

myPresto 4.2

- LigandBox -

USER MANUAL

Version 1.0

Copyright (C) 2006-2010 National Institute of Advanced Industrial Science and Technology (AIST)

Copyright (C) 2006-2010 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 4.0* 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 4.0* 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	1
2 . LigandBox システムインストール.....	2
2-1. インストールに必要な環境	2
2-2. ディレクトリ構成	2
2-3. C_PROG 配下のツールについて	3
2-3. プログラムのコンパイル	4
2-3-1.perl コマンドパスの確認	4
2-3-2.perl スクリプトの実行権の付与	4
2-3-3.tools 配下のツールのコンパイル	4
3 . 各種ツールの使用法.....	6
3-1. データ作成ツール(data_convert)の使用法	6
3-1-1.regssystem.conf 設定ファイルの設定	6
3-1-2.データの作成(data_convert の使用)	9
3-2. データ登録ツール(data_regist)の使用法	10
3-2-1.regssystem.conf 設定ファイルの設定	10
3-2-2.dbsysytem.conf 設定ファイルの設定	10
3-2-3.source_list 設定ファイルの設定	12
3-2-4.データの登録(data_regist の使用)	12
3-3. 同一分子グループ化ツール(data_link)の使用法	14
3-4. 異性体登録ツール(crt_structural isomer)の使用法	14
3-5.DB 検索ツール(data_search)の使用法	15
3-5-1.search_list ファイルの設定	15
3-5-2.DB 検索(data_search の使用)	19
4 . サブプログラムの使用法.....	20
4-1.SplitMolFile	20
4-2.GetMolCharge	22
4-3.CheckAtomCoord	23
4-4.GetHomoLumo	24
4-5.GetAtomicCharge	25
4-6.GetMolProperty	26
4-7.CreateBondMatrix	27

1. 概要

本システムは、低分子化合物の2次元構造を3次元化し、座標、分子物性情報をDBに登録し、又、DBデータから検索条件に従い、MOL2 ファイルを出力するシステムである。

本システムは、大きく分けて以下の5種類のツールで構成している。

MDL sdf データ(2次元)から Sybyl MOL2 データ(3次元)を作成(ツール data_convert)

Sybyl MOL2 データをDBに登録(ツール data_regist)

同一分子のグループ化(ツール data_link)

異性体の登録(ツール crt_structural_isomer)

データの検索(MOL2 ファイル出力)(ツール data_search)

は2次元で与えられた分子構造を3次元に立ち上げるとともに、各種分子物性値の計算を行う。計算結果は Sybyl MOL2 形式で出力する。

は で作成した Sybyl MOL2 データを postgresSQL データベースに登録する。

は登録されたデータ間で同一分子がある場合には、それらの分子のグループ化処理を行う。

は で登録された座標データをもとに、光学異性体、配座異性体の作成を行い、作成データをデータベースに登録する。

は各種検索条件に基づいてデータベース検索を行い、条件に一致した分子構造を Sybyl MOL2 形式で出力する。

【参考文献】

Yoshifumi Fukunishi, Yuusuke Sugihara, Yoshiaki Mikami, Kohta Sakai, Hiroshi Kusudo, Haruki Nakamura, "Advanced in-silico drug screening to achieve high hit ratio-development of 3D-compound database", *Synthesiology*, 2009, 2, 60-68.

Fukunishi, Yoshifumi; Lintuluoto, Masami, Development of Chemical Compound Libraries for In Silico Drug Screening, *Current Computer - Aided Drug Design*, Volume 6, Number 2, June 2010, pp. 90-102

2 . LigandBox システムインストール

2-1. インストールに必要な環境

本システムを使用するには以下の環境が必要となります。

- UNIX(LINUX)環境 : 本システムの実行環境です。
- PostgreSQL(ver.8.1以降) : DB 登録、検索に必要な環境。
- Perl(ver.5.8以降) : 本システムの実行に必要な環境。
- Perl DBI パッケージ(ver.1.46以降) : Perl で DB アクセスするのに必要となります。
- Perl DBD::Pg パッケージ(ver.1.22以降) : Perl で DB アクセスするのに必要となります。
- C コンパイラ : perl スクリプト内で呼び出している各種ツールのコンパイルに使用します。
- myPresto tplgeneL : 3次元座標データ作成時に使用します。
- myPresto cosgene : 3次元座標データ作成時に使用します。
- mopac(ver.7) : 分子物性を計算する際に使用します。

【注意事項】

- PostgreSQL について、事前にデータベースを作成し、サービスが起動している事を前提に説明を行います。
- myPresto の各ツールはインストールされている事を前提に説明を行います。
myPresto ツールのインストールは myPresto マニュアルを参考の事。
- mopac については、インストールされている事を前提に説明を行います。

2-2. ディレクトリ構成

MOLDB 以下のディレクトリをインストール先のディレクトリにコピーします。

MOLDB のディレクトリ構成は以下の様になっています。

- 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)
- cgi-bin : web 表示用ディレクトリ
- conf : DB 用 SQL 関連ファイル用ディレクトリ
- etc : 各種設定ファイル用ディレクトリ
- log : ログ出力ファイル用ディレクトリ
- package : perl モジュール用ディレクトリ

