About This Document

This document accompanies the “myPresto 4.2 USER MANUAL.” Copyrights, program licensing conditions, the authors, and references are as stated in the “myPresto 4.2 USER MANUAL.”

Acknowledgements

This work was supported by grants from New Energy and Industrial Technology Development Organization (NEDO) and Ministry of Economy, Trade and Industry (METI), JAPAN.

This software was developed as part of a research project advanced by the late Dr. Yoshimasa Kyogoku.
Table of Contents

1  Introduction .................................................. 6

2  System Structure ............................................ 7
   2.1  Overall structure ........................................ 7
   2.1.1  Interaction matrix preparation phase ................ 8
   2.1.2  Interaction matrix analysis phase ................... 9
   2.2  Directory structure ....................................... 10

3  Interaction Matrix Preparation Tools .......................... 14
   3.1  Grid data generation shell ................................ 14
   3.2  Docking job start shell ................................... 17
   3.3  Docking execution script .................................. 26
   3.4  Docking score totaling shell .............................. 27
   3.5  Score merge data conversion program .................... 29

4  Screening Tools ........................................... 31
   4.1  MTS Screening Program ...................................... 31
   4.2  MTS screening job start shell ............................ 34
   4.3  DSI screening program ...................................... 35
   4.4  DSI screening job start cell ............................... 37

5  In silico Screening Sample ................................. 38
   5.1  Overview of screening work ............................... 38
   5.2  Description of work ....................................... 41
   5.3  Sample work conditions and environment preparation .... 41
       5.3.1  Work conditions .................................... 41
       5.3.2  Preparation of screening environment ............... 42
       5.3.3  Preparation of screening tools ..................... 42
   5.4  Interaction matrix data file preparation ................ 43
   5.5  Screening work ........................................... 48
Screening by the MTS method

........................................ 53
1 Introduction

The in silico screening component of the myPresto system stores docking scores from the docking engine (sievgene) and creates interaction matrix data. Referencing the interaction matrix data, in silico performs screening using proteins and low-molecular compounds.

This system performs docking calculations of combinations of proteins and low-molecular compounds, and stores the data. Considering the number of combinations that are possible using 182 types of proteins and 2 million low molecules, docking must be performed more than 182 x 2 million = 360 million times. For this reason, we examined our data generation and totaling techniques, and improved work efficiency by automating the majority of the data storage tasks. We have also automated the screening process to allow users to perform screening using any desired protein and low molecule.

Screening methods used are MTS (Multiple Target Screening) and DSI (Docking Score Index).

- **MTS**: For a target protein, this method examines the order of docking scores of low-molecular compounds and selects compounds that rank high with that target protein.

- **DSI**: For each low-molecular compound, this method examines the patterns of the protein docking scores and selects low molecules whose patterns resemble the docking score pattern of the specified low-molecular compound.


在定義確保するためには、各々のデータについて保存されるとどうなるか、どのようにしてデータを管理するのか、システム全体のデータ構造を理解しておくことが重要です。そのため、各データの詳細、管理方法、システム全体のデータ構造についての理解が必要です。
## System Structure

### 2.1 Overall structure

The in silico screening component of the myPresto system consists of the "interaction matrix preparation phase", which creates and updates (adds, deletes) interaction matrices, and the "interaction matrix analysis phase", which analyzes the matrices and performs screening.

**Interaction matrix preparation phase**
- Preparation of target proteins and low-molecular compounds
- Docking by means of the docking engine (sievgene) (score calculation)
- Creation of new matrices from the docking scores
- Addition to existing interaction matrices

**Interaction matrix analysis phase**
- Screening by MTS (Multiple Target Screening) method
- Screening by DSI (Docking Score Index) method

![Figure 2.1. Overview of the in silico screening system](image)

Create interaction matrix

<table>
<thead>
<tr>
<th>Interaction matrix</th>
<th>Analyze interaction matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>Results of analysis</td>
<td></td>
</tr>
</tbody>
</table>

- Protein
- Low-molecular compound data
Interaction matrix preparation phase

This phase prepares the proteins and ligands, and performs docking calculation. The purpose of this phase is to perform docking of the target proteins and target low molecules using the docking engine (sievgene), total the output docking scores, and obtain a matrix for protein - low molecule interaction.

The structure of the interaction matrix preparation phase is shown in Fig. 2.2.

The interaction matrix preparation phase is performed in the following order:

1. Prepare protein data and ligand data for docking.
2. Start the shell that creates the sievgene grid file from the protein data. (To increase calculation speed, this is created as a grid file that indicates the potential at the docking parts of each protein.)
3. Prepare the protein data list and ligand data list for docking.
4. Start the shell that inputs the docking calculations into the cluster machine.
5. Start the shell that totals the docking score file.

Figure 2.2. Structure of interaction matrix preparation phase

Combination data

Results

PDB, topology, pocket information, mol2 data, etc.

Docking

Docking

Docking

Docking

Docking

Docking

Docking job in

Automated

Data

preparation

Arrangement

Preparation,
In this phase, the results of the interaction matrix preparation phase are merged with the interaction matrix database prepared in the system, and screening is performed.

"LigandBOX", the compound database in the myPresto system, contains the interaction matrices of 182 proteins x 1 million ligands from the 2004 version of the compound database, and the interaction matrices of 182 proteins x 2 million ligands from the 2005 version of the database. This compound data is compiled into units of 10,000 compounds each so that screening is performed using data in units of 182 proteins x 10,000 compounds. The data of 182 proteins x 10,000 compounds can also be merged with data prepared by the user to perform screening with added compounds. By repeating this process 100 times, the results of screening with 1 million compounds can be obtained. By repeating this process 200 times, the results of screening with 2 million compounds can be obtained.

