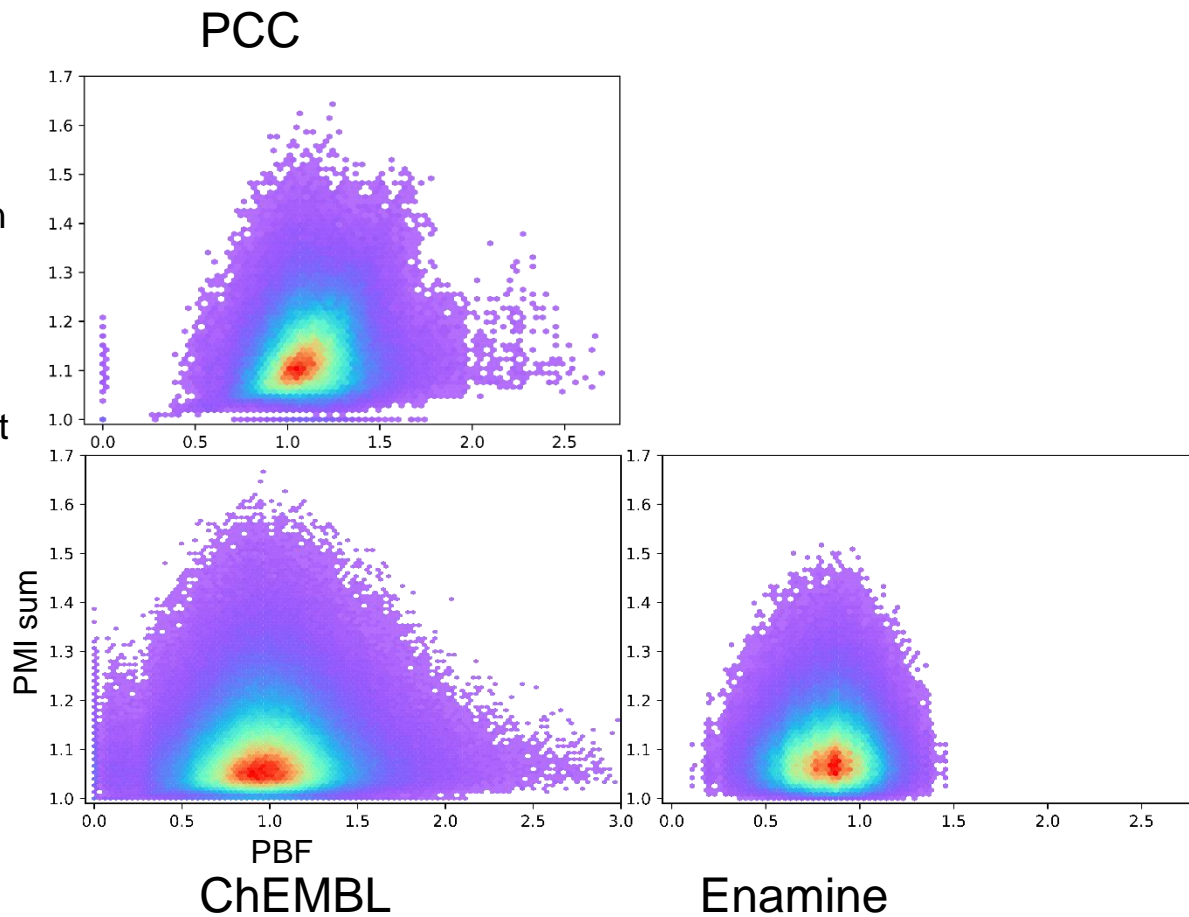




# PCC: PBF

The Plane of Best Fit provides an additional probe of the three-dimensionality of libraries.

Here the PBF is arranged against the sum of PMI values to fine tune the thresholds to use in library analysis on the two planes.

