

# MolSite

## Ligand-binding pocket prediction

方式設計書

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)

## 目次

|       |                             |    |
|-------|-----------------------------|----|
| 1     | システムの概要                     | 1  |
| 2     | システム構成                      | 2  |
| 3     | 外部仕様                        | 4  |
| 3.1   | ディレクトリ構成                    | 4  |
| 3.2   | make_pre_point.pl           | 8  |
| 3.2.1 | 機能                          | 8  |
| 3.2.2 | 入出力                         | 8  |
| 3.2.3 | 起動方法                        | 9  |
| 3.3   | make_grid.csh               | 9  |
| 3.3.1 | 機能                          | 9  |
| 3.3.2 | 入出力                         | 9  |
| 3.3.3 | 起動方法                        | 9  |
| 3.4   | make_docking_socre_ps.csh   | 9  |
| 3.4.1 | 機能                          | 9  |
| 3.4.2 | 入出力                         | 9  |
| 3.4.3 | 起動方法                        | 10 |
| 3.5   | make_docking_score_multi.pl | 10 |
| 3.5.1 | 機能                          | 10 |
| 3.5.2 | 入出力                         | 10 |
| 3.5.3 | 起動方法                        | 11 |
| 3.6   | make_score_data_ps.pl       | 11 |
| 3.6.1 | 機能                          | 11 |
| 3.6.2 | 入出力                         | 11 |
| 3.6.3 | 起動方法                        | 11 |
| 3.7   | make_score_data_multi_ps.pl | 11 |
| 3.7.1 | 機能                          | 11 |
| 3.7.2 | 入出力                         | 12 |
| 3.7.3 | 起動方法                        | 12 |
| 3.8   | make_scoring_grid.pl        | 12 |
| 3.8.1 | 機能                          | 12 |
| 3.8.2 | 入出力                         | 12 |
| 3.8.3 | 起動方法                        | 13 |

|        |                                   |    |
|--------|-----------------------------------|----|
| 3.9    | make_docking_pocket_coordinate.pl | 13 |
| 3.9.1  | 機能                                | 13 |
| 3.9.2  | 入出力                               | 14 |
| 3.9.3  | 起動方法                              | 15 |
| 3.10   | verify_point.pl                   | 15 |
| 3.10.1 | 機能                                | 15 |
| 3.10.2 | 入出力                               | 15 |
| 3.10.3 | 起動方法                              | 15 |
| 3.11   | sievgene の制御ファイル                  | 16 |
| 4      | プログラム詳細                           | 16 |
| 4.1    | make_pre_point.pl                 | 16 |
| 4.1.1  | 機能                                | 16 |
| 4.1.2  | 引数                                | 16 |
| 4.1.3  | 出力                                | 16 |
| 4.1.4  | 処理フロー                             | 18 |
| 4.2    | create_pre_point.f90              | 19 |
| 4.2.1  | 機能                                | 19 |
| 4.2.2  | 標準入力                              | 19 |
| 4.2.3  | 出力                                | 20 |
| 4.2.4  | サブルーチン                            | 20 |
| 4.2.5  | 処理フロー                             | 21 |
| 4.3    | make_docking_socre2.csh           | 21 |
| 4.3.1  | 機能                                | 21 |
| 4.3.2  | 変更箇所                              | 22 |
| 4.3.3  | 引数                                | 22 |
| 4.3.4  | 出力                                | 22 |
| 4.4    | RUN_docking_ps.pl                 | 24 |
| 4.4.1  | 機能                                | 24 |
| 4.4.2  | 変更箇所                              | 24 |
| 4.4.3  | 引数                                | 25 |
| 4.5    | make_score_data_ps.pl             | 26 |
| 4.5.1  | 機能                                | 26 |
| 4.5.2  | 変更箇所                              | 26 |
| 4.5.3  | 引数                                | 26 |
| 4.6    | make_score_data_multi_ps.pl       | 26 |
| 4.6.1  | 機能                                | 26 |

|        |                                   |    |
|--------|-----------------------------------|----|
| 4.6.2  | 変更箇所                              | 27 |
| 4.6.3  | 引数                                | 27 |
| 4.7    | make_scoring_grid.pl              | 27 |
| 4.7.1  | 機能                                | 27 |
| 4.7.2  | 引数                                | 28 |
| 4.7.3  | 出力                                | 28 |
| 4.7.4  | 処理フロー                             | 29 |
| 4.8    | calc_sigma.f90                    | 29 |
| 4.8.1  | 機能                                | 29 |
| 4.8.2  | 標準入力                              | 30 |
| 4.8.3  | 出力                                | 30 |
| 4.8.4  | サブルーチン                            | 30 |
| 4.8.5  | 処理フロー                             | 30 |
| 4.9    | make_docking_pocket_coordinate.pl | 31 |
| 4.9.1  | 機能                                | 31 |
| 4.9.2  | 引数                                | 31 |
| 4.9.3  | 出力                                | 31 |
| 4.9.4  | 処理フロー                             | 32 |
| 4.10   | calc_average_coord.f90            | 33 |
| 4.10.1 | 機能                                | 33 |
| 4.10.2 | 標準入力                              | 33 |
| 4.10.3 | 出力                                | 33 |
| 4.10.4 | サブルーチン                            | 33 |
| 4.10.5 | 処理フロー                             | 34 |
| 4.11   | calc_center_coord.f90             | 34 |
| 4.11.1 | 機能                                | 34 |
| 4.11.2 | 標準入力                              | 34 |
| 4.11.3 | 出力                                | 34 |
| 4.11.4 | サブルーチン                            | 34 |
| 4.11.5 | 処理フロー                             | 34 |
| 4.12   | verify_point.pl                   | 35 |
| 4.12.1 | 機能                                | 35 |
| 4.12.2 | 引数                                | 35 |
| 4.12.3 | 出力                                | 35 |
| 4.12.4 | 処理フロー                             | 36 |
| 4.13   | calc_verification_data.f90        | 36 |

|        |              |    |
|--------|--------------|----|
| 4.13.1 | 機能 .....     | 36 |
| 4.13.2 | 標準入力 .....   | 36 |
| 4.13.3 | 出力 .....     | 36 |
| 4.13.4 | サブルーチン ..... | 36 |
| 4.13.5 | 処理フロー .....  | 36 |

## 1 システムの概要

蛋白質-化合物ドッキングにおいて、ドッキングポケットの位置はドッキングスコアの算出に大きな影響を与えるため、重要な要素である。従来は、ドッキングポケットを実験的に求めるか、ホロ酵素の非蛋白質部分をドッキングポケットとするなどしていた。

本開発では、蛋白質の主鎖周辺の原子それぞれをドッキングポケットとしてドッキングスコアを求め、最適と考えられるドッキングポケットを定量的に決定するシステム(以下、ポケット探索システムとする)を開発する。

ポケット探索システムは、ドッキングポケットを探索したい蛋白質を用意し、蛋白質の主鎖に属する中心原子(C、N、O等)の空間座標をドッキングポケットとし、ドッキングポケットごとに、sievgene による既知化合物とのドッキングジョブを実行してドッキングスコアを求める。

ドッキングスコアは、最良のドッキングポケットほど良い値となるが、ドッキング化合物によってばらつきが発生する。そのため、ドッキングポケットごとにスコアの標準偏差を求め、より高スコアで、かつ、ばらつきが少ないスコアを算出しているドッキングポケットを最適なドッキングポケット候補として採用する。

次に、各ドッキング化合物の原子の中心座標を求め、さらにドッキング化合物全体の中心の平均座標を求めることで、最終的なドッキングポケットの中心座標を決定する。

## 2 システム構成

本システムは、ドッキングポケットの作成、ドッキングジョブの実行、ドッキングポケット座標の決定の3段階に分けられる。ドッキングポケットを決定したい蛋白質が複数に対しても、1つの場合と同等の処理でそれぞれの蛋白質のドッキングポケットの中心座標を決定するシステムを考慮する。図 3.2.1-1 と図 3.2.1-2 エラー! 参照元が見つかりません。に、1蛋白質のドッキングポケットを決定するまでの動作概要を示す。

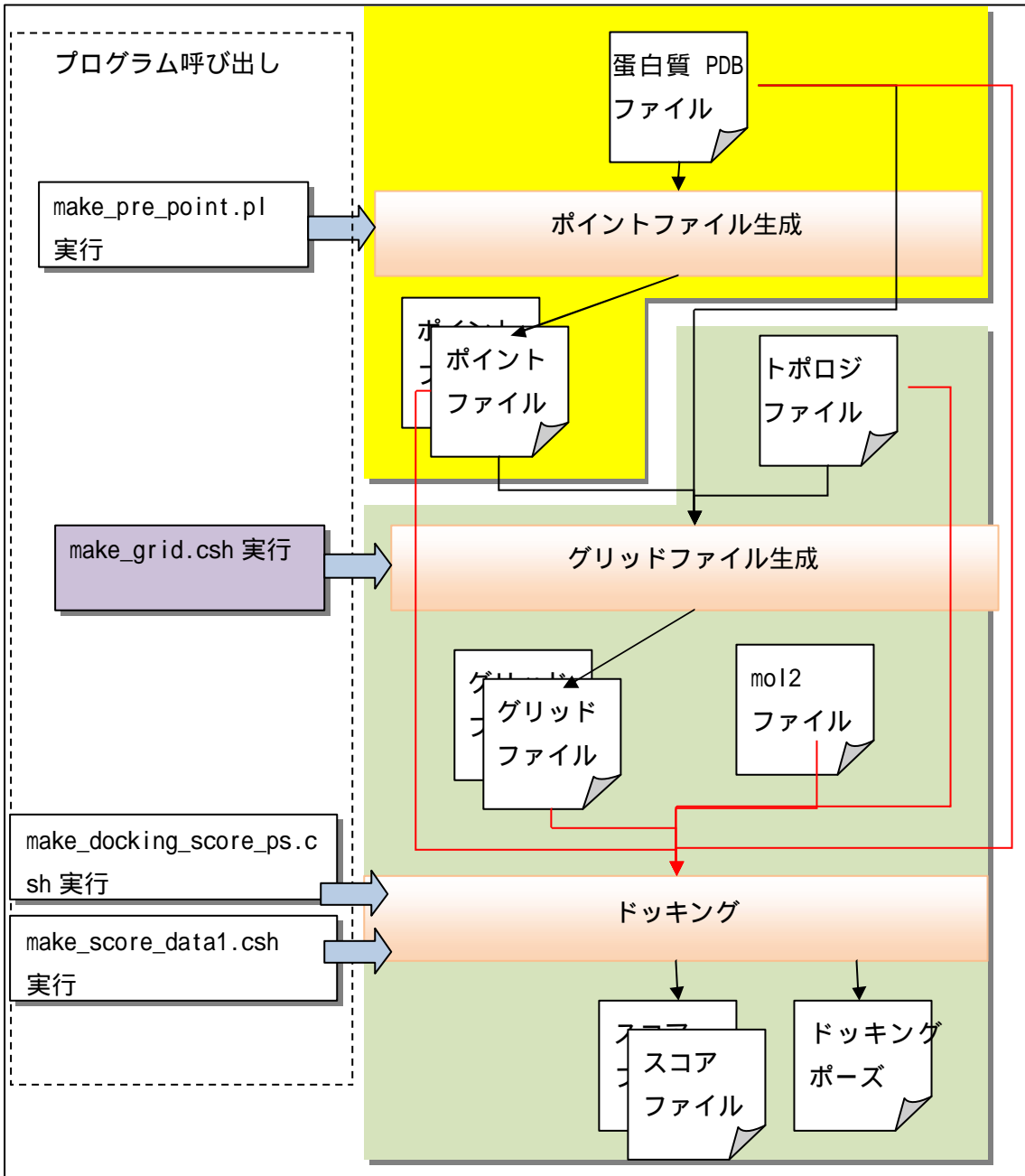


図 3.2.1-1 システムの動作概要(1)

