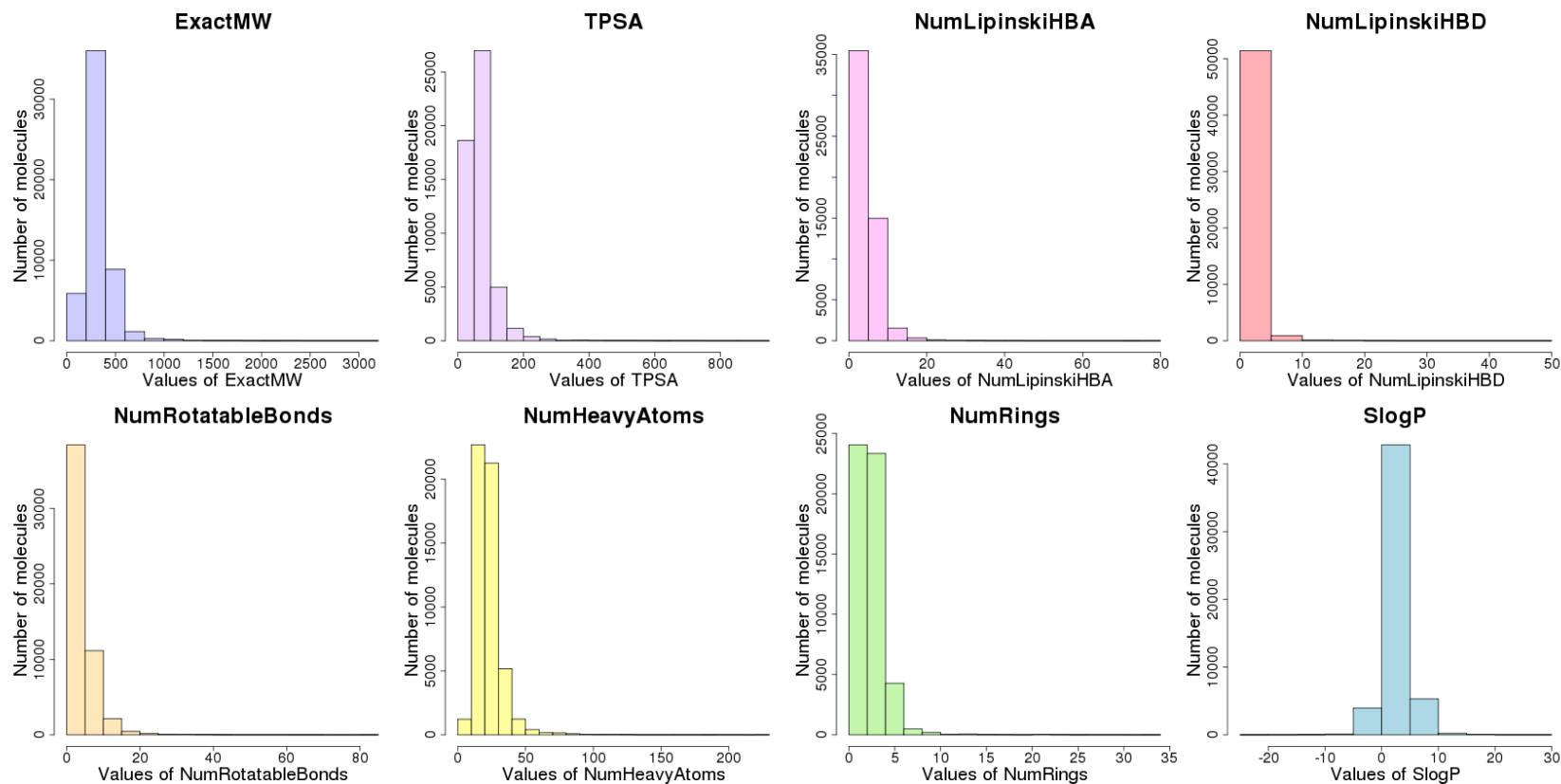


CN (55837 molecules)