Figure 2.1. Structure of interaction matrix analysis phase

The interaction matrix analysis phase is performed in the following order.

‡@ Prepare the data created in the interaction matrix preparation phase and the interaction matrix data in the system.
‡A Configure screening settings.
‡B Start the shell that inputs the screening job.
‡C Analyze the output results.
The directory structure of the in silico screening system of the myPresto system is shown in Fig. 2.4.

<table>
<thead>
<tr>
<th>Directory</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Home directory]</td>
<td></td>
</tr>
<tr>
<td>&quot;¥ protein/</td>
<td>Protein data directory</td>
</tr>
<tr>
<td>&quot;¥ [Protein ID]/</td>
<td>Protein directories (multiple)</td>
</tr>
<tr>
<td>&quot;¥ Pro.pdb</td>
<td>Protein PDB file</td>
</tr>
<tr>
<td>&quot;¥ Pro.tpl</td>
<td>Topology file (ASCII format)</td>
</tr>
<tr>
<td>&quot;¥ ProB.tpl</td>
<td>Topology file (binary format)</td>
</tr>
<tr>
<td>&quot;¥ point.pdb</td>
<td>Pocket file</td>
</tr>
<tr>
<td>&quot;¥ ligand/</td>
<td>Mlow molecular data directory</td>
</tr>
<tr>
<td>&quot;¥ [Low molecule group name]/</td>
<td>Low molecule group directories (multiple)</td>
</tr>
<tr>
<td>&quot;¥ [Low molecule name].mol2</td>
<td>Low molecule data files (multiple)</td>
</tr>
<tr>
<td>&quot;¥ v2005</td>
<td>Version directory</td>
</tr>
<tr>
<td>&quot;¥ bin/</td>
<td>Execution program directory</td>
</tr>
<tr>
<td>&quot;¥ [sievgene execution program]</td>
<td></td>
</tr>
<tr>
<td>&quot;¥ make_grid.csh</td>
<td>Grid generation shell</td>
</tr>
<tr>
<td>&quot;¥ make_docking_score.csh</td>
<td>Docking job start shell</td>
</tr>
<tr>
<td>&quot;¥ make_score_data.csh</td>
<td>Docking score totaling shell</td>
</tr>
<tr>
<td>&quot;¥ Run_docking.pl</td>
<td>Docking execution script</td>
</tr>
<tr>
<td>&quot;¥ DataMaker</td>
<td>Score extraction program</td>
</tr>
<tr>
<td>&quot;¥ input/</td>
<td>Control file directory</td>
</tr>
<tr>
<td>&quot;¥ sievgene control file template</td>
<td></td>
</tr>
<tr>
<td>&quot;¥ Job registration shell template</td>
<td></td>
</tr>
<tr>
<td>&quot;¥ list/</td>
<td>List file directory</td>
</tr>
<tr>
<td>&quot;¥ grid/</td>
<td>Grid file directory</td>
</tr>
<tr>
<td>&quot;¥ [Protein name]/</td>
<td>Protein directory (multiple)</td>
</tr>
<tr>
<td>&quot;¥ grid.file</td>
<td>Grid file</td>
</tr>
<tr>
<td>&quot;¥ work/</td>
<td>Work directory</td>
</tr>
<tr>
<td>&quot;¥ [Protein name]/</td>
<td>Directories for each protein (multiple)</td>
</tr>
<tr>
<td>&quot;¥ [Low molecule group name]/</td>
<td>Low molecule group directory (multiple)</td>
</tr>
<tr>
<td>&quot;¥ Work file</td>
<td></td>
</tr>
<tr>
<td>&quot;¥ result/</td>
<td>Execution result output directory</td>
</tr>
<tr>
<td>&quot;¥ [Protein name]/</td>
<td>Directories for each protein (multiple)</td>
</tr>
<tr>
<td>&quot;¥ [Low molecule group name]/</td>
<td>Low molecule group directory (multiple)</td>
</tr>
<tr>
<td>&quot;¥ [Low molecule group name].scores</td>
<td>Docking score merge file</td>
</tr>
<tr>
<td>&quot;¥ score/</td>
<td>Score totaling directory</td>
</tr>
<tr>
<td>&quot;¥ [Score totaling file name].dat</td>
<td>Interaction matrix data file</td>
</tr>
</tbody>
</table>
Figure 2.4. Screening system directory structure
1) Protein data directory
Protein data directory is where the protein data is stored, and the name of the directory is "protein". Each sub-directory contains the following types of data:

- PDB file of protein (Pro.pdb)
- Coordinate data of protein
- ASCII-format topology file (Pro.tpl)
- Binary-format topology file (ProB.tpl)
- Pocket file (point.pdb)
- Coordinate data of docking target

2) Ligand molecular data directory
This directory contains low molecule data, and has the name "ligand". Low molecules are grouped into units of 10,000 molecules, and each has a directory with the group name. The file format of the low molecule data is mol2.

3) Version directory
A version directory is created for each docking version, and the initial version is "v1". The shell that executes each phase of the in silico screening system is executed in this directory.

The structure of the version directory is as follows:

- "bin": Directory for execution programs
  - Sievgene and in silico screening shells are placed here.
- "input": Control file directory
  - Sievgene control file template and job registration shell template are placed here.
- "list": Directory for list files
  - Lists of the docking target proteins and low molecule groups are placed here.
- "grid": Directory for grid files
  - This directory contains grid files. When the grid data generation shell "make_grid.csh" is started, this directory is automatically created.
- "work": Work directory
  - Temporary files and logs are output to this directory. When the docking job input shell "make_docking_score.csh" is started, this directory is automatically created.
Execution result output directory ("result")

The score file containing the results of docking is output to this directory. When the docking job input shell "make_docking_score.csh" is started, this directory is automatically created.

