

myPresto 5.0

- Screening with TGS method-

USER MANUAL

2018/01/12

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	TGS 法スクリーニングシステムの概要	4
2	システムのフェーズ構成	4
3	ディレクトリ構成	5
4	インストールと準備	6
5	テストプログラムの実行	6
6	操作手順	7
7	サンプルデータを使った実行例	9

1 TGS 法スクリーニングシステムの概要

TGS (Topology Graph Similarity) 法は、指定した化合物と似た化合物をデータベースの中から抽出する方法です。この方法では、分子の共有結合をエッジとした分子グラフをエッジ行列表示とし、その行列固有値を指標とし、化合物の類似性を探索します。分子の構造情報は、実数値のベクトルへと変換され、ベクトルの距離から類似性が計算されます。非常に高速ですが、光学異性体、配座を区別することはできません。

検索対象化合物(以下「クエリ化合物」とします)の立体構造を準備し、スクリーニング対象化合物を用いてスクリーニング計算を行うことによって、数百万個のスクリーニング対象化合物に対してクエリ化合物の探索を行うことができます。

手法の詳細については、文献を参照してください。

“A similarity search using molecular topological graphs”, Y. Fukunishi, H. Nakamura, Journal of Biomedicine and Biotechnology, Volume 2009 (2009), Article ID 231780