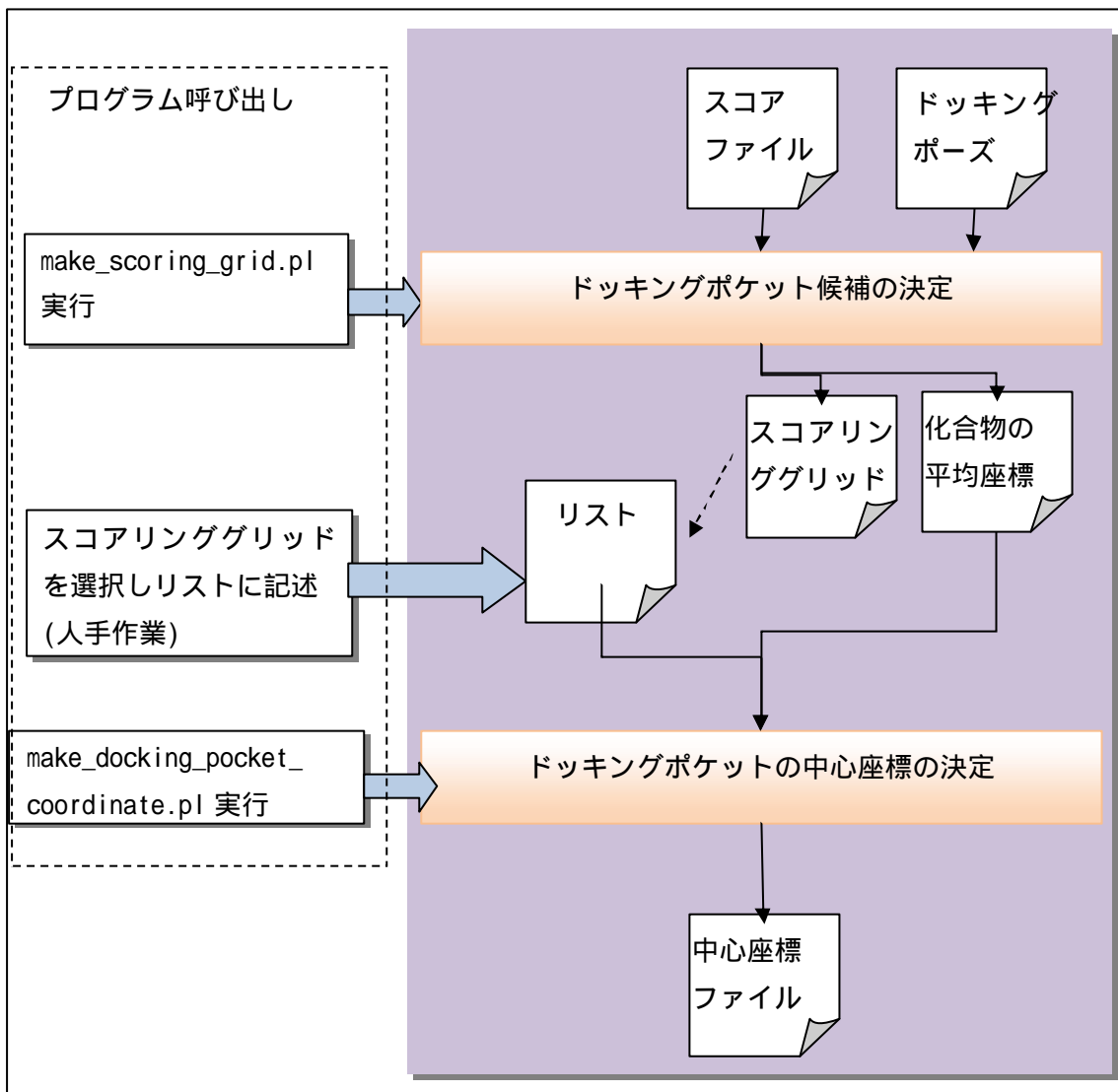


図 3.2.1-2 システムの動作概要(2)

は、ドッキングポケット位置を求めたい蛋白質から複数のポイントファイルを生成するフェーズである。生成されたポイントファイルの数だけ、ディレクトリが作成される。そして、蛋白質ファイル、トポロジファイル、ポイントファイルが作成されたディレクトリに配置され、後のドッキングは別々の蛋白質データとして使用される。

は、 で用意された蛋白質データを用いてドッキングジョブを行う。ドッキングにより、ドッキングスコアと、ドッキングポーズのファイルが作成される。

は、 で作成されたドッキングスコアとドッキングポーズを用いてドッキングポケット位置の候補を求め、ドッキングポケットの中心座標を決定する。



### 3 外部仕様

#### 3.1 ディレクトリ構成

初期のディレクトリ構成について図 3.2.1-1 に示す。

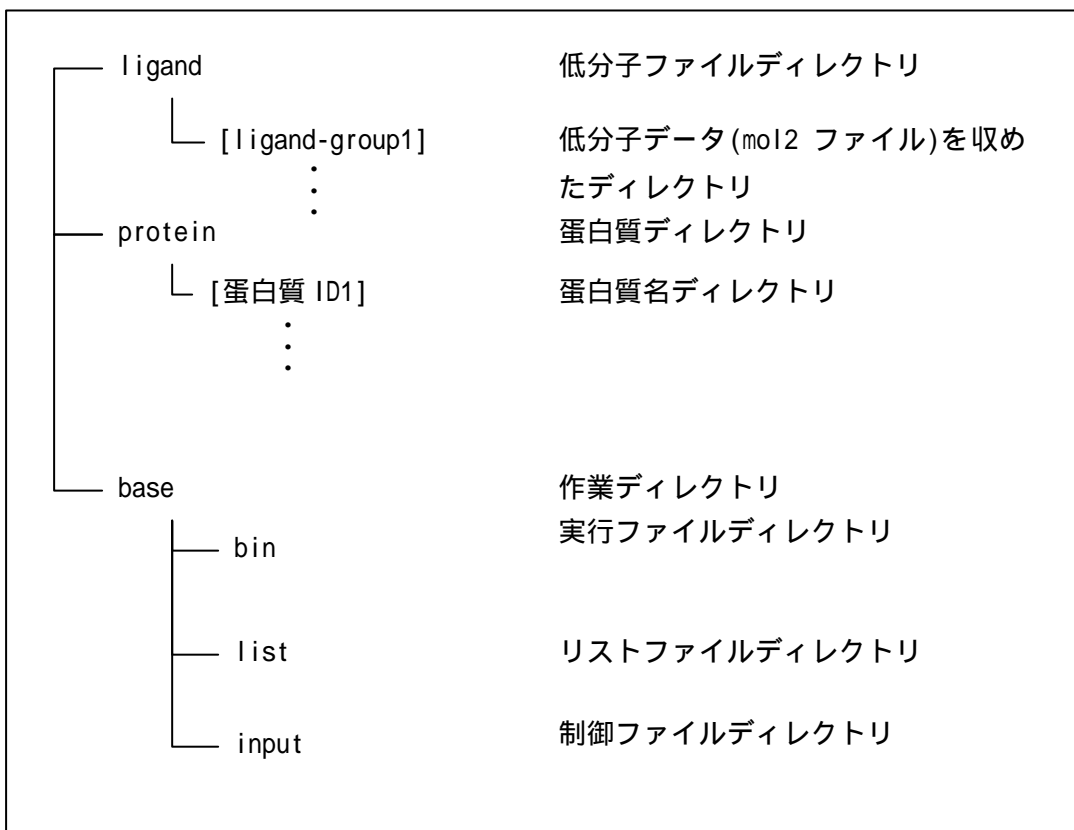


図 3.2.1-1 初期ディレクトリ構成

初期のディレクトリ構成は、ligand ディレクトリ、protein ディレクトリ、base ディレクトリである。

ligand ディレクトリには、ligand グループディレクトリがあり、そのディレクトリに mol2 ファイルを格納する。

base ディレクトリは、作業ディレクトリであり、実行プログラムが収められた bin ディレクトリ、ligand と protein のリストを収めた list ディレクトリ、sievgene 実行に用いる制御ファイルを格納した input ディレクトリからなる。

bin ディレクトリに格納する実行プログラムについて、表 3.2.1-1 に示す。

表 3.2.1-1 実行プログラム一覧

| #  | プログラム名                            | 説明                          |
|----|-----------------------------------|-----------------------------|
| 1  | make_pre_point.pl                 | ポイントファイル作成プログラム実行スクリプト      |
| 2  | create_pre_point                  | ポイントファイル作成プログラム             |
| 3  | make_grid.csh                     | グリッドファイル作成プログラム実行スクリプト      |
| 4  | make_docking_socre2.csh           | ドッキングジョブ実行スクリプト             |
| 5  | make_docking_score_multi.pl       | ドッキングジョブ実行スクリプト(マルチ mol2 用) |
| 6  | RUN_docking_ps.pl                 | ドッキングジョブバッチスクリプト            |
| 7  | sievgene                          | ドッキングジョブ実行及びグリッドファイル作成プログラム |
| 8  | make_score_data_ps.pl             | 相互作用行列作成実行スクリプト             |
| 9  | make_score_data_multi_ps.pl       | 相互作用行列作成実行スクリプト(マルチ mol2 用) |
| 10 | DataMaker                         | 相互作用行列作成用プログラム              |
| 11 | make_scoring_grid.pl              | スコアリンググリッド計算実行スクリプト         |
| 12 | calc_sigma                        | スコアリンググリッド計算プログラム           |
| 13 | calc_average_coord                | 低分子化合物の平均座標計算プログラム          |
| 14 | make_docking_pocket_coordinate.pl | ドッキングポケット中心座標計算実行スクリプト      |
| 15 | calc_center_coord                 | ドッキングポケット中心座標計算プログラム        |

表 3.2.1-1 の内、create\_pre\_point、calc\_sigma、calc\_average\_coord、calc\_center\_coord は、それぞれの実行用スクリプトから起動されるため、後章のプログラム詳細で説明する。また、sievgene と DataMaker は、相互作用行列作成等で使用するプログラムであるため説明を省略するが、sievgene の制御ファイルは本章で説明する。

