

myPresto 5.0
- TOOLS-

USER MANUAL

2018/1/14

Copyright (C) 2006-2018 Next Generation Natural Product Chemistry (N²PC)
Copyright (C) 2006-2018 National Institute of Advanced Industrial Science and Technology (AIST)
Copyright (C) 2006-2018 Japan Biological Informatics Consortium (JBIC)

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、新エネルギー・産業技術総合開発機構 (NEDO)、及び、経済産業省 (METI) 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・京極好正博士の推進する研究プロジェクトで開発されました。

目次

1.	概要	5
2.	使用方法	7
3.	setwater	8
4.	mergetpl	12
5.	SHAKEinp	14
6.	RIGIDinp	15
7.	GBSAinp	17
8.	自由エネルギー計算 (Filling potential 法 + WHAM 法) 解析.....	18
8.1.	Generate_NextFP.....	18
8.2.	Extract_Atom.....	21
8.3.	Wham_Analysis.....	22
9.	拡張アンサンブル向け解析ツール.....	24
9.1.	reweightFB	25
9.2.	reweightST	26
9.3.	reweightGST.....	27
9.4.	selection	28
9.5.	clustering	30
10.	存在確率 (Potential Mean Force) 解析ツール.....	34
10.1.	pmf	34
10.2.	contour	36
11.	pca	37
12.	Gamess2tplinp.....	39
13.	Gauss2tplinp.....	40
14.	tpl2mol2	41
15.	add_ion	42
16.	confgene	45

17.	confgeneC	47
18.	自由エネルギー摂動法（開発中）	49
18.1.	計算方式	49
18.2.	vdw パラメータおよび電荷のスケーリング機能（cosgene）	51
18.3.	analyze	53
18.4.	FEP	54
19.	Hgene：詳細は別冊 Hgene マニュアルを参照.....	56
20.	MV0	61
21.	sptool	63
21.1.	Descriptor	63
21.2.	Solubility	65
21.3.	FreqMaker	67
21.4.	wln	67
21.5.	pls	68
21.6.	mlr	68
21.7.	bys	69
21.8.	trans_code	70
22.	tpl2capbc	73

1. 概要

本マニュアルは、myPresto に含まれるツールプログラムについて説明しています。
ツールプログラムには、以下のプログラムがあります。

- set_water
- mergetpl
- SHAKEinp
- RIGIDinp
- GBSAinp
- Generate_NextFP
- Extract_Atom
- Wham_Analysis
- reweightFB
- reweightST
- reweightGST
- selection
- clustering
- pmf
- contour
- pca
- Gamess2tplinp
- Gauss2tplinp
- tpl2mol2
- add_ion
- configene
- configeneC
- analyze
- FEP
- Hgene
- MVO
- Descriptor
- Solubility
- FreqMaker
- Wln
- Pls
- Mlr

- Bys
- trans_code
- tp12capbc

2. 使用方法

myPresto に含まるツールプログラムは、toolsYYMMDD.tar.gz として提供されています。(YYMMDD は年月日を示す数字が入ります。) 他にも、同じものが siegene_pack や cosgene_pack に含まれています。ここでは、toolsYYMMDD.tar.gz に含まれるプログラムの使用方法について説明します。まず、toolsYYMMDD.tar.gz を、書き込み可能なディレクトリに配置して、次のコマンドで展開します。

```
% tar -xzf toolsYYMMDD.tar.gz
```

展開した後のディレクトリ構造は以下のようになっています。

```
toolsYYMMDD/
├── OREADME/
├── doc/
├── (tool-1)/ (tool-1 というのは、ツールプログラムの一つを指します。)
│   ├── OREADME
│   ├── bin/
│   │   ├── install.sh
│   │   ├── test_*.sh (*の部分は、ツールプログラムによって異なります。)
│   │   └── (tool-1) (tool-1 のバイナリファイル。install.sh 実行後に出現します。)
│   ├── sample/ (サンプルファイルが用意されていないものもあります。)
│   └── src/
│       └── (tool-1)/
│           ├── src/
│           └── bin/
├── (tool-2)/ (tool-2 というのは、ツールプログラムの一つを指します。)
│   ├── OREADME
│   ├── bin/
│   │   ├── install.sh
│   │   ├── test_*.sh (*の部分は、ツールプログラムによって異なります。)
│   │   └── (tool-2) (tool-1 のバイナリファイル。install.sh 実行後に出現します。)
│   ├── sample/ (サンプルファイルが用意されていないものもあります。)
│   └── src/
│       └── (tool-2)/
│           ├── src/
│           └── bin/
```

tools の下のディレクトリが、それぞれ1つのツールに対応しています。多くのツールプログラムに対して、bin/install.sh と sample/を用意していますが、ツールプログラムによっては、install.sh がないもの、sample/がないものもあります。

install.sh が用意されているツールプログラムについては、そのディレクトリに移動して、次のコマンドを実行することにより、実行バイナリファイルが作成されます。

どちらか一方を実行します。

```
% bin/install.sh (intel コンパイラを使用しない場合、GNU コンパイラもしくはシステムの cc を使用します。)
```

または、

```
% bin/install.sh intel (intel コンパイラを使用します。)
```

実行バイナリファイルは、(tools)/bin/の下にコピーされます。テスト実行用プログラムが用意されているものは、bin/test_*.sh(*は、ツールによって異なります)を実行することができます。

3. setwater

蛋白質などの系に水を付加します。作成できる水分子のモデルは、TIP3P または TIP4P です。使用前に、水分子の座標データ（システム添付：tools/setwater/tip3_base.pdb および tip4_base.pdb）を作業ディレクトリにコピーしておきます。

また、結晶水も付加することができます。結晶水を付加する場合は、結晶水だけを抜き出した PDB 形式の座標を準備してください。この場合の書式は、レコード名“HETATM”、残基名“HOH”としてください。

結晶水の例

HETATM	1	0	HOH	1	-7.948	-7.948	-7.948
HETATM	2	0	HOH	2	-7.948	-7.948	-4.844
HETATM	3	0	HOH	3	-7.948	-7.948	-1.741

【注意】 setwater は、事前に周期境界条件、温度 300K、密度 1g/cm³ で平衡化した水の座標をもとに、溶媒水の座標を発生させます。使用前に、水分子の座標データ（システム添付：tools/setwater/tip3_base.pdb および tip4_base.pdb）を作業ディレクトリにコピーしておきます。

【注意】 結晶水は PDB などの水素の省略された形（酸素だけ）を想定しています。また、結晶水の付加では、酸素に水素を付加するとき、水素は一定方向にのみ配向します。

入力データ

- (1) 蛋白質など水を付加する系の PDB ファイル名
- (2) 結晶水の PDB ファイルの使用の有無
- (3) 結晶水の PDB ファイル名（結晶水の PDB ファイルを使用する場合）
- (4) 結果の出力の PDB ファイル名
- (5) 水を入れるセルの形状
- (6) 水を入れるセルの半径、セルの辺の長さ：値が正であれば、半径や辺の長さ。値が負であるとき（-r ないし、-x, -y, -z）、蛋白質の一番外側の原子からセルの壁までの最短距離が r となる半径、または、セルの辺の長さを自動計算する。セルの半径、長さは、出力 PDB ファイル末尾に、出力される（ver4.2 より）
- (7) 水を入れるセルの中心の種別（系の中心で指定、任意の座標で指定）

- (8) 水を入れるセルの中心の座標 (セルの中心を座標で指定する場合)
- (9) 付加する水分子の密度の係数 (通常は 1.0)
- (10) van der Waals 半径のダンピングファクター (通常は 1.0)
- (11) 水分子のモデル (TIP3P または TIP4P)

■使用例 その1: セルの中心、セルの辺の長さを全て自分で設定する場合。

```

% setwater
--- setwater ---
Input file name (PDB of target molecule) ?
protein.pdb (1)
-> none.pdb
Do you use crystal water file (Y or N) ?
Y (2)
Input file name (PDB of crystal water) ?
crystal_water.pdb (3)
-> box.pdb
Input file name (output) ?
new_water.pdb (4)
-> res
Input cell type (sphere="S", ellipsoid="E", cube="C", parallelepiped="P") ?
E (5)
Input length (A, B, C) ?
10.0 20.0 30.0 (6)
-> ellipsoid : 10.0000000000 20.0000000000 30.0000000000
Input center of water (mass center="C", 3D-coordinate="D") ?
D (7)
Input coordinate (X, Y, Z) ?
10.0 0.0 0.0 (8)
-> coordinate : 10.0000000000 0.0000000000 0.0000000000
Input density of water (usually 1.0) ?
1.0 (9)
-> 1.0000000000
Input vdW damping factor (usually 1.0) ?
1.0 (10)
-> 1.0000000000
Input water model (TIP3P="3", TIP4P="4") ?
3 (11)
-> TIP3P
%

```

■使用例 その2 : 周期系で、蛋白質からセル境界までの距離を 10 Åにする。

```
% setwater
--- setwater ---
Input file name (PDB of target molecule) ?
protein.pdb (1)
-> none.pdb
Do you use crystal water file (Y or N) ?
N (2)
Input file name (output) ?
new_water.pdb (4)
-> res
Input cell type (sphere="S", ellipsoid="E", cube="C", parallelepiped="P") ?
C (5)
Input length (A, B, C) ?
-10.0 -10.0 -10.0 (6)
-> ellipsoid : 10.0000000000 20.0000000000 30.0000000000
Input center of water (mass center="C", 3D-coordinate="D") ?
C (7)
Input density of water (usually 1.0) ?
1.0 (9)
-> 1.0000000000
Input vdW damping factor (usually 1.0) ?
1.0 (10)
-> 1.0000000000
Input water model (TIP3P="3", TIP4P="4") ?
3 (11)
-> TIP3P
%
```

■使用例 その3 : 孤立系で、蛋白質からセル境界までの距離を 10 Åにする。

```
% setwater
--- setwater ---
Input file name (PDB of target molecule) ?
protein.pdb (1)
-> none.pdb
Do you use crystal water file (Y or N) ?
N (2)
Input file name (output) ?
new_water.pdb (4)
-> res
Input cell type (sphere="S", ellipsoid="E", cube="C", parallelepiped="P") ?
S (5)
Input length (A, B, C) ?
-10.0 (6)
-> ellipsoid : 10.0000000000 20.0000000000 30.0000000000
Input center of water (mass center="C", 3D-coordinate="D") ?
C (7)
Input density of water (usually 1.0) ?
1.0 (9)
-> 1.0000000000
Input vdW damping factor (usually 1.0) ?
1.0 (10)
-> 1.0000000000
Input water model (TIP3P="3", TIP4P="4") ?
3 (11)
-> TIP3P
%
```

4. mergetpl

複数のトポロジーファイルを1つのトポロジーファイルにマージします。

蛋白質-化合物-溶媒系など複数分子を含む系の場合に用います。Ver4.2より、tplgene で、複数分子を含む系を一括で扱えるようになったので、mergetpl を使う場面は、少なくなっています。

入力データ

- (1) マージ対象のトポロジーファイル名 (1 個目)
- (2) マージ対象のトポロジーファイル名 (2 個目)
- (3) マージ対象のトポロジーファイル名 (3~10 個目)
- (4) 結果の出力のトポロジーファイル名

【注意】ポテンシャル関数指定記述 ("TPL> FUNC"行) および非結合相互作用指定記述 ("TPL> NONBOND"行) の内容が異なる場合、マージした結果のトポロジーファイルは正しくありません。結果のトポロジーファイルには、最初に入力指定したトポロジーファイルの "TPL> FUNC"行 および "TPL> NONBOND"行 の内容が出力されます。

【注意】上記の問題は、tplgene で作成した蛋白質のトポロジーファイルと tplgeneL で作成した低分子のトポロジーファイルをマージする場合、剛体で取り扱う TIP4P モデルの水分子のトポロジーファイルをマージする場合、また、人手で改変したトポロジーファイルを取り扱う場合などで発生しやすいので、特に注意が必要です。

■使用例

```
% mergetpl
--- mergetpl ---
  Input file name ? ( end: RETURN )
aa. tpl (1)
  Input file name ? ( end: RETURN )
bb. tpl (2)
  Input file name ? ( end: RETURN )
cc. tpl (3)
  Input file name ? ( end: RETURN )

  Output file name ?
output. tpl (4)
--- done ---
%
```

◆水分子のトポロジーファイルの扱い

水分子のトポロジーファイルとして、TIP3P モデルおよび TIP4P モデルをシステムに添付しています (tools/common/tip3p. tp1 および tip4p. tp1)。これらの TIP3P と TIP4P のトポロジーファイルは、“OW”原子に対応する、非結合相互作用指定記述 (“TPL> NONBOND”行) の “相互作用タイプ”の内容が異なります。