2 システムのフェーズ構成

TGS 法スクリーニングシステムは下記二つのフェーズで構成されています。

(1) 固有値 DB 作成フェーズ

スクリーニング対象化合物の分子グラフ固有値を計算し、テキスト形式の固有値 DB に化合物名と固有値を格納します。

本フェーズは一度行えば、スクリーニング対象化合物を変更・追加・削除しない限り再実行する必要はありません。

(2) スクリーニングフェーズ

クエリ化合物の分子グラフ固有値とスクリーニング対象化合物の固有値を比較し、スクリーニング対象上位化合物を検索します。

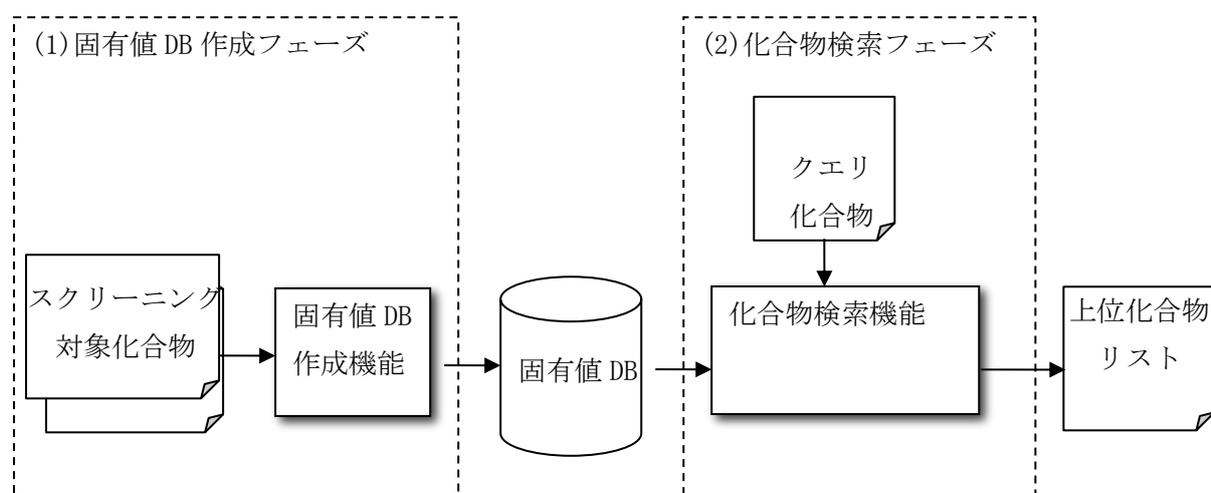


図 1. フェーズ構成

3 ディレクトリ構成

TGS法スクリーニング環境は図2のようにTGS法スクリーニング環境ディレクトリ以下、4つのディレクトリと5つのサブディレクトリから構成されています。

TGSyymmdd/	TGS法スクリーニング環境ディレクトリ
└database/	スクリーニング化合物ディレクトリ
└*/	スクリーニング化合物グループディレクトリ
└src/	ソースディレクトリ
└TGS/	TGS法プログラムディレクトリ
└bin/	実行ファイルディレクトリ
└screening/	スクリーニング作業ディレクトリ
└eigenDB/	固有値DB作成ディレクトリ
└query/	クエリ化合物ディレクトリ
└upper/	二次スクリーニングディレクトリ

図2. ディレクトリ構成

各ディレクトリの内容について説明します。

- (1) スクリーニング化合物ディレクトリ (database)
スクリーニング対象化合物を保存します。
- (2) スクリーニング化合物グループディレクトリ (*)
スクリーニング対象化合物グループを保存します。
- (3) ソースディレクトリ (src)
プログラムのソースを保存します。
- (4) TGS法プログラムディレクトリ (src/TGS)
本システムで使用するプログラムのソースを保存します。
- (5) 実行ファイルディレクトリ (bin)
本システムで使用する実行ファイルを保存します。
- (6) スクリーニング作業ディレクトリ (screening)
スクリーニング計算のためのコマンドを実行するディレクトリです。スクリーニング計算結果はすべてこのディレクトリの下に出力します。
- (7) 固有値DB作成ディレクトリ (eigenDB)
スクリーニング対象化合物の分子グラフ固有値を計算し、スクリーニング対象化合物名と固有値を保存します。
- (8) クエリ化合物ディレクトリ (query)
クエリ化合物(電荷情報付き MOL2 形式)を保存します。
- (9) 二次スクリーニングディレクトリ (upper)

二次スクリーニングの結果を保存します。

4 インストールと準備

TGS 法スクリーニング環境の準備手順を以下に示します。本システムの環境設定をするためには、GNU の FORTRAN コンパイラ(gfortran)、もしくは、intel の FORTRAN コンパイラ (ifort)が必要です。

(1) システムのインストール

本システムをインストールしたいディレクトリに移動します。“TGSyymmdd.tar.gz”をこのディレクトリに移動させ、以下のコマンドを実行します。(yymmdd には年月日を示す数字が入ります。)

```
% tar -xzvf TGSyymmdd.tar.gz
% cd TGS
```

以下のコマンドは、どちらか一方のみを実行します。

```
% bin/install.sh          (gfortran を使う場合)
% bin/install.sh intel    (ifort を使う場合)
```

install.sh コマンド実行後、“TGSyymmdd/bin/”に、“createDB”、“selectTGS”の実行プログラムが配置されます。

(2) スクリーニング対象化合物の配置

“TGSyymmdd/database/”ディレクトリの下にスクリーニング対象化合物(mol2 形式)を配置します。サンプルとして、“lig_sample/”ディレクトリが用意されています。実際に、使用するためには、このディレクトリを削除して、スクリーニング対象化合物を含めたディレクトリを用意してください。

5 テストプログラムの実行

サンプルデータを用いたテスト実行用プログラムを用意しています。このプログラムを実行することによって、インストールがうまくいっているか、結果ファイルが出力されるか等を確認できます。以下のプログラムで、テストプログラムを実行します。

```
% bin/test_TGS.sh
```

6 操作手順

TGS 法スクリーニング計算は以下の手順で実行します。

(1) クエリ化合物のデータ準備

"TGSyymmdd/screening/query/"ディレクトリにクエリ化合物(電荷情報付き MOL2 形式)を配置します。ファイル名は"query.mol2"とします。

(2) スクリーニング対象化合物の固有値 DB 作成

"TGSyymmdd/screening/"ディレクトリで、以下のコマンドを実行します。
(カレントディレクトリが"TGS"の場合)

```
% cd screening
% ../bin/createDB.sh
```

