

# ***myPresto 5.0***

*- 3DdataConstruction -*

USER MANUAL  
2018/01/11

## 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. 3DdataConstruction システムとは .....	1
1-1. 3DdataConstruction システムとは .....	1
1-2. 各ツールのディレクトリ構成.....	2
1-3. インストール方法.....	4
1-4. テスト計算プログラムの実行方法.....	4
2. 分子構造の3次元化処理ツール(create3d).....	5
2-1. ディレクトリ構成.....	5
2-2. TOOL 配下のツールについて .....	5
2-3. 使用方法 .....	6
2-3-1. create3d.pl.....	6
2-3-2. setLigandBoxId.pl.....	9
2-3-3. pickUpAnnotation.pl.....	10
3. 医薬品として不適切な分子のフィルタリングツール(Filtering).....	13
3-1. ディレクトリ構成.....	13
3-2. 使用方法 .....	13
3-2-1. Filtering.csh.....	13
3-2-2. SortByMW.csh.....	14
4. 類似分子のクラスタリングツール(Clustering).....	16
4-1. ディレクトリ構成.....	16
4-2. 使用方法 .....	16

---

---

## 1. 3DdataConstruction システムとは

### 1-1. 3DdataConstruction システムとは

3DdataConstruction システムは、低分子化合物の 2 次元構造を 3 次元化し、各種、分子物性情報をマルチ MOL2 ファイル形式で出力する機能を持ったツール群です。このツール群は、myPresto の一部として提供されている化合物データベース LigandBOX の作成に使われています。

本システムは以下のツールで構成されています。

#### ①分子構造の 3 次元化処理ツール(create3d)

MDL sdf データ (2 次元) から Sybyl MOL2 データ (3 次元) を作成します。また、各種分子物性値や光学異性体、幾何異性体を併せて作成します。

#### ②LigandBox ID 付与ツール

①で作成したマルチ MOL2 ファイルに対して、ユニークになる様に LigandBox ID を付与します。

#### ③医薬品として不適切な分子のフィルタリングツール(Filterring.csh)

医薬品として不適切な分子をフィルタリングし、医薬品としての候補化合物の MOL2 データを作成します。

#### ④類似分子のクラスタリングツール(Clustering.csh)

対象化合物データに類似分子が含まれる場合、代表 1 件のみを採用し、他の類似データを不採用とし、一定数の化合物データセットを作成します。

## 1-2. 各ツールのディレクトリ構成

以下に、各ツールのディレクトリ構成を示します。

```

3DdataConstructionYYMMDD (YYMMDD には年月日を示す数字が入ります。)
--README
--bin/
  |--install.sh
  |--Makefile
  |--Makefile.intel
  |--test_3DdataConstruction.sh
--CLUSTERING/
  |--Makefile
  |--Makefile.intel
  |--TGS/
    |--bin/
      |--(createDB) (install.sh 実行後に出現)
      |--(selectTGS) (install.sh 実行後に出現)
    --database/
    --screening/
      |--Adjustment_clustering.pl
      |--Clustering.csh
      |--command.sh
      |--eigenDB
      |--Judge_SelectMol.pl
      |--PickUp_data.pl
      |--query
      |--RESULT
      |--SelectMol2_1st.pl
      |--upper
    --src/
      |--TGS/
--doc/
--FILTERING/
  |--ExcludeMol/
    |--DB/
    |--README.txt
    |--src/
      |--(FindExcludeMol) (install.sh 実行後に出現)
    --TOOL/
  |--Filtering_agri.csh
  |--Filtering.csh
  |--FilteringSortMW.pl
  |--Makefile
  |--MoveMol2toDir.pl
  |--PickUpAndSetMol2OnDir.pl
  |--ReConst2MW.pl
  |--SetRandom.pl
  |--SortByMW.csh
--MOLDB/
  |--bin/
    |--create3d.pl
    |--pickUpAnnotation.pl
    |--setLigandBoxId.pl
  |--Makefile
  |--Makefile.intel
  |--TOOL/
    |--confgeneC
    |--confgeneC_YYMMDD/
      |--bin/
      |--src/
        ____ (confgeneC) (install.sh 実行後に出現)
    |--cosgene_multiPDB_Mol2_v1/
    |--cosgene_multiPDB_Mol2_v1_YYMMDD/ (次のページへつづく)