【TIP3P】

```
18 0 1 1.76830 0.152000 0.833333 0.500
```

【TIP4P】

```
18 0 1 1.7699 0.155000 0.833333 0.500
```

よって、TIP4P のトポロジーファイルを使う場合は、マージの際に“TPL> NONBOND”行の「OW」に相当する部分の値を TIP4P 用に修正する必要があります (TIP3P との混在はない)。

(省略)
:
TPL> NONBONDS
:NUMBER OF TYPE= 39
1 0 1 1.90800 0.086000 0.833333 0.500; c
2 0 1 1.90800 0.109400 0.833333 0.500; c3
3 0 1 0.60000 0.015700 0.833333 0.500; h
4 0 1 0.00000 0.000000 0.833333 0.500; ho
5 0 1 0.60000 0.015700 0.833333 0.500; hs
6 0 1 1.48700 0.015700 0.833333 0.500; hc
7 0 1 1.38700 0.015700 0.833333 0.500; h1
8 0 1 1.28700 0.015700 0.833333 0.500; h2
9 0 1 1.18700 0.015700 0.833333 0.500; h3
10 0 1 1.10000 0.015700 0.833333 0.500; hx
11 0 1 1.45900 0.015000 0.833333 0.500; ha
12 0 1 1.40900 0.015000 0.833333 0.500; h4
13 0 1 1.35900 0.015000 0.833333 0.500; h5
14 0 1 0.00000 0.000000 0.833333 0.500; hw
15 0 1 1.82400 0.170000 0.833333 0.500; n
16 0 1 1.66120 0.210000 0.833333 0.500; o
17 0 1 1.66120 0.210000 0.833333 0.500; o2
18 0 1 <u>1.76990</u> <u>0.155000</u> 0.833333 0.500; ow TIP4P
:
(省略)

5. SHAKEinp

トポロジーファイルと PDB ファイルから、対象原子の原子番号と拘束距離を指定した SHAKE ファイルを作成します。このツールで水分子 TIP3P モデルを指定するためには、TIP3P モデルの SHAKE ファイル（システム添付：tools/SHAKEinp/tip3_shk.model）が必要です。使用前に、このファイルを作業ディレクトリにコピーしておきます。SHAKE ファイルが作業ディレクトリの存在しない場合にはシステム内のデータを用いて TIP3P 情報を出力します。

入力データ

- (1) SHAKE ファイルを作成する系のトポロジーファイル名
- (2) SHAKE ファイルを作成する系の PDB 名
- (3) 出力 SHAKE ファイル名
- (4) 水分子 TIP3P モデルを使用するか否か（水分子が含まれる場合のみ）

オプション

-itpl	<tpl_file>	トポロジーファイル名を <tpl_file> に指定します。
-ipdb	<pdb_file>	PDB ファイル名を <pdb_file> に指定します。
-oshk	<shk_file>	SHAKE ファイル名を <shk_file> に指定します。
-h		SHAKEinp の使用方法を表示します。

コマンドライン・オプションを用いて指定した項目は、対話入力での入力がスキップされます。オプションで指定しなかったものだけを対話的に入力することになります。

■使用例

```
% SHAKEinp
Please input TPL filename.
indo_tip3p.tpl (1)
Please input PDB filename.
indo_tip3p.pdb (2)
Please input SHAKE filename.
indo_tip3p.shk (3)

INFORMATION>
  H2O was detected.
  Do you want to use TIP3P model?[yes/no]
yes (4)

INFORMATION> toolWriteTip3p
  The file "tip3_shk.model" is found.
  Information given by this file is used for the Tip3p model.

%% Program is done. %%
%% This program is normal end. %%
```

6. RIGIDinp

トポロジーファイルから剛体モデル指定ファイルを作成します。このプログラムではトポロジーファイルから、水素と結合している原子の情報を取得し、その結合原子グループを剛体とみなし、拘束を行います。

水分子の TIP3P、TIP4P モデルを剛体に指定する場合には、TIP3P、TIP4P モデルの剛体指定ファイル（システム添付：tools/RIGIDinp/tip3_rig.model、tip4_rig.model）が必要となります。プログラム実行前に、これらのファイルを作業ディレクトリにコピーしておきます。また、任意のフラグメントの拘束を行う場合には、フラグメント DB ファイル（システム添付：tools/RIGIDinp/fragment.db）にフラグメント情報を記述する必要があります。プログラム実行前に、このファイルを作業ディレクトリに置いて下さい。

Cosgene 自体は、自動で剛体モデル指定ファイルを作成するオプションを内蔵しています。

入力データ

- (1) 剛体モデル指定ファイルを作成する系のトポロジーファイル名
- (2) 剛体モデルの指定レベル
 - (i) 水素原子との結合のみを剛体として指定
 - (ii) (i) に加え、任意のフラグメントを剛体として指定

オプション

- i <tpl_file>
トポロジーファイル名を<tpl_file>に指定します
- l [allH | fr]
剛体モデルの指定レベルを指定します
- | | |
|----------------------|--------|
| (i) 水素原子との結合のみを拘束 | ⇒ allH |
| (ii) (i) + フラグメントを拘束 | ⇒ fr |

※オプション”-l”を指定しない場合には、水素原子との結合のみを拘束します。

【注意】 出力する剛体モデルファイル名は XXX.rig となります。

(XXX はトポロジーファイル名から拡張子を除いたもの)

■ 使用例

```
% RIGIDinp -i indo.tpl
```

また、オプション“-h”または、“-help”の指定により、RIGIDinp の使用方法を見ることができます。

```
% RIGIDinp -h
```

または、

```
% RIGIDinp -help
```

7. GBSAinp

トポロジーファイルから、GB/SA 用パラメータ指定ファイルを作成します。mkGBSAin.pl で GB/SA 用パラメータ指定ファイルを作成するためには、GB/SA 用のパラメータ DB ファイル（システム添付：tools/GBSAinp/gb_sa.db）が必要です。使用前に、このファイルを作業ディレクトリにコピーしておきます。

【注意】 入力に PDB ファイルを指定した場合、内部で connectAtomPDB.x を利用します。その為、connectAtomPDB.x を mkGBSAin.pl と同じディレクトリに配置します。

入力データ

- (1) GB/SA 用のパラメータ DB ファイル名（システム添付：gb_sa.db）
- (2) GB/SA 用パラメータファイルを作成する系のトポロジーまたは PDB ファイル名
- (3) 出力 GB/SA 用パラメータ指定ファイル名

■使用例

```
% mkGBSAin.pl
%% INPUT DB FILE NAME. %%
gb_sa.db
%% SELECT INPUT FILE BY THE NEXT NUMBER. %%
  1 : PDB FILE
  2 : TPL FILE
2
%% INPUT FILE NAME. %%
vas-dih.tpl
%% INPUT OUTPUT FILE NAME. %%
vas-dih.sol
```

8. 自由エネルギー計算 (Filling potential 法 + WHAM 法) 解析

Filled Potential 計算における反発ポテンシャルと求心ポテンシャルを示す Umbrella Potential ファイルを作成し、座標トラジェクトリ群と Umbrella Potential の解析結果から自由エネルギーのヒストグラムを解析します。

8.1. Generate_NextFP

MD の座標トラジェクトリ、この MD での Umbrella Potential ファイル、およびユーザ指定を読み込み、新規の Umbrella Potential ファイルを作成します。

入力データ

- (1) 制御ファイル
 - (1-1) 前回の MD での Umbrella Potential 指定ファイル名
 - (1-2) 出力 Umbrella Potential 指定ファイル名
 - (1-3) 初期座標 PDB ファイル名
 - (1-4) 前回の MD での座標トラジェクトリファイル名
 - (1-5) 座標トラジェクトリの読み飛ばし回数
 - (1-6) 座標トラジェクトリ読み込み回数
 - (1-7) 座標トラジェクトリファイル形式 ("single" | "double")
 - (1-8) PDB ファイルの画面表示オプション ("yes" | "no")
 - (1-9) 求心関数の種別
 - (1-10) 温度
 - (1-11) ガウス型反発関数の高さ
 - (1-12) ガウス型ポテンシャル中心座標の更新間隔を制御する範囲
 - (1-13) ガウス型反発関数の幅
 - (1-14) 求心関数の高さ
 - (1-15) 求心関数の幅
 - (1-16) 目標とする最終座標
 - (1-17) 掃引開始回数、終了回数
- (2) 初期座標 PDB ファイル
- (3) 前回の MD での座標トラジェクトリファイル
- (4) 前回の MD 入力の Umbrella Potential 指定ファイル

■ 使用例

```
% Generate_NextFP < genefp.inp
```

■制御ファイル例

```

newopt_fp      ; 入力 Umbrella Potential 指定ファイル
newopt_fp2    ; 出力 Umbrella Potential 指定ファイル
initial.pdb   ; 初期座標
xx_traject.cor ; 前回の MD のトラジェクトリ
-1000        ; トラジェクトリの読み飛ばし回数
2000         ; トラジェクトリ読み込み回数
s            ; トラジェクトリファイル形式
y            ; PDB ファイルの原子表示
HAR2         ; 求心関数種別 (HAR1 | HAR2 | LIN1 | LIN2)
300.0        ; 対象の温度
0.5          ; ガウス型反発関数の高さ
2.5 6.0      ; ガウス型ポテンシャルの中心座標の更新間隔を制御する範囲
3.0         ; ガウス型反発関数の幅
5.0         ; 求心関数の高さ
1.0         ; 求心関数の幅
ATOM 4131 0 WAT 839 0.000 0.000 -8.000 15.00 -0.83 ; 目標焦点座標 1
ATOM 4131 0 WAT 839 0.000 0.000 -8.000 17.00 -0.83 ; 目標焦点座標 2
1 50        ; 掃引開始回数、終了回数

```

出力データ

(1) Umbrella Potential ファイル

■出力 Umbrella Potential ファイル例

```

FILL> GAUS
      2      1 ; DIMENSION      NUMBER OF ATOMS
      6      ; ATOM ID

      0.000000 ; WEIGHT DIM=      1
      0.030000 ; RADIUS DIM=      1

ATOM                                0.000  0.000 -2.000 ; CENTER-1 ATOM=      6

      0.500000 ; WEIGHT DIM=      2
      3.000000 ; RADIUS DIM=      2

ATOM                                0.283  0.269 -2.312 ; CENTER-1 ATOM=      6

FILL> HAR1
      1      1 ; DIMENSION      NUMBER OF ATOMS
      6      ; ATOM ID

      0.500000 ; WEIGHT DIM=      1
      3.000000 ; RADIUS DIM=      1

ATOM                                0.283  0.269 -2.312 ; CENTER-1 ATOM=      6

```

8.2. Extract_Atom

MDの座標トラジェクトリ、このMDでのUmbrella Potentialファイル、およびユーザ指定を読み込み、Umbrella Potential対象原子の座標のみ切り出したトラジェクトリファイルを生じます。

入力データ

- (1) 制御ファイル
 - (1-1) Umbrella Potential 指定ファイル名
 - (1-2) 座標トラジェクトリファイルの原子数
 - (1-3) 座標トラジェクトリファイル数
 - (1-4) 入力、出力座標トラジェクトリファイル名
 - (1-5) 座標トラジェクトリの読み飛ばし回数
 - (1-6) 座標トラジェクトリ読み込み回数
 - (1-7) 座標トラジェクトリファイル形式 ("s"ingle | "d"ouble)
- (2) 座標トラジェクトリファイル群
- (3) 前回のMD入力のUmbrella Potentialファイル

■制御ファイル例

```
newopt_fp          ; 入力 Umbrella Potential 指定ファイル
1023               ; トラジェクトリファイルの原子数
4                 ; 入力トラジェクトリファイル数
xx_trj1.cor      w_1.cor      ; 入力/出力トラジェクトリファイル名  - 1
xx_trj2.cor      w_2.cor      ; 入力/出力トラジェクトリファイル名  - 2
xx_trj3.cor      w_3.cor      ; 入力/出力トラジェクトリファイル名  - 3
xx_trj4.cor      w_4.cor      ; 入力/出力トラジェクトリファイル名  - 4
-1000            ; トラジェクトリの読み飛ばし回数
2000             ; トラジェクトリ読み込み回数
s                ; トラジェクトリファイル形式
```

■使用例

```
% Extract_Atom < extract.inp
```

8.3. Wham_Analysis

複数 MD の座標トラジェクトリと、Umbrella Potential 指定 ファイルから、各トラジェクトリでの自由エネルギーを計算します。

入力データ

- (1) 制御ファイル
 - (1-1) 最終世代の Umbrella Potential 指定ファイル名
 - (1-2) 座標トラジェクトリの読み飛ばし回数
 - (1-3) 座標トラジェクトリ読み込み回数
 - (1-4) 座標トラジェクトリファイル形式 ("s"ingle | "d"ouble)
 - (1-5) 自由エネルギー計算の平均値計算のサンプリング距離
 - (1-6) WHAM 解析の収束ループ数
 - (1-7) 計算温度
 - (1-8) WHAM 解析時のメモリ指定 ("m"emory | "s"peed)
 - (1-9) 座標トラジェクトリファイル数
 - (1-10) 入力座標トラジェクトリファイルと各世代 Umbrella Potential ファイル
- (2) 各世代の座標トラジェクトリファイル
- (3) 各世代の Umbrella Potential ファイル