The result directory also contains the "Protein name" directory indicating the target protein of docking, and the "Low molecule group name" directory indicating the target low molecule group of docking. The docking score files of the target protein and the target low molecule are respectively output to these directories.

Score totaling directory ("score")

The docking result is totaled and output as the interaction matrix data file to this directory. When the docking score totaling shell "make_score_data.csh" is started, this directory is automatically created.

Note: The content provided is a natural language representation of the document. The text has been translated and reformatted for clarity and readability. The original source language is Korean.
3 Interaction Matrix Preparation Tools

The following five shells or programs are used to prepare the interaction matrix.

‡@ Grid data generation shell (make_grid.sh) in Figure 2.2 is automated.

‡A Docking job start shell (make_docking_score.csh) in Figure 2.2 is automated.

‡B Docking execution script (Run_docking.pl) This is called from make_docking_score.csh, and it executes sievgene.

‡C Docking score totaling shell (make_score_data.csh) in Figure 2.2 is automated.

‡D Score merge data conversion program (DataMaker) This is called from the docking score totaling shell (make_score_data.csh), and it converts the score data into the interaction matrix data format.

3.1 Grid data generation shell

This automates the task of creating a protein grid file for sievgene. The specifications of the grid data generation shell "make_grid.csh" are shown below.

<table>
<thead>
<tr>
<th>Input file</th>
<th>Program used</th>
<th>Output file</th>
<th>Function</th>
<th>Starting method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control file template (s0grid.inp) for sievgene, see Figure 3.2.</td>
<td>Docking execution script (Run_docking.pl)</td>
<td>Control file template (protein name).inp same format as in Figure 3.2.</td>
<td>Checks whether or not the protein grid file exists. If not, creates a grid file output directory and a sievgene control file, and executes sievgene to create a protein grid file.</td>
<td>Execute the following command in the version directory: <code>bin/make_grid.csh</code></td>
</tr>
</tbody>
</table>

myPresto 4.2
The processing flow of the grid data generation shell is shown in Figure 3.1. When the grid data generation shell is started, the sievgene control file (s0grid.inp) is edited and output (s1grid.inp). Sievgene is then executed using this file to generate grid data for each protein. The generated grid data are output as grid files to the sub-directories named for each protein in the grid directory.

Editing of s0grid.inp replaces the following text strings:

- `#NAMETO#` - Replaced with the topology file name
- `#NAMECO_in#` - Replaced with the protein PDB file name
- `#NAMEPO#` - Replaced with the pocket data file name
- `#GRIDGR#` - Replaced with the grid output file name

**Default settings**

- Check if directory and execution program exist
- Obtain protein list
- Check work directory; if none, create
- Execute `sievgene`
- Repeat as many times as there are proteins
- Check if protein grid file exists; if not, create grid file output directory and sievgene control file. If none, process the next protein.
- Register job for which `sievgene` is executed
- End

**Protein list**

- `[Protein name].inp`
```
PHASE> INPUT
TOPOLO  =    FORM
NAMETO  =    #NAMETO#    ; topology file    (protein)
COORDI  =    PDB
NAMECO  =    #NAMECO_in# ; coordinate file  (protein)
POINTC  =    PDB
NAMEPO  =    #NAMEPO#    ; probe point file (protein)
SETTAR = NORE
DAMPPA = 1.0d0
QUIT

; grid generation and Hash table generation

PHASE> GRID
GRIDPOtential    = NORE        ; Grid file reading SW (NORE/ASCI/BINA)
NAMEGRid         = #NAMEGR#    ; Grid file
OUTGRIdpotential = BINA        ; Grid file writing SW (NOWR/ASCI/BINA)
PROBDIst = 6.5  ; search margin
MARGIN   = 6.5  ; search margin
ITERAT   = 3    ; iteration of Grid potential smoothing
RADVDW   = 0.6  ; vDW boundary
RADELE   = 0.6  ; coulomb boundary
RADMESH  = 1.4  ; probe radius
DAMPVW = 0.99d0
USEPBG = NO     ; not use PB
QUIT
```

Fig. 3.2. Contents of s0grid.inp (for myPresto version 3 & 4)
This shell automates the task of docking the proteins and the low molecules.

**Input files**
- sievgene control file (s0.inp See Fig. 3.6.)
- Protein list file (file listing protein IDs)
- Low molecule list file name (file listing the low molecule files)

**Program used**
- Run_docking.pl

**Output file**
- sievgene control file (same format as s1.inp in Fig. 3.6)

**Function**
Taking the specified protein list file name and low molecule list file name as arguments, this shell automatically performs execution of sievgene using the proteins and low molecules in the list files.

**Starting method**
Execute the following command in the version directory.

```
>> bin/make_docking_score.csh protein list file name low molecule list file name
```
Description of processing

The processing flow for the docking job start shell and the docking execution script is shown in Fig. 3.5. When the docking job start shell is started, the specified protein list file is opened and the following processing is repeated for each protein:

- The sievgene control file is edited and output as s1.inp.
- The following command is executed using the low molecule list file name.(¢ indicates a space.

```bash
>> Run_docking.pl {low molecule list file name}
```
Fig. 3.5. Processing flow of docking score generation

Repeat as many times as there are proteins

Default settings

Check output directory; if none, create

Read in protein list file

Check input file directory; if none, error and end

End

Pro_md.pdb

ProB.tpl

_tmp_s1.inp

s0.inp

Score merge

file

Cluster node

Create sievgene control file templates for each protein. Do not create if file already exists.