```

```
    |--bin/
    |--src_mol2/
    |   |--(cosgene)      (install.sh 実行後に出現)
    |   |--test/
    |--Hgene/
    |--HgeneYYMMDD/
    |   |--bin/
    |   |--sample/
    |   |--src/
    |--split_mols/
    |--split_mols_YYMMDD/
    |   |--(split_mols)  (install.sh 実行後に出現)
    |--tplgeneL/
    |--tplgeneLYMMDD/
    |   |--tplgeneL
    |   |   |--bin/
    |   |   |--DB/
    |   |   |--src/
    |   |   |--(tplgeneL) (install.sh 実行後に出現)
    |--sample/
    |   |--CLUSTERING/
    |   |   |--c00001/
    |   |--FILTERING/
    |   |   |--command.sh
    |   |   |--test001_conf_id.mol2.gz
    |--MOLDB
    |   |--anot_list.txt
    |   |--command.sh
    |   |--mol2_list.txt
    |   |--test001.sdf.gz
```

### 1-3. インストール方法

3DdataConstructionYYMMDD.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。(YYMMDD には年月日を示す数字が入ります。) インストールには、C コンパイラと、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

```
% tar -xzf 3DdataConstructionYYMMDD.tar.gz
% cd 3DdataConstructionYYMMDD
```

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)
```

### 1-4. テスト計算プログラムの実行方法

次のコマンドで 3DdataConstruction に含まれるツール群のテスト実行ができます。

```
% bin/test_3DdataConstruction.sh
```

このプログラムの出力先は、3DdataConstructionYYMMDD/test\_3DdataConstruction です。このコマンドを実行することにより、3DdataConstruction に含まれるツール群が適切に動作することを確認することができます。

## 2. 分子構造の3次元化処理ツール(create3d)

### 2-1. ディレクトリ構成

以下に、3DdataConstructionYYMMDD/MOLDB 配下のディレクトリ構成を示します。

- bin : 化合物 DB 構築用各種ツール用ディレクトリ (perl スクリプト)
- TOOL : 化合物 DB 構築 perl スクリプト内で呼び出されるツール群ディレクトリ

### 2-2. TOOL 配下のツールについて

bin ディレクトリ配下にある各種ツールはプログラム内部で他のツールを呼び出します。これらのツールは TOOL ディレクトリ内に置いています。各ディレクトリの構成、及び、各ツールの処理概要は以下の通りです。

- Hgene : 分子座標フォーマット変換、水素付加ツール  
分子座標フォーマットの変換(sdf -> MOL2、MOL2 -> pdb)を行う。
- confgeneC : 環構造に関するコンフォーマー生成プログラム  
幾何、光学異性体を生成する。
- cosgene\_multiPDB\_Mol2\_v1 : 分子動力学プログラム  
エネルギー最小化処理を行い、最適化構造を計算する。
- split\_mol : マルチファイルを1分子ずつのファイルに分解する。
- tplgeneL : トポロジー生成プログラム  
低分子のトポロジーファイルを作成する。



## 2-3. 使用方法

用意されたサンプルファイルを利用する場合には、  
3DdataConstructionYYMMDD/sample/MOLDB に移動してください。

### 2-3-1. create3d.pl

create3d.pl を実行して、3次元化処理を行います。

create3d.pl は 3DdataConstructionYYMMDD/MOLDB/bin/ に配置されています。

#### 【プログラム名】

create3d.pl

#### 【機能】

入力の MDL sdf ファイルの 3次元化、物性値等の計算、光学・幾何異性体の生成を行います。物性値等として分子式、分子量、分子総電荷、水素ドナー数、水素アクセプタ数、HOMO エネルギー値、LUMO エネルギー値、LogS 値、LogP 値を計算、出力します。

#### 【コマンド】