■制御ファイル例

```
w_4.option ; 最終世代の入力 Umbrella Potential 指定ファイル名
0          ; トラジェクトリの読み飛ばし回数
2000      ; トラジェクトリ読み込み回数
s         ; トラジェクトリファイル形式
0.5       ; 自由エネルギー計算の平均値計算のサンプリング距離
1000      ; WHAM 解析収束ループ
310       ; 計算温度
m         ; メモリ優先
4         ; MD の世代数
w_1.cor w_1.option ; 1世代目のトラジェクトリファイルと Umbrella Potential ファイル
w_2.cor w_2.option ; 2世代目のトラジェクトリファイルと Umbrella Potential ファイル
w_3.cor w_3.option ; 3世代目のトラジェクトリファイルと Umbrella Potential ファイル
w_4.cor w_4.option ; 4世代目のトラジェクトリファイルと Umbrella Potential ファイル
```

■使用例

```
% Wham_Analysis < wham.inp
```

出力データ

(1) 自由エネルギー計算結果

標準出力の最後に下記のようなトラジェクトリ番号、RMSD、自由エネルギーの表を出力します。指定された範囲に座標が存在しない場合は自由エネルギーの欄に“-----”を出力します。

■出力ファイル例

```
FILL> GAUS
      2      1 ; DIMENSION      NUMBER OF ATOMS
      6      ; ATOM ID

INFORMATION> WHAM ANALYSIS RESULT
      EXP-ID R. M. S. D (A)  AVERAGE      FREE-ENERGY
          1  4.000000      0.000000115  0.984069810E+01
          2  4.513558      0.000000103  0.990723128E+01
          3  4.090083      0.000000228  0.942155513E+01
          4  3.778652      0.000000000  -----
```

9. 拡張アンサンブル向け解析ツール

拡張アンサンブルでの結果を解析するためのツールです。

各種の拡張アンサンブル MD でサンプリングした構造をクラスタリングし、代表的な構造を PDB ファイルとして出力します。

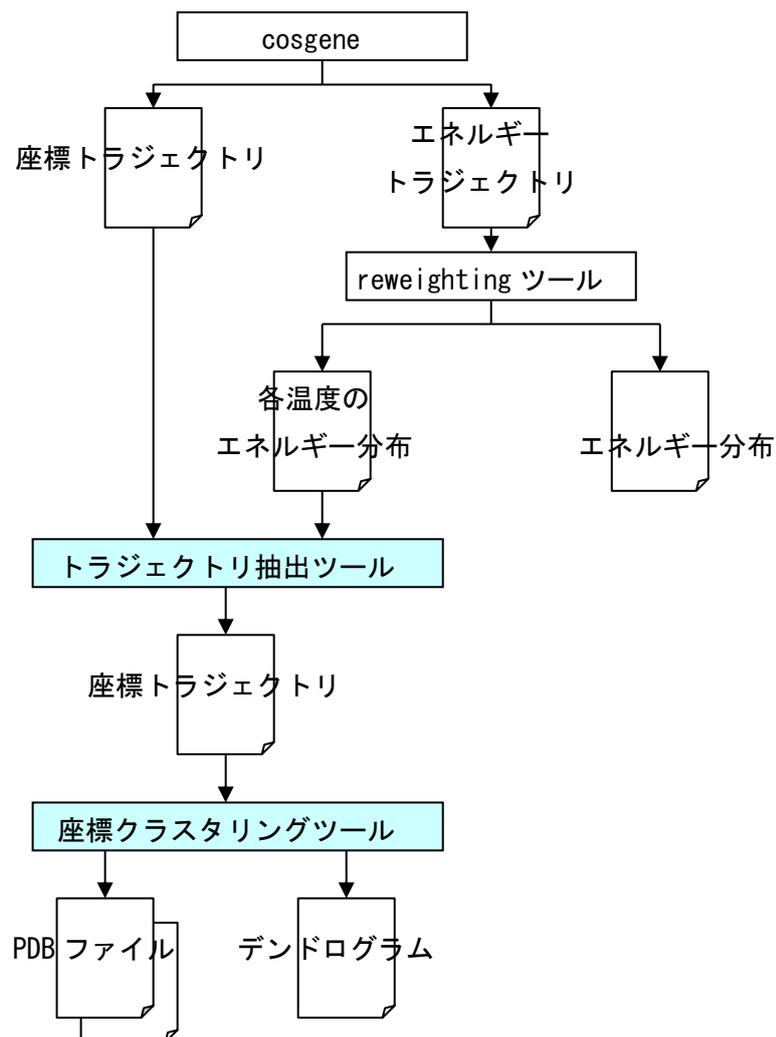
本ツールは座標抽出ツールと座標クラスタリングツールの二つで構成されています。

(1) トラジェクトリ抽出ツール

reweighting ツールが出力したエネルギー確率分布を元に座標トラジェクトリを抽出するツール

(2) 座標クラスタリングツール

座標トラジェクトリをクラスタリングするツール



9. 1. reweightFB

Force-biased McMD のエネルギートラジェクトリファイル、およびユーザ指定を読み込み、新規の canonical 分布を示すファイルを作成します。

入力データ

- (1) 制御ファイル
 - (1-1) Force-biased McMD 法での MD エネルギートラジェクトリファイル名
 - (1-2) 出力ファイル名
 - (a) 収束ループでのループ回数と relative partition function 値
 - (b) エネルギー、密度および総合確率
 - (c) 各温度でのエネルギーの canonical 分布
 - (d) 温度、平均エネルギー
 - (1-3) ヒストグラムパラメータ
 - (a) bin のサイズ(KCAL/MOL) (MD の値とあわせる)
 - (b) bin の個数(MD の値とあわせる)
 - (c) エネルギートラジェクトリのサンプリング範囲の下限
 - (d) エネルギートラジェクトリのサンプリング範囲の上限
 - (1-4) 温度、確率密度関数のパラメータ
 - (a) シミュレーション時の設定温度
 - (b) 出力する canonical 分布の温度下限(K)
 - (c) 出力する canonical 分布の温度上限(K)
 - (d) 出力する canonical 分布の間隔(K)
 - (e) canonical 分布を求める確率密度関数の下限値
- (2) Force-biased McMD 法での MD エネルギートラジェクトリファイル

■制御ファイル例

F. B. scale	:	Force-biased McMD のエネルギートラジェクトリファイル名
function. dat dencity. dat canonical. dat temperature. dat	:	出力ファイル名
1.0 351 1 44	:	bin サイズ bin 数
	:	エネルギートラジェクトリのサンプリング範囲の下限
	:	エネルギートラジェクトリのサンプリング範囲の上限
600 200 800 10 1. d-05	:	温度 温度下限 温度上限 温度ステップ 確率密度関数下限

■使用例

```
% reweightFB < rew. inp
```

9. 2. reweightST

Simulated-Tempering MCMD のエネルギートラジェクトリファイル、およびユーザ指定を読み込み、新規の canonical 分布を示すファイルを作成します。

入力データ

(1) 制御ファイル

(1-1) Simulated-Tempering 法での MD のエネルギートラジェクトリファイル名

(1-2) 出力ファイル名

(a) 各温度でのエネルギーの canonical 分布

(b) 各温度の平均エネルギー

(1-3) ヒストグラムパラメータ

(a) ポテンシャルの bin のサイズ (KCAL/MOL) (実行した MD の値とあわせる)

(b) 温度の bin のサイズ

(c) 分布を出力する温度の下限

(d) 分布を出力する温度の上限

(e) 温度の分割数

(1-4) サンプルング区間

(a) サンプルング区間の先頭

(b) サンプルング区間の最後

(1-5) 初期温度

(2) Simulated-Tempering 法での MD のエネルギートラジェクトリファイル

("S. T. energy")

■制御ファイル例

S. T. energy	:	Simulated-Tempering のエネルギートラジェクトリ
canonical average	:	出力ファイル名
1.0 100.0 200.0 700.0 6.0	:	ポテンシャル bin サイズ 温度 bin サイズ
	:	温度下限 温度上限 温度分割数
0 800	:	サンプルング区間先頭 サンプルング区間最後
600.0	:	初期温度

■使用例

```
% reweightST < rew.inp
```

9.3. reweightGST

Generalized Simulated-Tempering MCMD のエネルギートラジェクトリファイル、およびユーザ指定を読み込み、新規の canonical 分布を示すファイルを作成します。

入力データ

(1) 制御ファイル

(1-1) Generalized Simulated-Tempering 法での MD のエネルギートラジェクトリファイル名

(1-2) 出力ファイル名

- (a) 収束ループでのループ回数とポテンシャルの分配関数
- (b) エネルギー、密度および総合確率
- (c) 各温度でのエネルギーの canonical 分布
- (d) 温度、平均エネルギー

(1-3) ヒストグラムパラメータ

- (a) ポテンシャルの bin のサイズ (KCAL/MOL) (実行した MD の値とあわせる)
- (b) λ の下限 (実行した MD の値とあわせる)
- (c) λ の上限 (実行した MD の値とあわせる)
- (d) λ の分割数
- (e) G. S. T. でのエネルギー基準値 (実行した MD の値とあわせる)
- (f) η の値 (実行した MD の値とあわせる)

(1-4) サンプルング区間

- (a) サンプルング区間の先頭
- (b) サンプルング区間の最後

(1-5) 温度

- (a) 分布を出力する温度の下限
- (b) 分布を出力する温度の上限
- (c) 初期温度

(2) G. S. T. 法でのエネルギートラジェクトリファイル ("G. S. T. energy")

■制御ファイル例

G. S. T. energy	;	Generalized Simulated Tempering のエネルギートラジェクトリ
partition density canonical average	;	出力ファイル名
1.0 0.001 0.006 10 0.0 4.5	;	bin サイズ λ 下限 λ 上限 λ 分割数 エネルギー基準値 η 値
200 800 100	;	温度下限 温度上限 温度分割の幅

9.4. selection

トラジェクトリ抽出ツールは、入力したポテンシャルエネルギーの確率分布に従って、座標トラジェクトリの構造を抽出し、再構成するツールです。

入力データ

トラジェクトリ抽出ツールの入力を以下に示します。

(1) 座標トラジェクトリ

cosgene の出力トラジェクトリ

(2) エネルギー確率分布

reweighting ツールの出力ファイル

(3) トラジェクトリ抽出ツールの制御ファイル

(3-1) エネルギー確率分布ファイル名

(3-2) 座標トラジェクトリファイル名

(3-3) トラジェクトリファイルの型 (Single | Double)

(3-4) サンプルング区間先頭

(3-5) サンプルング区間最後

(3-6) 確率分布にかける係数 (抽出する座標数はこの係数に比例する)

(3-7) 出力トラジェクトリファイル名

(3-8) 原子数

制御ファイル例)

bestfit 対象原子指定ファイル例 (水素以外の蛋白質原子のみ bestfit する)

```
pdf. total
ala8. cor_ST
S
0
10000000
100.0
select. cor
32
```

標準出力例)

```
***** COORDINATE TRAJCTORY SELECT TOOL FOR COSGENE (2005/08/31) *****
FUNCTION : SELECT TRAJCTORY AND OUTPUT TRAJCTORY FILE

INPUT :
(1) ENERGY PROBABILITY DENCITY FUNCTION FILE NAME
(2) COSGENE TRAJCTORY FILE NAME
(3) TRAJCTORY FORMAT
(4) START LOOP NUMBER
(5) END LOOP NUMBER
(6) SELECTION RATE
(7) OUTPUT TRAJCTORY FILE NAME
OUTPUT :
(1) SELECTED TRAJCTORY
*****

INPUT ENERGY PROBABILITY DENCITY FUNCTION FILE NAME
INPUT TRAJCTORY FILE NAME
INPUT COORDINATE TRAJCTORY FORMAT ("S"ingle | "D"ouble)
INPUT START LOOP NUMBER
INPUT END LOOP NUMBER
SELECTION RATE (0.0 < RATE
OUTPUT NEW TRAJCTORY FILE NAME
***** SELECT TRAJCTORY RESULT *****

1) DISTRIBUTION
POTENTIAL-ENERGY  PROBABILITY(%)  TRAJCTORIES  SAMPLES  SAMPLE-RATE(%)
-0.55000E+01      0.933          0           0         -----
-0.45000E+01      1.000          1           1        100.000
-0.35000E+01      0.987          0           0         -----
-0.25000E+01      0.970          0           0         -----
-0.15000E+01      1.007          0           0         -----
-0.50000E+00      1.023          4           4        100.000
 0.50000E+00      1.039          2           2        100.000
 0.15000E+01      1.031          4           4        100.000
 0.25000E+01      1.022         10          10       100.000
 0.35000E+01      0.952         13          12        92.308
 0.45000E+01      0.905         10          10       100.000
 0.55000E+01      0.837         12           9        75.000
 0.65000E+01      0.796         15          11        73.333
 0.75000E+01      0.776         13          11        84.615
P. D. F. SUM=    99.9852000000000
STRUCT SUM=      2000
SAMPLE SUM=     1056

2) INPUT FILES
   TRAJCTORY FILE      :
   ala8.cor_ST

3) SELECTION
   TOTAL TRAJCTORY NUMBER :      2000
   SAMPLING BOUND        :      0 - 10000000
   SAMPLING NUMBER       :      2000
   RATE                   :    100.000000000000
   OUTPUT NUMBER         :      1056
   TRAJCTORY FILE       :
   select.cor

*****
```