- query : web 検索用クエリー保存ディレクトリ
- sbin : web 検索用バッチファイル保存ディレクトリ
- tmp : テンポラリ、エラーファイル出力用ディレクトリ
- C_PROG : 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

2-3.C_PROG 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(バイナリ形式、及び perl(.pl)形式)を呼び出します。これらのツールは tools 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様になっています。

- CheckAtomCoord : 分子構造チェックプログラム
エネルギー最小化を行った分子について結合角、二面角のチェックを行う。
- CreateBondMatrix : 結合行列固有値計算プログラム
同一分子の判定を行うのに必要な結合行列の固有値の計算を行う。
- GBSAinp : GB/SA 計算ファイル作成ツールディレクトリ
cosgene で GB/SA 計算を行う際に必要となる設定ファイルを作成する。
- GetAtomicCharge : 原子電荷出力ツールディレクトリ
MOPAC 出力から原子電荷の切り出しを行う。
- GetHomoLumo : HOMO/LUMO 値出力ツールディレクトリ
MOPAC 出力から HOMO/LUMO 値の切り出しを行う。
- GetMolCharge : 分子総電荷取得プログラム
- GetMolProperty : 分子物性計算プログラム
- Hgene : 分子座標フォーマット変換、水素付加ツールディレクトリ
分子座標フォーマットの変換(sdf -> MOL2、sdf -> mopac)を行う。
- ModifyAtomCoord : 初期座標の修正プログラム
cosgene 処理を行う前の座標の修正を行う。
- SplitMolFile : 分子分割プログラム
分子の分割、イオンのチェックを行う。
- 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。
- mop7w : MOPAC プログラム
電荷計算、HOMO、LUMO 値の計算を行う。
- 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. プログラムのコンパイル

2-3-1.perl コマンドパスの確認

bin ディレクトリ配下の perl スクリプト (crt_structuralisomer、data_convert、data_link、data_regist、data_search、) は perl 言語で記述されています。各ツールの 1 行目の perl コマンドパスを確認し、計算機環境に合わせて変更して下さい。

例：perl が "/usr/local/bin/perl" にインストールされている場合

```
#!/usr/local/bin/perl
```

2-3-2.perl スクリプトの実行権の付与

上記 perl スクリプトに実行権があるか否かを確認し、実行権がない場合には、実行権を付与して下さい。

例：ファイル AAA に実行権を付与する場合

```
# chmod 755 AAA
```

2-3-3.tools 配下のツールのコンパイル

perl スクリプト内で呼び出される各種ツール群は tools ディレクトリ配下に置いています。各 perl スクリプトを実行するには、これら tools 配下の各ツールがコンパイルされている事が必要となります。

- ・ GBSAin.pl

perl 言語で記述されています。2-3-1、2-3-2 に従い、コマンドパスの修正、実行権の付与を行って下さい。

- ・ CreateBondMatrix

- ・ GetAtomicCharge

- ・ GetHomoLumo

- ・ Hgene

- SplitMolFile
- ModifyAtomCoord
- CheckAtomCoord
- GetMolCharge
- GetMolProperty

これらのツールはC言語で記述されています。各ツールディレクトリ内の Makefile をエディタで開き、CC コンパイラのパスを計算機環境に合わせて変更して下さい。

変更後に make コマンドで実行ファイルを作成します。

例：CC コマンドパスが"/usr/local/bin"の場合

CC=/usr/local/bin/cc

3 . 各種ツールの使用法

MDL sdf データから DB 登録までの処理手順は以下の様になっています。これらの処理を完了(、 の処理は必須、 、 はオプション)後、 のデータ検索を行う事ができます。

MDL sdf データ(2次元)から SybylMOL2 データ(3次元)の作成(ツール data_convert)
Sybyl MOL2 データを DB に登録(ツール data_regist)
同一分子のグループ化(ツール data_link)
異性体の登録(ツール crt_structuralisomer)
データの検索(MOL2 ファイル出力)(ツール data_search)

3-1. データ作成ツール(data_convert)の使用法

data_convert ツールは、データ作成時に regsystem.conf ファイルの設定に従い処理を行います。計算機環境に従い、regsystem.conf ファイル内の各項目を変更する必要があります。

3-1-1. regsystem.conf 設定ファイルの設定

regsystem.conf ファイルでは各種コマンドのパス等の設定を行います。regsystem.conf ファイルは etc ディレクトリ配下にあります。

regsystem.conf 内の設定項目は以下の様になっています。

本ファイルは一度正しく設定したならば、それ以降、計算機環境に変更がない限り再設定の必要はありません。

設定項目一覧

• reglog_dir	: 登録ログファイルディレクトリ
• tmp_dir	: 登録用一時ディレクトリ
• sql_dir	: SQL の格納ディレクトリ
• gzip_path	: gzip コマンドパス
• gunzip_path	: gunzip コマンドパス
• zcat_path	: zcat コマンドパス
• tplgene_path	: tplgeneL コマンドパス
• tpldb_path	: tplgeneL DB ファイルパス