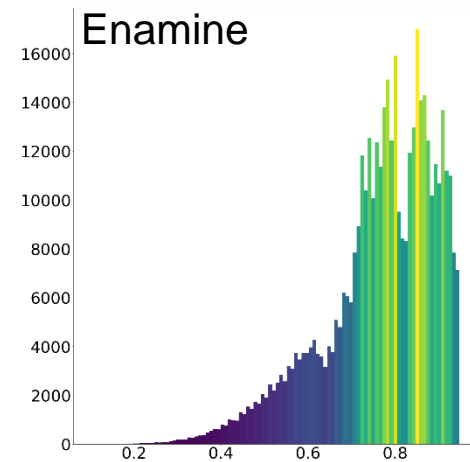
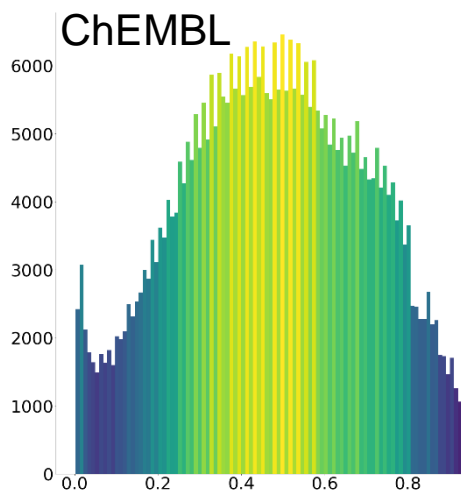
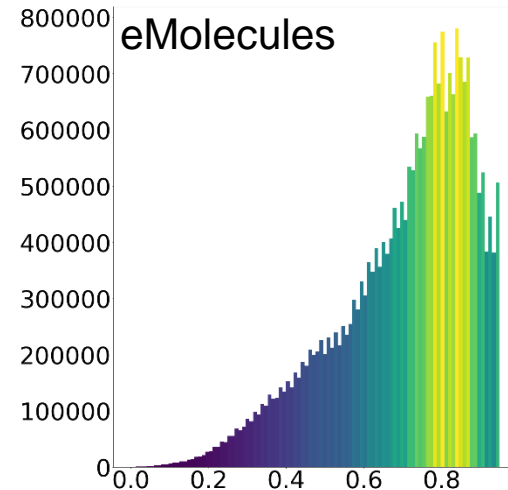
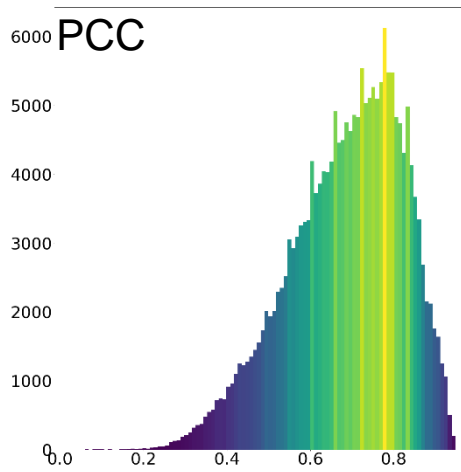




# PCC: QED

The Quantitative Estimate of Druglikeness is an aggregate property derived from functions that follow the property distributions of 771 orally available drugs. RDKit was used for the calculation with the following weights:

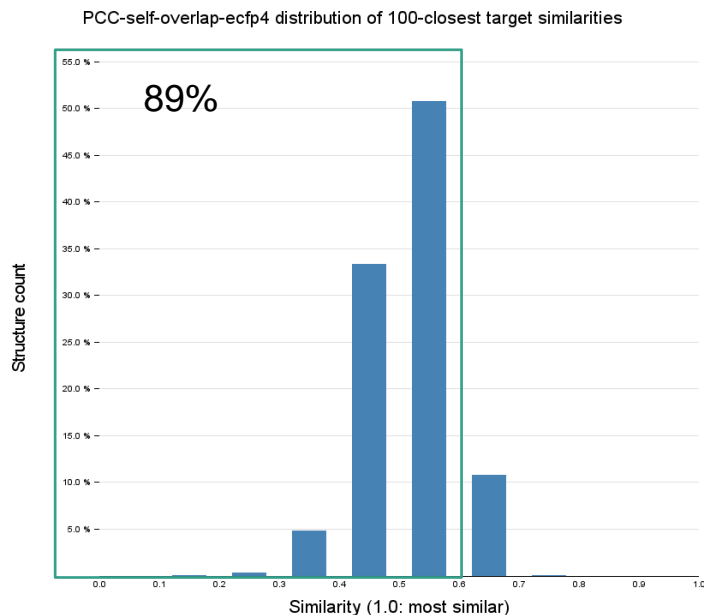
<i>MW</i> =0.66	<i>PSA</i> =0.06
<i>ALOGP</i> =0.46	<i>ROTB</i> =0.65
<i>HBA</i> =0.05	<i>AROM</i> =0.48
<i>HBD</i> =0.61	<i>ALERTS</i> =0.95



