

myPresto 5.0

– 部分構造探索 –

(substructure_search)

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. 部分構造探索システムの概要.....	4
2. システムの構成.....	4
3. ディレクトリ構成.....	5
4. システムの実行準備	6
5. 操作手順	7
6. サンプル実行方法.....	10

1. 部分構造探索システムの概要

部分構造探索システムは探索対象化合物 (mol2 形式) に対し、クエリ構造 (mol2 形式) と同じ構造を持つ化合物を検索するシステムです。分子は化学結合をエッジとするエッジ行列に変換され、部分構造の比較はウルマンの定理によって行われます。分子の配座、光学異性体は考慮することはできません。

2. システムの構成

部分構造探索システムは以下のプログラムで構成されています。

(1) 部分構造探索プログラム (match_x)

部分構造探索を行うプログラムです。下図のようにクエリ構造リストと探索対象化合物リストを指定し、リストに記載されたクエリ構造の mol2 ファイルと探索対象化合物の mol2 ファイルを読み込み、クエリ構造の部分構造を探索します。

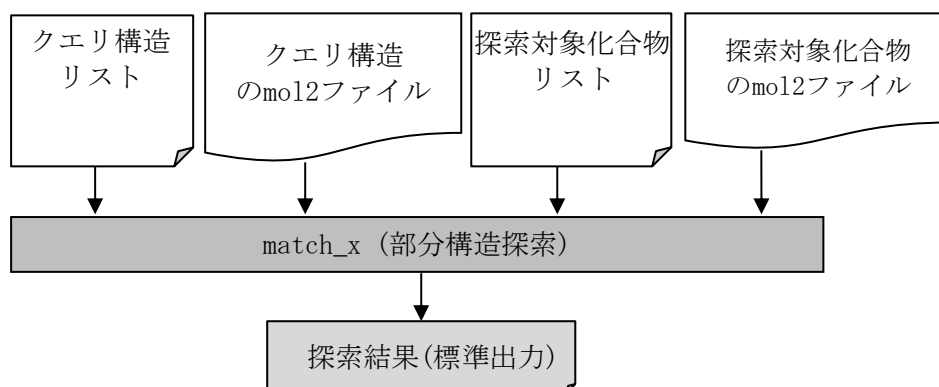


図 1. システム構成

3. ディレクトリ構成

部分構造探索システムは図2のように部分構造探索システムディレクトリ以下、8つのディレクトリで構成されています。

```
substructure_searchYYMMDD/
|--README
|--bin/
|   |--install.sh
|   |--check_binary.sh
|   |--clean_binary.sh
|   |--test_substructure_search.sh
|   |--(match_x) (install.sh 実行後に出現)
|--doc/
|   |--substructure_search_manual_v5.0_ja_151006.pdf
|--ligandDB//
|   |--lig_sample/
|--queryDB/
|   |--query_sample
|--search/
|--search_sample/
|   |--ligand.list
|   |--match_x.inp
|   |--query.list
|--src/
|   |--Makefile
|   |--makefie.intel
|   |--match_x
```

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) 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。

(2) ソースディレクトリ (src)

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。

(3) クエリ構造ディレクトリ (queryDB)

クエリ構造の mol2 ファイルを保存します。

(4) 探索対象化合物ディレクトリ (ligandDB)

探索対象化合物の mol2 ファイルを保存します。

(5) 部分構造探索作業ディレクトリ (search)

部分構造探索を実行するディレクトリです。

(6) 部分構造探索サンプル作業ディレクトリ (search_sample)

部分構造探索をサンプル実行するディレクトリです。

4. システムの実行準備

部分構造探索システムの準備手順を以下に示します。

1) システムのインストール

まず、substructure_searchYYMMDD.tar.gz を書き込み可能権限のあるディレクトリ（例えば、ホームディレクトリ）に配置して、次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。）インストールには、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。

```
% tar -xzf substructure_searchYYMMDD.tar.gz
```

"substructure_searchYYMMDD/"ディレクトリ内部のディレクトリ構成は図2のようになっています。以下図2に示したディレクトリ名で説明します。

次のコマンドを実行することにより、プログラムをインストールします。

```
% cd substructure_searchYYMMDD
```

次のコマンドは、どちらか一方を実行します。

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

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

2) テストプログラムの実行

次のコマンドで、テストプログラムを実行します。

```
% bin/test_substructure_search.sh
```

このテストプログラムの出力先は、substructure_searchYYMMDD/test_search_sample/です。このテストプログラムを実行することにより、プログラムが正常に動作しているかどうかを確認することができます。

(3) クエリ構造の mol2 ファイル配置

"substructure_searchYYMMDD/queryDB/"にクエリ構造(mol2 形式)を配置します。

(4) 探索対象化合物の mol2 ファイル配置

"substructure_searchYYMMDD/ligandDB/"に探索対象化合物(mol2 形式)を配置します。

5. 操作手順

部分構造探索システムは以下の手順で実行します。

(1) クエリ構造リストのファイル作成

"substructure_searchYMMDD/search/"にクエリ構造リストのファイルを作成します。リストは"substructure_searchYMMDD/queryDB/"以下のファイルパス名で指定します。

../queryDB/queryA/query001.mol2 ../queryDB/queryB/query002.mol2 ...	複数指定すると連続で探索できます。
---	-------------------