- cosgene_path : cosgene コマンドパス
- mopac_path : MOPAC コマンドパス
- molgene_path : molgene コマンドパス
- hgene_path : Hgene コマンドパス
- homolumo_path : getHomoLumo コマンドパス
- atomiccharge_path : getAtomicCharge コマンドパス
- mkgbsain_path : mkGBSAin.pl コマンドパス
- mkgbsaindb_path : mkGBSAin.pl 用パラメータパス
- reg_num : 一括登録の件数
- debug : デバッグフラグ

```
#-----  
# File : regsystem.conf  
# 機能 : 登録機能環境定義ファイル  
#-----  
# 登録ログファイルのディレクトリ  
#-----  
reglog_dir=./log  
#-----  
# 登録用一時ディレクトリ  
#-----  
tmp_dir=/home/molddb/MOLDB/tmp/  
#-----  
# SQL の格納ディレクトリ  
#-----  
sql_dir=./conf/sql  
#-----  
# gzip コマンドのパス  
#-----  
gzip_path=/bin/  
#-----  
# gunzip コマンドのパス  
#-----  
gunzip_path=/bin/  
#-----  
# zcat コマンドのパス  
#-----  
zcat_path=/bin/  
#-----  
# tplgeneL コマンドのパス  
# 力場パラメータのパス  
#-----  
tplgene_path=/home/molddb/bin/  
tpldb_path=/home/molddb/tplgeneL/DB/  
#-----  
# cosgene(Ver2.1) コマンドのパス  
#-----  
cosgene_path=/home/molddb/bin/
```

図 3 - 1 . regsystem.conf 設定例

```
#-----  
# mopac コマンドのパス  
#-----  
mopac_path=/home/moldb/bin/  
#-----  
# molgene コマンドのパス  
#-----  
molgene_path=/home/moldb/MOLDB/tools/molgene/  
#-----  
# Hgene コマンドのパス  
#-----  
hgene_path=/home/moldb/MOLDB/tools/Hgene/  
#-----  
# getHomoLumo コマンドのパス  
#-----  
homo_lumo_path=/home/moldb/MOLDB/tools/GetHomoLumo/  
#-----  
# getAtomicCharge コマンドのパス  
#-----  
atomiccharge_path=/home/moldb/MOLDB/tools/GetAtomicCharge/  
#-----  
# mkGBSAin.pl コマンドのパス  
# 参照パラメタのパス  
#-----  
mkgsain_path=/home/moldb/MOLDB/tools/GBSAinp/  
mkgsaindb_path=/home/moldb/MOLDB/tools/GBSAinp/  
#-----  
# 登録の一括件数  
#-----  
reg_num=10000  
#-----  
# デバッグフラグ  
#-----  
debug=on
```

図 3 - 1 . regsystem.conf 設定例 (続き)

3-1-2. データの作成 (data_convert の使用)

MDL sdf (2次元) から Sybyl MOL2 (3次元) を作成するには、MOLDB/bin 配下で以下のコマンドを入力します。

本ツールでは、1ファイル中に複数エントリの MOL2 データを持つ MOL2 ファイルを出力します。また、作成 MOL2 ファイルは圧縮形式で保存されます。

【コマンド】

data_convert

【オプション】

--input-sdffile(省略時: -i) SDDFILE 変換したいデータカタログの
sdf 形式のファイル(必須)
--dump-directory(省略時: -d) PREFIX 変換ファイルの出力先
ディレクトリ(必須)
--help(省略時: -h) コマンドヘルプを表示

例: sdf ファイル "/home/molddb/MOLDB/data/test.sdf.gz" を入力し、Sybyl MOL2 データ出力ディレクトリを "/home/molddb/MOLDB/tmp" に指定する場合

```
# cd bin
# ./data_convert -i /home/molddb/MOLDB/data/test.sdf.gz -d /home/molddb/MOLDB/tmp
```

入力する MDL sdf ファイルは圧縮されたもの (XXX.sdf.gz)、及び、圧縮されていないもの (XXX.sdf) のいずれの形式にも対応しています。

データ作成処理時のログは 3-1-1 で指定した reglog_dir ディレクトリ配下に出力されます。

データ作成エラー分子は 3-1-1 で指定した tmp_dir/error 配下に MDL sdf (2次元) ファイル形式でコピーされます。

3-2. データ登録ツール(data_regist)の使用法

data_regist ツールは、データ作成時に regsystem.conf、dbsystem.conf、及び、source_list ファイルの内容に従い処理を行います。計算機環境、データベース環境、DB登録対象により、各項目を変更する必要があります。

3-2-1. regsystem.conf 設定ファイルの設定

3-1-1 で正しく設定されていれば、本項で設定する必要はありません。ツール起動時にエラーになる場合には 3-1-1 を参照して再設定を行って下さい。

3-2-2. dbsystem.conf 設定ファイルの設定

dbsystem.conf ファイルでは、DB に関する項目について設定を行います。dbsystem.conf ファイルは etc ディレクトリ配下にあります。設定項目は以下の様になっています。