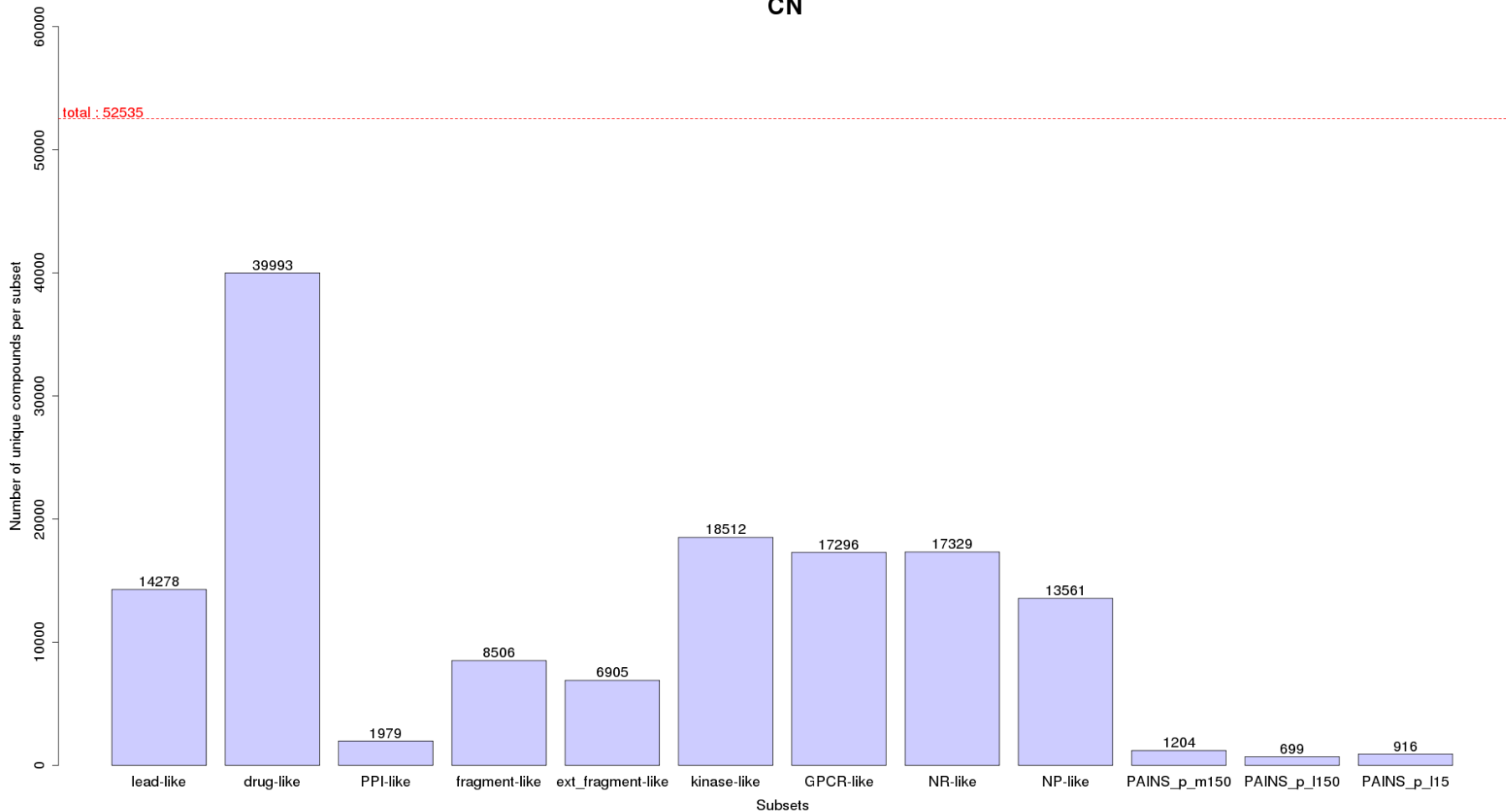
Distribution of RDKit molecular descriptors CN



CN (55837 molecules)