Execute Run_docking.pl

Create sievgene control file

Read in low molecule list file

Repeat as many times as there are low molecules

s1.inp

Execute sievgene

make_docking_score.csh

RUN_docking.pl

Low molecule data

grid.file

Low molecule list

Protein list
Contents of sievgene input data (s0.inp) for myPresto version 3

PHASE> INPUT
LIGAND  =    MOL2
NAMELI  =    #NAMELI#  ; coordinate file (ligand)
REFERE  =    MOL2
NAMERE  =    #NAMERE#  ; reference file (ligand)
TOPOLO  =    BINA
NAMETO  =    #NAMETO#    ; topology file (protein)
COORDI  =    PDB
NAMECO  =    #NAMECO_in# ; coordinate file (protein)
POINTC  =    PDB
NAMEPO  =    #NAMEPO#    ; probe point file (protein)
SETTAR = NORE
DAMPPA = 1.0d0
QUIT

; grid generation and Hash table generation

PHASE> GRID
GRIDPOtential    = BINA        ; Grid file reading SW (NORE/ASCI/BINA)
NAMEGRid         = #NAMEGR#    ; Grid file
OUTGRIdpotential = NOWR        ; Grid file writing SW (NOWR/ASCI/BINA)
PROBDIst = 6.5  ;
MARGIN   = 6.5  ; search margin
ITERAT   = 3    ; iteration of Grid potential smoothing
RADVDW   = 0.6  ; vDW boundary
RADELE   = 0.6  ; coulomb boundary
RADMESh  = 1.4  ; probe radius
DAMPVW = 0.99d0
USEPBG = NO     ; not use PB
QUIT
myPresto 4.2

; conformer generation

ATMDL = UNIT

CONFLImit = 100000

CONFORmernumber = 100

SORTATom = YES

DAMPINg = 0.7

PHASETorsion = 3

ROTTER = YES

QUIT

PHASE> DOCK

METHOD = FLEX

GENERAtion = 1

NUMCONFomer = 1000

MATCHING = 3

LOWMIN = 2.5

LOWMAX = 3.5

UPRMIN = 5.0

UPRMAX = 12.0

RADIUS = 6.0

WETVDW = 1.0

WETASA = 1.0

WETELE = 1.0

WETHYD = 1.0

EVALHB = NO

WETANH = 1.0d0

ROTLOH = NO

ROTPSC = NO

MOVNUM = 10

CANDID = 10

DOCKSP = FAST
QUIT

EXE> MIN

METHOD= STEEP       CPUTIM  =  360000.0
UPRATE= 1.0         DOWNRATE= 0.3
LOOPLI= 100         UPDATE  =  100
MONITO= 100         CONVGR  =  0.1D0
CUTMET= RESA        CUTLEN  =  22.0D0
DIEFUN= DIST        DIEVAL  =  4.0D0
LOGFOR= SHOR
QUIT

PHASE> OUTPUT

COORDInate     = #COORDI#    ; coordinate file type
NAMECOordinate = tmp.cor     ; coordinate file
NAMESCore      = #NAMESC#    ; score file
CANDIDatenumber= 0           ; number of PDB
SCORENumber    = 1           ; number of score
QUIT

EXE> SIEV

Fig. 3.6.a Contents of sievgene input data (s0.inp) for version 3
Contents of sievgene input data (s0.inp) for myPresto version 4
; input file for sievgene version 4

PHASE> INPUT

LIGAND  =    MOL2
NAMELI  =    #NAMELI#  ; coordinate file  (ligand)
REFERE  =    NORE
NAMERE  =    #NAMERE#  ; reference file   (ligand)
TOPOLO  =    FORM
NAMETO  =    #NAMETO#    ; topology file    (protein)
COORDI  =    PDB
NAMECO  =    #NAMECO_in# ; coordinate file  (protein)

myPresto 4.2
POINTC  =    PDB
NAMEPO  =    #NAMEPO#    ; probe point file (protein)
SETTAR = NORE
DAMPPA = 1.0d0
QUIT

; gird generation and Hash table generation

PHASE> GRID
GRIDPOtential    = ASCI        ; Grid file reading SW (NORE/ASCI/BINA)
NAMEGRid         = #NAMEGR#    ; Grid file
OUTGRIdpotential = NOWR        ; Grid file writing SW (NOWR/ASCI/BINA)
PROBDIst = 6.5  ;
MARGIN   = 6.5  ; search margin
ITERAT   = 3    ; iteration of Grid potential smoothing
RADVDW   = 0.6  ; vDW boundary
RADELE   = 0.6  ; coulomb boundary
RADMESh  = 1.4  ; probe radius
DAMPVW = 0.99d0
USEPBG = NO     ; not use PB
QUIT

; conformer generation

PHASE> CONF
ATMMDL           = UNIT     ; united atom model
CONFLImit        = 100000   ;
CONFORmernumber  = 100      ;
SORTATom         = YES      ;
DAMPINg          = 0.7      ;
PHASETorsion     = 3        ;
ROTTER           = YES


QUIT

PHASE> DOCK

METHOD = FLEX
PROSUR = HYDR
GENERAtion = 1
NUMCONFomer = 10000
MATCHING = 3
LOWMIN = 2.5
LOWMAX = 3.5
UPRMIN = 5.0
UPRMAX = 12.0
RADIUS = 6.0
WETVDW = 1.0
WETASA = 1.0
WETELE = 1.0
WETHYD = 1.0
EVALHB = NO
WETANH = 1.0d0
ROTLOH = NO
ROTPSC = NO
MOVNUM = 100
CANDID = 10
QUIT