設定項目一覧

- db_type : データベースのタイプ(本システムでは固定 : postgres)
- databasehome : PostgreSQL インストールディレクトリパス
- db_name : 登録するデータベース名
- user : postgres ユーザ名
- pass : postgres ユーザパスワード
- port : ポート番号
- pghost : データベースホスト
- pgdata : データディレクトリ

db_name を変更する事により、複数のデータベースを作成する事ができます。また、異なるデータについて同じ db_name 名で登録すると、以前に登録していたデータベース情報が上書きされることがありますので、細心の注意を払って行って下さい。

```
#-----  
# File : dbssystem.conf  
# 機能 : データベース環境定義ファイル  
#-----  
  
#-----  
# データベースのタイプ (postgres 固定)  
#-----  
db_type=postgres  
  
#-----  
# コマンド ($home/bin/psql が)存在するように設定)  
#-----  
databasehome=/home/mol/db/postgresql8.1/  
  
#-----  
# 使用データベース環境  
# db_name : データベース名  
# user   : データベースユーザ名  
# pass   : パスワード  
# port   : ポート番号  
# pghost : データベースホスト  
# pgdata : データディレクトリ  
#-----  
db_name=testdb  
user=db3dmgr  
pass=db3dmgr  
port=5432  
pghost=localhost  
pgdata=/data/postgres/DATA
```

図 3 - 2 . dbssystem.conf 設定例

3-2-3. source_list 設定ファイルの設定

source_list ファイルでは、DB に登録する Sybyl MOL2 ファイルパスの設定を行います。source_list ファイルは etc ディレクトリ配下にあります。

本ファイルでは、登録したい MOL2 ファイルをパス付きで指定します。複数件の MOL2 ファイルを登録する場合には、その数分のファイルパスを指定します(5 つの MOL2 ファイルの登録を行う際には 5 行分指定を行います)。

source_list ファイル中には空白行を入れない様にして下さい。

```
/home/moldb/MOLDB/data/test1.mol2.gz
/home/moldb/MOLDB/data/test2.mol2.gz
/home/moldb/MOLDB/data/test3.mol2.gz
/home/moldb/MOLDB/data/test4.mol2.gz
/home/moldb/MOLDB/data/test5.mol2.gz
```

図 3 - 3 . source_list ファイル設定例

3-2-4. データの登録(data_regist の使用)

3-1-2 で作成した Sybyl MOL2(3 次元)を DB 登録する際には、MOLDB/bin 配下で以下のコマンドを入力します。DB 登録するには登録対象分子を source_list ファイルに記述する必要があります(3-2-3 参照)。

本ツールでは、1 ファイル中に複数エントリの MOL2 データを持つ MOL2 ファイルを出力します。また、作成 MOL2 ファイルは圧縮形式で保存されます。

【コマンド】

data_regist

【オプション】

--input-listfile(省略時:-i) FILE 登録したい MOL2 ファイル情報が記述されたリストファイル (必須)
--tablecreate(省略時:-t) テーブルを作成する場合のオプション (初期の登録 / 登録をやり直したい場合に指定する)
--help(省略時:-h) コマンドヘルプを表示

例：/home/molddb/MOLDB/etc/source_list を指定して DB 登録を行う場合

```
# cd bin
# ./data_regist -t -i /home/molddb/MOLDB/etc/source_list
```

データ作成処理時のログは 3-1-1 で指定した reglog_dir ディレクトリ配下に出力されます。

3-3. 同一分子グループ化ツール(data_link)の使用法

3-2-4で登録したDBデータについて、同一分子のグループ化処理を行う際には、MOLDB/bin配下で以下のコマンドを入力します。

【コマンド】

data_link

【オプション】

- resetup(省略時:-r) 一度作成したグループを無効化して再度グループを作成しなおします。
- alias(省略時:-a) Namiki NOTE 情報を使用してリンクを行う。
- chemical_formula(省略時:-c) 分子式等でリンクを行う。
- help(省略時:-h) コマンドヘルプを表示

例：登録したDBデータに対して、NOTE情報を用いてグループ化を行う場合

```
# cd bin
# ./data_link -r -a
```

3-4. 異性体登録ツール(crt_structuralisomer)の使用法

3-2-4で登録したDBデータをもとに、異性体座標を作成し、DBに登録する際には、MOLDB/binで以下のコマンドを入力します。本ツールで作成、登録される異性体は、光学異性体、及び、配座異性体です。

【コマンド】

crt_structuralisomer

【オプション】

- resetup(省略時:-r) 一度作成した異性体を無効化して再度作成しなおします。
- help(省略時:-h) コマンドヘルプを表示

例：異性体の登録を行う場合

```
# cd bin
```

```
# ./crt_structuralisomer
```

3-5.DB 検索ツール(data_search)の使用法

登録された DB データを検索する際には、DB 検索ツール(data_search)を使用します。DB 検索は検索条件設定ファイル(search_list)ファイルの条件をもとに行います。

3-5-1.search_list ファイルの設定

search_list ファイルには検索条件を記述します。search_list ファイルは etc ディレクトリ配下に保存しています。本ファイル内指定できる検索条件は以下の様になっています。

以下の検索条件は複数の条件を指定して検索する事もできます。

検索条件

原子数(SATOM_NUM、EATOM_NUM)

SATOM_NUM 以上、EATOM_NUM 以下の範囲を検索する。コメントアウト(行頭が#)は検索対象としない、他の検索条件も同様)