9.5. clustering

各種の拡張アンサンブル計算でサンプリングした構造をクラスタリングし、代表的な構造を PDB ファイルとして出力します。

入力データ

- (1) 制御ファイル
 - (1-1) トポロジーファイル名
 - (1-2) ベストフィットの適用 ("Y" | "N")
 - (1-2-1) ベストフィット対象原子指定ファイル名
 - (1-3) RMSD 計算対象原子の指定 ("Y" | "N")
 - (1-3-1) RMSD 計算対象原子指定ファイル名
 - (1-4) サンプル数
 - (1-5) クラスタ数
 - (1-6) サンプリング区間先頭
 - (1-7) サンプリング区間最後
 - (1-8) 入力トラジェクトリファイル名
 - (1-9) トラジェクトリファイルの型 ("S" | "D")
 - (1-10) クラスタリング方法 ("nearest " | "furthest" | "median " | "centroid"
| "average " | "flexible" | "ward ")
 - (1-10-1) flexible 指定時の β 値
 - (1-11) 出力 PDB 名先頭
 - (1-12) デンドログラムファイル名
- (2) cosgene の入力トポロジーファイル
- (3) cosgene の出力トラジェクトリファイル
- (4) ベストフィット対象原子指定ファイル (ベストフィット対象原子指定時)
ベストフィット対象原子を cosgene の「系の重心合わせ指定用ファイル」と
同じ書式で指定します。(p. 161「A. 2. 11 系の重心合わせ指定用ファイル」を参照)
- (5) RMSD 計算対象原子指定ファイル (RMSD 計算対象原子指定時)
RMSD 計算対象原子を cosgene の「系の重心合わせ指定用ファイル」と同じ書式で
指定します。(p. 161「A. 2. 11 系の重心合わせ指定用ファイル」を参照)

【注意】

メモリ使用量の観点から、サンプリングする構造は 1000 個以内が望ましいです。

【注意】

サンプリングする構造を 1000 個以上に指定することは可能です。メモリ確保に失敗したときは以下のエラーメッセージを出力し、プログラムの実行を停止します。

“CANNOT ALLOCATE MEMORY, DECREASE SAMPLING NUMBER”

■制御ファイル例

```
ala8. tpl      ; トポロジーファイル名
n              ; ベストフィットの適用
n              ; RMSD 計算対象原子の指定
10            ; サンプル数
10            ; クラスタ数
10            ; サンプリング区間先頭
40            ; サンプリング区間最後
ala8. cor_ST  ; 入力トラジェクトリファイル名
S             ; トラジェクトリファイルの型
nearest       ; クラスタリング方法
ala8. cls     ; 出力 PDB 名先頭
ala8. tree    ; デンドログラムファイル名
```

■ベストフィット対象原子指定ファイル例

(水素以外の蛋白質原子のみベストフィットする例)

```
SETBST> LIST
FIX 1   1  1 32 H* YES ; 蛋白(チェーン1)の1~32 残基の“H*”は bestfit 対象外
FIX 2   2  1  1 *  YES ; リガンド(チェーン2)の全原子は bestfit 対象外
FIX 3 1000 1  1 *  YES ; 水分子(チェーン3~1000)の全原子は bestfit 対象外
```

■RMSD 計算対象指定ファイル

(水素以外のリガンド原子のみ RMSD 計算する例)

```
SETBST> LIST
FIX 1   1  1 32 *  YES ; 蛋白(チェーン1)の1~32 残基の全原子は bestfit 対象外
FIX 2   2  1  1 H* YES ; リガンド(チェーン2)の“H*”は bestfit 対象外
FIX 3 1000 1  1 *  YES ; 水分子(チェーン3~1000)の全原子は bestfit 対象外
```

出力データ

(1) ログ (標準出力に出力)

(1-1) ツールの使用方法

(1-2) データ入力問い合わせ

(1-3) クラスタリング条件

(1-4) 入力トポロジーファイル情報

(1-5) ベストフィット対象原子一覧 (ベストフィット対象原子指定ファイルに表示指定がある場合)

(1-6) RMSD 計算対象原子一覧 (RMSD 計算対象指定ファイルに表示指定がある場合)

(1-7) クラスタリング進行状況

(1-8) 出力 PDB ファイル名

(2) 代表構造の PDB ファイル

クラスタ番号、構造数、エネルギーおよびループ回数をコメント出力し、原子情報を出力します。出力ファイル名は「"出力 PDB 名先頭"+". "+ループ回数」となります。

(3) デンドログラムファイル

ループ回数とポテンシャルを葉の名称としたデンドログラムを出力します。

■代表構造の PDB ファイル例

```
REMARK CLUSTER : 1
REMARK STRUCTURE NUMBER: 140
REMARK LOOP : 10000
REMARK POTENTIAL : 176.955627441406
ATOM 1 CA ACE 1 2.508 1.314 -3.948 12
ATOM 2 HH31 ACE 1 2.771 1.634 -4.954 1.01 0.11
ATOM 3 HH32 ACE 1 2.166 0.280 -3.974 1.01 0.11
ATOM 4 HH33 ACE 1 1.718 1.947 -3.546 1.01 0.11
ATOM 5 C ACE 1 3.771 1.408 -3.102 12.01 0.60
```

■デンドログラムファイル例

```
(
(
"10000 176.96 KCAL/MOL"
,
(
"13000 174.52 KCAL/MOL "
,
"16000 184.61 KCAL/MOL "
)
)
```

(つづく)

(つづき)

```
)  
.  
(  
(  
"19000 163.18 KCAL/MOL "  
.  
(  
"22000 162.05 KCAL/MOL "  
;  
"28000 147.56 KCAL/MOL "  
)  
)  
;  
"25000 146.70 KCAL/MOL "  
)  
)  
;
```

10. 存在確率 (Potential Mean Force) 解析ツール

PMF 解析ツールは、1次元あるいは2次元のデータと存在確率を入力し、各データが存在する確率を出力します。カノニカルアンサンブルおよびマルチカノニカルアンサンブル計算に対応しています。

2次元データの解析結果はPMF ツールの解析結果を等高線作成ツールで、excel 用の等高線のグラフに加工することができます。

10.1. pmf

モニター指定トラジェクトリファイル、エネルギートラジェクトリファイル、エネルギー確率分布ファイル (マルチカノニカルアンサンブルに対しては必須) およびユーザ指定を入力し、各データの存在確率を出力します。

入力データ

- (1) 制御ファイル
 - (1-1) MD の形式 ("C"anonical | "M"ulti-canonical)
 - (1-2) データの次元数 (1 | 2)
 - (1-3) トラジェクトリデータの型 ("S"ingle | "D"ouble)
 - (1-4) モニター指定トラジェクトリファイル名
 - (1-5) エネルギートラジェクトリファイル名
 - (1-6) データの上限、加減 (2次元の場合は計4つ)
 - (1-7) データの bin の数 (2次元の場合は計2つ)
 - (1-8) 出力ファイルのフォーマット ("N" | "S" | "C")
 - N: 全ヒストグラムデータ
 - S: スキャッタープロット形式
 - C: 等高線データ
 - (1-9) 出力ファイル名
 - (1-10) エネルギー確率分布ファイル ((1-1) が "M"の場合は必須)
 - (1-11) スキャッタープロットのサンプル数 ((1-8) が "S"の場合は必須)
 - (1-12) 出力データ種別 ("P"robability | "E"nergy)
 - (1-13) エネルギー換算のための温度 ((1-12) が "E"の場合は必須)
- (2) モニター指定トラジェクトリファイル
- (3) エネルギートラジェクトリファイル

(4) エネルギー確率分布ファイル ((1 - 1) が "M" の場合は必須)

■制御ファイル例

C	;	実行した MD の種別 "C" anonical "M" ulti-canonical
2	;	構造の数 (1 or 2)
S	;	モニター指定・エネルギートラジェクトリの型 ("S" ingle "D" ouble)
aa. tra	;	モニター指定トラジェクトリファイル名
aa. ene	;	エネルギートラジェクトリファイル名
-180.0 180.0 -180.0 180.0	;	構造 1 の上下限 構造 2 の上下限
30 30	;	構造の分割数
C	;	出力形式 ("N" ormal "S" catter-plot "C" ontour-map)
cont. data	;	出力ファイル名
P	;	確率を出力

出力データ

入力データ (1 - 8) の指定が "N", "S", "C" の場合、それぞれ以下のデータを出力します。

(1) "N" の場合

入力データに従ったヒストグラムを作成し、csv ファイル形式でヒストグラムデータを作成します。

各行は、"構造 1 の下限, [構造 2 の下限,] 確率" で構成されます。

■出力例

-0.1800000E+03, -0.1800000E+03, 0.2000000E-02
-0.1680000E+03, -0.1800000E+03, 0.5200000E-02
-0.1560000E+03, -0.1800000E+03, 0.1160000E-01

(2) "S" の場合

確率分布に従い、ユーザが指定した数だけ代表点を作成します。

データ形式は (1) と同じ csv ファイル形式です。

(3) "C"の場合

先頭行がデータの行列数、データの下限值、データの範囲、最大値、モニター指定トラジエクトリ名で、次の行から確率分布を行列で表した csv ファイル形式のデータを作成します。

このファイルは等高線作成ツールで、等高線データに変換することができます。

■出力例

```
SIZE= 8 8 LOWER= -180.0 -180.0 BOUND= 360.0 360.0 MAX = 0.312E-01 FILE=aa.tra  
0.200E-02, 0.520E-02, 0.116E-01, 0.144E-01, 0.240E-02, 0.160E-02, 0.320E-02, 0.200E-02  
0.120E-02, 0.120E-02, 0.200E-02, 0.400E-02, 0.360E-02, 0.800E-03, 0.800E-03, 0.120E-02  
0.800E-03, 0.400E-03, 0.360E-02, 0.400E-02, 0.280E-02, 0.800E-03, 0.800E-03, 0.400E-03
```

10. 2. contour

pmf ツールで作成した等高線データファイルを入力し、CSV ファイル形式の等高線データを作成します。

入力データ

- (1) 制御ファイル
 - (1-1) 入力等高線データファイル名
 - (1-2) 等高線の数
 - (1-3) 等高線の値
 - (1-4) 出力等高線ファイル名

出力データ

- (1) 等高線ファイル

11. pca

主成分分析により、指定した原子の座標のクラスタリングを行い、代表構造と主成分分析結果を出力します。

入力データ

- (1) 制御ファイル
 - (1-1) トポロジーファイル名
 - (1-2) 構造の重ね合わせの適用 ("Y" | "N")
 - (1-3) 重ね合せ対象原子指定ファイル名 ((1-2) が "Y" の場合に必須)
 - (1-4) RMSD 計算対象原子の特定 ("Y" | "N")
 - (1-5) RMSD 計算対象原子指定ファイル ((1-4) が "Y" の場合に必須)
 - (1-6) サンプリングする構造数
 - (1-7) クラスタの個数
 - (1-8) サンプリングの開始
 - (1-9) サンプリングの最後
 - (1-10) 座標トラジェクトリファイル名
 - (1-11) 座標トラジェクトリファイルの型 ("S" | "D")
 - (1-12) クラスタリング手法
("nearest" | "furthest" | "median" | "centroid" |
"average" | "flexible" | "ward")
 - (1-13) flexible での β 値 ((1-12) が "flexible" の場合に必須)
 - (1-14) 主成分のスケーリング適用 ("Y" | "N")
 - (1-15) プロットデータの1軸
 - (1-16) プロットデータの2軸
 - (1-17) kmax の値
 - (1-18) プロットデータのファイル名
- (2) 重ね合せ対象原子指定ファイル ((1-2) が "Y" の場合)
- (3) RMSD 計算対象原子指定ファイル ((1-4) が "Y" の場合)
- (4) 座標トラジェクトリファイル

