**How are the Physicochemical Properties in ChEMBL Calculated?**

All properties are calculated on the parent form of the molecule i.e after any salts have been removed. The exception is the FULL\_WT, which, where applicable, is the molecular weight of the salt, plus any present hydrates.

Properties are only calculated on single component compounds (not mixtures) with molecular weight<1000 and containing only atoms H, C, N, O, S, P, F, Cl, Br and I. The exception is MW\_freebase which is calculated for all molecules.

**Section 1**

These calculations are performed using algorithms available in RDKit.

**MW\_freebase**

Molecular weight of the parent form of the molecule

**MW\_Monoisotopic**

The monoisotopic mass of the compound calculated as the sum of the masses of the most abundant isotopes in the compound.

**AlogP**

Calculated value for the lipophilicity of a molecule expressed as log (octanol/water partition coefficient). Method used for the calculation is as described in:

Prediction of Physicochemical Parameters by Atomic Contributions

Scott A. Wildman and and Gordon M. Crippen, Journal of Chemical Information and Computer Sciences 1999 39 (5), 868-873. DOI: 10.1021/ci990307l

**PSA**

Polar surface area is calculated by the method by P Ertl.

Fast calculation of molecular polar surface area as a sum of fragment based contributions and its application to the prediction of drug transport properties, Ertl, P., Rohde, B., Selzer, P., J. Med. Chem. 2000, 43, 3714-3717.

**HBA**

Count of the number of Hydrogen Bond Acceptors. Based on matching these SMARTS patterns:

[$([N;!H0;v3]),$([N;!H0;+1;v4]),$([O,S;H1;+0]),$([n;H1;+0])]

**HBD**

Count of the number of Hydrogen Bond Donors. Based on matching these SMARTS patterns:

[$([O,S;H1;v2]-[!$(\*=[O,N,P,S])]),$([O,S;H0;v2]),$([O,S;-]),$([N;v3;!$(N-\*=!@[O,N,P,S])]),$([nH0,o,s;+0])]

**HBA\_Lipinski**

Count of nitrogen and oxygen atoms in the molecule

**HBD\_Lipinski**

Count of hydrogens attached to nitrogen or oxygen atoms

**RTB**

Number of rotatable bonds in the molecule. Based on matching this SMARTS pattern:

[!$(\*#\*)&!D1]-&!@[!$(\*#\*)&!D1]

**Num\_RO5\_Violations**

Number of properties defined in Lipinski’s Rule of 5 (RO5) that the compound fails. Conditions which violate the RO5 are:

Molecular weight>500

AlogP>5

HBD>5

HBA>=10

**Num\_Lipinski\_RO5\_Violations**

As above except used HBA\_Lipinski and HBD\_Lipinski instead of HBA and HBD

Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. Adv. Drug Deliv. Rev., 1997, 23, 3-25.

**RO3\_Pass**

Rule of 3 passes. It is suggested that compounds that pass all these criteria are more likely to be hits in fragment screening.

molecular weight <=300,

number of hydrogen bond donors <=3,

number of hydrogen bond acceptors <=3

AlogP <=3.

RTB <=3

PSA<=60

A ‘Rule of Three’ for fragment-based lead discovery? Miles Congreve, Robin Carr,

Chris Murray and Harren Jhoti. Drug Discovery Today, 2003,8(19), 876-877

**Aromatic\_Rings**

The number of aromatic rings in the molecule

**Heavy\_Atoms**

The number of non-hydrogen atoms in the molecule

**QED\_Weighted**

This is the quantitative estimate of drug-likeness as described in:

“Quantifying the chemical beauty of drugs”

G. Richard Bickerton, Gaia V. Paolini, Jeremy Besnard, Sorel Muresan and Andrew L. Hopkins. Nature Chemistry, 2012, 4, 90-98

The values range from 0 -1 where 1 is the most drug-like and 0 the least drug-like.

**Section 2**

These properties are calculated using ChemAxon tools (ChEMBL\_26 onwards).

CX\_LogP is the calculated Octanol/Water Partition Coefficient.

CX\_LogD is the calculated Octanol/Water Distribution Coefficient at pH7.4. This is defined as the ratio of concentrations of all molecular species (neutral & ionized) in octanol divided by the concentration of all species in aqueous media at the pH specified.

pKa is defined as -log10 Ka, where Ka is the dissociation constant:

HA H+ + A-

BH+ B + H+

Ka = [H+][A-]/[HA]

Ka = [H+][B]/[BH+]

CX\_MOST\_APKA

Acidic pKa is the pKa for the most acidic group of the molecule

CX\_MOST\_BPKA

Basic pKa is the pKa for the most basic group of the molecule

**Molecular Species**

An approximation of the species occurring at pH7.4 and can be ACID, BASE, NEUTRAL or ZWITTERION

These are defined according to the definitions:

Acid(A) ACD\_MOST\_ApKa <6.5 and ACD\_MOST\_BpKa<8.5

Base (B) ACD\_MOST\_ApKa >6.5 and ACD\_MOST\_BpKa>8.5

Neutral (N) ACD\_MOST\_ApKa >6.5 and ACD\_MOST\_BpKa<8.5

Zwitterion (ZW) ACD\_MOST\_ApKa <6.5 and ACD\_MOST\_BpKa>8.5

The molecular species is an approximation. This does not use absolute pKa and considers both most acidic and most basic pKa; compounds may be polyprotic. The calculation of pKa is temperature-dependant; further details on the ChemAxon pKa calculations can be found here - https://docs.chemaxon.com/display/docs/calculators\_pka-calculation.md