分子量

SMOL_WEIGHT 以上、EMOL_WEIGHT 以下の範囲を検索する。

分子総電荷

SMOL_CHARGE 以上、EMOL_CHARGE 以下の範囲を検索する。

ドナー原子数

SDONOR_NUM 以上、EDONOR_NUM 以下の範囲を検索する。

アクセプタ原子数

SACCEPTOR_NUM 以上、EACCEPTOR_NUM 以下の範囲を検索する。

HOMO 値

SHOMO 以上、EHOMO 以下の範囲を検索する。

LUMO 値

SLUMO 以上、ELUMO 以下の範囲を検索する。

G 値

SDELTA_G 以上、EDELTA_G 以下の範囲を検索する。

G/原子数

SATOMIC_DELTA_G 以上、EATOMIC_DELTA_G 以下の範囲を検索する。

分子式

CHEMICAL_FOMULA で指定された文字列の検索を行う。

先頭一致の場合は^

中間を LIKE 検索する場合は*

末尾一致の場合は\$

記号を使用する事ができる。

溶解度

SSOLUBILITY 以上、ESOLUBILITY 以下の範囲を検索する。

logP 値

SLOGP 以上、ELOGP 以下の範囲を検索する。

その他の指定条件

・1 ディレクトリ内に出力する MOL2 ファイル数(DATA_LIMIT)

例：原子数が 30 以上、50 以下の分子の検索を行う場合の指定

SATOM_NUM:30

EATOM_NUM:50

例：原子数が 20 以上、40 以下で、分子量が 100 以上、200 以下の検索を行う場合の指定

SATOM_NUM:20

EATOM_NUM:40

SMOL_WEIGHT:100

EMOL_WEIGHT:200

例：分子量が 300 以上、500 以下で、分子総電荷が 0 の検索を行う場合の指定

SMOL_WEIGHT:300

EMOL_WEIGHT:500

SMOL_CHARGE:0

EMOL_CHARGE:0

```
#=====
# search_list:検索用設定ファイル
#=====

#-----
#DATA_LIMIT:対象ディレクトリのトリミング数を指定
#           (指定しない場合は 100 が自動設定)
#-----
DATA_LIMIT:1000

#-----
#SATOM_NUM : 検索対象範囲の原子数。S*が以上,E*が以下
#EATOM_NUM  (=の場合は両方向じ数値を設定)
#           (指定しない場合は検索対象としない)
#-----
SATOM_NUM:0
EATOM_NUM:300

#-----
#SMOL_WEIGHT : 検索対象範囲の分子量。S*が以上,E*が以下
#EMOL_WEIGHT (=の場合は両方向じ数値を設定)
#           (指定しない場合は検索対象としない)
#-----
#SMOL_WEIGHT:1
#EMOL_WEIGHT:9999.999

#-----
#SMOL_CHARGE : 検索対象範囲の分子総電荷。S*が以上,E*が以下
#EMOL_CHARGE (=の場合は両方向じ数値を設定)
#           (指定しない場合は検索対象としない)
#-----
#SMOL_CHARGE:-1
#EMOL_CHARGE:1

#-----
#SDONOR_NUM : 検索対象範囲のドナー原子数。S*が以上,E*が以下
#EDONOR_NUM  (=の場合は両方向じ数値を設定)
#           (指定しない場合は検索対象としない)
#-----
#SDONOR_NUM:0
#EDONOR_NUM:999

#-----
#SACCEPTOR_NUM : 検索対象範囲のアクセプタ原子数。S*が以上,E*が以下
#EACCEPTOR_NUM (=の場合は両方向じ数値を設定)
#           (指定しない場合は検索対象としない)
#-----
#SACCEPTOR_NUM:0
#EACCEPTOR_NUM:99

#-----
#SHOMO : 検索対象範囲の最高被占分子軌道。S*が以上,E*が以下
#EHOMO  (=の場合は両方向じ数値を設定)
#           (指定しない場合は検索対象としない)
#-----
#SHOMO:-999.999
#EHOMO:999.999
```

図 3 - 3 . search_list ファイル設定例

```

#-----
#SLUMO : 検索対象範囲の最低空分子軌道。S*が以上,E*が以下
#ELUMO  (=の場合は両方向じ数値を設定)
#        (指定しない場合は検索対象としない)
#-----
#SLUMO:-999.999
#ELUMO:999.999

#-----
#SDELTAG : 検索対象範囲の G 値。S*が以上,E*が以下
#EDELTAG (=の場合は両方向じ数値を設定)
#        (指定しない場合は検索対象としない)
#-----
#SDELTAG:-99.9999
#EDELTAG:99.9999

#-----
#SATOMIC_DELTAG : 検索対象範囲の G/原子数値。S*が以上,E*が以下
#EATOMIC_DELTAG (=の場合は両方向じ数値を設定)
#        (指定しない場合は検索対象としない)
#-----
#SATOMIC_DELTAG:-99.9999
#EATOMIC_DELTAG:99.9999

#-----
#CHEMICAL_FOMULA : 検索対象の分子式
#                  以下の正規表現で検索文を作成する。
#                  ^:先頭一致
#                  *:中間を LIKE 検索
#                  $:末尾一致
#                  (^,$を使わない場合は先頭/末尾は LIKE 検索になる。)
#                  (指定しない場合は検索対象としない)
#-----
#CHEMICAL_FOMULA:^H11C8N5S2$

#-----
#SOLUBILITY : 検索対象の溶解度。S*が以上,E*が以下
#            (=の場合は両方向じ数値を設定)
#            (指定しない場合は検索対象としない)
#-----
#SSOLUBILITY:-10
#ESOLUBILITY:10

#-----
#LOGP : 検索対象の logP 値。S*が以上,E*が以下
#      (=の場合は両方向じ数値を設定)
#      (指定しない場合は検索対象としない)
#-----
#SLOGP:0.4
#ELOGP:0.8