実行すると、"TGS/screening/eigenDB/"ディレクトリにスクリーニング対象化合物の固有値 DB が以下の形式で作成されます。

```
c001/0000014-01.mol2                14  18.524
-0.372  -0.391  -0.445  -0.504  -0.559  -0.695  -0.880  -1.099  -1.288  -1.674
-1.914  -3.193  -5.510   2.164   1.943   1.596   1.364   1.206   1.003   0.483
 0.375   0.112  -0.180  -0.353  -0.612  -0.949  -1.154
```

図 3. 固有値 DB ファイル例

(3) 1 次スクリーニングの実行

"TGSyymmdd/screening/"ディレクトリで、以下のコマンドを実行します。

```
% ../bin/1stSelect.sh
```

実行すると、スクリーニング対象化合物の上位 50 化合物の固有値 DB が "TGSyymmdd/screening/*/eigen.db" に作成されます。

```
c001/0112483-01.mol2                22  34.794
-0.377  -0.388  -0.399  -0.409  -0.415  -0.420  -0.611  -0.622  -0.745  -0.769
-0.937  -0.961  -1.006  -1.134  -1.598  -1.832  -1.884  -2.208  -2.335  -5.154
-10.589   2.421   2.141   2.027   1.760   1.473   1.370   1.355   1.198   0.972
 0.763   0.523   0.522   0.302   0.206  -0.055  -0.281  -0.717  -0.807  -0.831
-0.856  -1.143  -1.343
```

図 4. 1 次スクリーニング結果ファイル例

(4) 2次スクリーニングの実行

"TGSyymmdd/screening/"ディレクトリで、以下のコマンドを実行します。

```
% ../bin/2ndSelect.sh
```

実行すると、スクリーニング対象化合物の上位 1000 化合物の TGS スコアファイルが "TGSyymmdd/screening/upper/top.list" に作成されます。

1	c001/3336907-01.mo12	0.0000000000
2	c027/2842248-01.mo12	0.0000002398
3	c081/3003480-01.mo12	0.0000002513
4	c043/2332762-01.mo12	0.0000003463
5	c098/1484094-01.mo12	0.0000003463
6	c005/1202868-02.mo12	0.0000004023
7	c017/1202868-01.mo12	0.0000004023
8	c035/2368156-02.mo12	0.0000004058
9	c082/2368156-01.mo12	0.0000004058
10	c003/0401987-01.mo12	0.0000004105

図 5. 2次スクリーニング結果ファイル例(上位 10 化合物)

左から順位、化合物名、スコアが記載されています。スコアは類似性が高くなると小さい値を取り、同一化合物の場合は 0.0 の値となります。

7 サンプルデータを使った実行例

以下の手順でサンプルデータを用いて動作確認を行うことができます。サンプルの実行には上記手順の 1.4-(2) までの準備が必要です。スクリーニング対象化合物として "TGSyymmdd/database/lig_sample/" ディレクトリに 296 化合物データを配置しています。また、"TGSyymmdd/screening/query" ディレクトリにクエリ化合物データの "query.mol2" を配置しています。

(1) スクリーニング対象化合物の固有値 DB 作成

"TGSyymmdd/screening/" ディレクトリで、以下のコマンドを実行します。

```
% ../bin/createDB.sh
```

(2) 1 次スクリーニングの実行

"TGSyymmdd/screening/" ディレクトリで、以下のコマンドを実行します。

```
% ../bin/1stSelect.sh
```

(3) 2 次スクリーニングの実行

"TGSyymmdd/screening/" ディレクトリで、以下のコマンドを実行します。

```
% ../bin/2ndSelect.sh
```

実行後、スクリーニング対象化合物の上位 50 化合物の TGS スコアファイルが "TGSyymmdd/screening/upper/top.list" に作成されます。

以上

myPresto 5.0
- screening with TGS method -