

Proof of concept

Intersystems, Ajusteo and OPM

Sujet

Calculer des propriétés physico-chimiques d'un composé à partir de sa représentation en une dimension (SMILES). Convertir sa formule chimique en « *structured data file* » (SD file) et vérifier dans les deux sens de conversion, qu'à partir de l'un on puisse retrouver l'autre.

Documentation

Le format SMILES est un langage symbolique de description de la structure des molécules chimiques sous forme de courtes chaînes de caractères ASCII.

Les formats de données structurées utilisés en chimie pour la représentation par l'interprétation d'un algorithme de modélisation des composés sont appelés « *chemical table file* » (CT file). Il en existe plusieurs et certains sont détaillés [ici](#).

Evaluation

Afin de valider la bonne transformation des structures de données, il sera nécessaire de produire pour chaque composé :

- une image représentant deux molécules (l'une issue du SMILES et l'autre générée à partir du SD file produit) dont les parties communes seront surlignées ;
- pour chaque ligne, nous devons retrouver les propriétés suivantes :
 - o IUPAC name
 - o Formula
 - o MW
 - o SMILES
 - o ClogP
 - o ClogD
 - o TPSA
 - o pKa Basic 1st strongest
 - o H donor
 - o H acceptor
 - o Heavy atom count
 - o Rotatable bonds
 - o mpo
 - o mpo_clogP
 - o mpo_clogD
 - o mpo_PSA
 - o mpo_pka
 - o mpo_HBD
 - o mpo_MW
- le SD file généré

Contraintes techniques

Réaliser les essais à l'aide de la bibliothèque « *rdkit* » disponible en Python ([ici](#)) et/ou du plugin **Instant JChem** de l'éditeur Chemaxon.