表 3.2.1-2 に、ユーザが準備する入力ファイル一覧を示す。

表 3.2.1-2 ユーザが準備する入力ファイル一覧

| # | プログラム名            | 格納先ディレクトリ            | 説明                                   |
|---|-------------------|----------------------|--------------------------------------|
| 1 | Pro_md.pdb        | protein/[蛋白質名]ディレクトリ | 蛋白質の PDB ファイル                        |
| 2 | Pro.tpl           | protein/[蛋白質名]ディレクトリ | 蛋白質のトポロジファイル                         |
| 3 | protein_point.inp | base/input ディレクトリ    | 蛋白質制御ファイル                            |
| 3 | s0grid.inp        | base/input ディレクトリ    | グリッドファイル作成用制御ファイル                    |
| 4 | s0.inp            | base/input ディレクトリ    | ドッキングジョブ実行用制御ファイル                    |
| 5 | 低分子化合物リスト         | base/list ディレクトリ     | 低分子化合物を列挙したリスト                       |
| 6 | ポイントリスト           | base/list ディレクトリ     | ドッキングポケットの候補となる(蛋白質名 ID+ ポケット番号)のリスト |

表 3.2.1-2 の#1~5 は、初期状態でユーザが作成可能なファイルである。また、#6 については、make\_scoring\_grid.pl を実行した後に、ユーザが作成する。ポイントリストについては 3.9.2 で説明する。

本システム実行により、ドッキングポケットの中心の平均座標を求めた後のディレクトリ構成を図 3.2.1-2 に示す。

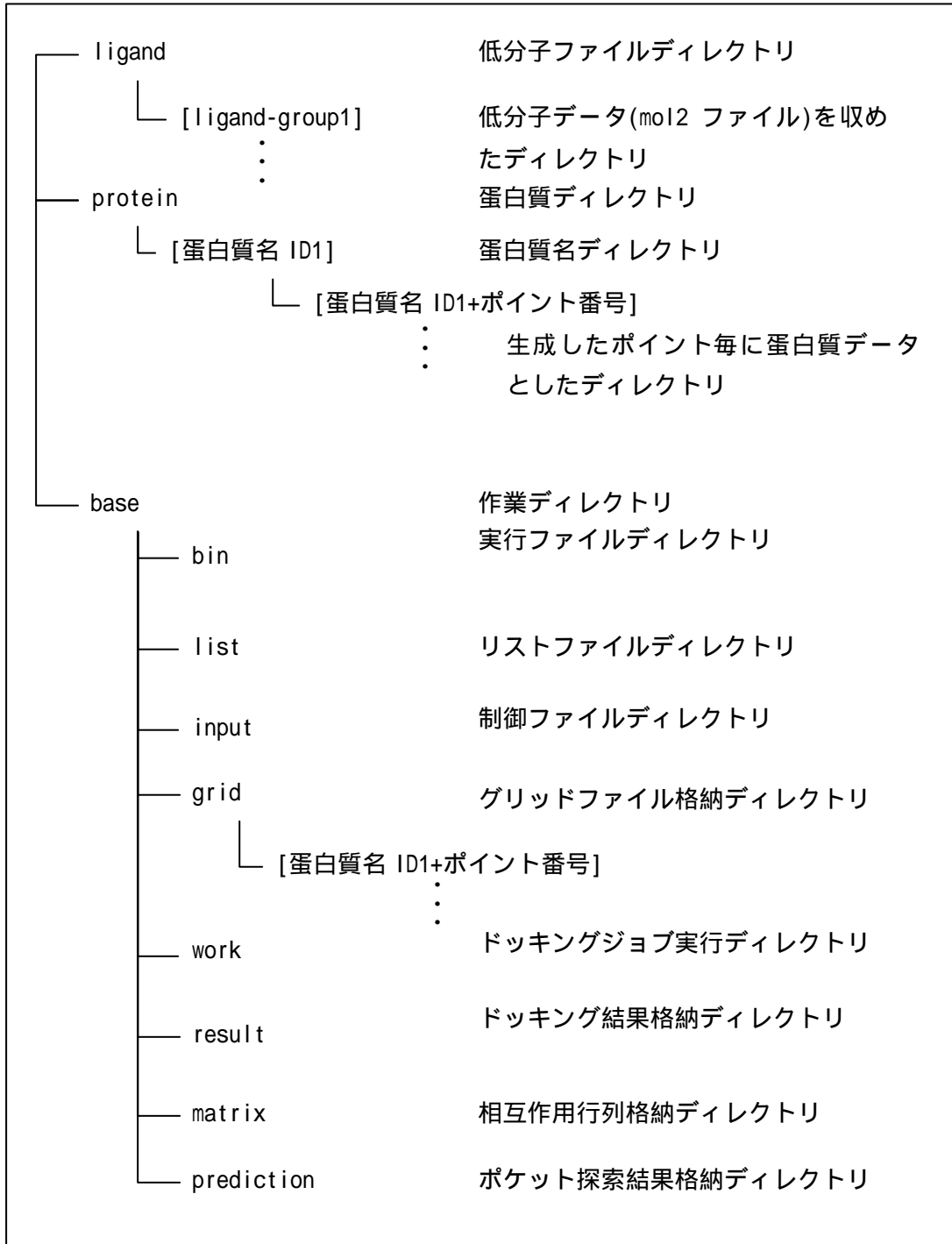


図 3.2.1-2 本システム実行後のディレクトリ構成

## 3.2 make\_pre\_point.pl

### 3.2.1 機能

ポイントファイルを作成する。制御ファイルを読み込み、protein ディレクトリに(蛋白質 ID+ ポイント番号)のディレクトリを作成する。次に、蛋白質名と原子名を引数として create\_pre\_point を実行する。蛋白質名と原子名の組が複数行ある場合は、create\_pre\_point が繰り返し実行される。

### 3.2.2 入出力

入出力ファイル一覧を表 3.2.2-1 に示す。

表 3.2.2-1 make\_pre\_point.pl の入出力一覧

| # | ファイル名             | 入力/出力 | 説明                                  |
|---|-------------------|-------|-------------------------------------|
| 1 | protein_point.inp | 入力    | 蛋白質制御ファイル<br>ファイル名は任意               |
| 2 | point.pdb         | 出力    | 生成したポイントファイル<br>一つの蛋白質について一つ以上作成される |
| 3 | protein.list      | 出力    | 蛋白質リストファイル                          |

蛋白質制御ファイルのフォーマットを以下に示す。

|             |  |
|-------------|--|
| 蛋白質 ID1 原子名 | all_point.pdbを生成する蛋白質 蛋白質主鎖からポケットの中心として選択する原子名 |
| 蛋白質 ID2 原子名 |  |
| 蛋白質 ID3 原子名 |  |
| ...(以下略)... |  |

point.pdb は1座標分のデータとなる。point.pdb のフォーマット例を以下に示す。

|      |   |   |     |         |        |        |
|------|---|---|-----|---------|--------|--------|
| ATOM | 1 | C | PIN | -32.179 | -7.029 | 36.907 |
|------|---|---|-----|---------|--------|--------|

protein.list は、list ディレクトリに作成される。相互作用行列作成時に make\_docking\_score.csh 実行で使用する蛋白質リストと同じフォーマットとなる。フォーマット例を以下に示す。

|                    |
|--------------------|
| (蛋白質 ID1 + ポイント番号) |
| (蛋白質 ID2 + ポイント番号) |

…(以下略)…

### 3.2.3 起動方法

make\_pre\_point.pl は base ディレクトリで、以下のコマンドで実行する。

```
make_pre_point.pl point.inp
```

## 3.3 make\_grid.csh

### 3.3.1 機能

グリッドファイルを作成する。実行すると、base ディレクトリに grid ディレクトリが作成し、protein ディレクトリにある(蛋白質 ID+ポイント番号)ディレクトリの蛋白質データを用いてグリッドファイルを作成する。グリッドファイルは、grid ディレクトリに格納される。相互作業行列作成で使用する make\_grid.csh と同じプログラムを使用する。

### 3.3.2 入出力

入出力ファイル一覧を表 3.3.2-1 に示す。

表 3.3.2-1 make\_grid.csh の入出力ファイル一覧

| # | ファイル名           | 入力/出力 | 説明           |
|---|-----------------|-------|--------------|
| 1 | (蛋白質 ID+ポイント番号) | 入力    | 蛋白質データディレクトリ |
| 2 | grid.file       | 出力    | グリッドファイル     |

### 3.3.3 起動方法

make\_grid.csh は base ディレクトリで、以下のコマンドで実行する。

```
make_grid.csh
```

## 3.4 make\_docking\_socre\_ps.csh

### 3.4.1 機能

ドッキングジョブを投入する。内部では、蛋白質リストと低分子化合物リストを用いて、ドッキングジョブのバッチスクリプトを実行する RUN\_docking\_ps.pl を呼び出す。

### 3.4.2 入出力

入出力ファイル一覧を表 3.4.2-1 に示す。

表 3.4.2-1 make\_docking\_score\_ps.csh の入出力ファイル一覧

| # | ファイル名 | 入力/出力 | 説明 |
|---|-------|-------|----|
|---|-------|-------|----|

|   |        |    |                      |
|---|--------|----|----------------------|
| 1 | 蛋白質リスト | 入力 | ドッキング対象の蛋白質名のリスト     |
| 2 | 化合物リスト | 入力 | ドッキング対象の化合物ファイル名のリスト |

蛋白質リストは、make\_pre\_point.pl 実行で作成された蛋白質リストを指定する。

化合物リストは、低分子グループディレクトリ名をファイル名として、低分子グループディレクトリに格納されている mol2 ファイルを 1 行に 1 つずつ記入する。フォーマット例を以下に示す。

```
lig_001.mol2
lig_002.mol2
lig_003.mol2
... (以下略) ...
```

蛋白質リスト及び化合物リストのファイル名は任意であるが、#は後述のプログラムで識別用のトークンとして使用するため、本システムにおいてファイル名に#を使用してはならない。

### 3.4.3 起動方法

make\_docking\_score\_ps.csh は base ディレクトリで、以下のコマンドで実行する。

蛋白質リストと低分子リストは、パス指定せずにファイル名を直接指定する。

```
make_docking_score_ps.csh 蛋白質リスト 化合物リスト
```

## 3.5 make\_docking\_score\_multi.pl

### 3.5.1 機能

ドッキングジョブを投入する。マルチ mol2 ファイルを使用する場合に make\_docking\_score\_ps.csh の代わりに実行する。相互作業行列作成で使用する make\_docking\_score\_multi.pl と同じプログラムを使用する。

### 3.5.2 入出力

入出力ファイル一覧を表 3.5.2-1 に示す。

表 3.5.2-1 make\_docking\_score\_multi2.pl の入出力ファイル一覧

| # | ファイル名  | 入力/出力 | 説明                   |
|---|--------|-------|----------------------|
| 1 | 蛋白質リスト | 入力    | ドッキング対象の蛋白質名のリスト     |
| 2 | 化合物リスト | 入力    | ドッキング対象の化合物ファイル名のリスト |

蛋白質リストと化合物リストは、make\_docking\_socre\_ps.csh で指定するリストと同様で

ある。

### 3.5.3 起動方法

make\_docking\_score\_multi.pl は base ディレクトリで、以下のコマンドで実行する。  
蛋白質リストと低分子リストは、パス指定せずにファイル名を直接指定する。

```
make_docking_score_multi.pl 蛋白質リスト 化合物リスト
```

