

# ***myPresto 5.0***

- 合成容易性予測ツール -

(*Synthetic Accessibility*)

USER MANUAL

2018/1/11

Copyright (C) 2006-2018 Next Generation Natural Product Chemistry (N<sup>2</sup>PC)

## 本ドキュメントについて

本ドキュメントは、「**myPresto 5.0 USER MANUAL**」の別冊です。コピーライト、プログラム使用許諾条件、著者および引用文献については、「**myPresto 5.0 USER MANUAL**」の記述に準じます。

## 謝辞

本ソフトウェアの研究開発は、国立研究開発法人日本医療研究開発機構(AMED)の援助によって行われました。ここに感謝の意を記します。

本ソフトウェアは、故・京極好正博士の始められた研究の中で開発されました。

# 目次

1. SyntheticAccessibility の概要 .....	4
2. SyntheticAccessibility のインストールとテスト計算の実行 .....	5
2.1. インストール方法 .....	5
2.2. テストプログラムの実行 .....	5
3. SyntheticAccessibility の実行 .....	6
4. データベースの作成方法 .....	7
4.1. データベースの作成 .....	7
4.2. 合成容易性の予測 .....	8
5. サンプルの実行 .....	9
5.1. データベースの作成 .....	9
5.2. SA 値の予測 .....	9
6. Reference .....	10

## 1. Synthetic Accessibility の概要

与えられた化合物の合成容易性を、分子の複雑度、光学活性中心、対称性から予測するプログラムです。合成容易性は、0(簡単)～10（困難）の数字で表示します。

予測は次の 2 つのステップで行います。

(1) データベースの生成

これは、化合物データベース含まれている分子の断片の出現頻度のデータベースです。化合物データベースは、利用可能な化合物から構築するがよいでしょう。

(2) クエリー分子に対する合成容易性予測

予測は、最初のステップで生成したデータベースに基づいて行われます。

## 2. SyntheticAccessibility のインストールとテスト計算の実行

SyntheticAccessibilityYYMMDD.tar.gz を Linux 計算機にインストールする手順を説明します。(YYMMDD には年月日を示す数字が入ります。)

### 2.1. インストール方法

SyntheticAccessibilityYYMMDD.tar.gz を、ユーザーが書き込み可能なディレクトリに配置してから、以下のコマンドを実行してください。インストールには、GNU の FORTAN コンパイラ(gfortran)もしくは、Intel の FORTRAN コンパイラ(ifort)が必要です

```
% tar -xzvf SyntheticAccessibilityYYMMDD.tar.gz  
% cd SyntheticAccessibilityYYMMDD
```

次のコマンドは、どちらか一方を実行します。

```
% bin/install.sh          (GNU のコンパイラを使用する場合)  
% bin/install.sh intel   (Intel のコンパイラを使用する場合)
```

### 2.2. テストプログラムの実行

次のコマンドで SyntheticAccessibility のテストプログラムを実行できます。

```
% bin/test_SA.sh
```

このコマンドを実行すると、サンプルデータについて、テスト計算を開始します。テストプログラムの出力先は、SyntheticAccessibilityYYMMDD/test\_SyntheticAccessibility/です。このコマンドは、データベースの作成と合成容易性予測の両方についてテストを実行します。

### 3. SyntheticAccessibility の実行

まず、作成済みのデータベースを使用する場合における SyntheticAccessibility の実行方法について説明します。テスト用の実行データベースは既に作成されています。3つのデータベース、atom\_3.db、atom\_4.db、bond\_2.db が用意されています。それぞれ、FA3、FA4、FB2 モデルに対応しています。FA4 モデルがいちばん詳細で、次が FA3、FB2 と荒くなります。データベースのサイズは、FA4 > FA3 > FB2 となります。データベースサイズが大きくなるにつれ、予測が正確になっていきます。

test\_a4 ディレクトリに、探索したい分子を input.mol2 というファイル名で置いてください。ファイルは、Sybyl mol2 フォーマットである必要があります。SA 予測をするには、SyntheticAccessibility プログラムを次のように起動してください。

```
% ./bin/SyntheticAccessibility -i input.mol2  
-d ../sampleDB/atom_4.db -cc 0.25856 -cg -1.6566E-03  
-ca -2.9625E-02 -ci 3.527E-02 -cb 0.0
```