出力データ

- (1) プロットデータファイル
- (2) 樹形図ファイル (ファイル名は "pca.tree")
- (3) 代表構造 PDB ファイル (ファイル名は "pca.*")

■制御ファイル例

```
ala_ala.tpl ; topology file name
y           ; use bestfit ("y" | "n")
ala_ala.bst ; bestfit file name, when use bestfit
y           ; restrict rmsd target ("y" | "n")
ala_ala.rmsd ; rmsd target file name, when restrict rmsd target
200         ; sampling number of coordinate
4           ; delegate structure count
0           ; sampling start number
1000        ; sampling last number
select.cor  ; coordinate trajectory file name
s           ; coordinate trajectory file format ( "s" | "d" )
average     ; clustering method name
pca         ; result pdb file prefix
n           ; scale principle component ("y" | "n")
2           ; 1-axis plot data dimension
3           ; 2-axis plot data dimension
30          ; number of clustering elements
pca.plot    ; plot data
```

12. Gamess2tplinp

量子化学計算プログラム GAMESS の出力ファイルから tplgeneL の入力ファイルを作成します。

入力データ

- (1) GAMESS の出力ファイル名

出力データ

- (1) 電荷情報ファイル(XXX. charge)
- (2) 結合次数情報ファイル(XXX. bond)
- (3) Z-matrix 情報ファイル(XXX. zmat)

【注意】 上記の XXX は GAMESS の出力ファイル名から拡張子を除いたものです。

■ 使用法

```
% Gamess2tplinp methanol.log
```

13. Gauss2tplinp

量子化学計算プログラム Gaussian98 の出力ファイルから tplgeneL の入力ファイルを作成します。

入力データ

- (1) Gaussian98 の出力ファイル名

出力データ

- (1) 電荷情報ファイル(XXX. charge)
- (2) 結合次数情報ファイル(XXX. bond)
- (3) Z-matrix 情報ファイル(XXX. zmat)

【注意】 上記の XXX は Gaussian98 の出力ファイル名から拡張子を除いたものです。

■ 使用法

```
% Gauss2tplinp methanol.out
```

14. tpl2mol2

トポロジーファイルと PDB ファイルから、MDL mol または Sybyl mol2 形式のファイルを作成して出力します。以下のオプションを用いて、入出力のファイル形式を指定します。

オプション

- ipdb <pdbfile>
PDB ファイル<pdbfile>を入力ファイルとします
- itpl <tplfile>
トポロジーファイル<tplfile>を入力ファイルとします
- omol2 <mol2file>
出力ファイルを Sybyl mol2 ファイル<mol2file>とします
- omdl <mdlfile>
出力ファイルを SD ファイル<mdlfile>とします
- h, -help
ヘルプメッセージを表示します

【注意】 トポロジーファイルおよび PDB ファイルの指定 (オプション: -ipdb, -itpl) は必須です。

■ 使用法

```
% tpl2mol2 -ipdb 2ala.pdb -itpl 2ala.tpl -omol2 2ala.mol2 -omdl 2ala.mol
```

15. add_ion

溶質（溶媒水分子以外の分子）の作り出す電場を、各溶媒水分子の座標について距離依存誘電率（ $\epsilon \propto r$ ）で計算し、もっとも電位の高いところ、ないし低いところにある水分子をカウンターイオンで置換します。カウンターイオンで置換後、同様の計算を繰り返し、指定した数のカウンターイオンを全て配置するまで作業を繰り返します。この時、次のカウンターイオンは、前回までに配置したカウンターイオンより一定距離以上離れた場所に配置します。

入力データ

1行目：入力ファイル名：溶媒水分子の付加された全系の座標ファイル名。

2行目：出力ファイル名：水をカウンターイオンに置き換えた全系の座標ファイル名。

3行目：イオンの付加方法；

1：カウンターイオンの数を直接入力

2：系の電荷を中和する必要最低限の数を自動的に付加する。

3：イオンの濃度（イオン mol/水 mol）を指定し、系の電荷を中和するイオン数を自動的に付加する。生理食塩水の場合は、0.00277 です。

4：生理食塩水濃度でのイオン濃度

イオン付加方法=1の場合

4行目：Na⁺イオンの数

5行目：Cl⁻イオンの数

6行目：カウンターイオンを順次付加したとき、次のカウンターイオンを以前付加したカウンターイオンより一定距離内に配置しない。その半径（Å）。

7行目：水とイオンの交換確率(0.0~1.0の値)

イオン付加方法=2の場合

4行目：カウンターイオンを順次付加したとき、次のカウンターイオンを以前付加したカウンターイオンより一定距離内に配置しない。その半径（Å）。

5行目：水とイオンの交換確率(0.0~1.0の値)

イオン付加方法=3の場合

4行目：イオンの濃度（イオン mol/水 mol）

5行目：カウンターイオンを順次付加したとき、次のカウンターイオンを以前付加したカウンターイオンより一定距離内に配置しない。その半径（Å）。6~8Åが適切です。

6行目：水とイオンの交換確率(0.0~1.0の値)

イオン付加方法 = 4 の場合

4 行目 : カウンターイオンを順次付加したとき、次のカウンターイオンを以前付加したカウンターイオンより一定距離内に配置しない。その半径 (Å)。6~8Å が適切です。

5 行目 : 水とイオンの交換確率 (0.0~1.0 の値)

■ 入力例 : Na/Cl イオンの数を全部指定する場合

```
zifcmp.pdb_vac  
zifcmx.pdb  
1  
80  
72  
6.0  
0.2
```

■ 入力例 : 最も良く使う生理食塩水を使う場合

```
zifcmp.pdb_vac  
zifcmx.pdb  
3  
0.00277  
6.0  
0.2
```

■ 使用例

add_ion とタイプする。標準入力から入力します。入力例は、ion.input としています。

```
% add_ion < ion.input
```

【注意】 `add_ion` によって付加されたカウンターイオンの配置は、エネルギー的に安定ではありません。従って、全系の MD 計算に移る前に、蛋白質・DNA の座標を固定した状態で、溶媒水とカウンターイオンだけの溶媒部分に対し MD 計算を行い、溶媒部分を十分平衡状態に近づけておきます。

【注意】 蛋白質、DNA、溶媒水、カウンターイオンの並びは、トポロジーファイルの MOLECULES 欄と PDB 上で同じ順番でなければなりません。

16. confgene

Sybyl mol2 形式のファイルで記述された入力分子の配座を発生し、PDB 形式、Z-matrix、RESP 入力ファイル、Gaussian 入力ファイルを作成して出力します。配座の発生は、ランダムサーチにより、環以外の部分についてのみ行います。電荷情報は、ユーザ自身が手入力で指定できますが、自動計算も可能です。

【注意】電荷の自動計算は、完全ではありません。また、結合次数の情報が正確でなければ、電荷の計算は正しく行われません。原子の順番は、並べ替えられます。PDB 出力、RESP 出力、Gaussian 入力ファイルなどの出力ファイルでの原子の並びは、相互に矛盾無く出力されます。

入力データ

- (1) 配座を発生させたい分子の mol2 ファイル名
- (2) 発生させたい配座の総数。

配座数が原理的に存在しない場合やランダムサーチの一定の試行回数において原子間衝突のない配座が得られない場合、指定した数よりも生成する配座の数が少なくなる場合があります。

- (3) 回転角の指定 (N)

回転可能な二面角を $(360 \div N)$ 度ずつ回転させることで配座を発生させます。

- (4) 原子の総電荷。

自動計算の場合は「a」と入力。数値は、左詰めで、空白を数字の前に入れないこと。

- (5) 原子を並べ替えるときの、出発点とする原子の番号。

分子に含まれる原子数以下の任意の数字。通常 1 で良い。この原子を出発点に、グラフ上で近い原子に、若い番号を割り当てるように原子の並べ替えを行ないます。

出力ファイル

(以下、生成した配座の数が N 個の場合)

- (1) conf1.pdb ~ confN.pdb :

発生した配座。conf1.pdb は、入力座標と同じです。メチル基など等価な原子が回転した場合も、異なる配座と数えるため、化学的には同じ配座が含まれる場合があります。

- (2) conf1.com ~ confN.com :

発生した配座 (conf1.pdb~confN.pdb) に対応する Gaussian 入力ファイル。実行オプションは AM1 での構造最適化になっていますが、コメント行には RESP 計算用の電場グリッド発

生オプションが記載されます。

(3) resp.in :

RESP 入力ファイル

(4) qin :

RESP 用初期電荷ファイル。全て電荷は 0 と設定しています。

(5) zmat.dat :

分子の初期配座での Z-matrix です。

■ 使用法

```
% confgene
Input File name (mol2 file)
ligand.mol2 (1)
File =ligand.mol2
Input number of conformers
2 (2)
no_conf= 2
Input number of rotation phase(=6:60 deg, =3:120 deg)
3 (3)
no_phase= 3
Input total charge of the molecule(a=auto calc)
-2 (4)
charge =-2
readmol3=ligand.mol2
numatom 41, 40
Input start atom number
1 (5)
```

17. confgeneC

Sybyl mol2 形式のファイルで記述された入力分子の配座を発生し、Sybyl mol2、MDL mol 形式、PDB 形式ファイルを作成して出力します。配座の発生は、4 員環以上の環構造の部分について生成します。また、分子内にキラル中心が存在する場合には、光学異性体も同時に生成して出力します。

入力データ

必須項目

- (1) 配座を発生させたい分子のファイル名
- (2) 入力分子のファイルフォーマット(1 : Sybyl mol2、2 : MDL mol、3 : PDB)
- (3) 発生させたい配座の総数

採り得る全ての配座を生成させたい場合には「a」を指定します。

配座数が原理的に存在しない場合や配座生成時に原子間衝突のない配座が得られない場合、指定した数よりも生成する配座の数が少なくなる場合があります。また、「a」を指定した場合で異性体が多数ある場合には、最大 999 件のデータを出力します。

オプション(対話形式では指定できません)

- (4) 回転角の指定 (N)
回転可能な二面角を $(360 \div N)$ 度ずつ回転させることで配座を発生させます。指定しない場合には $N=6$ として処理を行います。
- (5) 原子間チェックオプション
チェックオプションを指定した場合には、原子間距離が近い構造を作成した場合には、構造が重ならない様に座標の修正を行います。

出力ファイル

(以下、生成した配座の数が N 個で、出力ファイル形式で Sybyl mol2 を指定した場合)

- (1) confXXX.mol2 : XXX は 1 から N の 3 桁の数字
発生した配座座標を出力したもの。出力ファイル形式で MDL mol、PDB を指定した場合には、拡張子がそれぞれ mol、pdb となります。
- (2) confXXXc.mol2 : XXX は 1 から N の 3 桁の数字
ファイル名の数字の後ろに c の文字がある場合には、光学異性体ファイルである事を示します。出力ファイル形式で MDL mol、PDB を指定した場合には、拡張子がそれぞれ mol、pdb となります。

■ 使用法

```
% confgeneC
Please select Input File Format by the next number!
  1 : Sybyl mol2 (*.mol2)
  2 : MDL mol (*.mol)
1 (1)

INFORMATION> toolGetFilename
  Sybyl mol2 input file was selected.

Please select Output File Format by the next number!
  1 : Sybyl mol2 (*.mol2)
  2 : MDL mol (*.mol)
  3 : PDB pdb (*.pdb)
1 (2)

INFORMATION> toolGetFilename
  Sybyl mol2 output file was selected.

Please select Input File Name!
sample.mol2 (3)
Please input number of conformers(a=all pattern).
a (4)
  input file = sample.mol2

number of conformers that want to be created = 999
num of rotation phase = 6

INFORMATION> toolSetChiralFlg
  This molecule has 3 chiral center(s).

INFORMATION> toolCountCirc
  Circular structure(s) have found.
INFORMATION> toolCreateChiralMol
  New coordinates are generated for chiral center "C(3)"
INFORMATION> toolCreateChiralMol
  New coordinates are generated for chiral center "C(5)"
INFORMATION> toolCreateChiralMol
  New coordinates are generated for chiral center "C(6)"

This program creates 11 conformers.

Program is done normally.
```

18. 自由エネルギー摂動法（開発中）

以下の（１）～（３）の機能により、自由エネルギー摂動法による自由エネルギー計算を行います。（１）は `cosgene` の機能で、（２）および（３）はツールです。

- （１）指定した原子の vdW パラメータおよび電荷をスケールリングし、スケールリングによって新たに発生する出力トポロジーデータに加える「vdW パラメータおよび電荷のスケールリング機能」
- （２）`cosgene` 形式のトポロジーファイルおよび座標トラジェクトリファイルを入力し、それらのトポロジーデータおよび座標データを用いて各ステップごとのエネルギー計算を行う「`analyze` ツール」
- （３）`cosgene` 形式のエネルギートラジェクトリファイルを２つ入力し、それらのエネルギーデータから自由エネルギー計算を行う「`FEP` ツール」