## 3.6 make\_score\_data\_ps.pl

### 3.6.1 機能

化合物グループ名を付加した相互作用行列を作成する。

### 3.6.2 入出力

入出力ファイル一覧を表 3.6.2-1 に示す。

表 3.6.2-1 make\_score\_data\_ps.pl の入出力ファイル一覧

| # | ファイル名  | 入力/出力 | 説明  |
|---|--------|-------|---|
| 1 | 蛋白質リスト | 入力    | ドッキング対象の蛋白質名のリスト  |
| 2 | 化合物リスト | 入力    | ドッキング対象の化合物ファイル名のリスト                                    |
| 3 | 相互作用行列 | 出力    | ファイルの先頭行に化合物グループ名を付加した相互作用行列<br>base/matrix ディレクトリに出力する |

蛋白質リストと化合物リストは、make\_docking\_socre2.csh で指定するリストと同様である。作成された相互作用行列のファイル名は、以下の規則で決められる。

**蛋白質リスト\_#化合物リスト#.dat**

#で囲まれた化合物リスト名は、make\_scoring\_grid.pl で取得され使用される。

### 3.6.3 起動方法

make\_score\_data.csh は base ディレクトリで、以下のコマンドで実行する。

蛋白質リストと低分子リストは、パス指定せずにファイル名を直接指定する。

```
make_score_data.csh 蛋白質リスト 化合物リスト
```

## 3.7 make\_score\_data\_multi\_ps.pl

### 3.7.1 機能

化合物グループ名を付加した相互作用行列を作成する。マルチ mol2 ファイルを使用する場合に make\_score\_data\_ps.pl の代わりに実行する。



### 3.7.2 入出力

入出力一覧を表 3.7.2-1 に示す。

表 3.7.2-1 make\_score\_data\_multi\_ps.pl の入出力ファイル一覧

| # | ファイル名  | 入力/出力 | 説明  |
|---|--------|-------|---|
| 1 | 蛋白質リスト | 入力    | ドッキング対象の蛋白質名のリスト  |
| 2 | 化合物リスト | 入力    | ドッキング対象の化合物ファイル名のリスト                                    |
| 3 | 相互作用行列 | 出力    | ファイルの先頭行に化合物グループ名を付加した相互作用行列<br>base/matrix ディレクトリに出力する |

蛋白質リストと化合物リストは、make\_docking\_socre\_ps.csh で指定するリストと同様である。相互作用行列は make\_score\_data\_ps.pl と同様である。

### 3.7.3 起動方法

make\_score\_data\_multi.pl は base ディレクトリで、以下のコマンドで実行する。

蛋白質リストと低分子リストは、パス指定せずにファイル名を直接指定する。

```
make_score_data_multi.pl 蛋白質リスト 化合物リスト
```

## 3.8 make\_scoring\_grid.pl

### 3.8.1 機能

スコアリンググリッドを計算するための実行プログラムを呼び出す。

### 3.8.2 入出力

make\_scoring\_grid.pl の入出力ファイルを表 3.8.2-1 に示す。

表 3.8.2-1 make\_scoring\_grid.pl の入出力ファイル一覧

| # | ファイル名            | 入力/出力 | 説明                        |
|---|------------------|-------|---------------------------|
| 1 | 蛋白質名             | 入力    | 蛋白質名を指定する<br>ポケット番号は指定しない |
| 2 | 蛋白質名_sigma.score | 出力    | スコアリンググリッドの結果ファイル         |

蛋白質名\_sigma.score のフォーマットを以下に示す。

|                          |     |           |
|--------------------------|-----|-----------|
| average = (蛋白質名+ ポケット番号) | 平均値 | スコア平均値が最大 |
|--------------------------|-----|-----------|

|        |                               |                                     |
|--------|-------------------------------|-------------------------------------|
| top    | = (蛋白質名+ ポケット番号) スコア          | 最大スコア                               |
| sigma1 | = (蛋白質名+ ポケット番号) 平均値 + 標準偏差   | スコアの分布度数                            |
| sigma2 | = (蛋白質名+ ポケット番号) 平均値 + 標準偏差*2 | $\mu + \sigma * N(N:1 \sim 5)$ の最大値 |
| sigma3 | = (蛋白質名+ ポケット番号) 平均値 + 標準偏差*3 |                                     |
| sigma4 | = (蛋白質名+ ポケット番号) 平均値 + 標準偏差*4 | $\mu$ : スコアの平均値                     |
| sigma5 | = (蛋白質名+ ポケット番号) 平均値 + 標準偏差*5 | $\sigma$ : 標準偏差                     |

### 3.8.3 起動方法

まず、base/prediction ディレクトリを作成し、化合物グループリスト lig\_grp を作成する。make\_score\_data.csh 等で指定した化合物リスト名を 1 行に 1 つ記述したリストファイルである。化合物グループリストのフォーマット例を以下に示す。

```
c001
c002
c003
... (以下略)...
```

次に、base/prediction ディレクトリに相互作業行列データリストファイル matrix\_list を作成する。matrix\_list のフォーマットを以下に示す。

```
../matrix/pro01.list_##LIGAND_GROUP##.dat
../matrix/pro01.list_lig01.dat
... (以下略)...
```

#LIGAND\_GROUP#は、化合物グループリストに記述された化合物グループ名で置換される。置換された結果、pro01.list\_#c001#.dat(化合物グループ名が c001 の場合)となる。

base/prediction ディレクトリで、以下のコマンドで実行する。

```
make_scoring_grid.pl 蛋白質名
```

make\_scoring\_grid.pl を実行すると、lig\_grp と matrix\_list が内部で読み込まれる。

## 3.9 make\_docking\_pocket\_coordinate.pl

### 3.9.1 機能

ドッキング化合物の原子の平均座標を求める実行プログラムとドッキングポケットの中心の平均座標を計算する実行プログラムを呼び出す。出力ファイルは prediction ディレクトリに格納される。

### 3.9.2 入出力

入出力ファイルを表 3.9.2-1 に示す。

表 3.9.2-1 make\_docking\_pocket\_coordinate.pl の入出力ファイル一覧

| # | ファイル名                 | 入力/出力 | 説明                                   |
|---|-----------------------|-------|--------------------------------------|
| 1 | 蛋白質リスト                | 入力    | 最適なドッキングポイントの蛋白質名(蛋白質名+ポイント番号)のリスト   |
| 2 | 蛋白質名+ポケット番号_ave.coord | 出力    | 蛋白質名+ポケット番号とドッキングした化合物の X、Y、Z 座標の平均値 |
| 3 | 蛋白質名_center.coord     | 出力    | 蛋白質+ポケット番号の予測ドッキングポケットの座標            |

蛋白質リストは、利用者が蛋白質名\_sigma.score の内容を確認し、適切な(蛋白質+ポケット番号)を選択し、記述する。蛋白質リストのフォーマットを以下に示す。

|   |                                |
|---|--------------------------------|
| (蛋白質+ポケット番号)<br>(蛋白質+ポケット番号)<br>(蛋白質+ポケット番号)<br>…(以下略)… | (蛋白質+ポケット番号)が複数あるときは1行ごとに記述する。 |
|---|--------------------------------|

蛋白質名+ポケット番号\_ave.coord のフォーマットを以下に示す。

|                                       |              |              |
|---------------------------------------|--------------|--------------|
| 化合物の X 座標の平均                          | 化合物の Y 座標の平均 | 化合物の Z 座標の平均 |
| 化合物の X 座標の平均                          | 化合物の Y 座標の平均 | 化合物の Z 座標の平均 |
| 化合物の X 座標の平均                          | 化合物の Y 座標の平均 | 化合物の Z 座標の平均 |
| …(以下、蛋白質名+ポケット番号にドッキングさせた化合物の数表示される)… |              |              |

蛋白質名\_center.coord のフォーマットを以下に示す。

|                                      |           |
|--------------------------------------|-----------|
| P (蛋白質+ポケット番号)                       |           |
| @ave X 座標 Y 座標 Z 座標                  | ポケット中心座標  |
| @sigma X 座標の標準偏差 Y 座標の標準偏差 Z 座標の標準偏差 | 中心座標の標準偏差 |
| P (蛋白質+ポケット番号)                       |           |
| @ave X 座標 Y 座標 Z 座標                  |           |
| @sigma X 座標の標準偏差 Y 座標の標準偏差 Z 座標の標準偏差 |           |
| …(以下略)…                              |           |

### 3.9.3 起動方法

make\_docking\_pocket\_coordinate.pl は以下のコマンドで実行する。出力ファイルは prediction ディレクトリに格納される。

```
make_docking_pocket_coordinate.pl ポイントリスト
```

### 3.10 verify\_point.pl

#### 3.10.1 機能

make\_docking\_pocket\_coordinate.pl によって得られた *蛋白質名\_center.coord* のドッキングポケット座標の検証を行う。ドッキングポケット位置が既知の場合に使用できる。

#### 3.10.2 入出力

入出力ファイルを表 3.10.2-1 に示す。

表 3.10.2-1 verify\_point.pl の入出力ファイル一覧

| # | ファイル名                    | 入力/出力 | 説明                    |
|---|--------------------------|-------|-----------------------|
| 1 | PDB ファイル                 | 入力    | 既知のドッキングポケットのポイントファイル |
| 2 | <i>蛋白質名_center.coord</i> | 入力    | 予測ドッキングポケットの計算結果      |
| 2 | <i>蛋白質.verify</i>        | 出力    | 検証結果                  |

*蛋白質.verify* のファイルフォーマットを以下に示す。

```
P (蛋白質 + ポイント番号)
@center-to-center distance: 10.128763
@minimum distance : 5.89981
```

@center-to-center distance は、既知のドッキングポケットの中心座標と、*蛋白質名\_center.coord* で得られた予測ドッキングポケット座標との距離を表す。

@minimum distance は、既知ドッキングポケットの原子と *蛋白質名\_center.coord* で得られた予測ドッキングポケット座標との最短距離を表す。

#### 3.10.3 起動方法

verify\_point.pl は以下のコマンドで実行する。出力ファイルは実行ディレクトリに格納される。

```
verify_point.pl PDB ファイル 蛋白質名_center.coord
```

### 3.11 sievgene の制御ファイル

sievgene の制御ファイルは、以下のパラメータを設定する。

|                 |          |
|-----------------|----------|
| COORDInate      | PDB      |
| NAMECOordinate  | pose.pdb |
| CANDIDatenumber | 1        |
| SCORENumber     | 1        |

この条件を設定してドッキングジョブを実行すると、ドッキングジョブのワークディレクトリに pose.pdb ファイルが作成され、ドッキングが行われた分のドッキングポーズが出力される。

## 4 プログラム詳細

### 4.1 make\_pre\_point.pl

#### 4.1.1 機能

make\_pre\_point.pl は、蛋白質制御ファイルを読み込み、ポイントファイルを生成する蛋白質名とポイントとする原子名の対をメモリに格納する。次に、蛋白質名の数だけ後述の create\_pre\_point を呼び出して中間ファイル all\_point.pdb を作成させる。作成された all\_point.pdb を読み込み、ポイント情報を取得する。all\_point.pdb の各原子をポケットとするため、ポケットの数だけ(蛋白質名+ポケット番号)としてディレクトリを作成し、Pro\_md.pdb と Pro.tpl および 1 原子分の point.pdb を格納する。作成した(蛋白質名+ポケット番号)のリスト protein.list を list ディレクトリに作成する。