この三行を改行せずに、一行で入力してから、リターンキーを押してください。パラメータは、FA4 モデルに最適化しています。

test\_a3、test\_a4、test\_b2 ディレクトリに用意した SyntheticAccessibility.sh ファイルには、それぞれ最適化されたパラメータが設定されています。

test\_a3、test\_a4、test\_b2 ディレクトリはそれぞれ、FA3、FA4、FB2 モデルに対応しています。

## 4. データベースの作成方法

合成容易性予測に使用するデータベースを自ら作ることができます。

### 4.1. データベースの作成

データベースを作成するには、利用可能な化合物を Sybyl mol2 フォーマットで用意してください。

create\_sampleDB.sh にデータベース作成コマンドのサンプルがあります。

createdb の入力は次のとおりです。

1 行目 : データベースとする mol2 ファイル名のリストのファイル名

2 行目 : モデルレベル

(1, 2, 3, 4 が FA1/FB1, FA2/FB2, FA3/FB3, FA4/FB4 に対応)

3 行目 : フラグメント化のタイプ

原子(ATOM)ベースか結合(BOND)ベースか

4 行目 : 出力のタイプ(TEXT または BIN)

5 行目 : 生成するデータベースの名前

入力のサンプル

mol2.list

3

ATOM

BIN

atom\_3.db

create\_sampleDB.sh のサンプルのように入力を与える方法の他に、

% ..../bin/createdb

として起動し、対話的にパラメータを入力することもできます。

## 4.2. 合成容易性の予測

合成容易性予測をするには、**SyntheticAccessibility** プログラムを次の様に起動します。

```
% ./bin/SyntheticAccessibility -i $mol -d ../sampleDB/atom_4.db -cc  
0.25856 -cg -1.6566E-03 -ca -2.9625E-02 -ci 3.527E-02 -cb 0.0
```

このコマンドは、1行で入力してリターンキーを押してください。

この項目は **SyntheticAccessibility** の実行と同様です。

## 5. サンプルの実行

### 5.1. データベースの作成

sampleDB ディレクトリに入り、atom\_3.db、atom\_4.db、bond\_2.db をコピーして別の場所に保管しておいてください。

その後に、次のコマンドを実行してください。

```
% ./create_sampleDB.sh
```

(単に分子の SA 予測をしたいだけの場合は、前述のコピーを忘れないでください。  
このコマンドで、これらのファイルは上書きされてしまいます。)

### 5.2. SA 値の予測

FA4 モデルを適用するには、test\_a4 ディレクトリに入り

```
% ./SyntheticAccessibility.sh
```

を実行してください。

このコマンドは、x.mol2 に対する SA 予測をします。

出力例

```
*****
*   SYNTHETIC ACCESSIBILITY      V1.0      *
*                                         2015/01/16  *
*****
INFORMATION>
 1) QUERY MOLECULE      : x.mol2
 2) SUBSTRUCTURE DB     : .../sampleDB/atom_4.db
(中略)
@ SyntheticAccessibility =      7.781

# 104937          MULT    1.2993652991851706E-288
# 104937          LOG     -287.88626873567256
# 104937          MIN     0.0000000000000000

INFORMATION> DB READ TIME = 8.99999961E-03 s
INFORMATION> ESTIMATE TIME = 7.00000022E-03 s
```

## 6. Reference

プログラムを使用する際は次の論文を参照ください。

Yoshifumi Fukunishi, Takashi Kurosawa, Yoshiaki Mikami, Haruki Nakamura.

Prediction of Synthetic Accessibility Based on Commercially Available Compound Databases.

Journal of Chemical Information and Modeling. (2014), 54 (12), 3259-3267.