```

図 3 - 3 . search_list ファイル設定例(続き)

3-5-2.DB 検索(data_search の使用)

3-2-4 で登録した DB データをもとに、DB 検索する際には、MOLDB/bin で以下のコマンドを入力します。

検索でヒットした構造は、指定したディレクトリ配下に保存します。出力する Sybyl MOL2 ファイルは 1 件毎に 1 ファイルで出力し、3-5-1 で指定した件数毎に保存用ディレクトリを作成します。

また、同ディレクトリ内に検索ログ、及び、ヒットした構造と同一分子のグループリストを出力します。

【コマンド】

data_search

【オプション】

--input-file(省略時:-i) FILE 検索条件が記述されたファイル (必須)
--dump-directory(省略時:-d) PREFIX 検索結果の出力先ディレクトリ(必須)
--help(省略時:-h) コマンドヘルプを表示

例: "/home/molddb/MOLDB/etc/search_list"ファイル内の条件で DB 検索を行い、出力を "/home/molddb/MOLDB/OUT"ディレクトリ配下に保存する場合

```
# cd bin
```

```
# ./data_search -i /home/molddb/MOLDB/etc/search_list -d /home/molddb/MOLDB/OUT
```

4 . サブプログラムの使用法

4-1.SplitMolFile

【概要】本プログラムは1つの mol ファイル中に複数分子が存在する場合に、最大原子数の分子を切り出すプログラムである。

【入出力形式】

入力：mol ファイル

出力：mol ファイル

【使用法】

SplitMolFile <入力ファイル名> <出力ファイル名>

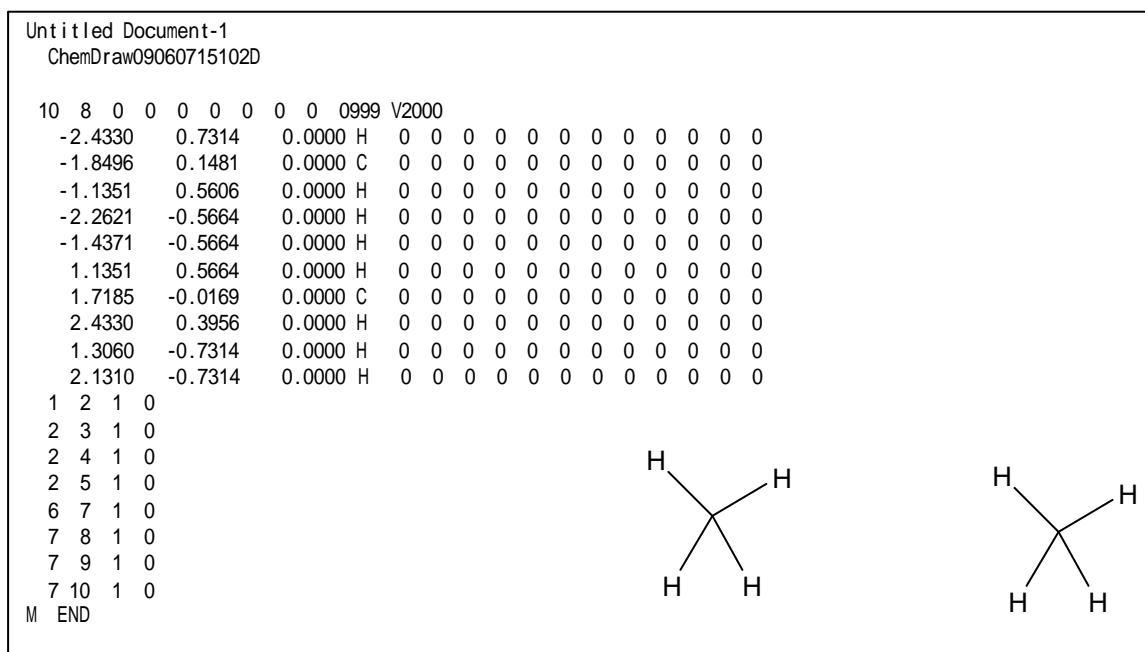
【使用例】

2つのメタン分子を含む mol ファイルから1つの mol ファイルを抽出する。

(#はコマンドプロンプトを示す。以下同様)

```
# SplitMolFile 2methane.mol mehtane.mol
```

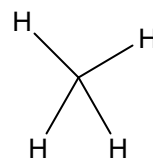
入力ファイル(2methane.mol)



出力ファイル(methane.mol)

```
methane.mol  
molgene
```

```
5 4 0 0 0 0  
-2.4330 0.7314 0.0000 H 0 0 0 0 0 0  
-1.8496 0.1481 0.0000 C 0 0 0 0 0 0  
-1.1351 0.5606 0.0000 H 0 0 0 0 0 0  
-2.2621 -0.5664 0.0000 H 0 0 0 0 0 0  
-1.4371 -0.5664 0.0000 H 0 0 0 0 0 0  
1 2 1 0 0 0  
2 3 1 0 0 0  
2 4 1 0 0 0  
2 5 1 0 0 0
```