EXE> MIN

METHOD = STEEP
CPUTIM = 360000.0
UPRATE = 1.0
DOWNRATE = 0.3
LOOPLI = 50
UPDATE = 100
MONITO = 50
CONVGR = 0.1d0
CUTMET = RESA
CUTLEN = 22.0d0
DIEFUN = DIST
DIEVAL = 4.0d0
LOGFOR = SHOR
QUIT
PHASE> OUTPUT

COORDInate = #COORDI#    ; coordinate file type
NAMECOordinate = tmp.cor     ; coordinate file
NAMESCore      = #NAMESC#    ; score file
CANDIDatenumber= 0           ; number of PDB
SCORENumber    = 1           ; number of score
QUIT

EXE> SIEV

Fig. 3.6.b Contents of sievgene input data (s0.inp) for version 4
Editing of s0.inp consists of replacement of the following text strings:

#NAMETO#FTopology file name
#NAMECO_in#FProtein PDB file name
#NAMEPO#FPocket data file name
#GRID_DIR#FGrid file name
#NAMELI#FLow molecule file name
#NAMERE#FLow molecule coordinate file name
#NAMESC#FOutput score file name
Docking execution script

This automates the task of reconfiguring sievgene settings and executing sievgene for each low molecule. This is repeated as many times as there are low molecules in the low molecule list.

**Input files**
- sievgene control file (s1.inp) (same format as in Fig.
- Low molecule list file (see Fig.)

**Program used**
sievgene load module

**Output files**
- Score merge files that merge by group the score files resulting from docking of the specified proteins and low molecules in sievgene ({low molecule group name}.scores)

**Function**
Executes docking calculation in sievgene for each low molecule indicated in the low molecule list file, and stores the scores output by sievgene.

**Starting method**
>> Run_docking.pl low molecule list file name
- The script is started automatically from make_docking_score.csh.

**Description of processing**
The low molecule list file is opened and the processing below is repeated for each low molecule.
- See Fig. 3.5.
- The sievgene control file (s1.inp) is edited.
- The created control file is input into sievgene and sievgene is started.
- The score files output from sievgene are added to the score merge file.
This shell totals the score files that are output as the result of docking of proteins and small molecules.

**Input files**
- Protein list file (see Fig. 3.3)
- Low molecule list file (see Fig. 3.4)

**Programs used**
- DataMaker

**Output file**
- Interaction matrix data file (see Fig. 3.7)

| Protein ID | Low molecule name | Docking score
|------------|-------------------|----------------
| P [Protein ID1] | [Low molecule name 1] | -2.79, -6.87, -179.04, -476.18 |
| P [Protein ID2] | [Low molecule name 1] | -2.58, -6.99, -160.11, -420.55 |

Fig. 3.7. Interaction matrix data file

The protein name is entered in the 2nd column of lines that have P in the 1st column. The low molecule name is entered in the 3rd column of lines that have # in the 1st column. The docking score is entered in the 3rd column of lines that have @ in the 1st column.

**Function**
- Creates an interaction matrix file using the protein list and low molecule list.
【スタート方法】
Execute the following command in the version directory (indicates a space).

```
>> bin/make_score_data.csh
```

【処理の流れ】
The processing flow for docking score totaling is shown in Fig. 3.8.
When the docking score totaling shell is started, the protein list is read and the following processing is repeated for each protein.

- The protein name is added to the interaction matrix data file in the format `protein name`.
- The DataMaker input file is generated.
- The following command is executed.

```
>> ./DataMaker < {DataMaker input file}
```
- Add the output result to the interaction matrix file.

**Fig. 3.8 Processing flow for docking score totaling**
Repeat as many times as there are proteins.

**Default settings**
- Check output directory; if none, create
- Read in protein list file
- Read in low molecule list file
- Read in score merge file
- Interaction matrix data file
- `make_score_data.csh`
- Output protein name
- Low molecule list
- Score merge file
- DataMaker
- Output in interaction matrix data format

myPresto 4.2
```
{ low molecule list file name }
{ score merge file name }
{ output file name }

Fig. 3.9 Control file format

# 1 [Low molecule name 1]
@ 1 -2.79 -6.87 -179.04 -476.18
# 2 [Low molecule name 2]
@ 1 -3.13 -8.17 -200.26 -526.67

Fig. 3.10. DataMaker output format
```

This program reads the score merge file, extracts the score data, and uses the low molecule list to convert the data to interaction matrix format.

```
>>./DataMaker < {DataMaker control file}
```

The program is started automatically from make_score_data.csh.
Description of processing

The low molecule list file is read and the following processing is repeated for each low molecule.

- The "#" line is output using the low molecule name (see Fig. 3.10).
- The file name column is extracted from the score merge file.
- The program checks whether or not the low molecule name in the low molecule list matches the file name in the score merge file. If the names match, the scores on the "@" line of the score merge file are read and output up to the 5th column. If the names do not match, the program returns to repeated processing.
## 4 Screening Tools

### 4.1 MTS Screening Program

This program uses the MTS method to order low molecular compounds. Three types of ordering are output: ordering using only the MTS method, ordering using MASC scores, and ordering that merges both.

**Input files**
- Control file (see Fig. 4.1)
- Protein list file (see Fig. 3.3)
- Hit list file (see Fig. 4.2)
- Interaction matrix data list file (see Fig. 4.3)

The values at left are entered on each line.