#### 4.1.2 引数

引数を表 4.1.2-1 に示す。

表 4.1.2-1 引数一覧

| # | 引数名       | 説明                    |
|---|-----------|-----------------------|
| 1 | 蛋白質制御ファイル | 制御ファイルの相対(絶対)パスを指定する。 |

#### 4.1.3 出力

出力メッセージを表 4.1.3-1 に示す。

表 4.1.3-1 出力メッセージ一覧

| # | メッセージ                         | 説明            |
|---|-------------------------------|---------------|
| 1 | Error!! protein_point.inp not | 蛋白質制御ファイルがない。 |

|    |  |   |
|----|--|---|
|    | found.   |   |
| 2  | Error!! cannot open protein_point.inp. [システムメッセージ]                         | 蛋白質制御ファイルが開けない。   |
| 3  | Error!! protein directory not found.                                       | protein ディレクトリがない。  |
| 4  | Error!! protein_point.inp format is invalid.                               | 蛋白質制御ファイルのフォーマットが不正。  |
| 5  | Error!! PROTEIN directory not found.                                       | 蛋白質制御ファイルに記述された PROTEIN 蛋白質名のディレクトリが protein ディレクトリにない。<br>(PROTEIN 蛋白質名のディレクトリは絶対パスで表示する。) |
| 6  | Error!! Pro_md.pdb not found in PROTEIN directory.                         | PROTEIN 蛋白質名のディレクトリに Pro_md.pdb がない。  |
| 7  | Error!! Pro.tpl not found in PROTEIN directory.                            | PROTEIN 蛋白質名のディレクトリに Pro.tpl が ない。  |
| 8  | Error!! create_pre_point not found.  | create_pre_point がない  |
| 9  | Error!! cannot open __tmp_control.inp. [システムメッセージ]                         | 一時ファイル__tmp_control.inp が作成できない   |
| 10 | [システムメッセージ]<br>Error!! cannot execute create_pre_point.                    | create_pre_point が実行できない  |
| 11 | Error!! cannot change to PROTEIN directory . [システムメッセージ]                   | PROTEIN 蛋白質名ディレクトリに移動できない   |
| 12 | Error!! cannot create PROTEIN+NO directory.                                | PROTEIN+NO ディレクトリが作成できない。   |
| 13 | Error!! all_point.inp not found.   | all_point.pdb ファイルがない   |
| 14 | Error!! cannot open all_point.inp. [システムメッセージ]                             | all_point.pdb ファイルが開けない   |
| 15 | Error!! cannot create symbolic link of Pro_md.pdb to PROTEIN+NO directory. | PROTEIN ディレクトリの Pro_md.pdb のシンボリックリンクを PROTEIN+NO ディレクトリに作成できない。                            |

|    |   |   |
|----|---|---|
| 16 | Error!! cannot create symbolic link of Pro.tpl to PROTEIN+NO directory.   | PROTEIN ディレクトリの Pro.tpl のシンボリックリンクを PROTEIN+NO ディレクトリに作成できない。 |
| 17 | Error!! cannot create point.pdb to PROTEIN+NO directory.  | point.pdb を PROTEIN+NO ディレクトリに作成できない。                         |
| 18 | Usage.           make_pre_point.pl<br>[protein_point.inp].<br>protein_point.inp will be<br>created by user and file name is<br>your disposal. | 引数がない場合に Usage を表示する。   |

#### 4.1.4 処理フロー

処理フローを図 4.1.4-1 に示す。

```

・蛋白質制御ファイル読込
・蛋白質名保存
・原子名保存
  蛋白質名の数 > 0
|  protein/蛋白質名ディレクトリがある
| |  ・create_pre_point 標準入力用ファイルを作成する
| |  ・protein/蛋白質名ディレクトリで create_pre_point 実行
|
|  ・all_point.pdb 読込
|  ポイント数 > 0
| |  ・(protein/蛋白質名+ポイント番号)ディレクトリを作成
| |  ・(protein/蛋白質名+ポイント番号)ディレクトリに Pro_md.pdb と Pro.tpl の
| |  シンボリックリンクを作成
| |  ・(protein/蛋白質名+ポイント番号)ディレクトリに point.pdb を作成する。
| |  ・list/protein.list に(protein/蛋白質名+ポイント番号)を追記する。
| |  ・ポイント数      ポイント数 - 1
|
|  ・蛋白質名の数      蛋白質名の数  ? 1
|  ・all_point.pdb を削除する

```

図 4.1.4-1 make\_pre\_point.pl の処理フロー

## 4.2 create\_pre\_point.f90

### 4.2.1 機能

create\_pre\_point.f90 は、Pro\_md.pdb の蛋白質主鎖にあたる ATOM 行の内、指定した原子の行を選択し、互いに一定距離以上離れた原子群を all\_point.pdb として標準出力する。原子と原子間距離は外部プログラム make\_pro\_point.pl により指定される。

### 4.2.2 標準入力

create\_pre\_point.f90 の標準入力を表 4.2.2-1 に示す。

表 4.2.2-1 create\_pre\_point.f90 の標準入力

| # | 引数名        | 説明               |
|---|------------|------------------|
| 1 | Pro_md.pdb | 蛋白質 PDB ファイル     |
| 2 | 原子名        | 元素記号で指定する(C、N、O) |



|   |         |  |
|---|---------|--|
| 3 | 原子間距離 1 | 指定した原子周りに存在する原子数を数えるときに設定する、最大原子間距離<br>この値より離れた原子は数えない |
| 4 | 原子間距離 2 | 指定した原子同士の距離<br>この値より近い原子は採用しない                         |

#### 4.2.3 出力

all\_point.pdb ファイルを出力する。

#### 4.2.4 サブルーチン

read\_pdb

処理:

Pro\_md.pdb から全 ATOM 行を読み込み、配列に格納する。さらに、指定した原子の座標を別の配列に格納する。

呼び出し形式:

```
read_pdb(fileName, allAtomNum, atomName, pointAtomNum, allCoord, pointCoord)
```

引数:

```
character*133, intent(in)::fileName           ! pdb ファイル名(Pro_md.pdb)
integer, intent(out)::allAtomNum             ! 全原子数
character*1, intent(in)::atomName           ! 指定した原子(元素記号)
integer, intent(out)::pointAtomNum          ! 指定した原子の原子数
real*4, dimension(3,*), intent(out)::allCoord ! 全原子の座標の配列
real*4, dimension(3,*), intent(out)::pointCoord ! 指定した原子の座標の配列
```

戻り値:

なし

count\_surrounded\_atom

処理:

指定した原子を中心に、一定範囲内に存在する原子数を数える。

呼び出し形式:

```
count_surrounded_atom(allAtomNum, pointAtomNum, allCoord, pointCoord, &
                      countAtom, distance1)
```

引数:

```
integer, intent(in)::allAtomNum             ! 全原子数
integer, intent(in)::pointAtomNum          ! 指定した原子の原子数
real*4, dimension(3,*), intent(in)::allCoord ! 全原子の座標の配列
real*4, dimension(3,*), intent(in)::pointCoord ! 指定した原子の座標配列
integer, dimension(*), intent(out)::countAtom ! 指定範囲内の原子数の配列
real*4, intent(in)::distance                ! 指定した原子同士の距離
```

戻り値:

なし

select\_conform\_point

処理：

指定した原子同士の距離が、一定距離以上離れている原子を配列に格納する。

呼び出し形式：

```
select_conform_point(pointAtomNum, pointCoord, countAtom, probeNum,&
                      probeCoord)
```

引数：

|   |                                     |
|---|-------------------------------------|
| integer, intent(in)::pointAtomNum               | ！ 指定した原子の原子数                        |
| real*4, dimension(3,*), intent(in)::pointCoord  | ！ 指定した原子の座標配列                       |
| integer, dimension(*), intent(in)::countAtom    | ！ 指定範囲内の原子数の配列                      |
| integer, intent(out)::probeNum                  | ！ 指定した原子同士で一定距離<br>離れている原子数         |
| real*4, dimension(3,*), intent(out)::probeCoord | ！ 指定した原子同士で一定<br>距離<br>離れている原子の座標配列 |
| real*4, intent(in)::distance                    | ！ 指定した原子同士の距離                       |

戻り値：

なし

#### 4.2.5 処理フロー

処理フローを図 4.2.5-1 に示す。

- 標準入力読込
- read\_pdb()
- count\_surrounded\_atom()
- select\_conform\_point()
- all\_point.pdb 出力

図 4.2.5-1 create\_pre\_point.f90 の処理フロー

#### 4.3 make\_docking\_socre2.csh

##### 4.3.1 機能

make\_docking\_score\_ps.csh は、既存の make\_docking\_score.csh と同様、ドッキングジョブを投入するためのスクリプトであるが、内部プログラムに対し、蛋白質に対する全化合物のドッキングポーズを保存するための処理として、マルチスレッドプログラムを実行

するように修正を行っている。マルチスレッドプログラムは、sievgene と、sievgene が出力する 1 化合物分のドッキングポーズのファイルを別ファイルに追記するためのプロセスを別々のスレッドとして実行する。2 つのスレッドが終了することにより、マルチスレッドプログラムは終了する。ここでは、make\_docking\_score.csh から make\_docking\_score\_ps.csh への修正箇所について記述する。

#### 4.3.2 変更箇所

make\_docking\_score.csh を、以下の行について修正する。

```
set CMD = "../../../bin/RUN_docking.pl ../../../list/$LIGAND_LIST"
```

修正前の make\_docking\_score.csh(47 行目)

```
set CMD = "../../../bin/RUN_docking_ps.pl ../../../list/$LIGAND_LIST"
```

修正後(make\_docking\_score\_ps.csh(47 行目))

#### 4.3.3 引数

引数を表 4.3.3-1 に示す。

表 4.3.3-1 引数一覧

| # | ファイル名  | 説明                   |
|---|--------|----------------------|
| 1 | 蛋白質リスト | ドッキング対象の蛋白質名のリスト     |
| 2 | 化合物リスト | ドッキング対象の化合物ファイル名のリスト |

