Development of a Machine Learning model for predicting atomic 3D contributions to the logP

Carlos Cruz Marín[1,2]; Enric Herrero[1]; Javier Vázquez[1]; Francisco Javier Luque[2]

[1] Departament de Recerca a Pharmacelera
[2] Dept. de Nutrició, Ciències de l'Alimentació i Gastronomia, Institut de Química Teòrica i Computacional, Universitat de Barcelona

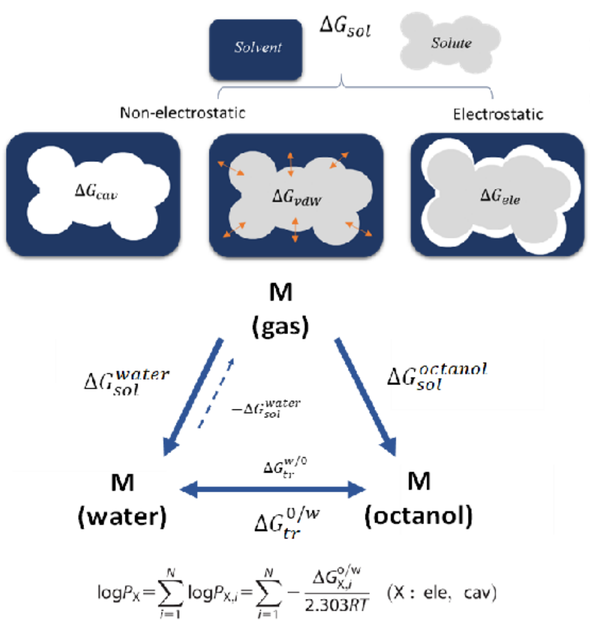
Introduction

- The octanol-water partition coefficient (**logP**) is a fundamental property in the drug design process. It is widely used for stablishing predictive models of the compound's pharmacokinetics characterized by the **ADME** properties (Adsorption, Distribution, Metabolism and Excretion). From another side, it gives important information about the (de)solvation, which determines the affinity of a potential drug with the receptor, guided by the 3D distribution of hydrophobic and hydrophilic regions and its complementary with the ligand.

- The experimental measurement of the logP has a significant cost, which makes necessary the usage of predictive models, specially considering the exponential growth of the chemical space, with libraries that contain billions of compounds.

- In this context, the development of Artificial Intelligence (**AI**) models gives a computational efficient solution to the problem, allowing the 3D hydrophobic description of the bioactive specie with a reduced computational cost, being applicable to virtual screening campaigns.

$$\Delta G_{sol} = \Delta G_{ele} + \Delta G_{cav} + \Delta G_{vW}$$



Objectives

- Maintain the accuracy of the method
- Increase the computation time

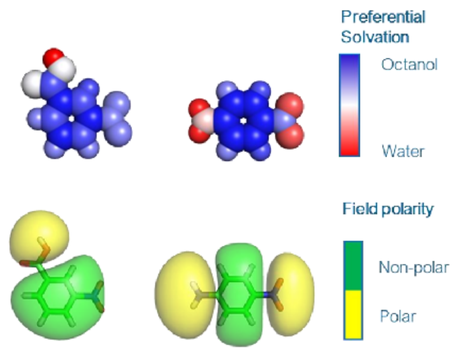
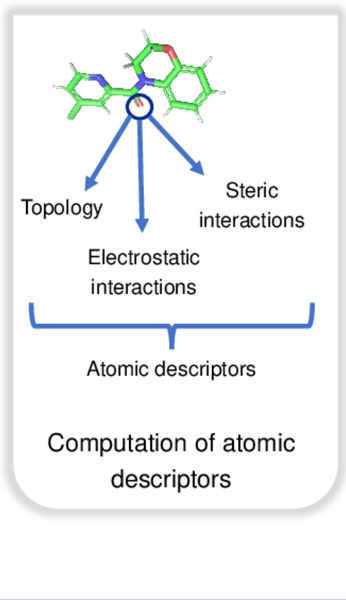
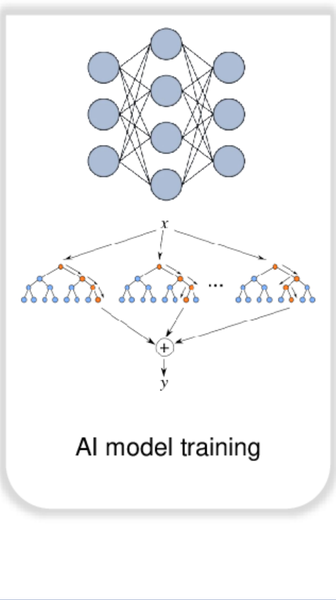


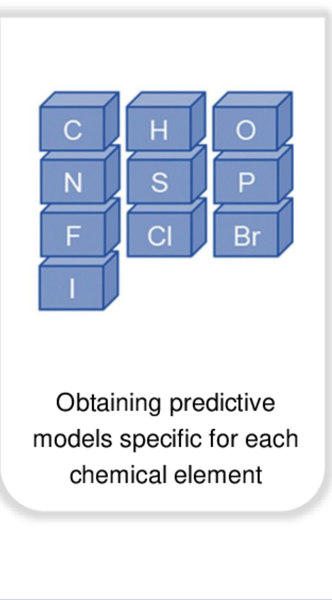
Figure 1. IEFPCM/MST model for solvation energy calculation.