Number of compounds per subset

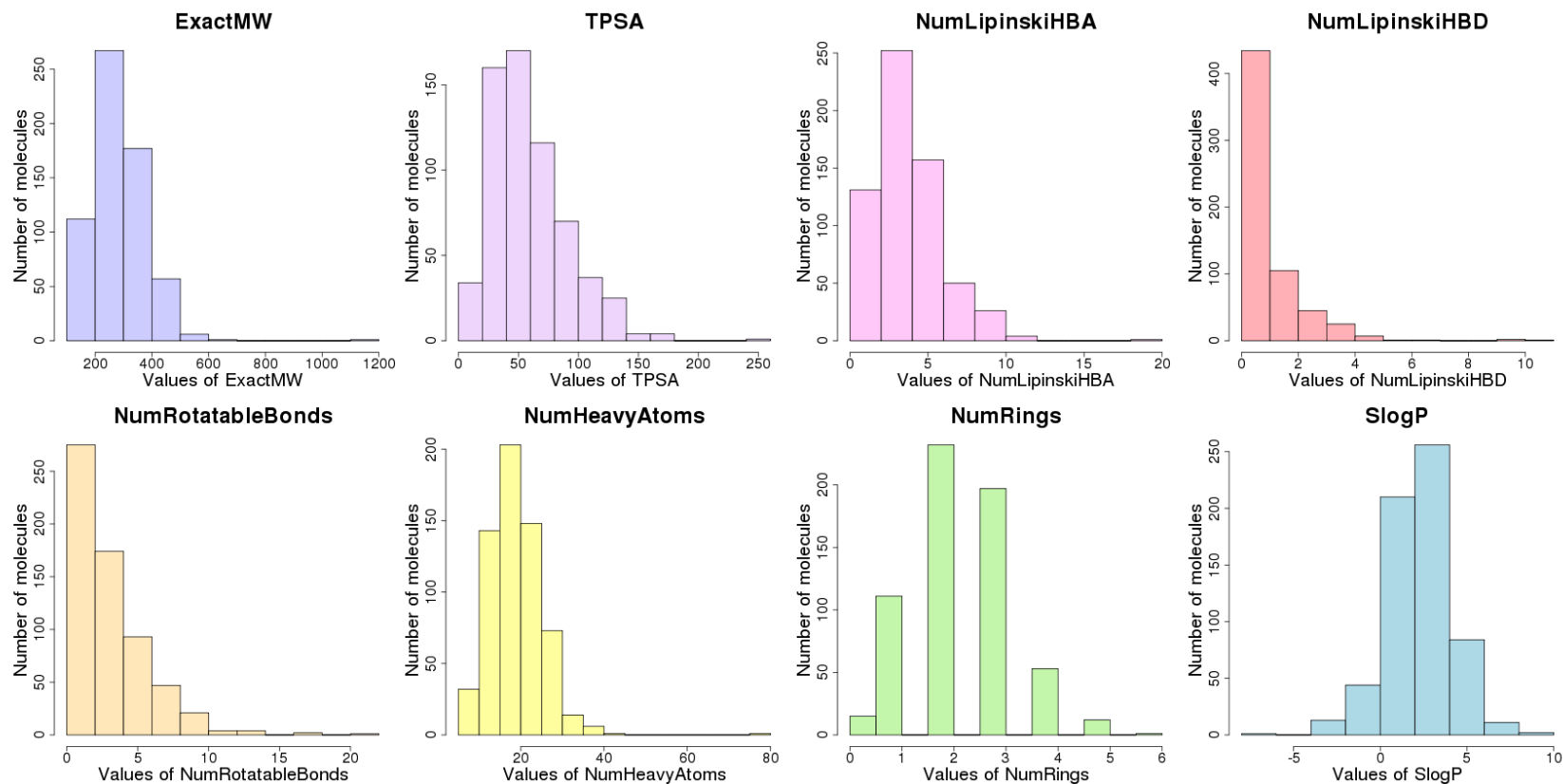
CN



Few molecules removed during the preparation of the molecules because they contain mixtures, bad atom types, inorganic atoms, etc.