#### 4.3.4 出力

sievgene の実行ディレクトリに pose.pdb を出力する。

pose.pdb のフォーマット例を以下に示す。sievgene が出力する 1 化合物のドッキングポーズを複数追記したものとなっている。

```
REMARK TimTec-ST074559.mol2 1 of 1
REMARK SCORE = -157.5265
REMARK dG-SCORE = -2.9107
REMARK HIT-OPTIMIZED SCORE = -93.8035
REMARK MTS SCORE = -197.4631
REMARK ASA = -95.3750
REMARK ELE = -13.8821
REMARK HYD = -14.7741
REMARK VDW = 0.0000
```

| REMARK   |        |                       |     | RMSD = | 81.8508   |         |         |        |
|----------|--------|-----------------------|-----|--------|-----------|---------|---------|--------|
| ATOM     | 1      | C                     | LGD | 1      | -55.204   | -19.368 | 43.976  |        |
| ATOM     | 2      | O                     | LGD | 1      | -54.652   | -18.328 | 43.537  |        |
| ATOM     | 3      | O                     | LGD | 1      | -54.675   | -20.488 | 43.790  |        |
| ATOM     | 4      | C                     | LGD | 1      | -56.064   | -19.238 | 45.177  |        |
| ATOM     | 5      | H                     | LGD | 1      | -56.891   | -18.540 | 44.967  |        |
| ATOM     | 6      | N                     | LGD | 1      | -56.579   | -20.627 | 45.206  |        |
| ATOM     | 7      | H                     | LGD | 1      | -55.751   | -21.241 | 45.232  |        |
| ATOM     | 8      | H                     | LGD | 1      | -57.048   | -20.789 | 44.304  |        |
| ATOM     | 9      | H                     | LGD | 1      | -57.203   | -20.778 | 46.006  |        |
| ATOM     | 10     | C                     | LGD | 1      | -55.301   | -18.837 | 46.451  |        |
| ATOM     | 11     | H                     | LGD | 1      | -54.915   | -17.825 | 46.323  |        |
| ATOM     | 12     | H                     | LGD | 1      | -54.441   | -19.493 | 46.590  |        |
| ATOM     | 13     | C                     | LGD | 1      | -56.164   | -18.876 | 47.697  |        |
| ATOM     | 14     | C                     | LGD | 1      | -55.995   | -19.893 | 48.646  |        |
| ATOM     | 15     | H                     | LGD | 1      | -55.236   | -20.649 | 48.507  |        |
| ATOM     | 16     | C                     | LGD | 1      | -56.792   | -19.925 | 49.793  |        |
| ATOM     | 17     | H                     | LGD | 1      | -56.661   | -20.710 | 50.524  |        |
| ATOM     | 18     | C                     | LGD | 1      | -57.758   | -18.937 | 49.986  |        |
| ATOM     | 19     | Cl                    | LGD | 1      | -58.741   | -18.975 | 51.398  |        |
| ATOM     | 20     | C                     | LGD | 1      | -57.937   | -17.918 | 49.048  |        |
| ATOM     | 21     | H                     | LGD | 1      | -58.688   | -17.157 | 49.206  |        |
| ATOM     | 22     | C                     | LGD | 1      | -57.138   | -17.890 | 47.903  |        |
| ATOM     | 23     | H                     | LGD | 1      | -57.276   | -17.101 | 47.177  |        |
| ENDMDL   |        |                       |     |        |           |         |         |        |
| 4.3.4.1  | REMARK | Princeton-OSSK_52780  |     |        | 1         | of      | 1       |        |
| 4.3.4.2  | REMARK | SCORE =               |     |        | -203.7131 |         |         |        |
| 4.3.4.3  | REMARK | dG-SCORE =            |     |        | -4.7042   |         |         |        |
| 4.3.4.4  | REMARK | HIT-OPTIMIZED SCORE = |     |        | -127.8900 |         |         |        |
| 4.3.4.5  | REMARK | MTS SCORE =           |     |        | -320.9366 |         |         |        |
| 4.3.4.6  | REMARK | ASA =                 |     |        | -176.0026 |         |         |        |
| 4.3.4.7  | REMARK | ELE =                 |     |        | -4.0711   |         |         |        |
| 4.3.4.8  | REMARK | HYD =                 |     |        | -11.0711  |         |         |        |
| 4.3.4.9  | REMARK | VDW =                 |     |        | -1.3975   |         |         |        |
| 4.3.4.10 | REMARK |                       |     |        | RMSD =    | 68.2562 |         |        |
| 4.3.4.11 | ATOM   | 1                     | C   | LGD    | 1         | -44.390 | -17.986 | 43.377 |
| 4.3.4.12 | ATOM   | 2                     | O   | LGD    | 1         | -44.868 | -17.069 | 44.081 |
| 4.3.4.13 | ATOM   | 3                     | O   | LGD    | 1         | -43.752 | -17.790 | 42.323 |
| 4.3.4.14 | ATOM   | 4                     | C   | LGD    | 1         | -44.511 | -19.391 | 43.821 |
| 4.3.4.15 | ATOM   | 5                     | H   | LGD    | 1         | -43.525 | -19.868 | 43.706 |
| 4.3.4.16 | ATOM   | 6                     | N   | LGD    | 1         | -44.759 | -19.036 | 45.242 |
| 4.3.4.17 | ATOM   | 7                     | H   | LGD    | 1         | -43.934 | -18.508 | 45.560 |
| 4.3.4.18 | ATOM   | 8                     | H   | LGD    | 1         | -44.774 | -19.923 | 45.763 |
| 4.3.4.19 | ATOM   | 9                     | H   | LGD    | 1         | -45.616 | -18.498 | 45.357 |
| 4.3.4.20 | ATOM   | 10                    | C   | LGD    | 1         | -45.586 | -20.206 | 43.089 |
| 4.3.4.21 | ATOM   | 11                    | C   | LGD    | 1         | -46.497 | -21.085 | 43.754 |
| 4.3.4.22 | ATOM   | 12                    | H   | LGD    | 1         | -46.450 | -21.233 | 44.821 |
| 4.3.4.23 | ATOM   | 13                    | C   | LGD    | 1         | -47.519 | -21.748 | 43.024 |

|          |      |    |   |     |   |         |         |        |
|----------|------|----|---|-----|---|---------|---------|--------|
| 4.3.4.24 | ATOM | 14 | H | LGD | 1 | -48.234 | -22.371 | 43.538 |
| 4.3.4.25 | ATOM | 15 | C | LGD | 1 | -47.607 | -21.586 | 41.617 |
| 4.3.4.26 | ATOM | 16 | H | LGD | 1 | -48.391 | -22.114 | 41.097 |
| 4.3.4.27 | ATOM | 17 | C | LGD | 1 | -46.685 | -20.747 | 40.919 |
| 4.3.4.28 | ATOM | 18 | C | LGD | 1 | -45.675 | -20.101 | 41.674 |
| 4.3.4.29 | ATOM | 19 | H | LGD | 1 | -44.955 | -19.508 | 41.165 |
| 4.3.4.30 | ATOM | 20 | O | LGD | 1 | -46.694 | -20.490 | 39.547 |
| 4.3.4.31 | ATOM | 21 | C | LGD | 1 | -47.742 | -21.079 | 38.741 |
| 4.3.4.32 | ATOM | 22 | H | LGD | 1 | -48.701 | -20.708 | 39.075 |
| 4.3.4.33 | ATOM | 23 | H | LGD | 1 | -47.716 | -22.166 | 38.837 |
| 4.3.4.34 | ATOM | 24 | C | LGD | 1 | -47.553 | -20.677 | 37.287 |
| 4.3.4.35 | ATOM | 25 | H | LGD | 1 | -46.602 | -21.056 | 36.916 |
| 4.3.4.36 | ATOM | 26 | H | LGD | 1 | -47.574 | -19.593 | 37.200 |
| 4.3.4.37 | ATOM | 27 | H | LGD | 1 | -48.366 | -21.101 | 36.694 |
| ENDMDL   |      |    |   |     |   |         |         |        |
| …(以下略)…  |      |    |   |     |   |         |         |        |

#### 4.4 RUN\_docking\_ps.pl

##### 4.4.1 機能

make\_docking\_score\_ps.csh から呼び出される。RUN\_docking\_ps.pl は、RUN\_docking.pl とほぼ同様の動作を行うが、sievgene をバックグラウンドで実行してから、ドッキングポーズファイル pose.pdb の出力を一時ファイルにリダイレクトするプロセスを実行するように RUN\_docking.pl の以下の箇所を変更する。一時ファイルは、sievgene が終了した後に pose.pdb にリネームされる。

##### 4.4.2 変更箇所

RUN\_docking.pl の、以下の行を変更する。

```
my $NAMECO      = "";
my $NAMEESC     = "../.../result/";
my $COORDI     = "PDB"
```

修正前の RUN\_docking.pl (40 ~ 42 行目)

```
my $NAMECO      = "";
my $NAMEESC     = "../.../result/";
my $COORDI     = "PDB"
my $pose_file   = "pose.pdb"; # 追加行
my $tmp_pose_file = "tmp_pose.pdb"; # 追加行
```

修正後の RUN\_docking\_ps.pl

```
#
# run sievgene
```

```
#
system("$cmd_sievgene < $control_file >> $log_file");
system("cat $NAMEESC >> $score_file");
unlink $NAMEESC;
```

修正前の RUN\_docking.pl (142 ~ 147 行目)

```
#
# run sievgene
#
system("mkfifo $pose_file"); # 追加行
system("$cmd_sievgene < $control_file >> $log_file &");
system("cat $pose_file | perl -ane 'print ¥"¥@F¥¥n¥";' >> $tmp_pose_file");#追加
行
unlink($pose_file); # 追加行
system("cat $NAMEESC >> $score_file");
unlink $NAMEESC;
```

修正後の RUN\_docking\_ps.pl

```
}
system("gzip -f $log_file");
close(CTRL_SKEL);
close(LIGAND_LIST);
```

修正前の RUN\_docking.pl (157 ~ 161 行目)

```
}
rename($tmp_pose_file, $pose_file); # 追加行

system("gzip -f $log_file");
close(CTRL_SKEL);
close(LIGAND_LIST);
```

修正後の RUN\_docking\_ps.pl

#### 4.4.3 引数

引数を表 4.4.3-1 に示す。

表 4.4.3-1 RUN\_docking\_ps.pl の引数一覧

| # | ファイル名  | 説明             |
|---|--------|----------------|
| 1 | 低分子リスト | ドッキング対象の低分子リスト |

## 4.5 make\_score\_data\_ps.pl

### 4.5.1 機能

make\_score\_data\_ps.pl は、相互作用行列を作成する。make\_score\_data\_ps.pl は、make\_score\_data.csh とほぼ同様の動作を行うが、出力する相互作用行列ファイルの先頭行に化合物グループ名を出力する点異なる。

### 4.5.2 変更箇所

make\_score\_data.csh の、以下の行を変更する。

```
#
# 4) loop by protein list
#
foreach protein_name (`cat $LIST_DIR/$PROTEIN_LIST`)
  echo $protein_name" data making"
```

修正前の make\_score\_data.csh(61 ~ 65 行目)

```
#
# 4) loop by protein list
#
echo "LG $LIGAND_LIST" >> $score_data_file          # 追加行
foreach protein_name (`cat $LIST_DIR/$PROTEIN_LIST`)
  echo $protein_name" data making"
```

修正後の make\_score\_data\_ps.pl

### 4.5.3 引数

引数を、表 4.5.3-1 に示す。

表 4.5.3-1 make\_score\_data\_ps.pl の引数一覧

| # | ファイル名  | 説明                   |
|---|--------|----------------------|
| 1 | 蛋白質リスト | ドッキング対象の蛋白質名のリスト     |
| 2 | 化合物リスト | ドッキング対象の化合物ファイル名のリスト |

## 4.6 make\_score\_data\_multi\_ps.pl

### 4.6.1 機能

make\_score\_data\_multi\_ps.pl は、相互作用行列を作成する。make\_score\_data\_multi\_ps.pl は、make\_score\_data\_multi.pl とほぼ同様の動作を行うが、出力する相互作用行列ファイルの先頭行に化合物グループ名を出力する点異なる。マルチ mol2 ファイルを使用する場合に、make\_score\_data\_ps.pl の代わりに使用する。