```
% perl (path)/create3d.pl --isdf <入力 sdf(.gz)> --wkdir <出力ディレクトリ名>
--source <データソース名> --supplier <サプライヤ名> --idtag <ID タグ名>
--multi [on|off]
```

このコマンドは、途中で改行せずに、1行で入力してください。

また、<出力ディレクトリ名>と--source の間、<ID タグ名>と--multi の間にスペースが必要であることに注意してください。

#### 【引数】

--isdf : 入力 sdf ファイル名を指定します。圧縮形式(\*.sdf.gz)を指定する事もできます。

例えば、XXX.sdf(.gz)を指定した場合、XXX.mol2.gz ファイルを出力します。

--wkdir : 出力ファイルを作成するディレクトリを指定します。本ディレクトリ内に  
シングル mol、シングル mol2、各種計算中間ファイルを保存します。

--source : データソース名を指定します。指定した文字列が、Sybyl mol2 ファイルのコメント項目、SOURCE 欄に出力されます。

--supplier : 入力 sdf にサプライヤ名タグ(SUPPLIER)が存在しない場合に指定します。  
本内容が mol2 ファイルのコメント行 SUPPLIER に記載されます。

--idtag : 入力 SDF に IDNUMBER タグが存在しない場合に設定します。指定した文字列のタグを ID のタグと見なし、mol2 ファイルのコメント行 IDNUMBER に記載されます。

--multi : 異性体を一括で生成する場合は on を、生成しない場合は off を指定します。

## 【出力ファイル】

--isdf オプションで指定したファイル名の拡張子を".mol2.gz" に変更したファイルを生成します。マルチ mol2 形式(1 つのファイル中に複数の分子情報が記載されている形式、COMMENT 項目から BOND 項目が複数ある形式)で出力されます。

## 【コマンド使用例】

```
% perl ../../MOLDB/bin/create3d.pl --isdf test001.sdf.gz --wkdir OUT --source TEST
--supplier TEST --idtag NAMIKI_ID --multi on
```

このコマンドは、途中で改行せずに、1 行で入力してください。

また、TEST と -supplier の間にスペースが必要であることを注意してください。

## 【出力ファイル例】

```
@<TRIPOS>COMMENT
SOURCE = TEST
SUPPLIER = ACB_Blocks(Fragment)
IDNUMBER = 2IA-0011
MOLECULAR_FORMULA = C16H17N2
MOLECULAR_WEIGHT = 237.326
MOLECULAR_CHARGE = 1
NUM_OF_DONOR = 4
NUM_OF_ACCEPTOR = 0
HOMO = -11.1204
LUMO = -4.0972
NUM_OF_CHIRAL_ATOMS = 1
LOGS =
LOGP =
NOTE = ACB_Blocks(Fragment)_2IA-0011;

@<TRIPOS>MOLECULE
00000000.mol2
35 37 0 0 0
SMALL
USER_CHARGES

@<TRIPOS>ATOM
1 C1          9.5966   -1.6527   0.0641 C.ar    1 LGD        -0.1385
```

2 N1	8.8181	-2.7739	0.2459 N. ar	1 LGD	-0.1959
3 C2	8.8054	-0.5072	-0.0869 C. ar	1 LGD	-0.1495
4 C3	7.5103	-2.4078	0.2089 C. ar	1 LGD	0.0322
5 C4	7.4191	-0.9731	0.0055 C. ar	1 LGD	-0.0995

(略)

32 H14	3.9929	-0.7053	-0.0140 H	1 LGD	0.1526
33 H15	11.1430	-7.1011	-2.6232 H	1 LGD	0.1527
34 H16	9.9765	-3.4171	-4.6635 H	1 LGD	0.1527
35 H17	10.1661	-5.8839	-4.6701 H	1 LGD	0.1539

@<TRIPOS>BOND

1	1	2	1
2	1	3	2
3	2	4	1
4	3	5	1
5	1	6	1

(略)

35	16	33	1
36	17	34	1
37	18	35	1

### 2-3-2. setLigandBoxId.pl

setLigandBoxId.pl を実行して、各ファイルの個々の分子について LigandBox ID を付与します。

setLigandBoxId.pl は 3DdataConstructionYYMMDD/MOLDB/bin/ に配置されています。

#### 【プログラム名】

setLigandBoxId.pl

#### 【機能】

create3d.pl を使用して作成したマルチ mol2 ファイルの一覧リスト、データセットを区別する為の名称を入力することにより、LigandBoxID を付与したマルチ mol2 ファイルを作成します。

#### 【コマンド】