Methodology

- The developed AI models have been trained using data obtained from calculations done using the IEFPCM/MST model at the semiempiric RM1 level[1,2] applied to the molecules from the database.
- The predictive capacity of the obtained models has been tested using different architectures, such as Random Forest Regressor or Artificial Neural Networks. For both cases, the optimization of the parameters have been carried out using a Bayesian Optimization process. The partition among the test and the train set is the standard 0.2-0.8.
- Models are developed for each element[3] individually.
- We used a set of 50.000 molecules (from Enamine HTS) as training (80%) and test sets (20%). An extra external validation set from Enamine HTS was also used, containing 50.000 molecules.







Atomic Descriptors

> Topology

Bonds and group of connections which form the 2D enviornment of the atom.

> Electrostatic interactions: D_{ele}

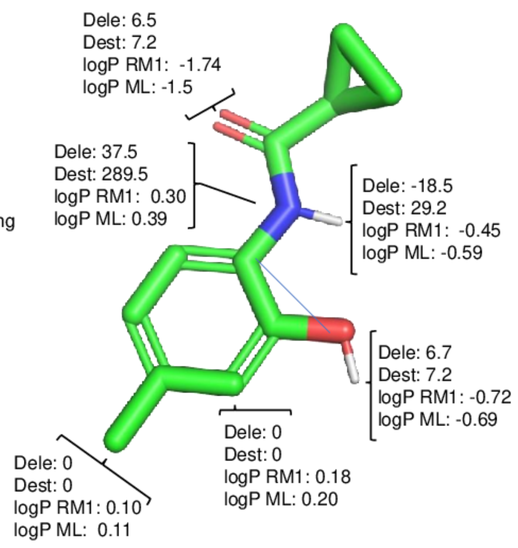
Electron-withdrawing environment — Electron-donating environment

- +

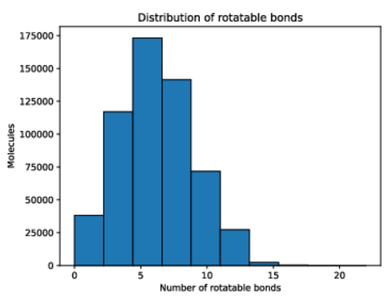
> Steric interactions: D_{est}

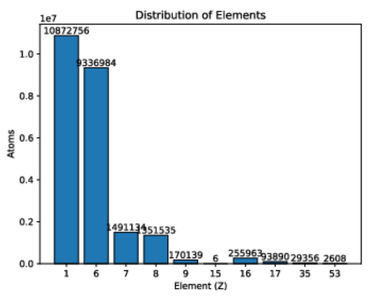
Sterically accessible — Sterically hindered