#### 4.6.2 変更箇所

make\_score\_data\_multi.pl の、以下の行を変更する。

```
print "matrix -> matrix/$matrix_name\n";  
  
#  
# 3) make protein_part of matrix  
#
```

変更前の make\_score\_data\_multi.pl (93~97 行目)

```
print "matrix -> matrix/$matrix_name\n";  
  
print OUT "LG $lig_file_list\n";      # 追加行  
#  
# 3) make protein_part of matrix  
#
```

変更前の make\_score\_data\_multi\_ps.pl

#### 4.6.3 引数

引数を、表 4.6.3-1 に示す。

表 4.6.3-1 make\_score\_data\_multi\_ps.pl の引数一覧

| # | ファイル名  | 説明                   |
|---|--------|----------------------|
| 1 | 蛋白質リスト | ドッキング対象の蛋白質名のリスト     |
| 2 | 化合物リスト | ドッキング対象の化合物ファイル名のリスト |

#### 4.7 make\_scoring\_grid.pl

##### 4.7.1 機能

make\_scoring\_grid.pl は、作業ディレクトリ prediction ディレクトリを作成する。相互作用行列を読み込み、蛋白質 ID 1 + ポイント番号ごとに、スコアリングされた化合物のリスト(以下、スコアリングリスト)蛋白質\_score.list を作成し、calc\_sigma を起動する。

スコア出力が成功した化合物のスコアのみ出力されるため、相互作用行列に出力された順番で、かつ、スコアリングリストに出力された化合物のスコアのみを読み取るようにする。出力されるファイルは prediction ディレクトリに格納される。

スコアリングリストのフォーマット例を以下に示す。

|       |      |    |                        |
|-------|------|----|------------------------|
| -1.02 | c001 | 1  | スコア 化合物グループ名 出力番号      |
| -2.00 | c001 | 2  |                        |
| NONE  | c001 | -3 | 3 番目の化合物はスコアが出力されなかった) |
| -2.87 | c001 | 4  |                        |



|                          |                             |
|--------------------------|-----------------------------|
| -1.60 c001 +5<br>…(以下略)… | 5番目の化合物は他の蛋白質でスコアが出ない場合があった |
|--------------------------|-----------------------------|

出力番号は、相互作用行列の 1 蛋白質における順番を示している。+ が付与された番号は、他の蛋白質ではスコアが出ない場合があったこと化合物であることを意味する番号である。- が付与された番号は、当該蛋白質とその化合物の組み合わせでスコアが出力されなかったことを意味する番号である。後に、ドッキングポーズファイルから、全蛋白質においてスコアが出力されている化合物のみのドッキングポーズを取得するが、ドッキングポーズファイルには、当該化合物ではスコア出力された場合のポーズも含まれている。そのため、何番目のポーズが不要かを判断するために ± 符号が使われる。

#### 4.7.2 引数

引数を表 4.7.2-1 に示す。

表 4.7.2-1 make\_scoring\_grid.pl の引数一覧

| # | 引数名  | 説明                        |
|---|------|---------------------------|
| 1 | 蛋白質名 | 蛋白質名を指定する<br>ポケット番号は指定しない |

#### 4.7.3 出力

出力を表 4.7.3-1 に示す。

表 4.7.3-1 出力メッセージ一覧

| # | メッセージ                                       | 説明                             |
|---|---|--------------------------------|
| 1 | Error!! matrix_list not found.              | matrix_list がない。               |
| 2 | Error!! cannot open matrix_list.[システムメッセージ] | matrix_list が開けない。             |
| 3 | Error!! lig_grp not found.                  | lig_grp がない。                   |
| 4 | Error!! cannot open lig_grp.[システムメッセージ]     | lig_grp が開けない。                 |
| 5 | Error!! 相互作用行列 format is invalid.           | 相互作用行列ファイルが不正。                 |
| 6 | Error!! cannot open __tmp_control.inp.[!]   | 一時ファイル__tmp_control.inp が開けない。 |
| 7 | Error!! cannot open 蛋白質                     | 蛋白質_score.list を作成できない。        |

|    |   |  |
|----|---|--|
|    | _score.list.[システムメッセージ]                   |  |
| 8  | Error!! calc_sigma not found.             | calc_sigma 実行モジュールが見つからない。                   |
| 9  | Error!! 蛋白質名_sigma.score not found.       | calc_sigma によって作成されたはずの蛋白質名_sigma.score がない。 |
| 10 | Usage. make_scoring_grid.pl [matrix_file] | 引数が指定されなかったときに Usage を表示する。                  |

#### 4.7.4 処理フロー

処理フローを、図 4.7.4-1 に示す。

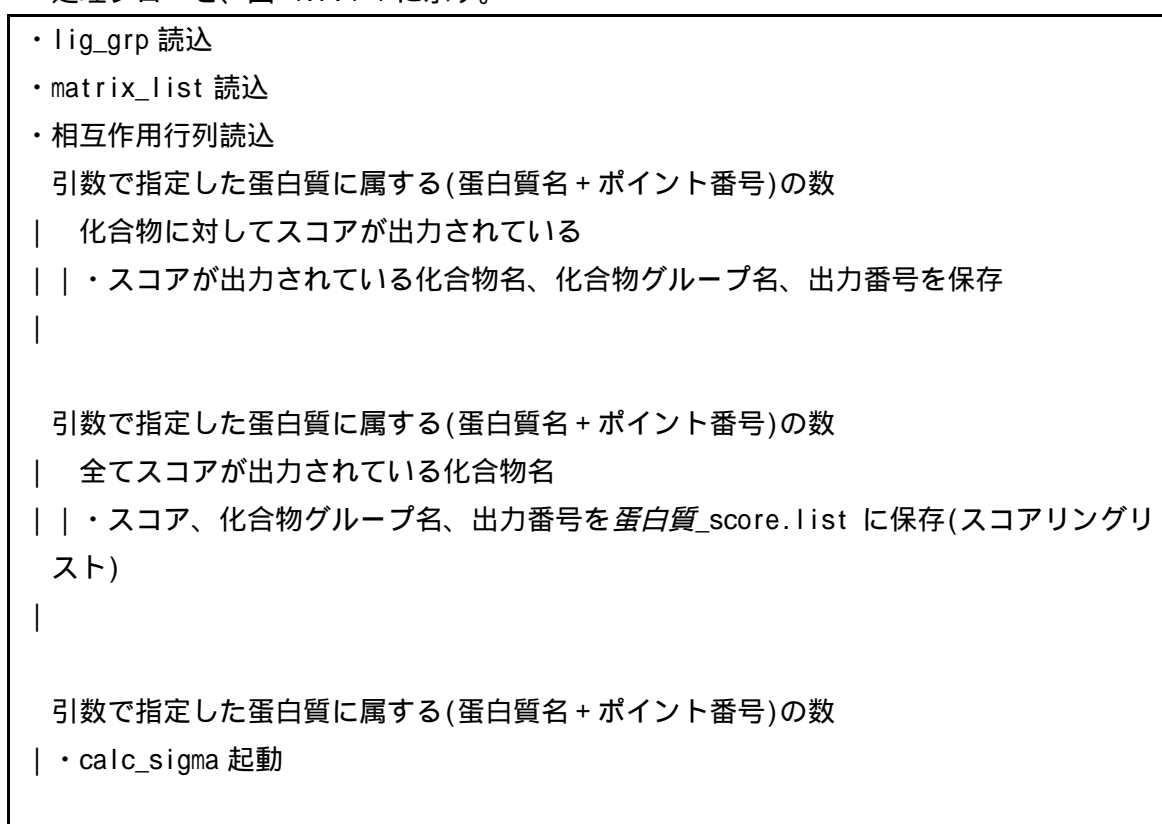


図 4.7.4-1 make\_scoring\_grid.pl の処理フロー

#### 4.8 calc\_sigma.f90

##### 4.8.1 機能

calc\_sigma.f90 は、スコアリングリストを読み込み、各(蛋白質名 + ポケット番号)のスコアリンググリッドの平均値とスコアの標準偏差を求め、結果を蛋白質名\_sigma.score に出力する。

#### 4.8.2 標準入力

標準入力を、表 4.8.2-1 に示す。

表 4.8.2-1 calc\_sigma.f90 の標準入力

| # | 引数名       | 説明                                  |
|---|-----------|-------------------------------------|
| 1 | スコアリングリスト | make_scoring_grid.pl が作成したスコアリングリスト |
| 2 | 標準偏差範囲数   | $\mu + *N$ における N                   |

#### 4.8.3 出力

蛋白質名\_sigma.score ファイルを出力する。

#### 4.8.4 サブルーチン

メインルーチンのみ

#### 4.8.5 処理フロー

処理フローを図 4.8.5-1 に示す。

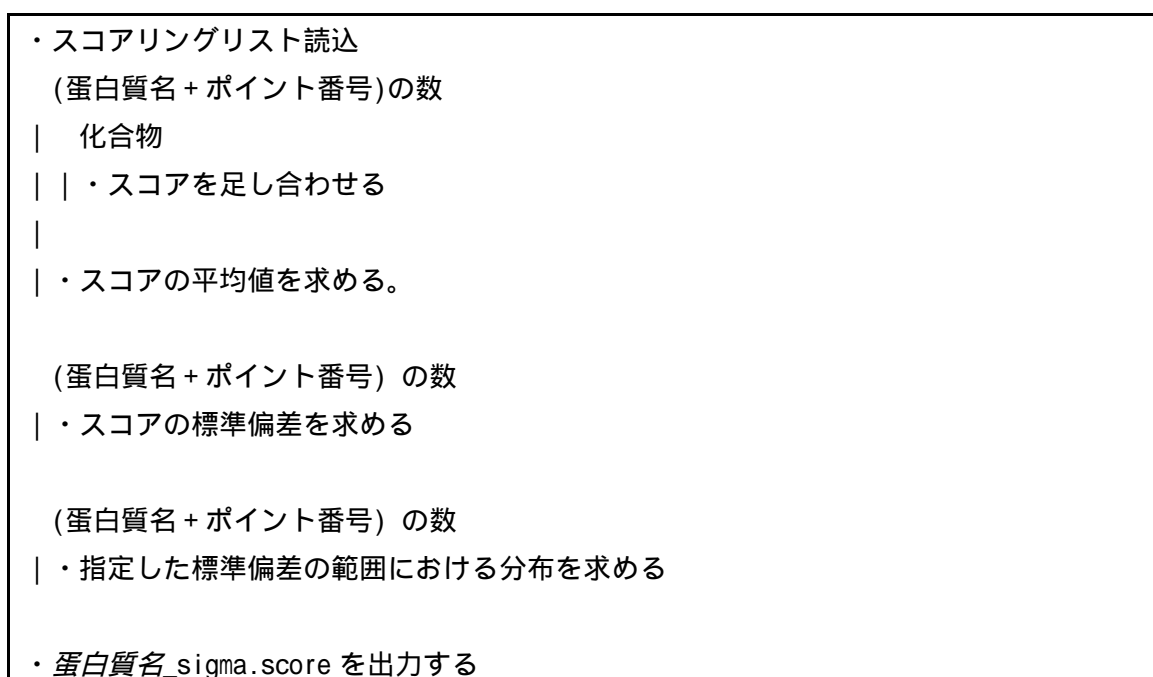


図 4.8.5-1 calc\_sigma.f90 の処理フロー

## 4.9 make\_docking\_pocket\_coordinate.pl

### 4.9.1 機能

make\_docking\_pocket\_coordinate.pl は、蛋白質リストに記述された(蛋白質名+ ポケット番号)について、蛋白質\_score.list を用いてドッキングポーズファイル pose.pdb を(蛋白質名+ ポケット番号) ごとに一つのファイルにまとめ、calc\_average\_coord.f90 を実行して蛋白質名\_ave.coord を出力する。次に、calc\_center\_coord.f90 を呼び出してドッキング中心座標を求める。