```
% perl (path)/setLigandBoxId.pl <input mol2 list> <base name>
```

#### 【引数】

input mol2 list : マルチ mol2 ファイルの一覧ファイル

base name : データセットを区別する為の名称

例えば、” TEST2017” と入力した場合の LigandBox ID は” TEST2017-xxxxxxx-yy” (x, y は数字) と出力されます。

#### 【出力ファイル】

LigandBox ID が付与されたマルチ mol2 ファイルが出力されます。

#### 【コマンド使用例】

```
% perl ../../MOLDB/bin/setLigandBoxId.pl mol2_list.txt TEST1508
```

mol2\_list.txt の記述例 :

test001\_conf.mol2.gz

#### 【出力ファイル】

マルチ mol2 ファイルの一覧ファイルで指定された各ファイルについて、対応する mol2 が作成される。ファイル名は、

もとのファイル名(拡張子除く) + “\_id”.mol2.gz

となる。(例えば、201404HTS\_01\_01\_conf\_id.mol2.gz など)

### 2-3-3. pickUpAnnotation.pl

pickUpAnnotation.pl を実行して、mol2 ファイル中の各分子の物性値を抽出し、対応する RDF ファイルを出力します。

pickUpAnnotation.pl は 3DdataConstructionYYMMDD/MOLDB/bin/に配置されています。

#### 【プログラム名】

pickUpAnnotation.pl

#### 【機能】

入力のマルチ mol2 ファイルの各分子の物性値の抽出と、対応する RDF ファイルを出力します。

#### 【コマンド】

```
% perl (path)/pickUpAnnotation.pl --imol2 <入力 mol2> --anolist <アノテーションリストファイル> --otxt <RDF ファイル名>
```

このコマンドは、1行で入力してください。

#### 【引数】

--imol2 : 入力 mol2 ファイル名を指定します。圧縮形式(\*.mol2.gz)を指定する事もできます。

--anolist : RDF ファイルに出力する物性値項目と、mol2 ファイル中で記載されている物性値項目名の対応を示す情報が記載されたファイルのファイル名をしています。

--otxt : 出力する RDF ファイル名を指定します。

## 【アノテーションリストファイルについて】

タブ区切りテキストとします。先頭文字が#の場合はコメント行とし、各行の1項目目がRDFファイルに出力する項目名、2項目目がmol2ファイルで定義されている項目名を示します。以下に、ファイル例を示します。

```
# property tag_name_of_files
ligandbox_id    LIGANDBOX_ID
source_name     SOURCE
source_id      SOURCE_ID
supplier_name   SUPPLIER
supplier_id     IDNUMBER
molecular_formula    MOLECULAR_FORMULA
molecular_weight    MOLECULAR_WEIGHT
molecular_charge    MOLECULAR_CHARGE
num_of_donor    NUM_OF_DONOR
num_of_acceptor NUM_OF_ACCEPTOR
homo    HOMO
lumo    LUMO
num_of_chiral_atoms    NUM_OF_CHIRAL_ATOMS
LogP    LOGP
LogS    LOGS
note    NOTE
```

**【出力ファイル】**

--otxt オプションで指定したファイル名の RDF ファイルを出力します。

**【コマンド使用例】**

```
% perl ../../MOLDB/bin/pickUpAnnotation.pl --imol2 test001_conf_id.mol2.gz  
--anolist anot_list.txt --otxt test001_conf_id.rdf
```

このコマンドは、途中で改行を入れずに 1 行で入力してください。

また、test001\_conf\_id.mol2.gz と --anolist の間にスペースが必要であることを注意してください。

**【出力ファイル例】 (一部抜粋)**