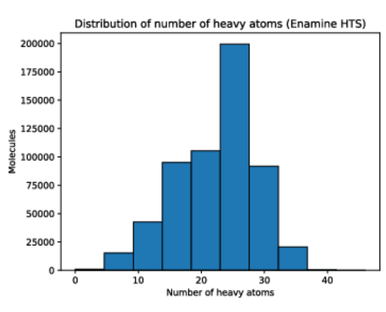
0 +

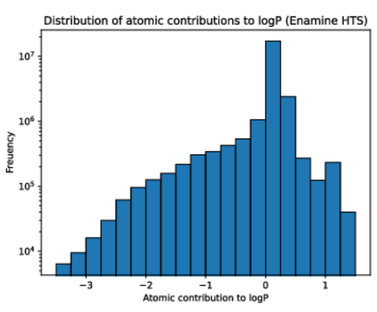


Sample characterization



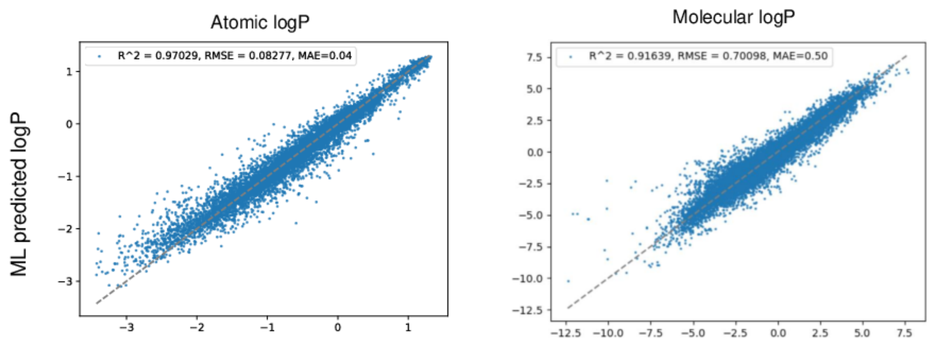






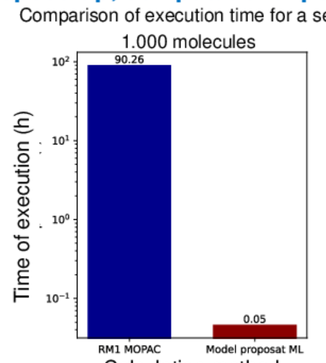
Results

Correlation with reference model

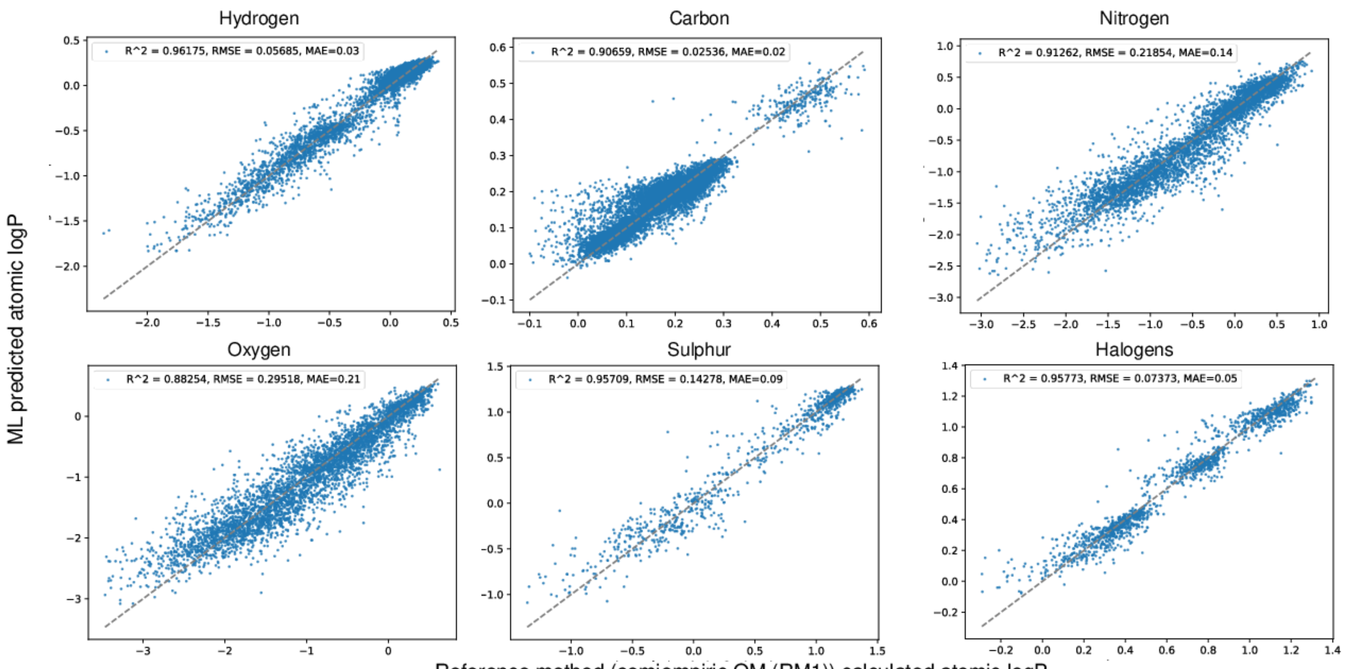


Speed up; computational power

Comparison of execution time for a set of 1.000 molecules



Specific results by element



Conclusions

- The balance between predictive accuracy (around 90%) and its computational cost is very promising, indicating that the computational strategy is effective for its application in virtual screening of chemical libraries.
- A 2000/2500-fold times faster calculation was obtained in comparison with the reference method (IEFPCM/MST-RM1) using 1 AWS c5.12xlarge, providing a cost-effectives methodology for calculating atomic and molecular logP, a key parameter in drug discovery.
- These results gives us a solid base for approaching to the development of methods based on more accurate parametrizations of the IEFPCM/MST model, for instance at the Density Functional Theory (DFT), increasing the quality of the description of the logP.

REFERENCES

[1] F. J. Luque, Barril X., Orozco M., Fractional Description of free energies of solvation, J. Comput. Aided Mol. Des. ,1998.
[2] Gerd B. Rocha, R. O. Freire, A. M. Simas, J. J. P. Stewart, A reparametrization of AM1 for H, C, N, O, P, S, F, Cl, Br and I, J. Comput. Chem.,2005.
[3] Bleiziffer P., Schaller K., Riniker S., Machine Learning of Partial Charges Derived from High-Quality Quantum-Mechanical Calculations, JCIM., 2018

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Cofinanciado por la Unión Europea
AGENCIA ESTATAL DE INVESTIGACIÓN
PDI2020-117646RB-I00 MICIU/AEI /10.13039/501100011033