![Fig. 0.1. Format of MTS control file](image1)

X
lig_001
lig_002

X is entered at the beginning of the group. One ligand file name is entered on each line.

(g.mol2 is not entered)

![Fig. 0.2. Format of hit list file](image2)

matrix_file1
matrix_file2
matrix_file3

Interaction matrix data file names to be read are entered on each line.

![Fig 4.3. Format of interaction matrix data list file](image3)
The program reads the interaction matrices entered in the interaction matrix data list, performs ordering by the MTS method using the protein names in the protein list, the target protein name, and the low molecules in the hit list, and outputs the number of orders specified in "number of orders output" in descending order.

The control file is read.
The interaction matrix data list file is read.
The entered interaction matrix data is read and the score is stored in one matrix.
The protein data entered in the protein list is selected and the matrix reconfigured.
A conversion matrix is randomly created, and the task of repeated ordering is performed repeatedly to leave a matrix with the highest level of database enrichment.
The order produced by the MTS method using the conversion matrix with the highest database enrichment that is left, the order produced by the MASC method, and the order produced by merging both methods are output as comp_list_MTS, comp_list_RS, and comp_list_UN. The number of orders output is the number specified in `{number of orders output}`.

An interaction matrix consisting only of the low molecules indicated in comp_list_MTS, comp_list_RS, and comp_list_UN is output.
myPresto 4.2

MTS screening job start shell

**Input files**
- Low molecule group list file (see Fig. 3.4)
- Interaction matrix data list file template (see Fig. 4.4)
  - ../../matrix/pro.list_#LIGAND_GROUP#.dat
  - ../../matrix/pro.list_vc01.dat
  - ../../matrix/pgd2.list_#LIGAND_GROUP#.dat
  - ../../matrix/pgd2.list_vc01.dat

**Program used**
- MTS screening program

**Output file**
- Interaction matrix data file list

**Function**
Executes MTS screening for each low molecule group indicated in the low molecule group list file.

**Start method**
./run_mts.sh

**Description of processing**
The low molecule group list file is read and the following processing is repeated for each low molecule group.

- A directory with the low molecule group name is created in the startup directory.
- Specific text strings in the interaction matrix data file list template are replaced with the low molecule group name, and the file is stored in the created directory.
- The MTS screening program is executed from the created directory.
This orders low molecular compounds using the DSI method.

### Input files
- DSI control file (see Fig. 4.5)
- Protein list file (see Fig. 3.3)
- Hit list file (see Fig. 4.2)
- Interaction matrix data list file (see Fig. 4.3)

```
../pro.list
../hit_list
merge_list
```

### Output files
- comp_list_PCA (DSI screening order)
- out_log_PCA (same format as Fig. 3.7)
- final.db (Database enrichment result)

### Function
This program reads the interaction matrices indicated in the interaction matrix data list, uses the proteins in the protein list, the target protein name, and the query low molecules in the hit list to perform ordering by means of the DSI method, and outputs the number of orders specified in "number of orders output" in descending order.

### Starting method
```
>> ./selectDSI < {DSI control file}
```
Description of processing

- The control file is read.
- The interaction matrix data list file is read.
- All entered interaction matrix data is read, and the scores are stored in one matrix.
- Only protein data appearing in the protein list is selected, and the interaction matrix is reconfigured.
- A conversion matrix is randomly created, and the task of repeated ordering is performed repeatedly to store a matrix with the highest level of database enrichment.
- The order produced by the DSI method using the conversion matrix with the highest database enrichment that is left is output as the comp_list_PCA file. The number of orders output is the number specified in {number of orders output}.
- An interaction matrix consisting only of the low molecules indicated in comp_list_PCA is output.
DSI screening job start cell

**Input files**
- Low molecule group list file
- Interaction matrix data file list template

**Program used**
- DSI screening program

**Output files**
- Interaction matrix data file list

**Function**
Starts a DSI screening job for each low molecule group indicated in the low molecule group list file.

**Start method**
`./run_dsi.sh`

**Description of processing**
The low molecule group list file is read, and the following processing is repeated for each low molecule group.

1. A directory with the same name as the name of the low molecule group is created in the startup directory.
2. Specific text strings in the interaction matrix data file list template are replaced with the low molecule group name, and the file is stored in the created directory.
3. The DSI screening program is executed from the created directory.
5 In silico Screening Sample

5.1 Overview of screening work

In this sample, low molecule screening is performed using the DSI method and the MTS method. Both methods use a docking score, which is a value that evaluates the ease of bonding of a low molecule to the protein.

**DSI method:**
For each low-molecular compound, this method examines the patterns of the protein docking scores and selects low molecules that resemble the docking score pattern of the specified low-molecular compound.

**MTS method:**
For a target protein, this method examines the order of the docking scores of low molecules and selects compounds that rank high with that target protein.

Screening consists of the two tasks below.

- **Screening**
The low molecules are ordered and a list of names of the high-ranking low molecules is created. The interaction matrix data file that indicates the interaction matrix (see the next page) is used as input data.

- **Preparation of an interaction matrix data file**
The low molecule and protein data are combined and docking calculation is performed, and a new interaction matrix data file is created. This task is performed when new proteins or new low molecules not included in an existing interaction matrix are used. During screening, the newly created interaction matrix data file and the existing interaction matrix data file are used to perform screening that includes the new proteins and low molecules.

The relation between these two tasks is shown in Fig. 5.1.
The interaction matrix contains the docking score of each of the combinations of proteins and low molecules (see Fig. 5.2).