```
<?xml version="1.0"?>  
<rdf:RDF xmlns:ligboxo="file://localhost/c:/work/document/ligandbox.owl#"  
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"  
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">  
  <rdf:Description rdf:about="TEST1508-00000001-01">  
    <ligboxo:ligandbox_id>TEST1508-00000001-01</ligboxo:ligandbox_id>  
    <ligboxo:source_name>TEST</ligboxo:source_name>  
    <ligboxo:supplier_name>ACB_Blocks(Fragment)</ligboxo:supplier_name>  
    <ligboxo:supplier_id>2IA-0011</ligboxo:supplier_id>  
    <ligboxo:molecular_formula>C16H17N2</ligboxo:molecular_formula>  
    <ligboxo:molecular_weight>237.326</ligboxo:molecular_weight>  
    <ligboxo:molecular_charge>1</ligboxo:molecular_charge>  
    <ligboxo:num_of_donor>4</ligboxo:num_of_donor>  
    <ligboxo:num_of_acceptor>0</ligboxo:num_of_acceptor>  
    <ligboxo:homo>-11.1204</ligboxo:homo>  
    <ligboxo:lumo>-4.0972</ligboxo:lumo>  
    <ligboxo:num_of_chiral_atoms>1</ligboxo:num_of_chiral_atoms>  
    <ligboxo:LogS></ligboxo:LogS>  
    <ligboxo:LogP></ligboxo:LogP>  
    <ligboxo:same_molecule>ACB_Blocks(Fragment)_2IA-0011</ligboxo:same_molecule>  
  
    (略)  
  
  </rdf:Description>  
</rdf:RDF>
```

### 3. 医薬品として不適切な分子のフィルタリングツール(Filtering)

本システムでは、医薬品として不適切な分子のフィルタリングを行い、医薬品候補となる化合物の mol2 ファイルを生成します。

#### 3-1. ディレクトリ構成

以下に、3DdataConstruction/FILTERING 配下のディレクトリ構成を示します。

- \*.csh / \*.pl ファイル：実行ファイル Filtering.csh、及び、関連プログラム保存ディレクトリまた、フィルタリング後のマルチ mol2 ファイルを低分子量側から順次ディレクトリに保存する SortByMW.csh プログラム、及び関連プログラムも保存しています。
- ExcludeMol：フィルタリングプログラム保存ディレクトリ

#### 3-2. 使用方法

用意されたサンプルファイルを利用する場合には、3DdataConstructionYYMMDD/sample/FILTERING に移動してください。

##### 3-2-1. Filtering.csh

Filtering.csh を実行して、フィルタリング対象リストの作成、フィルタリング対象分子の除去を行います。

Filtering.csh は 3DdataConstructionYYMMDD/FILTERING/ に配置されています。

##### 【プログラム名】

Filtering.csh

##### 【機能】

指定したマルチ mol2 ファイルを入力として、医薬品として不適切な分子を除去したマルチ mol2 ファイルを生成します。内部処理で以下のツールを呼び出します。

FindExcludeMol：指定したマルチ mol2 内の分子の構造情報を計算し、また、データベースファイルとの部分構造マッチングを行い、医薬品として不適切な分子の一覧を作成します。

SetExcludeInfoMol2.pl：入力マルチ mol2 ファイルと FindExcludeMol で作成したリストファイルから、医薬品として不適切な分子を除去したマルチ mol2 ファイルを生成します。



**【コマンド】**

```
% csh (path)/Filtering.csh <multi-mol2 file>
```

**【引数】**

multi-mol2 file : 入力マルチ mol2 ファイル名を指定します。圧縮形式(\*.mol2.gz)を指定する事もできます。例えば、XXX.mol2(.gz)を指定した場合、XXX\_logs.mol2.gz ファイルを出力します。

**【出力ファイル】**

指定した入力ファイル名(XXXX.mol2(.gz))を XXXX\_exclude.mol2.gz に変更したマルチ mol2 ファイルを生成します。

**【コマンド使用例】**

```
% csh ../../FILTERING/Filtering.csh test001_conf_id.mol2.gz
```

**3-2-2. SortByMW.csh**

SortByMW.csh コマンドを実行して、マルチ mol2 ファイルを分割し、シングル mol2 ファイルを生成します。その際に、低分子量側から順次ディレクトリに保存します。

SortByMW.csh は 3DdataConstructionYYMMDD/FILTERING/に配置されています。

**【プログラム名】**

SortByMW.csh

**【機能】**

マルチ mol2 ファイルを分割し、シングル mol2 ファイルを生成します。その際に、低分子量側から順次ディレクトリに保存します。

**【コマンド】**