4-2. GetMolCharge

【概要】本プログラムは入力した mol ファイル中の分子の分子総電荷を計算する。計算結果を標準出力に出力する。

【入出力形式】

入力：mol ファイル

出力：標準出力

【使用法】

GetMolCharge <入力ファイル名>

【使用例】

酢酸の分子総電荷を計算する。コマンドを実行すると分子電荷値-1を返す。

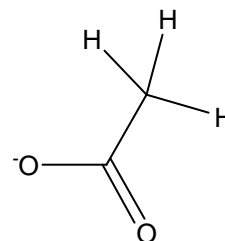
```
# cat acetic_acid

Untitled document-1
ChemDraw09060715212D

 7 6 0 0 0 0 0 0 0 0999 V2000
-0.1922 -0.3984 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0.2203 0.3160 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0.2203 -1.1129 0.0000 O 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-1.0172 -0.3984 0.0000 O 0 5 0 0 0 0 0 0 0 0 0 0 0 0 0 0
-0.3631 0.8994 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 0.4338 1.1129 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 1.0172 0.1025 0.0000 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
 1 2 1 0
 1 3 2 0
 1 4 1 0
 2 5 1 0
 2 6 1 0
 2 7 1 0
M CHG 1 4 -1
M END

# GetMolCharge acetic_acid.mol
Molecular charge = -1

%% Program is done. %%
%% This program ended normally. %%
```



4-3. CheckAtomCoord

【概要】本プログラムは入力した mol2 ファイル中の分子の結合角、二面角が妥当な値になっているかをチェックする。計算結果を標準出力に出力する。

【入出力形式】

入力：mol2 ファイル

出力：標準出力

【使用法】

CheckAtomCoord <入力ファイル名>

【使用例 1】

```
# checkAtomCoord good.mol2
===== good.mol2 =====
All angle and torsion are good.

# CheckAtomCoord bad.mol2
===== bad.mol2 =====
ERROR> toolChecAtomAngle
      7- 8- 11 = 151.5666
```