蛋白質リストは、蛋白質名\_sigma.score から最良のスコアリンググリッドを選択し、ファイルに記述したものである。複数蛋白質がある場合は、1行ごとに記述する。

### 4.9.2 引数

引数を表 4.9.2-1 に示す。

表 4.9.2-1 make\_docking\_pocket\_coordinate.pl の引数一覧

| # | 引数名    | 説明                      |
|---|--------|-------------------------|
| 1 | 蛋白質リスト | (蛋白質名+ ポケット番号)が記述されたリスト |

### 4.9.3 出力

出力を表 4.9.3-1 に示す。

表 4.9.3-1 make\_docking\_pocket\_coordinate.pl の出力メッセージ一覧

| # | メッセージ   | 説明                                 |
|---|---|------------------------------------|
| 1 | Error!! 蛋白質リスト not found.                               | 蛋白質リストがない。                         |
| 2 | Error!! cannot open 蛋白質リスト.[システムメッセージ]                  | 蛋白質リストが開けない。                       |
| 3 | Error!! calc_average_coord not found.                   | calc_average_coord 実行モジュールが見つからない。 |
| 4 | Error!! cannot execute calc_average_coord. .[システムメッセージ] | calc_average_coord が実行できない。        |
| 5 | Error!! 蛋白質_score.list not found.                       | 蛋白質_score.list がない。                |
| 6 | Error!! cannot open 蛋白質_score.list. .[システムメッセージ]        | 蛋白質_score.list が開けない。              |
| 7 | Error!! 蛋白質-低分子グループ                                     | 蛋白質-低分子グループの pose.pdb がない。         |

|    |   |   |
|----|---|---|
|    | pose.pdb not found.                                     |   |
| 8  | Error!! cannot open 蛋白質-低分子グループ pose.pdb. [システムメッセージ]   | 蛋白質-低分子グループ pose.pdb が開けない。                         |
| 9  | Error!! cannot open 蛋白質-低分子グループ pose1.pdb               | 蛋白質-低分子グループ pose1.pdb が開けない。                        |
| 10 | Error!! cannot open __tmp_control.inp.[\$!]             | 一時ファイル__tmp_control.inpが開けない。                       |
| 11 | Error! 蛋白質名_ave.coord not found.                        | calc_average_coord によって作成されたはずの蛋白質名_ave.coordがない。   |
| 12 | Error!! calc_center_coord not found.                    | calc_center_coord 実行モジュールが見つからない。                   |
| 13 | Error!! cannot execute calc_center_coord. [システムメッセージ]   | calc_center_coord が実行できない。                          |
| 14 | Error!! 蛋白質名_center.coord not found.                    | calc_center_coord によって作成されたはずの蛋白質名_center.coordがない。 |
| 15 | Usage.<br>make_docking_pocket_coordinate.pl<br>[蛋白質リスト] | 引数がなかった場合に出力される Usage。                              |

#### 4.9.4 処理フロー

処理フローを図 4.9.4-1 に示す。

```

・蛋白質リスト読込
  引数で指定した蛋白質に属する(蛋白質名 + ポイント番号)の数
| ・スコアリングリストの化合物グループの pose.pdb を読込
| ・出力番号に該当するポーズを一時ファイル pose1.pdb に書き込み
| ・calc_average_coord 起動
| ・pose1.pdb を削除

蛋白質名
| 蛋白質名_center.coord がある
| | ・calc_center_coord を実行
|

・スコアリングリスト削除

```

図 4.9.4-1 make\_docking\_pocket\_coordinate.pl の処理フロー

#### 4.10 calc\_average\_coord.f90

##### 4.10.1 機能

calc\_average\_coord.f90 は、化合物のドッキングポーズファイル pose.pdb を読み込み、化合物の主鎖となる C、N、O 原子の空間座標の平均値を計算する。計算結果は、*蛋白質名\_ave.coord* に追記する。

##### 4.10.2 標準入力

標準入力を、表 4.10.2-1 に示す。

表 4.10.2-1 calc\_average\_coord.f90 の標準入力

| # | 引数名       | 説明  |
|---|-----------|---|
| 1 | pose1.pdb | make_scoring_grid.pl が指定する化合物のドッキングポーズファイル。 |

##### 4.10.3 出力

*蛋白質名\_ave.coord* を出力する。

##### 4.10.4 サブルーチン

メインルーチンのみ

#### 4.10.5 処理フロー

処理フローを、図 4.10.5-1 に示す。

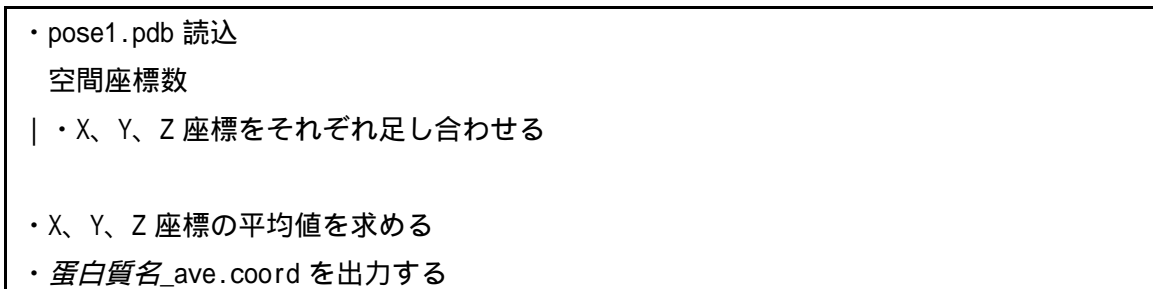


図 4.10.5-1 calc\_average\_coord.f90 の処理フロー

#### 4.11 calc\_center\_coord.f90

##### 4.11.1 機能

calc\_center\_coord.f90 は、*蛋白質名\_ave.coord* を読み込んで、蛋白質にドッキングした化合物全体の空間座標の平均値を計算する。

##### 4.11.2 標準入力

標準入力を表 4.11.2-1 に示す。

表 4.11.2-1 calc\_center\_coord.f90 の標準入力一覧

| # | 引数名                   | 説明                               |
|---|-----------------------|----------------------------------|
| 1 | <i>蛋白質名_ave.coord</i> | calc_average_coord.f90 が出力したファイル |

##### 4.11.3 出力

*蛋白質名\_center.coord* ファイルを出力する。

##### 4.11.4 サブルーチン

メインルーチンのみ

##### 4.11.5 処理フロー

処理フローを図 4.11.5-1 に示す。

- ・ 蛋白質名\_ave.coord 読込
- ・ 平均座標計算
- ・ 標準偏差計算
- ・ 結果を蛋白質名\_center.coord に出力する

図 4.11.5-1 calc\_center\_coord.f90 の処理フロー

## 4.12 verify\_point.pl

### 4.12.1 機能

verify\_point.pl は、既知のドッキングポケットが判明している場合に、make\_docking\_pocket\_coordinate.pl で求めた予測ドッキングポケット座標と、既知のドッキングポケットの距離の差を求め、求めた予測ドッキングポケット座標の検証を行う。

### 4.12.2 引数

引数を表 4.12.2-1 に示す。

表 4.12.2-1 verify\_point.pl の引数一覧

| # | ファイル名             | 説明                    |
|---|-------------------|-----------------------|
| 1 | PDB ファイル          | 既知のドッキングポケットのポイントファイル |
| 2 | 蛋白質名_center.coord | 予測ドッキングポケットの計算結果ファイル  |

### 4.12.3 出力

出力を表 4.12.3-1 に示す。

表 4.12.3-1 verify\_point.pl の出力メッセージ一覧

|   | メッセージ  | 説明                                |
|---|--|-----------------------------------|
| 1 | Usage: verify_point.pl FILE1 FILE2               | 引数不正の場合に出力される Usage               |
| 2 | Error!! PDB ファイル not found.                      | PDB ファイルがない                       |
| 3 | Error!! 蛋白質名_center.coord not found.             | 蛋白質名_center.coord がない             |
| 4 | Error!! ../bin/calc_verification_data not found. | ../bin/calc_verification_data がない |
| 5 | cannot open __tmp_control.inp. [システムメッセージ]       | 一時ファイル__tmp_control.inp が開けない     |
| 6 | cannot open _known_ligand.pdb. [システムメッセージ]       | 一時ファイル_known_ligand.pdb が開けない     |



|   |  |                                       |
|---|--|---------------------------------------|
| 7 | cannot open PDB ファイル.[システムメッセージ]                       | 既知ポイントファイルが開けない                       |
| 8 | Error!! cannot esexecute ../bin/calc_verification_data | ../bin/calc_verification_data が実行できない |
| 9 | Error!! 蛋白質.verify not found.                          | 蛋白質.verify がない                        |

#### 4.12.4 処理フロー

処理フローを図 4.12.4-1 に示す。

- ・既知 PDB ファイル読込
- ・既知 PDB ファイルの ATOM 行を一時ファイルに書き出す
- ・calc\_verification\_data を呼び出す

図 4.12.4-1 verify\_point.pl の処理フロー

#### 4.13 calc\_verification\_data.f90

##### 4.13.1 機能

calc\_verification\_data.f90 は、既知のドッキングポケット PDB ファイルと蛋白質名\_center.coord を読み込んで、蛋白質名\_center.coord で得られた予測ドッキングポケット座標と既知のポケット座標の差を計算する。

##### 4.13.2 標準入力

標準入力を表 4.13.2-1 に示す。

表 4.13.2-1 calc\_verification\_data.f90 の標準入力一覧

| # | 引数名               | 説明                              |
|---|-------------------|---------------------------------|
| 1 | PDB ファイル          | 蛋白質の既知ドッキングポケットファイル             |
| 1 | 蛋白質名_center.coord | calc_center_coord.f90 が出力したファイル |

##### 4.13.3 出力

蛋白質名.verify ファイルを出力する。

##### 4.13.4 サブルーチン

メインルーチンのみ

##### 4.13.5 処理フロー

処理フローを図 4.13.5-1 に示す。

- PDB ファイル読込
- 蛋白質名\_center.coord ファイル読込
- PDB ファイルの平均座標計算
- PDB ファイルの平均座標と蛋白質名\_center.coord ファイルの座標の最短距離を計算
- PDB ファイルの各原子と蛋白質名\_center.coord ファイルの座標の最短距離を計算
- 蛋白質.verify を出力する

図 4.13.5-1 calc\_verification\_data.f90 の処理フロー