```
% csh (path)/SortByMW.csh <multi-mol2 list file> <output directory>
```

**【引数】**

multi-mol2 file : 入力マルチ mol2 ファイル名のリストファイルを指定します。

output directory : 出力ファイルを作成するディレクトリを指定します。

**【出力ファイル】**

指定した出力ディレクトリに低分子量側の分子から順次シングル mol2 ファイルを生成します。分子は1ディレクトリ辺り 5000 件のデータを保存します。指定した出力ディレクトリ配下に cXXXXX(X は数字)のディレクトリを順次作成します。

**【コマンド使用例】**

```
% echo test001_conf_id.mol2.gz >> list_of_multi_mol2.txt  
% csh ../../FILTERING/SortByMW.csh list_of_multi_mol2.txt ./SORT_OUT
```

## 4. 類似分子のクラスタリングツール(Clustering)

### 4-1. ディレクトリ構成

以下に、3DdataConstructionYYMMDD/CLUSTERING 配下のディレクトリ構成を示します。

- TGS/bin : TGS ツールの実行形式ファイルが保存されています。
- TGS/screening : 実行ファイル Clustering.csh、及び、本 csh から呼び出される perl プログラム保存ディレクトリ
- TGS/database : TGS ツールに必要な database ファイルを保存するディレクトリ
- TGS/src : TGS ツールのソースプログラム保存ディレクトリ

### 4-2. 使用方法

Clustering.csh を実行して、対象ディレクトリ内の各分子間の分子の距離を計算し、距離が近い(類似性が高い)分子についてデータセット作成対象外とします。最終的に類似性が低い分子のデータセットを作成します。

※本ツールは CLUSTERING/TGS/screening ディレクトリ配下で実行して下さい。

サンプルとして利用できるデータセットは

3DdataConstruction170907r2/sample/CLUSTERING/c00001/に配置されています。

#### 【プログラム名】

Clustering.csh

#### 【機能】

対象ディレクトリ内の各分子間の分子の距離を計算し、距離が近い(類似性が高い)分子についてデータセット作成対象外とします。最終的に類似性が低い分子のデータセットを作成します。本ツールでは、指定ディレクトリ配下の cXXXXX(X は数字)ディレクトリ内にあるシングル mol2 ファイルを計算対象とし、1つのディレクトリ内にある分子間の距離を計算します。ディレクトリ間の分子の距離の計算は行いません。

なお、内部処理で以下のツールを呼び出します。

TGS 各種ツール：TGS 計算に使用します。詳細は myPresto マニュアルを参照の事。

PickUp\_data.pl：指定した件数の分子データを TMP ディレクトリにコピーします。閾値を決定する為に少数のデータをピックアップするのに使用します。

SelectMol2\_1st.pl：指定したディレクトリ内の cXXXXX (X は数字) ディレクトリ内の各分子の距離を計算し、閾値以下の分子をデータセット作成対象外とし、リストファイルに出力します。

Judge\_SelectMol.pl：少数データでの TGS 処理を行い、その結果をスケールアップした際の予想クラスタリングデータ数を計算し、閾値の調整を行います。予想クラスタリングデータ数が少ない場合は閾値を小さくし、データ数が多い場合は閾値を大きくします。

Adjustmet\_clustering.pl：全データを用いて TGS 処理を行った際の最終的なデータセット数が指定した値よりも多い、少ない場合のデータ数の調整を行います。

#### 【コマンド】

```
% csh (path)/Clustering.csh <pickup_num> <adjust_num> <input_num> <input_dir>
```

#### 【引数】

pickup\_num：TGS ツールの閾値を決定するのに使用する分子の件数を指定します。  
指定した分子数の分子を入力ディレクトリ内の各ディレクトリから抜き出し、クラスタリング処理を行い、結果により TGS 処理の閾値の調整を行います。

adjust\_num：最終的に作成したいデータセットの分子数を指定します。

input\_num：TGS 処理で入力する分子数を指定します。

input\_dir：入力ファイルが保存しているディレクトリを指定します。指定したディレクトリ配下にある cXXXXX ディレクトリ (X は数字) 内の mol2 ファイルをクラスタリング対象とします。

**【出力ファイル】**

TGS/screeing 配下に” RESULT/RESULT1/mol2” ディレクトリを作成し、その下に 1 万件のシングル mol2 ファイルを保存した cXXXX(X は数字)ディレクトリを作成します。

**【コマンド使用例】**

```
% csh Clustering.csh 50 200 500 $DIR/sample/CLUSTERING/
```