# PCC: Internal similarity

Distribution of the tanimoto similarity coefficient of the 100 closest neighbors

- Diversity analysis based on ECFP4 fingerprint
- 89% of chemical series of 100 compounds are highly dissimilar (coeff  $\leq 60\%$ )
- The majority of the PCC could be seen has cluster of 100 compounds

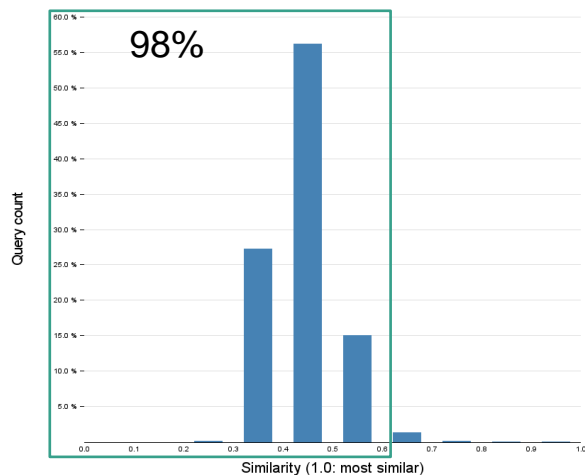


# PCC: Comparative similarity

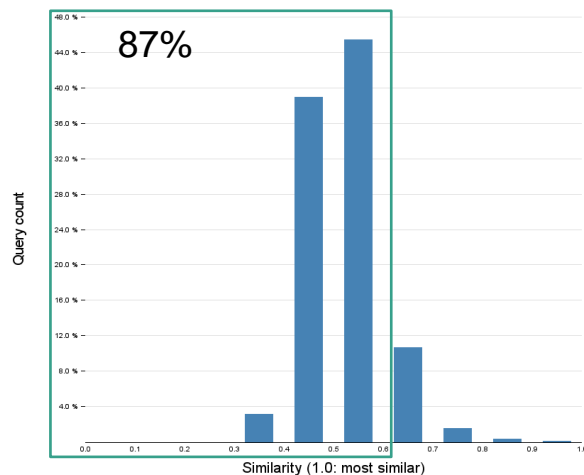
Distribution of the tanimoto similarity coefficient of the closest neighbors with public databases

- PCC show low similarity to very diverse ChEMBL database
- >85% of the PCC is dissimilar to the very large e-molecules database (16M)

