Consensus Diversity Plots: A Complete Diversity Analysis of Chemical Libraries for Drug Discovery

Mariana González-Medina,[1] Fernando D. Prieto-Martínez,[1] John R. Owen,[2] José L. Medina-Franco[1]

[1] Facultad de Química, Departamento de Farmacia, Universidad Nacional Autónoma de México (UNAM), Avenida Universidad 3000, Mexico City 04510, Mexico.

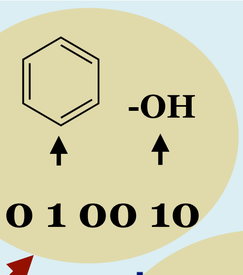
[2] High-Performance Computing Research Group, ECIT Institute, Northern Ireland Science Park, Queens Road, BT3 9DT Belfast, United Kingdom.

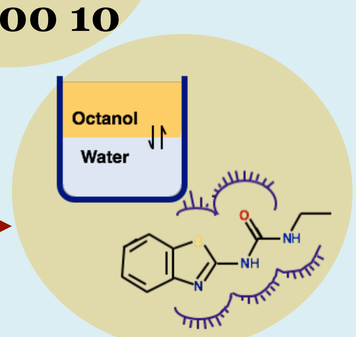
mgm_14392@comunidad.unam.mx, medinajl@unam.mx

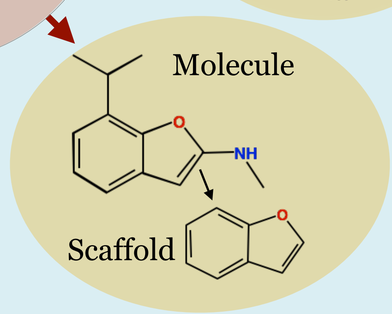
Molecular diversity

Measuring the diversity of compound data sets is relevant for drug discovery and many other areas of chemistry

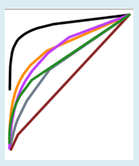
Molecular diversity depends on different molecular representations

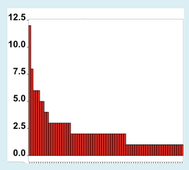


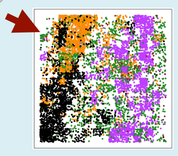




Cheminformatic analyses of diversity are performed using multiple criteria







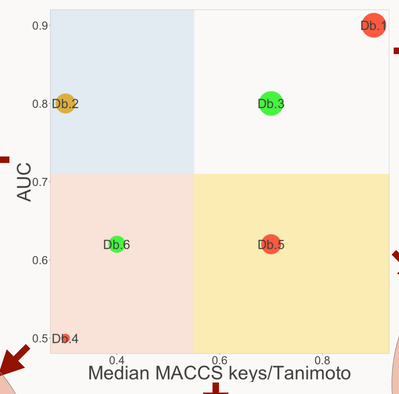
The assessment with each method is analyzed independently and is not straightforward to provide an evaluation of the"global diversity"

Goal

Visualize in few dimensions the diversity of several chemical libraries considering simultaneously multiple molecular representations

About the CDPs

It can be divided into four quadrants that classify data sets as high/low diverse considering both, fingerprints and scaffolds



Scaffold diversity plotted along the y axis

Each data point represents a data set and its size

Fingerprint diversity plotted along the x axis

Additional diversity criterion can be mapped with a categorical color

Molecular properties or complexity

Consensus Diversity Plots Version 2 (CDPs - V.2)

What can you do?

Scaffold diversity

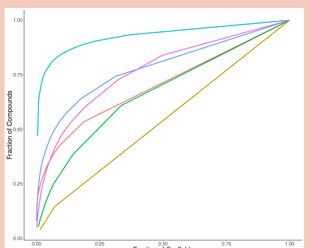
Input

Comma delimited file, with scaffold IDs

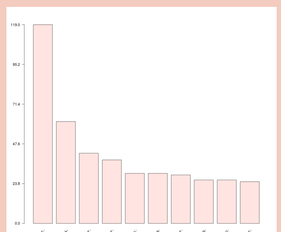
Cyclic System Recovery Curves

Scaled Shannon Entropy

Output



+ File with: AUC, F50, fraction of scaffolds and singletons



+ File with: SSE10 to SSE70

Output

Entire molecule diversity

Molecular fingerprints diversity

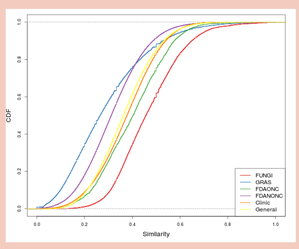
Input

Choose a fingerprint

Comma delimited file, with SMILES

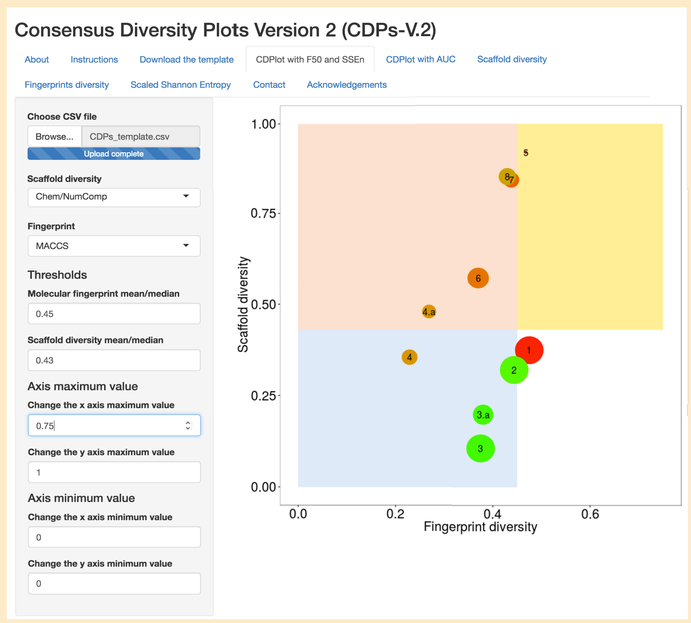
166 bits MACCS keys
ECFP4 and ECFP6
PubChem

Output



+ Tanimoto similarity summary

Consensus Diversity Plots



Perspectives

We want to keep improving this app! If you have suggestions or if you experience an error, please visit our website and send us an email

Add complexity as a property to be depicted by the color scale

Molecular properties diversity

Acknowledgments

UNAM: PAPIME PE200116 and PAIP 5000-9163

References

Owen JR.,et al. J Chem Inf Model. 2011, 51(7):1552–1563

Medina-Franco JL, et. al. QSAR Comb Sci. 2009, 28(11–12): 1551–1560

González-Medina M. et al. J Cheminform. 2016, 8, 63

www.difacquim.com