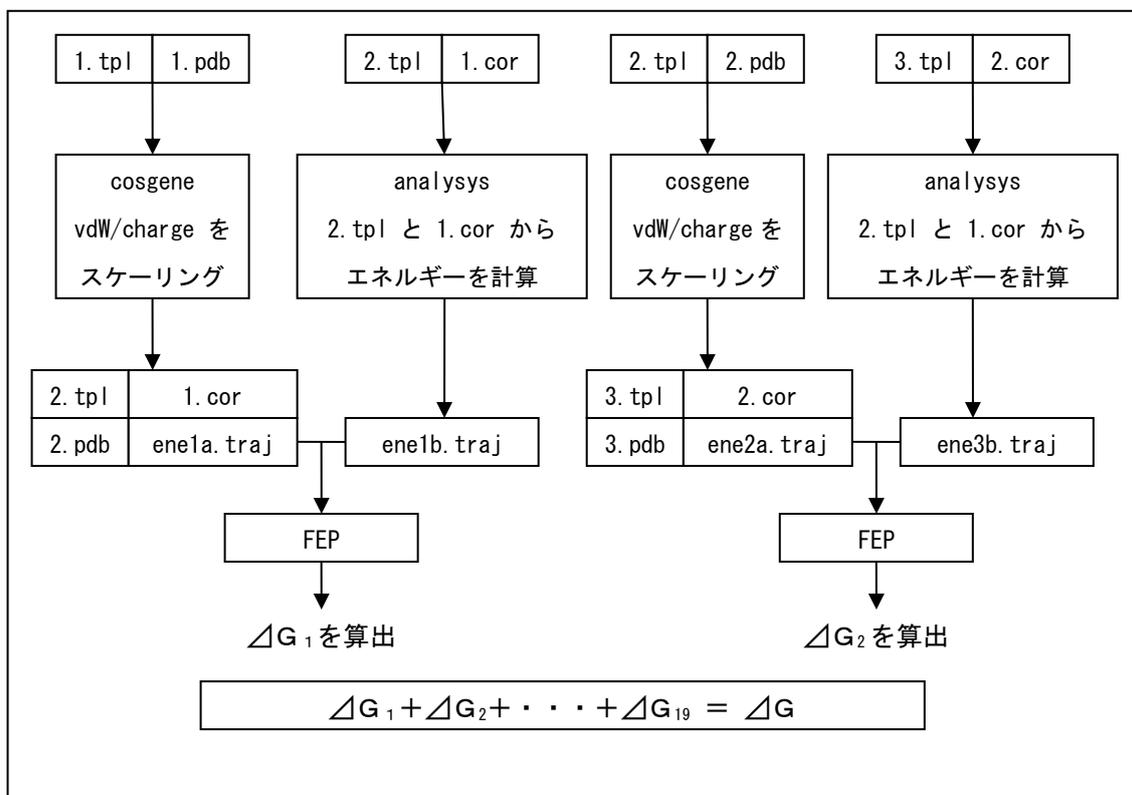
【注意】 自由エネルギー摂動法関連ツールは開発中であり、計算結果は保証しません。

18.1. 計算方式

以下の（１）～（４）手順で、自由エネルギー摂動法による自由エネルギー計算を行います。

- （１）`cosgene` による MD 計算を行い、PDB ファイル、座標トラジェクトリファイル、エネルギートラジェクトリファイル、スケールリングされたトポロジーファイルを出力する。
- （２）（１）で出力されたトポロジーファイルに基づき、（１）で出力された座標トラジェクトリファイルの座標についてエネルギー計算を行い、エネルギートラジェクトリファイルを出力する。
- （３）（１）および（２）で得られたエネルギートラジェクトリファイルを用いて自由エネルギー計算を行う。
- （４）（１）～（３）を繰り返し、それぞれの自由エネルギーの総和を算出することで、自由エネルギー摂動法による自由エネルギー計算を行う。

【 自由エネルギー摂動法イメージ図 】



自由エネルギー ΔG を算出する。

18.2. vdW パラメータおよび電荷のスケーリング機能 (cosgene)

cosgene に対して、スケーリング対象原子、vdW パラメータのスケーリングファクターおよび電荷のスケーリングファクターをファイル入力し、指定した原子の vdW パラメータおよび電荷のスケーリングを行います。スケーリングにより新たに発生するトポロジーデータは、出力トポロジーデータに加えられます。

使用方法

制御ファイルの OUTPUT フェーズでスケーリング指定を行い、INPUT フェーズで「スケーリングファイル」の入力指定を行います。対象原子およびスケーリングファクターの指定は「スケーリングファイル」で行います。

(1) スケーリング指定 (OUTPUT フェーズ)

項番	項目	キーワード	値	内容
#1	スケーリング指定	<u>TPLSCL</u>	選択型	VdW パラメータおよび電荷のスケーリング (<u>NO</u> YES)

(2) 「スケーリングファイル」入力指定 (INPUT フェーズ)

項番	項目	キーワード	値	内容
#1	スケーリング ファイル指定	<u>SCALIN</u>	選択型	「スケーリングファイル」指定 (<u>NORE</u> FORM)
#2		<u>UNITSC</u>	整数型	装置番号 (28)
#3		<u>NAMTSC</u>	文字列	ファイル名 (" ")

(3) 「スケーリングファイル」書式

対象フェーズ：OUTPUT フェーズ

用途：VdW パラメータおよび電荷のスケーリング対象原子およびスケーリングファクターを指定する。

書式：スケーリングファイルは以下の行で構成される。

[対象原子 ID vdW 半径スケーリングファクター 電荷スケーリングファクター]...
--

■使用例

1	0.95d0	0.95d0
6	0.90d0	0.90d0

18.3. analyze

トポロジーファイルおよび座標トラジェクトリファイルを入力し、各ステップのエネルギー計算を行います。計算結果の出力は、cosgene の書式による標準出力へのログ出力および cosgene のエネルギートラジェクトリファイル形式でのファイル出力を行います。

入力データ

(1) 制御ファイル

制御ファイルは以下のグループからなり、各グループは“QUIT”で終了します。

- **EXE> INPUT** グループ : 入力ファイル名を記載する。
- **EXE> MD** グループ : エネルギー計算条件を記載する。

※ 制御ファイルの書式は、cosgene の制御ファイルの書式を踏襲します。

※ cosgene の制御ファイルには **EXE> MIN** グループ、**EXE> ANALYZE** グループおよび **EXE> OUTPUT** グループがありますが、本ツールの制御ファイルにこれらの記述がある場合はそれらを読み飛ばします。(エラー処理は行いません。)

(2) トポロジーファイル (ASCII 形式のみ)

(3) 座標トラジェクトリファイル (ASCII, SINGLE, DOUBLE に対応)

座標トラジェクトリファイルの入力指定は、制御ファイルの INPUT フェーズで以下のように行います。

項番	項目	キーワード	値	内容
#1	座標トラジェクトリ ファイル指定	CRDTRJ	選択型	座標トラジェクトリファイル指定 (NORE ASCII SING DOUB)
#2		UNITCT	整数型	装置番号 (29)
#3		NAMTCT	文字列	ファイル名 (“ ”)

出力データ

(1) ログ (cosgene のログ書式で標準出力に出力)

(2) エネルギートラジェクトリファイル (ASCII, SINGLE, DOUBLE に対応)

■ 使用法

```
% analyze < analysis.inp > analysis.log
```

18. 4. FEP

2つのエネルギートラジェクトリファイルを入力し、それらのエネルギーデータから自由エネルギーを計算して標準出力に出力します。

入力データ

- (1) 制御ファイル
 - (1-1) 設定温度 [K]
 - (1-2) 温度の閾値 [K]
 - (1-3) 自由エネルギー計算ループ初期値
 - (1-4) トラジェクトリファイル名 (文字列、80字以内)
 - (1-5) (1-4) で指定したファイルのファイル形式
(“A” scii | “S” ingle | “D” ouble)
 - (1-6) トラジェクトリファイル名 (文字列、80字以内)
 - (1-7) (1-6) で指定したファイルのファイル形式
(“A” scii | “S” ingle | “D” ouble)
- (2) cosgene が出力するエネルギートラジェクトリファイル
(ASCII, SINGLE, DOUBLE に対応)
- (3) analyze が出力するエネルギートラジェクトリファイル
(ASCII, SINGLE, DOUBLE に対応)

出力データ

- (1) 自由エネルギー (標準出力に出力)

■ 使用法

```
% analyze < analysis.inp > analysis.log
```

■ FEP ツール制御ファイル例

```
300.0      : 設定温度
5.0        : 閾値
1          : ループ初期値
ini.trj    : トラジェクトリファイル名
D          : ini.trj のファイル形式 (“A” scii | “S” ingle | “D” ouble )
fin.trj    : トラジェクトリファイル名
D          : fin.trj のファイル形式 (“A” scii | “S” ingle | “D” ouble )
```


19. Hgene : 詳細は別冊 Hgene マニュアルを参照

化合物への H 原子付加/削除、原子電荷計算 (Gasteiger/MOPAC AM1 等) などを行います。入出力は PDB、MDL mol、Sybyl mol2、mmCIF または Mopac dat (出力のみ) 形式に対応しています。

用意されているオプションは以下の通りです (ヘルプ参照 “-H”)。

入力出力ファイルオプション (* .mdl, *.mol2/*.sm2, *.pdb)

<入力タイプ>

mdl : MDL mol ファイル
mol2 : Sybyl mol2 ファイル
pdb : PDB ファイル
cif : cif ファイル

<出力タイプ>

mdl : MDL mol ファイル
mol2 : Sybyl mol2 ファイル
pdb : PDB ファイル
mopcrt : mopac dat ファイル

※入力タイプ-icif(入力 cif)を指定した場合は、出力タイプ-opdb(出力 PDB)のみ対応しています。また、水素付加オプション等のオプションには対応していません。

ファイル名拡張子について

ファイル名には各フォーマットに対応した拡張子をつける必要があります。入出力タイプ毎に以下に示す拡張子以外を指定した場合にはエラーとなります。

-imdl -omdl : MDL mol ファイルの拡張子、入出力ファイルの拡張子は” .sdf” もしくは” .mol”

-imol2 -omol2 : Sybyl mol2 ファイルの拡張子、入出力ファイルの拡張子は” .sm2” もしくは” .mol2”

-ipdb -opdb : PDB ファイルの拡張子、入出力ファイルの拡張子は” .pdb”

-icif : cif ファイルの拡張子、入力ファイルの拡張子は” .cif”

-omopcrt : mopac dat ファイルの拡張子、出力ファイルの拡張子は” .dat”

処理内容オプション

各オプションをコマンドライン入力することにより、水素付加、水素付加方法選択、MOPAC 用電荷値取得等を行うことができます。

-h --hydrogen

: 出力データに、水素の原子情報、結合情報を付加します。後述の-p オプションと同時に指定した場合には、-p オプションが優先され、-p オプションの水素付加形式で水素付加を行います。

※入力ファイルが cif の場合は未対応

-d --delete-hydrogen

: 水素の原子情報、結合情報を削除してファイルを出力します。

-ch --charge

: コマンドラインで入力された-ch の次の引数をチャージ値として mopac dat ファイルの一行目に出力します。

※mopac dat ファイルを出力する場合のみ対応

-H --help

: Hgene のヘルプを出力します。

-dc --default-charge

: 入力データに含まれる電荷の値を mol2 ファイルに出力します。本オプションは出力タイプに Sybyl mol2 を指定した場合のみ対応しています。

(本オプションを指定しない場合(デフォルト)には、Gasteiger 電荷計算を行います)

※入力ファイルが Sybyl mol2 の場合、入力データの値をそのまま出力、

入力ファイルが MDL mol の場合、sd_charge の値を形式電荷に変換して出力します。

-p --ph

: 酸性/塩基性官能基が解離状態になるように水素の原子情報、結合情報を付加します。

-h オプションと同時に指定した場合には、-p オプションが優先され、-p オプションの水素付加形式で水素付加が行われます。

-m --metal

: コマンドラインで水素付加オプション(-h)、又は、解離状態水素付加オプション(-p)と同時に指定して金属配位型水素付加を行います。金属配位型水素付加は、カルボキ

シル基、リン酸基、スルホン酸基、ヒドロキサム酸、テトラゾール、キレート型構造のみとなります。それ以外の部位については、指定したオプション(-h、-p オプション)の水素付加方式に従います。

-bo --bondorder

: 指定した原子番号の結合について、指定した結合次数を割り当てます。複数個所の結合の結合次数を割り当てる場合は複数回指定します。

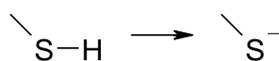
※入力が PDB ファイル(-ipdb オプション指定時)の時のみ有効となります。