**Fig. 5.2. Interaction matrix**

- **Low molecule data**
  - Docking program
- **Protein data**
  - Existing interaction matrix data file
- **Screening program**
  - Ordering result
- **Screening work**
  - Docking calculation
  - Screening (DSI, MTS)
- **Newly created interaction matrix data file**
  - Screened molecules:
    - Low molecule P
    - Low molecule 2
    - Low molecule 3
    - Low molecule x
  - Protein:
    - Protein 2
    - Protein 3
    - Protein 2.31
    - Protein 2.07
    - Protein 2.15
    - Protein 2.34
    - Protein 2.88
    - Protein 2.39
    - Protein 2.55
    - Protein 2.98
    - Protein 3.67
    - Protein 2.71

**Interaction matrix**

- **Docking score**
To perform screening with newly added low molecules or proteins, new interaction matrices $A$, $B$, and $C$ are created. There are three cases in which these matrices are created (see Fig. 5.3):

- Adding low molecules: Interaction matrix $A$
- Adding proteins: Interaction matrix $B$
- Adding molecules and proteins: Interaction matrices $A$, $B$, and $C$

In this sample, screening is performed with added low molecules (interaction matrix $A$ is added).

**Diagram:**

- Protein 1 to $y$
- Protein $P_1$ to $P_n$
- Low molecule $L_1$ to $L_m$
- Existing interaction matrix
- Added interaction matrix $A$
- Low molecule 1 to $x$
- Added interaction matrices
  - Added interaction matrix $B$
  - Added interaction matrix $C$
In this sample, screening work for the case of "Adding new low molecules" (see the previous page) is performed using new molecules. To perform screening using new low molecules, a file (in MOL2 format) containing the structural data of the low molecules is required. This file will be referred to as the "low molecule file." The low molecule file is used to create an interaction matrix equivalent to "Interaction matrix A" in Fig. 5.3. The work is performed in the order below. The detailed procedure is described in 5.3 to 5.5.

- **P** Environment preparation
  - Prepare the screening sample environment.

- **Q** Interaction matrix data file preparation
  - Create a new interaction matrix data file using the new low molecule file and the protein data included in the existing interaction matrix.

- **R** Screening work
  - Perform screening using the existing interaction matrix data file and the new interaction matrix data file created in (2). The DSI method and the MTS method are used for screening.

### 5.3 Sample work conditions and environment preparation

#### 5.3.1 Work conditions

The conditions of this screening sample are indicated below.

- For the existing interaction matrix, an interaction matrix data file of 182 proteins and 1000 low molecules is used.
- In the DSI method, the following three low molecule files are used as the specified low molecules.
  - 1pxxz.mol2, 4coxz.mol2, 6coxz.mol2
- The target protein of the MTS method is 4coxz in the existing interaction matrix.
5.3.2 Preparation of screening environment

Copy the compressed screening environment file `{screening_sample.tar.gz}` to the work directory and extract it. Extracting the file will create the `{screening_sample}` directory along with sub-directories containing all elements of the sample. In the following, the `{screening_sample}` directory will be referred to as `{HOME}`. The directories in `{HOME}` are indicated in Table 5.1.

Table 5.1. Description of directories in the screening environment

<table>
<thead>
<tr>
<th>Directory name</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>protein/</td>
<td>Protein data directory. Each protein has its own directory. Data for 182 proteins is stored.</td>
</tr>
<tr>
<td>ligand/</td>
<td>Low molecule file directory. Each low molecule group has its own sub-directory. Low molecule files are in MOL2 format. In this environment, 1000 low molecule files are stored in the lig0 sub-directory.</td>
</tr>
<tr>
<td>matrix/</td>
<td>Interaction matrix data file directory. In this environment, an interaction matrix data file created from 182 proteins and 1000 low molecules is stored.</td>
</tr>
<tr>
<td>grid/</td>
<td>Grid file directory. 182 files are stored, one for each protein.</td>
</tr>
<tr>
<td>screening1/</td>
<td>Sample screening work directory. Screening work is performed under this directory.</td>
</tr>
<tr>
<td>src/</td>
<td>Tool source directory. Program source directory for programs used for screening.</td>
</tr>
<tr>
<td>sample/</td>
<td>Sample directory. Directory for newly added low molecule samples. MOL2 data for 89 new low molecules and hit list files (see p. 9) are stored here.</td>
</tr>
</tbody>
</table>

The directory structure inside `{screening1}` is described in attachment 1.

5.3.3 Preparation of screening tools

The following tools are prepared in order to use the environment.

- `sievgene` Docking program
- `selectDSI` DSi screening program
- `selectMTS` MTS screening program
- `DataMaker` Interaction matrix generation program

Program sources are stored in the directories named for each program in the `{HOME}/src/` directory. Execute the `make` command in each directory and store the created execution programs using the above file names in the `{HOME}/screening1/bin/` directory.
Interaction matrix data file preparation

A new interaction matrix data file is prepared using a new low molecule file and the protein data included in an existing interaction matrix. This is performed in the following order.

M1. Preparation of low molecule file
M2. Preparation of low molecule list file
M3. Preparation of protein list file
M4. Execution of docking calculation shell
M5. Execution of totaling shell

M1. Preparation of low molecule files

Prepare the structural data of the new low molecules (this will be referred to as the low molecule file). The preparation method is as follows:

- Prepare the low molecule files in MOL2 format.
- Create a directory in the `{HOME}/ligand/` directory.
- Copy the low molecule files to the created directory.

In the case of this sample, the low molecule files that are used are stored in the `{HOME}/sample/lig_sample/` directory. The following three low molecules files are used:

1pxxz.mol2, 4coxz.mol2, 6coxz.mol2

Use the following command to create the `{HOME}/ligand/lig1/` directory:

```bash
>  mkdir {HOME}/ligand/lig1
```

Copy the above 3 low molecule files to the `{HOME}/ligand/lig1/` directory.