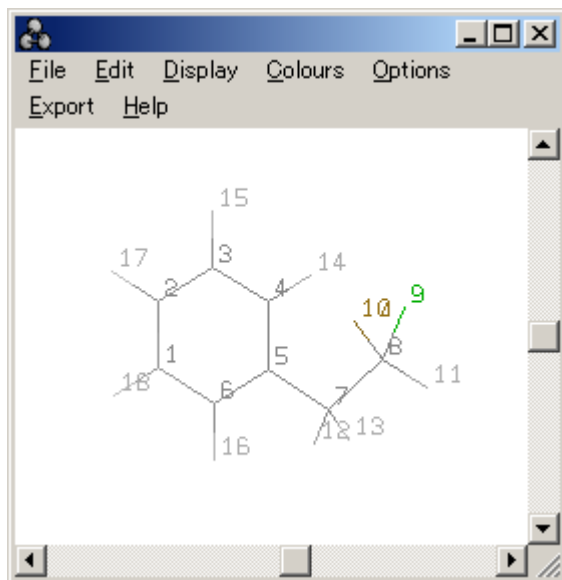


図 .good.mol2

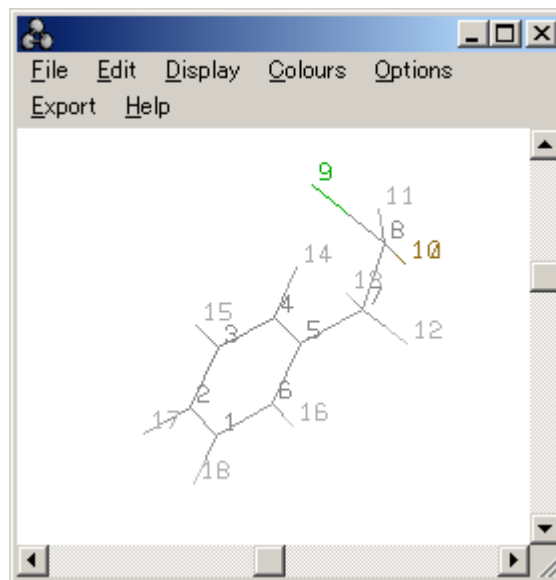


図 .bad.mol2

4-4. GetHomoLumo

【概要】本プログラムは MOPAC 出力ファイルから HOMO/LUMO 値を取得する。HOMO/LUMO 値は標準出力に出力する。

【入出力形式】

入力：mopac 出力ファイル

出力：標準出力

【使用法】

getHomoLumo <mopac 出力ファイル名>

【使用例】

以下に示す methane.dat について mopac 計算を行った後に、HOMO/LUMO 値の取得を行う。同時に原子電荷の値の取得を行う。

```
# cat methane.dat
AM1 BONDS XYZ VECTORS MMOK CHARGE=0

H -2.43300 1 0.73140 1 0.00000 1
C -1.84960 1 0.14810 1 0.00000 1
H -1.13510 1 0.56060 1 0.00000 1
H -2.26210 1 -0.56640 1 0.00000 1
H -1.43710 1 -0.56640 1 0.00000 1

# mopad7 mehtane
CYCLE: 1 TIME: 0.00 TIME LEFT: 3600.0 GRAD.: 415.400 HEAT: 215.2515
CYCLE: 2 TIME: 0.00 TIME LEFT: 3600.0 GRAD.: 204.356 HEAT: 141.7799
CYCLE: 3 TIME: 0.00 TIME LEFT: 3600.0 GRAD.: 97.203 HEAT: 112.0385
...
CYCLE: 16 TIME: 0.00 TIME LEFT: 3600.0 GRAD.: 8.983 HEAT:-8.501519
CYCLE: 17 TIME: 0.01 TIME LEFT: 3600.0 GRAD.: 6.749 HEAT:-8.686013
CYCLE: 18 TIME: 0.00 TIME LEFT: 3600.0 GRAD.: 2.918 HEAT:-8.756700
0.010u 0.010s 0:00.01 200.0% 0+0k 0+0io 279pf+0w

# getHomoLumo methane.out
HOMO = -13.2980
LUMO = 4.6540

# getAtomicCharge methane.out
CHARGE 1 0.0665
CHARGE 2 -0.2659
CHARGE 3 0.0666
CHARGE 4 0.0663
CHARGE 5 0.0665
```

4-5. GetAtomicCharge

【概要】本プログラムは MOPAC 出力ファイルから原子電荷(Mulliken 電荷)値を取得する。
原子電荷は標準出力に出力する。

【入出力形式】

入力：mopac 出力ファイル

出力：標準出力

【使用法】

getAtomicCharge <mopac 出力ファイル名>

【入出力形式】

入力：mopac 出力ファイル

出力：標準出力

【使用法】

getAtomicCharge <mopac 出力ファイル名>

【使用例】

前項 GetHomoLumo の項に記載

4-6. GetMolProperty

【概要】本プログラムは mol2 ファイルを入力し、その分子の分子式、分子量、水素ドナー/アクセプタ原子数を計算する。

【入出力形式】

入力：mol2 ファイル

出力：標準出力

【使用法】

GetMolProperty <mol2 ファイル>

【使用例】

メタン分子の各種物性値を計算する。

```
# cat methane.mol2

@<TRIPOS>MOLECULE
methane.mol2
  5 4 0 0 0
SMALL
GASTEIGER

@<TRIPOS>ATOM
  1 H1      -2.6380    0.9010    0.0170 H      1 UNK      0.00000
  2 C1      -1.8670    0.1320   -0.0110 C.3   1 UNK      0.00000
  3 H2      -0.9100    0.5720   -0.2870 H      1 UNK      0.00000
  4 H3      -2.1430   -0.6290   -0.7400 H      1 UNK      0.00000
  5 H4      -1.7800   -0.3240    0.9730 H      1 UNK      0.00000

@<TRIPOS>BOND
  1  1  2  1
  2  2  3  1
  3  2  4  1
  4  2  5  1

# GetMolProperty methane.mol2

INFORMATION> toolCountCirc
  No circular structures have found.
molecular formula = H4C
molecular weight  = 16.043
Num of Donor      = 0
Num of Acceptor   = 0
Num of chiral atoms = 0

%% Program is done. %%
%% This program ended normally. %%
```

4-7.CreateBondMatrix

【概要】本プログラムは、幾何異性体、光学異性体を構成する原子の結合に着目し、結合行列の固有値を計算する。この値を用いて同一分子の判定を行う。

【入出力形式】

入力：mopac 出力ファイル、フラグ(0:mol ファイル、1:mol2 ファイル)

出力：標準出力

【使用法】

CreateBondMatrix <入力 mol ファイル名> 0

CreateBondMatrix <入力 mol2 ファイル名> 1

【使用例】

```
# cat isomer.mol2
@<TRIPOS>MOLECULE
SYBYL Mol2.FTR
 12 11 0 0 0
SMALL
NO_CHARGES
@<TRIPOS>ATOM
 1  C   -1.494  -0.554  -0.336  C.2
 2  C   -0.256  -0.607  -0.821  C.2
 3  C    0.629   0.670  -0.785  C.3
 4  Cl  -2.615  -1.916  -0.308  Cl
 5  H   -1.869   0.416   0.091  H
 6  Br    0.484  -2.207  -1.580  Br
 7  C    1.171   0.965   0.637  C.3
 8  H    0.331   0.962   1.384  H
 9  F    1.806   2.315   0.657  F
10  Cl   2.378  -0.212   1.156  Cl
11  H    0.016   1.547  -1.132  H
12  H    1.488   0.570  -1.502  H
@<TRIPOS>BOND
 1  1  2  2
 2  1  4  1
 3  1  5  1
 4  2  3  1
 5  2  6  1
 6  3  7  1
 7  3 11  1
 8  3 12  1
 9  7  8  1
10  7  9  1
11  7 10  1

# CreateBondMatrix isomer.mol2 1
INFORMATION> toolCountCirc
  No circular structures have found.
Num of chiral atoms = 1
BondMatrix(geometrical) = 35.034547
BondMatrix**2(geometrical) = 1227.419488
BondMatrix(optical) = 35.034524
BondMatrix**2(optical) = 1227.417866
BondMatrix(total) = 35.034547
BondMatrix**2(total) = 1227.419488
```