図 3. 探索クエリ構造リストファイル

(2) 探索対象化合物リストのファイル作成

"substructure_searchYMMDD/search/"に探索対象化合物リストのファイルを作成します。リストは"substructure_searchYMMDD/ligandDB/"以下のファイル名パスで指定します。

../ligandDB/C001/00001.mol2 ../ligandDB/C002/00002.mol2 ../ligandDB/C003/00003.mol2 ...	複数指定すると連続で探索できます。 (相対パスで指定します。)
--	------------------------------------

図 4. 探索対象化合物リストファイル

(3) 制御ファイルの作成

"substructure_searchYMMDD/search/"に制御ファイルを作成します。制御ファイルの形式は以下の通りです。

query.list	クエリ構造リストのファイル名。
ligand.list	探索対象化合物リストのファイル名。
[探索オプション]	探索オプションを設定します。

図 5. 制御ファイル

探索オプションでは以下の設定ができます。記載されていない場合はデフォルトの設定で実行されます。

表 1. 探索オプション一覧表

#	オプション名	内容
1	DEBUG	デバッグ情報を表示します。
2	SHORT	クエリ構造を含む化合物の結果のみ表示します。指定しない場合は全探索対象化合物の結果を表示します。

<設定例>

オプションが複数ある場合、制御ファイルの[探索オプション]の行に並べて記載します。

DEBUG SHORT

(4) ワイルドカード原子の設定

クエリ構造原子の原子タイプを任意の原子タイプ(ワイルドカード原子)として探索を実行する場合は、クエリ構造の mol2 ファイルの編集が必要です。

クエリ構造の mol2 ファイルの COMMENT ブロックに“WILDCARD 原子 ID”の行を記述することでワイルドカード原子を指定します。

下記の例では、クエリ構造 NH2 に対して 1 番目の原子がワイルドカード指定されているため、NH2, CH2 等の二つの水素と結合する任意の原子が含まれる構造を持つ分子を探索します。

<pre>@<TRIPOS>MOLECULE NH2 3 2 0 0 0 SMALL NO_CHARGES @<TRIPOS>COMMENT WILDCARD 1 @<TRIPOS>ATOM 1 N -0.429 0.088 -0.339 N.4 2 H 0.354 -1.206 -0.521 H 3 H 0.377 1.054 0.518 H @<TRIPOS>BOND 1 1 2 1 2 1 3 1</pre>	<p>COMMENT ブロック 1 番原子(N)のワイルドカード指定</p>
---	--

図 6. ワイルドカード原子の設定例

(5) プログラム実行

"substructure_searchYYMMDD/search/"で以下のコマンドを実行します。

```
% ../bin/match_x < [制御ファイル]
```

出力は標準出力に結果を出力します。探索結果は以下の形式で出力します。

<pre>***** * SUBSTRUCTURE SEARCH PROGRAM V1.0 * * 2009/11/2 * ***** CONTROL FILE FORMAT: LINE 1: SUBSTRUCTURE LIST FILE (MOL-2 FILE NAMES) LINE 2: MOLECULDE LIST FILE (MOL-2 FILE NAMES) LINE 3: LOG FORMAT OPTION DEBUG = OUTPUT DEBUG INFORMATION SHORT = DISPLAY FINDING PATTERN ONLY INFORMATION> 1) STRUCUTRE FILE LIST : query.list 2) TARGET FILE LIST : ligand.list 3) DEBUG MODE : F 4) DISPLAY MODE : FINDING PATTERN ONLY INFORMATION> WILDCARD ATOMS SUBSTRUCTURE ATOMS 1 ../queryDB/query_sample/lig1.mo12 11 NO FILE_NAME ATOM SUB_NO PATTERN # 1 001.mo12 21 1 1 # 2 002.mo12 19 1 1 # 3 003.mo12 20 1 1 # 4 004.mo12 20 1 1 # 5 005.mo12 18 1 1 # 6 006.mo12 24 1 1 # 7 007.mo12 18 1 1 # 8 008.mo12 24 1 1 # 9 009.mo12 26 1 1 # 10 010.mo12 21 1 1 INFORMATION> WILDCARD ATOMS SUBSTRUCTURE ATOMS 2 ../queryDB/query_sample/lig2.mo12 12 NO FILE_NAME ATOM SUB_NO PATTERN # 5 005.mo12 18 2 1 # 6 006.mo12 24 2 1 # 7 007.mo12 18 2 1 # 8 008.mo12 24 2 1 # 9 009.mo12 26 2 1</pre>	<p>制御ファイルに記載した設定が表示されます。</p> <p>クエリに指定した部分構造の番号、ファイル名、原子数の順に表示されます。</p> <p>探索結果： 以下の順に表示されます。 内部番号、ファイル名、原子数、検索した部分構造の番号、見つかった部分構造の数</p> <p>クエリが複数指定された場合は、次の部分構造探索が開始されます。</p>
--	---

図 7. 部分構造探索出力例

6. サンプル実行方法

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 4.-(1) までの準備が必要です。

”substructure_searchYYMDD/search_sample/”で以下のコマンドを実行します。

```
% ../bin/match_x <match_x.inp > search.log
```

クエリ構造として query.list に記述されている”lig1.mol2”, ”lig2.mol2”が指定されます。探索対象化合物として ligand.list に記述されている”001.mol2”~”010.mol2”が指定され部分構造探索プログラムを実行します。実行後、”substructure_searchYYMDD/search_sample/”に部分構造探索結果である search.log が出力されます。

以上

myPresto 5.0
~~-部分構造探索-~~
substructure_search