PCC-ChEMBL-overlap-ecfp4 distribution of closest target similarities



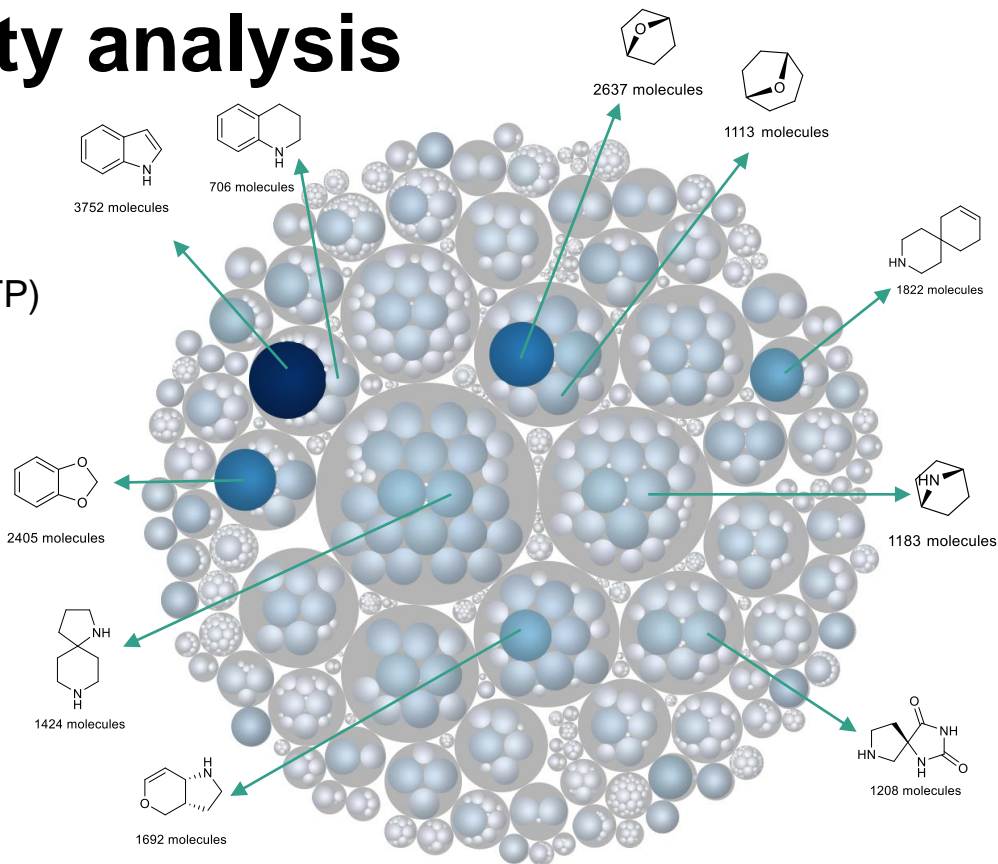
PCC-emolecules-overlap-ecfp4 distribution of closest target similarities



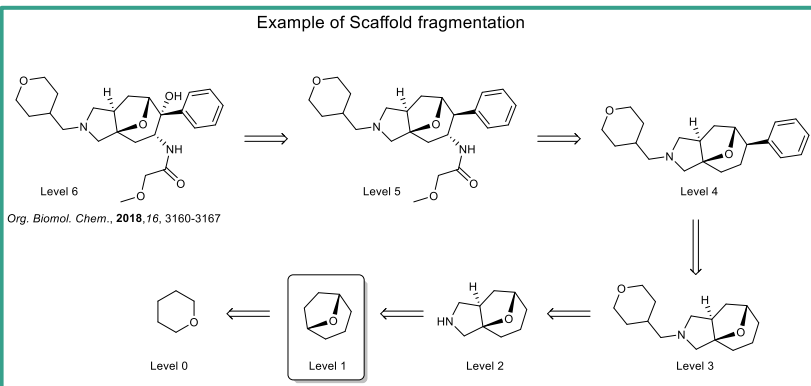
# PCC: Scaffold diversity analysis

## Tree maps analysis

- 2266 Level 1 Scaffolds
- 192 Clusters (60% similarity, SphereFP)
- No scaffold over represented
- High structural diversity



Example of Scaffold fragmentation



# Conclusion

The Public Compound Collection (PCC) synthesised by SMEs from 2013-2018 contains ~200 000 novel molecules

PCC explore novel chemical spaces (**High Fsp3, Saturated ring systems, novel cores**) not generally explored in standard HTS libraries, while still having attractive drug-like properties (**rule of 5, low logP, high QED**)



The ELF compound collection allow exploration of diverse chemical space and interpretable SAR in the screening output

# Acknowledgements

PCC Analysis:

- Chimed Jansen (Symeres)
- Hugues Lemoine (Edelris)



## THE ESCULAB PARTNERSHIP continues the ELF success story

A diverse partnership with a  
solid track record and great  
ambitions

