

合成容易性評価プログラムの改良

詳細設計書

2018年2月5日

Copyright (C) 2006-2018 Next Generation Natural Product Chemistry (N²PC)

目次

1. 目的	3
2. 合成容易性評価プログラムの改良の仕様.....	4
3. 開発機能.....	5
3. 1. 部分構造データベース生成プログラムの開発.....	5
3. 2. タンパク質・化合物の結合エネルギー実測値のフォーマットの整備と入力機能....	7

1. 目的

分子設計において、薬物の作用と副作用を予測しながら、合成容易性も推算して、効率よく分子を設計する手法を改良し、実用化に向けて開発を行う。

本設計書では「副作用予測手法の改良」に対する設計内容を記載する。

2. 合成容易性評価プログラムの改良の仕様

化合物の合成容易性を、今まで evalcomb /myPresto として開発してきた。当ソフトは、入力化合物に含まれる部分構造、キラル中心の数、分子グラフの graph complexity(グラフ複雑度)などの頻度を、既に合成された化合物のデータベースに対して、その出現頻度を調べる。この出現頻度、キラル中心の数、グラフ複雑度などを基に、合成化学者の考える合成容易性に対し、重線形回帰分析をかけて各項の重みを計算する。このようにして、出現頻度の高い要素で構成される分子は合成が容易であり、分子の構成要素が既存化合物データベースに滅多に見られない場合は合成困難と判定する。

昨年度までの、ソフトウェア仕様では、新しい入力分子が与えられるごとに、毎回、化合物データベースの解析を行うため、1分子当たりの合成容易性の推算に10分程度が必要だった。

1入力分子当たりの合成容易性推算を1秒未満とすることを目標とし、本年度では、代表的な化合物について、分子フラグメントの出現頻度をあらかじめ調べて、2次データベースを生成し、合成化学者の考える合成容易性に対し、重線形回帰分析をかけて各項の寄与率を求めておく。予測したい分子に対して、部分構造の分解などを行い、この2次データベースから迅速に出現頻度を求め、合成容易性を推算するソフトウェアを開発する。

3. 開発機能

3. 1. 部分構造データベース生成プログラムの開発

複数の mol2 ファイルを入力し、指定したレベルの部分構造を探索し。

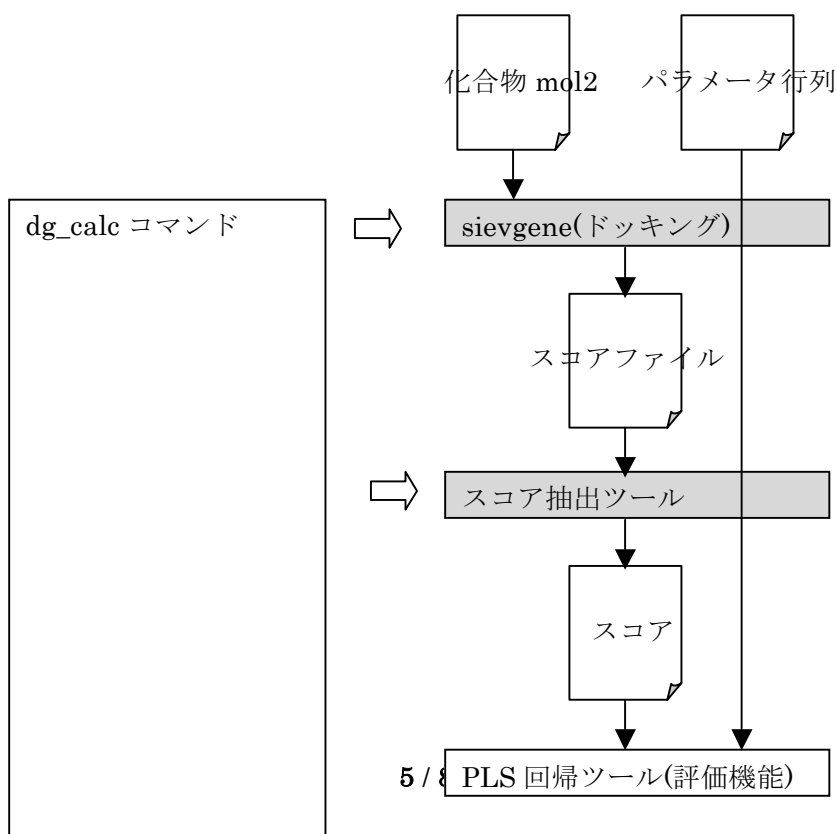
```
./createdb -i 対象 mol2 ファイルリスト -l 構造レベル -t [atom | bond] -o DB-name
```

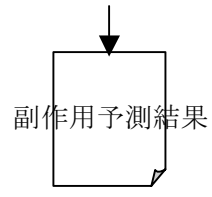
処理の高速化のため、下記の機能を実装する。

- (1) 部分構造を構成原子数ごとに分類する
- (2) 部分構造の構成原子名での比較を行い、一致しないものは結合グラフ比較を行わない
- (3)
- (4)

(1) システムの変更点

上記の dg_calc コマンドから、ドッキングプログラム、スコア抽出ツールを呼び出すことで当該化合物に対するドッキングスコアを抽出し、PLS 回帰による副作用予測結果を出力する。





副作用予測結果

3. 2. タンパク質・化合物の結合エネルギー実測値のフォーマットの整備と入力機能

4. ファイル形式

4. 1. ファイル

(1)原子数 : n

(2)原子タイプ : 5 文字*n

(3)結合行列 : $(n-1)*n/2$ *短精度実数

原子タイプは昇順に並べる。

Combine_Method.f90 をコンパイル対象ファイルに追加する。

以上