CNE 640 molecules

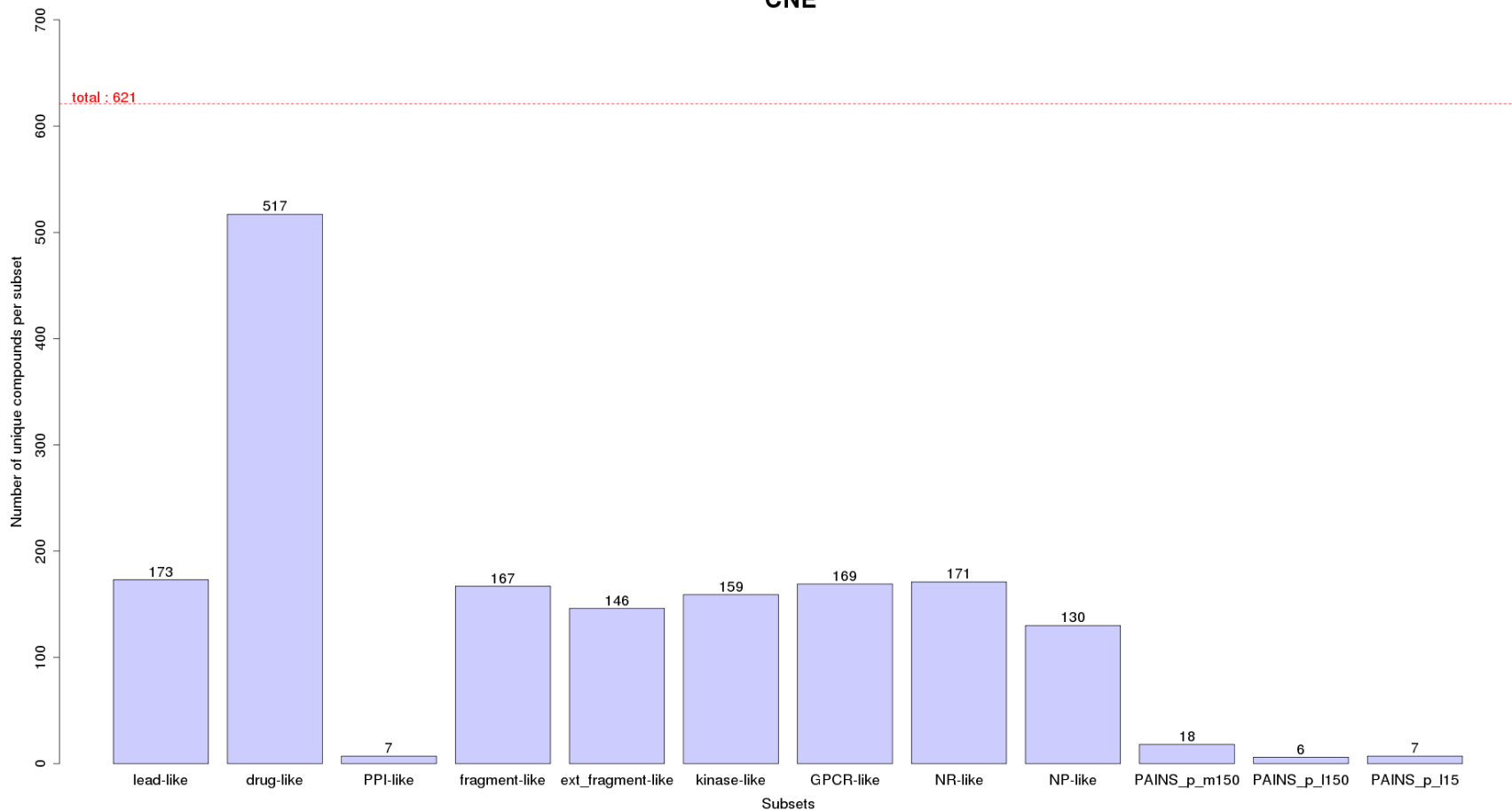
Distribution of RDKit molecular descriptors CNE



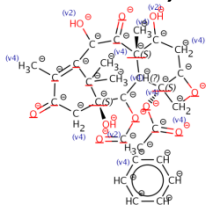
CNE 640 molecules

Number of compounds per subset

CNE



Few molecules removed during the preparation of the molecules because they contain mixtures, bad atom types, inorganic atoms, etc.



Molecular descriptors were computed using RDKit (<http://www.rdkit.org/>)
Lead-like, drug-like, PPI-like and fragment-like molecules were evaluated
based on published descriptor thresholds. (Lipinski, C. A. J. Pharm. Tox. Methods 44, 235–249
(2000); Teague, S. J. Ang. Chem. Int. Ed. 38, 3743–3748 (1999); Hamon, V. J R Soc Interface 11, 20130860–20130860
(2013); Congreve, M. Drug Discov. Today 8, 876–877 (2003))

Pains were computed by performing substructure search (Baell J.B. *et al.* J. Med. Chem.
53, 2719-2740 (2010))

NR-like (nuclear receptor), NP-like (natural products), GPCR-like and Kinase-
like molecules were calculated by similarity search of corresponding published
or in-house databases (Gatica E. A. *et al.* J. Chem. Inf. Model. 52, 1–6 (2011); Lagarde, N. *et al.* J. Med.
Chem. 57, 3117–3125(2014))

Calculations were performed by José-Manuel Gally, PhDstudent in the
Structural Bioinformatics and Chemoinformatics group of the ICOA, University
of Orléans, France.

For more information, please contact Prof. Pascal Bonnet:
pascal.bonnet@univ-orleans.fr