※入力が cif ファイル(-icif オプション指定時)は上記のオプションは無視されます。

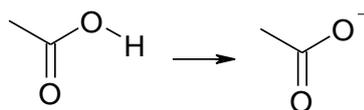
※-m --metal 指定時の水素付加形式について

以下に、-m --metal 指定時に対応する構造、及び、水素付加結果イメージを示します。

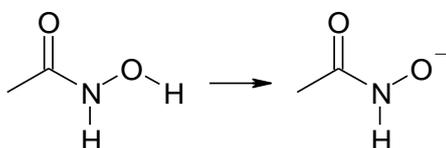
・チオール



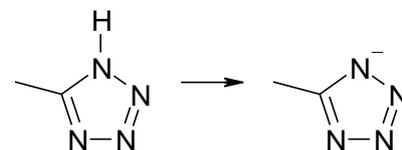
・カルボン酸



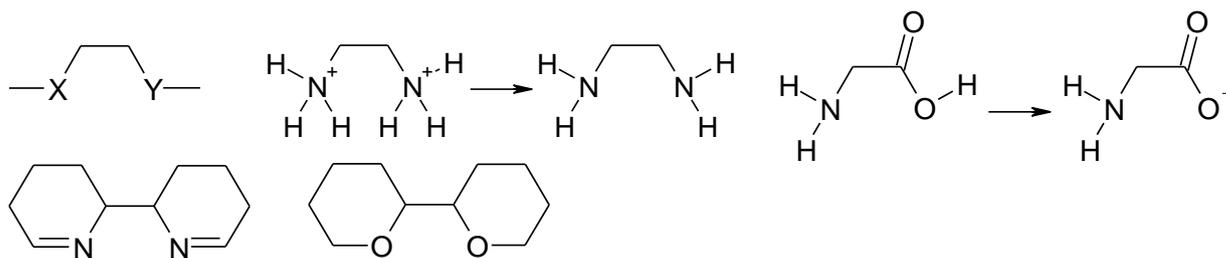
・ヒドロキサム酸



・テトラゾール



・キレート型(X, YはO又はN, Nは中性、Nの結合数3で平面構造を持つものは対象外)



-mop --mopac

: ハミルトニアンを指定して、MOPAC7 の計算を行います。指定方法は-mop キーワードの後に AM1、PM3 などのハミルトニアンを入力します。

出力ファイルに Sybyl mol2 または pdb ファイルを指定した場合には、MOPAC7 で計算した電荷を付与します。また、後述の-opt オプション指定時には構造最適化計算を行います。

※ 本機能は、Hgene コンパイル時に MOPAC7 を併せてコンパイルした場合のみ有効です。

-opt --optimization

: MOPAC7 の機能を利用して構造最適化計算を行います。
出力ファイルには最適化後の座標が出力されます。

※ 本機能は、Hgene コンパイル時に MOPAC7 を併せてコンパイルした場合のみ有効です。

-div -divide-molecule

: ファイル出力時に、1 つの分子を複数の残基に分割して出力します。最大 5 残基まで分割します。

-max -maximum-length

: 分割する残基の最大結合長を指定します。デフォルト値は 7 となっています。
-div オプション指定時のみ有効です。

-min -minimum-length

: 分割する残基の最大結合長を指定します。デフォルト値は 1 となっています。
-div オプション指定時のみ有効です。

-ligname -ligand-name

: 分割する残基の残基名を指定します。最初 2 文字を指定し、3 文字目は A から Z の文字を自動的に割り当てます。

■ 使用法

a) MDL mol → Sybyl mol2 変換 :

```
% Hgene -imdl sample.mol -omol2 sample.mol2
```

b) Sybyl mol2 → MDL mol 変換 :

```
% Hgene -imol2 sample.mol2 -omdl sample.mol
```

c) PDB → MDL mol 変換

```
% Hgene -ipdb sample.pdb -omdl sample.mol
```

d) 水素付加オプション (-h) :

```
% Hgene -imdl sample.mol -omol2 sample.mol2 -h
```

e) 水素削除オプション (-d) :

```
% Hgene -imdl sample.mol -omdl sample_out.mol -d
```

f) 電荷指定オプション (-ch) :

```
% Hgene -ipdb sample.pdb -omopert sample.dat -ch 0
```

g) 電荷入力値採用オプション (-dc) :

```
% Hgene -imdl sample.mol -omol2 sample.mol2 -dc
```

h) 解離型生成オプション (-p) :

```
% Hgene -imdl sample.mol -omol2 sample.mol2 -p
```

i) 結合次数指定オプション (-bo) :

```
% Hgene -ipdb sample.pdb -omdl sample.mol -p -bo 7 9 2
```

※原子 7 番と 9 番の結合に二重結合を割り当てる

20. MVO

MVO(Maximum Volume Overlap)法で化合物の重なり具合をスコアで評価し、最も重なりので大きい化合物の探索を行います。

入力化合物はマルチ mol2 ファイル形式に対応します。

探索対象化合物の mol2 ファイルに sievgene のスコアが付加されている場合は、sievgene のスコアと MVO スコアから得られるスコア値での評価ができます。

入力データ

(1) 制御ファイル

(1-1) 入力する参照化合物ファイル名 (mol2 ファイル)

(1-2) 入力する探索対象化合物ファイル名 (mol2 ファイル)

(1-3) 出力する探索結果の 1 位化合物ファイル名 (mol2 ファイル)

(1-4) MVO 計算対象となる二原子の電荷の差分絶対値の閾値

(1-5) sievgene スコアと MVO スコアの割合 C_{weight}

$$(\text{スコア} = \text{sievgene スコア} * (1 - C_{weight}) + \text{MVO スコア} * C_{weight})$$

出力データ

(1) 探索結果の 1 位化合物ファイル

ログ

標準出力に制御ファイルの設定条件、各化合物の MVO 計算結果、探索結果が表示されます。

```

*****
*
* selectMVO(Maximum Volume Overlap method) *
*
* Japan Biological Information Research Center *
* Aomi 2-41-6, Koto-ku, Tokyo 135-0064 *
* JAPAN *
*
* v1.0: May 26, 2010 *
*
*****

```

```

input reference file name [default=query.mol2]
input target file name [default=coordinate.mol2]
input best structure file name(output) [default=output.mol2]
input Q_THR value(threshold of atomic charge remainder) [default= 0.20000]
input MVO score weight(0.0<=x<=1.0) [default= 1.00000]

```

INFORMATION>

```

1) REFERENCE_FILE:
   ./test_old/lig_ref.mol2
      ATOMS      BONDS  MOLECULES
      48         52      2
2) TARGET_FILE:
   ./test_old/ex.cor
      ATOMS      BONDS  MOLECULES
      69         72      54
3) OUTPUT_FILE:
   output.mol2
4) Q_THR      :      0.20000
5) MVO_WEIGHT:      0.10000

```

MVO>

QUERY	TARGET	RMSD	SIEVGENE	MVO_SCORE	TOTAL_SCORE
1	1	29.01840	360.12521	338.34360	357.94702
1	2	29.69470	351.95981	341.00656	350.86447
1	3	29.18910	347.95001	292.69943	342.42493
:					
2	53	30.47210	317.28790	0.00000	285.55911
2	54	30.91340	331.87891	0.00000	298.69101

BEST (QUERY, TARGET) : 1 15

	QUERY	TARGET	RMSD	SIEVGENE	MVO	TOTAL_SCORE
@	1	15	29.81970	389.22760	140.72887	364.37772

■ 使用法

```
%selectMVO < mvo.inp > mvo.log
```

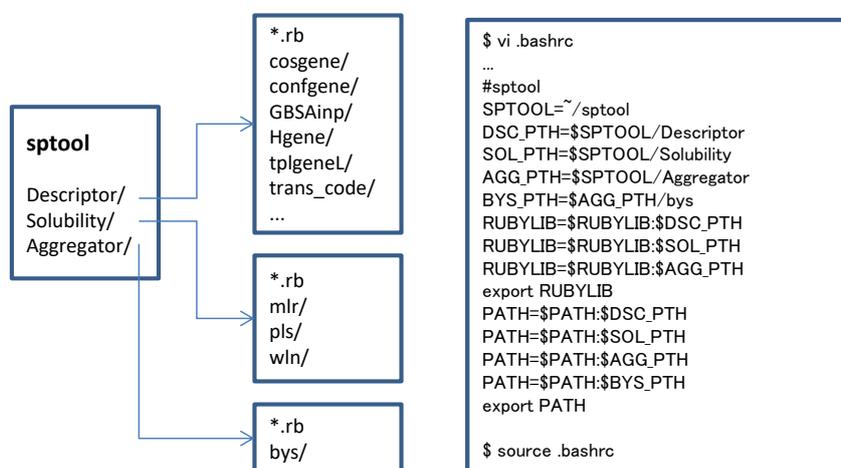
21. sptool

sptool は溶解度予測を行う為のツール群です。分子の構造情報をコード化し（記述子計算）、回帰計算により未知の分子の溶解度を推算します。また、推算した溶解度を使用してアグリゲータの予測も行います。

ツール群の機能は以下の3つ分かります。

名前	機能
Descriptor	記述子を計算する
Solubility	溶解度を推算する
Aggregator	アグリゲータを予測する

また、sptool は以下のディレクトリ構成となっています（下図左）。



プログラムの構築は、各種ツールのソースディレクトリ（src/等）に移動して適宜 make を実行して下さい。Ruby (ver1.8.5) スクリプトについては、起動ファイル (Descriptor.rb, Solubility.rb の SRC_DIR) にソースディレクトリの絶対パスを設定して下さい。また、環境変数の設定は必要に応じて行って下さい（上図右）。

21.1. Descriptor

分子の構造情報を読み込んで記述子を計算します。このツールは各記述子の計算に cosgene, confgene, Hgene, GBSAinp, tplgeneL, trans_code 等を利用します。インストール後、Descriptor.rb の SRC_DIR にソースディレクトリ (Descriptor/) を設定して下さい。

起動コマンド：

起動コマンドは以下の通りです。

Descriptor.rb -i <input_dir> -l <input_lst> -o <output_file>	
input_dir	single-mol2 ファイルを格納したディレクトリ名
input_lst	格納された single-mol2 ファイルのリスト

output_file	記述子ファイル名
-------------	----------

Mo12 ファイルリスト :

入力に必要な mo12 ファイルリストは以下の様に記述子します。

<code><id> <mo12_file> [<logS>]</code> (例 : 4 mo14.mo12 -1.590)	
id	mo12 ファイルの ID
mo12_file	mo12 ファイル名 (*.mo12)
logS	logS 値 (実験値、任意) logS 値は実験値が分かっている場合に設定 (データベースもしくは検証用) mo12 ファイル内の@<TRIPOS>COMMENT 以下に記述されている場合には必要ない

物理記述子 (Mo12 ファイル) :

mo12 ファイルの@<TRIPOS>COMMENT 以下に物理記述子を記述する事が出来ます。

記述は以下の書式行います。

<code>#<identifier> <value></code> (例 : #LogS -1.590)	
identifier	物理記述子 (LogS, DelO_H, AddN_H, ...)
value	設定値

記述可能な物理記述子は以下の通りです。

識別子	内容
LogS	logS 値
DelO_H	解離 H 原子数 (-OH)
AddN_H	付加 H 原子数 (-NH)
dG_o	オクタノール中の GB/SA 値
ddG_o	オクタノール中の GB/SA 値 (単位表面積当たり)
dASA_o	オクタノール中の 表面積 (単位体積当たり)
dG_w	水中の GB/SA 値
ddG_w	水中の GB/SA 値 (単位表面積当たり)
dASA_w	水中の 表面積 (単位体積当たり)
dG_wd	水中解離型の GB/SA 値
ddG_wd	水中解離型の GB/SA 値 (単位表面積当たり)
dASA_wd	水中解離型の 表面積 (単位体積当たり)

計算条件設定 :

指定可能な物理記述子の計算条件はそれぞれ指定個所が異なります。デフォルト値については各設定値を参照してください。

計算条件	指定個所
tplgeneL のパラメータ定義ファイル	Descriptor.rb 内の TPLLDBF
GBSAinp のパラメータ定義ファイル	Descriptor.rb 内の GBSADB

cosgene での最小化条件	min_template.erb
cosgene での GB/SA・ASA 計算条件	min_template_o.erb (オクタノール中) min_template_w.erb (水中)
confgene での発生配座数	Descriptor.rb 内の CNF_NUM
confgene での発生配座角度 (360/<n>)	Descriptor.rb 内の CNF_ANG

指定可能な構造記述子の計算条件も同様です。

計算条件	指定箇所
trans_code の記述子定義ファイル	Descriptor.rb 内の TRNSDB

21.2. Solubility

重み付き学習法による強化学習を伴った回帰計算を実行します。インストール後、Solubility.rb の SRC_DIR にソースディレクトリ (Solubility/) を設定して下さい。

起動コマンド：

起動コマンドは以下の通りです。

Solubility.rb -i <input_file> -o <output_file> -d <db_file> -m <w/n> <p/m>	
input_file	入力分子記述子ファイル名 (Descriptor.rb の output_file を想定)
output_file	回帰計算結果ファイル名
db_file	データベース記述子ファイル名 (Descriptor.rb の output_file を想定)
w/n	学習選択 w(eighted learning) / n(o learning)
m/p	回帰選択 m(lr) / p(ls)

回帰計算結果の内、推算結果は以下の様な書式になります。

推算結果：

項目	内容
ID	入力分子 ID
tchC1	データベースの回帰直線の係数 C1: $y = C1 * x + C2$
tchC2	データベースの回帰直線の係数 C2: $y = C1 * x + C2$
tchR2	データベースの決定係数
tchAveErr	データベースの平均エラー
tchMaxErr	データベースの最大エラー (絶対値が最大)
tchWL	重み付き学習時の追加デコイ数
tstExp	入力分子の logS (実験値)
tstPre	入力分子の logS (推算値)
tstErr	入力分子の推算エラー (絶対値)

推算精度：

項目	内容
tstC1	全入力分子の回帰直線の係数 C1: $y = C1 * x + C2$
tstC2	全入力分子の回帰直線の係数 C2: $y = C1 * x + C2$
tstR2	全入力分子の決定係数
tchAveErr	全入力分子の平均エラー
tchMaxErr	全入力分子の最大エラー (絶対値が最大)

21.3. FreqMaker

logS の頻度分布を作成します。

起動コマンド：

起動コマンドは以下の通りです。

FreqMaker.rb -i <input_file> -w <width> -o <output_file>	
input_file	入力 logS ファイル名 (Solubility.rb の output_file を想定)
width	ヒストグラム幅
output_file	ヒストグラムファイル名

ヒストグラムファイル：

出力するヒストグラムファイルは以下の様な書式になります。

<logS> <frequency> (例： -1.50 24)	
logS	logS 値 (<width>刻み)
frequency	頻度

21.4. wln

重み付き学習を行います。入力した各分子について強化学習データベースを用意します。

$\${FC}$ -o wln wln.f でコンパイルしてください (Fortran コンパイラ FC = ifort, g77 ...)。

起動コマンド：

起動コマンドは以下の通りです。

wln -t <input_file> -d <db_file> -b <bin> -c <decoy> -n <dsc_num>	
input_file	入力分子記述子ファイル名
db_file	データベース記述子ファイル名
bin	類似度分布の分散の bin 数
decoy	最大強化学習回数 (= decoy 数)
dsc_num	記述子列数 (拡張 Joback 記述子サイズ = 60 記述子, id 及び logS を除く)

OUTPUT	a) 入力分子記述子ファイル (<input_file>-<id>.dsc, <id>は入力分子の行番号) b) 強化学習データベース (<db_file>-<id>.dsc, <id>は入力分子の行番号) c) 類似度分布ファイル (*.dst) d) 強化学習分子リスト (*.lst)
--------	--

21.5. pls

PLS 回帰 (Partial Least Squares Regression) 分析を行います。\${FC} -o pls pls.f でコンパイルしてください (Fortran コンパイラ FC = ifort, g77 ...)。

起動コマンド:

起動コマンドは以下の通りです (5-1 と 5-2 または 5-3 のみ)。

(5-1) 回帰計算

<code>pls -d <db_file> -n <dsc_num></code>	
db_file	データベース記述子ファイル名
dsc_num	記述子列数 (拡張 Joback 記述子サイズ = 60 記述子, id 及び logS を除く)
OUTPUT	a) データベース回帰結果ファイル (*.rgf) b) 回帰係数ファイル (*.cff)

(5-2) 入力評価

<code>pls -t <input_file> -c <coeff_file> -n <dsc_num></code>	
input_file	入力分子記述子ファイル名
coeff_file	回帰係数ファイル名
dsc_num	記述子列数 (拡張 Joback 記述子サイズ = 60 記述子, id 及び logS を除く)
OUTPUT	a) 評価結果ファイル (*.cal)

(5-3) 回帰計算および入力評価

<code>pls -t <input_file> -d <db_file> -n <dsc_num></code>	
input_file	入力分子記述子ファイル名
db_file	データベース記述子ファイル名
dsc_num	記述子列数 (拡張 Joback 記述子サイズ = 60 記述子, id 及び logS を除く)
OUTPUT	a) データベース回帰結果ファイル (*.rgf) b) 評価結果ファイル (*.cal) c) 回帰係数ファイル (*.cff)

21.6. mlr

重回帰 (multiple liner regression) 分析を行います。\${FC} -o mlr mlr.f でコンパイルしてください (Fortran コンパイラ FC = ifort, g77 ...)。

起動コマンド：

起動コマンドは以下の通りです (6-1 と 6-2 または 6-3 のみ)。

(6-1) 回帰計算

<code>mlr -d <db_file> -n <dsc_num></code>	
db_file	データベース記述子ファイル名
dsc_num	記述子列数 (拡張 Joback 記述子サイズ = 60 記述子, id 及び logS を除く)
OUTPUT	a) データベース回帰結果ファイル (*.rgr) b) 回帰係数ファイル (*.cff)

(6-2) 入力評価

<code>mlr -t <input_file> -c <coeff_file> -n <dsc_num></code>	
input_file	入力分子記述子ファイル名
coeff_file	回帰係数ファイル名
dsc_num	記述子列数 (拡張 Joback 記述子サイズ = 60 記述子, id 及び logS を除く)
OUTPUT	a) 評価結果ファイル (*.cal)

(6-3) 回帰計算および入力評価

<code>mlr -t <input_file> -d <db_file> -n <dsc_num></code>	
input_file	入力分子記述子ファイル名
db_file	データベース記述子ファイル名
dsc_num	記述子列数 (拡張 Joback 記述子サイズ = 60 記述子, id 及び logS を除く)
OUTPUT	a) データベース回帰結果ファイル (*.rgr) b) 評価結果ファイル (*.cal) c) 回帰係数ファイル (*.cff)

21. 7. bys

ベイジアン解析を行います。ベイズ推定を行った後、シグモイド関数でフィッティングします。\${FC} -o bys bys.f でコンパイルしてください (Fortran コンパイラ FC = ifort, g77 ...)。フィッティングするシグモイド関数の定義は以下の通りです (c1=a, c2=b)。

$$P_{agg}(\log S) = \frac{1}{1 + e^{a(\log S - b)}}$$

起動コマンド：

起動コマンドは以下の通りです (7-1 と 7-2 または 7-3 のみ)。

(7-1) ベイジアン解析

<code>bys -d <db_file> -i <c1_value> <c2_value></code>
--

db_file	頻度分布ファイル名 (データベース)
c1_value	シグモイド関数の係数 C1 初期値
c2_value	シグモイド関数の係数 C2 初期値
OUTPUT	a) データベース解析結果ファイル (*.prb) b) シグモイド係数ファイル (*.cff)

(7-2) 入力評価

<code>bys -t <input_file> -n <dsc_num> <dsc_idx> -c <coeff_file></code>	
input_file	入力ファイル名 (回帰結果)
dsc_num	列サイズ (回帰結果列数 = 10 列)
dsc_idx	列番号 (logS の列番号 = 9、列番号 1~10)
coeff_file	シグモイド係数ファイル名
OUTPUT	a) 評価結果ファイル (*.est)

(7-3) ベイジアン解析および入力評価

<code>bys -t <input_file> -n <dsc_num> <dsc_idx> -d <db_file> -i <c1_value> <c2_value></code>	
input_file	入力ファイル名 (回帰結果)
dsc_num	列サイズ (回帰結果列数 = 10 列)
dsc_idx	列番号 (logS の列番号 = 9、列番号 1~10)
db_file	頻度分布ファイル名 (データベース)
c1_value	シグモイド関数の係数 C1 初期値
c2_value	シグモイド関数の係数 C2 初期値
OUTPUT	a) 評価結果ファイル (*.est) b) データベース解析結果ファイル (*.prb) c) シグモイド係数ファイル (*.cff)

頻度分布ファイル (データベース) :

入力する頻度分布ファイル (データベース) は以下の様な書式になります。

<code><marker> <state_frq > <non_state_frq></code> (例: -1.50 24 59)	
marker	logS 値
state_frq	アグリゲータ度数
non_state_frq	非アグリゲータ度数

21.8. trans_code

Ullmann の定理を利用して入力分子に含まれる官能基数をカウントします。官能基の定義は薬物候補低分子用の拡張 Joback 記述子です (定義ファイルは DB/に配置)。コンパイルは、ソースディレクトリ (src/) に移動した後、Makefile を適宜修正して make して下さい。

起動コマンド：

起動コマンドは以下の通りです。

(8-1) 分子構造から記述子に変換 (mol2 形式 DB)

<code>trans_code -i <input_file> -o <output_file> -d <db_file></code>	
input_file	分子構造ファイル名 (mol2 ファイル、multi-mol2 形式も可)
output_file	記述子ファイル名
db_file	官能基定義ファイル名 (multi-mol2 形式、DB 用)

(8-2) 分子構造から記述子に変換 (binary 形式 DB)

<code>trans_code -i <input_file> -o <output_file> -d <db_file> -b</code>	
input_file	分子構造ファイル名 (mol2 ファイル、multi-mol2 形式も可)
output_file	記述子ファイル名
db_file	官能基定義ファイル名 (binary 形式、DB 用)
“-b”	binary 利用スイッチ

(8-3) 分子構造から記述子に変換 (binary 形式 DB)

<code>trans_code -i <input_file> -o <output_file> -c</code>	
input_file	官能基定義ファイル (multi-mol2 形式、DB 用)
output_file	官能基定義ファイル名 (binary 形式、DB 用)
“-c”	binary 生成スイッチ

物理記述子 (Mol2 ファイル)：

mol2 ファイルの@<TRIPOS>COMMENT に以下の記述があれば、trans_code より出力される記述子ファイルに以下の物理記述子の項目が含まれるようになります。

記述は以下の様に行います。

<code>#<identifier> <value></code> (例：#LogS -1.590)	
identifier	物理記述子 (LogS, De10_H, AddN_H, …)
value	設定値

記述可能な物理記述子は以下の通りです。

識別子	内容
LogS	logS 値
De10_H	解離 H 原子数 (-OH)
AddN_H	付加 H 原子数 (-NH)
dG_o	オクタノール中の GB/SA 値
ddG_o	オクタノール中の GB/SA 値 (単位表面積当たり)
dASA_o	オクタノール中の 表面積 (単位体積当たり)
dG_w	水中の GB/SA 値
ddG_w	水中の GB/SA 値 (単位表面積当たり)

dASA_w	水中の 表面積 (単位体積当たり)
dG_wd	水中解離型の GB/SA 値
ddG_wd	水中解離型の GB/SA 値 (単位表面積当たり)
dASA_wd	水中解離型の 表面積 (単位体積当たり)

22. tp|2capbc

全系のトポロジーファイルと、`setwater` で生成した水分子ファイルを読み込み、孤立系での CAP 拘束指定ファイルと位置拘束ファイルを作成します。

入力データ

- (1) 全系のトポロジーファイル名 (1 個目)
- (2) `setwater` で生成した水分子ファイル名 (2 個目)
- (3) CAP 拘束指定ファイル名 (3 個目)
- (4) 位置拘束ファイル名 (4 個目)

分子名と、その個数を全系のトポロジーファイルから取得し、水の半径と中心を `setwater` で生成した水分子ファイルから読み込み、CAP 拘束指定ファイルと位置拘束ファイルを作成します。

CAP 拘束指定ファイルには、全分子が含まれますが、位置拘束ファイルは蛋白質に対してのみ作成します。

■ 使用例

```
% tpl2capbc
  input TPL file name
Pro_1.tpl (1)
  input WAT file name
wat.pdb (2)
  input CAP file name
capbc_file (3)
BOUND> INCLUDE
pro          1          1 YES
Lig         1          1 YES
tip3p_water 1        12801 YES
Cl          1          35 YES
Na          1          38 YES

BOUND> CENTER
COORDINATEs 13.1886    58.0962    52.1595
BOUND> RADIUS
46.9530
  input POS file name
posres_file (4)
GROUP> LIST
  1    0    1  2000 CA * 1.0 MASS YES
  1    0    1  2000 N  * 1.0 MASS YES
  1    0    1  2000 C  * 1.0 MASS YES
  1    0    1  2000 O  * 1.0 MASS YES

END
GROUP> STOP
```

(余白)

myPresto 5.0