Execute the following commands.

```bash
>  cd {HOME}/sample/lig_sample
>  cp 1pxxz.mol2 4coxz.mol2 6coxz.mol2 ../../ligand/lig1
```
Preparation of low molecule list file

Prepare the low molecule list file, which contains a list of the low molecule files prepared in "M1. Preparation of low molecule files".

Create the \{HOME\}/screening1/list/ directory. The format of the low molecule list file is shown in Fig. 5.4. One low molecule file name is indicated on each line. The name of the list file should be the same name as the directory created in "M1. Preparation of low molecule files".

```
ligand1.mol2
ligand2.mol2
ligand3.mol2
```

Fig. 5.4. Format of low molecule list file

In the case of this sample, the low molecule files were created in the \{HOME\}/ligand/lig1/ directory in "M1. Preparation of low molecule files", and thus the name of the low molecule list file should be "lig1". Use the following file without editing.

```
{HOME}/screening1/list/lig1
```

This file contains the following entries:

```
1pxxz.mol2
4coxz.mol2
6coxz.mol2
```
M3. Preparation of protein list file

Prepare a protein list to perform docking calculation with the low molecules prepared in M1. Preparation of low molecule files.

Create the {HOME}/screening1/list/ directory. The format of the protein list file is shown in Fig. 5.5. One protein name is entered on each line. Any name can be used for the protein list file name. Proteins included in existing interaction matrices are listed in {HOME}/screening1/list/pro.list.

```
<table>
<thead>
<tr>
<th>Protein Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1a28</td>
</tr>
<tr>
<td>1ai5</td>
</tr>
</tbody>
</table>
```

Fig. 5.5. Format of protein list file

In the case of this sample, use the following file without editing:

{HOME}/screening1/list/pro.list

M4. Execution of docking calculation shell

Generate the docking scores to be stored in the interaction matrix data file. Perform docking of the existing 182 proteins and the three low molecules prepared in M1. Preparation of low molecule files. Execute the command in {HOME}/screening1 indicated below. For the low molecule list file, use the file prepared in M2. Preparation of low molecule list file. The protein list file is stored as the {HOME}/screening1/pro.list file. If you do not intend to add or delete proteins, use this file.

```
> bin/make_docking_score.csh {Protein list file} {Low molecule list file} -
```

In the case of this sample, the protein list file is created as the pro.list file in the {home}/screening1/list/ directory. The name lig1 was assigned to the low molecule list file in M2. Preparation of low molecule list file, therefore execute the following command:

```
> cd {HOME}/screening1
```

myPresto 4.2
> bin/make_docking_score.csh pro.list lig1
Using the docking results, create an interaction matrix to be used for screening.

Execute the command in `{HOME}/screening1/` below.

```
> bin/make_score_data.csh {Protein list file} {Low molecule list file}
```

The created interaction matrix is stored in the `{HOME}/screening1/matrix/` directory.

The file name format is as follows:

```
{Protein list file}_{Low molecule list file}.dat
```

If a file of the same name already exists, an error will occur and the interaction matrix data file will not be created.

In the case of this sample

Execute the following command:

```
> cd `{HOME}/screening1`
> bin/make_score_data.csh pro.list lig1
```

The `pro.list_lig1.dat` interaction data file will be created in the `{HOME}/screening1/matrix/` directory.
Screening work

Screening is performed using the existing interaction matrix data file and the new interaction matrix data file created in "M5. Execution of totaling shell". In the DSI method, work is performed in the `{HOME}/screening1/dsi/` directory. In the MTS method, work is performed in the `{HOME}/screening1/mts/` directory.

Screening work consists of the following steps S1 to S6 for both the DSI method and the MTS method.

S1. Preparation of the protein list file
S2. Preparation of the hit list file
S3. Preparation of the interaction matrix list file
S4. Preparation of the screening program settings file
S5. Execution of the screening program
S6. Verification of the results

5.5.1. Screening by the DSI method

The DSI screening program uses the 4 files below.

- Protein list file
- Hit list file
- Interaction matrix list file
- Screening program settings file (for DSI)

After the above files have been prepared, screening is performed by executing the DSI screening program.

S1. Preparation of the protein list file

The protein list file indicates the proteins used in screening. The file format is shown in Fig. 5.6. One protein name is entered on each line.

```
{HOME}/screening1/DSI/pro.list
```

is a sample protein list file that lists 182 proteins.

```
1a28 (Protein name)
1ai5 (Protein name)
```

Fig. 5.6. Format of protein list file

In the case of this sample

Use the following file without editing:

```
{HOME}/screening1/DSI/pro.list
```
Preparation of hit list file

The hit list file lists the names of the low molecules (excluding the extension ".mol2") to be specified in DSI screening. The low molecule names are separated into two groups, upper and lower. The upper group of lower molecule data are used as a standard for performing screening. The lower group of low molecule names are used to study database enrichment, and can be omitted. If omitted, only database enrichment of the upper group is output.

The format of the hit list file is shown in Fig. 5.7. An "ARCH" must be entered at the beginning of each group. A sample hit list file is stored in the "{HOME}/sample/hit_list/" directory.

```
ARCH
ARCH
•••{ARCHED}••

ARCH
ARCH
•••{ARCHED}••
```

In the case of this sample, use the following file without editing:
```
{HOME}/screening1/dsi/hit_list
```

This file contains entries similar to the following:
```
1pxxz
coxz
coxz
```

myPresto 4.2
S3. Preparation of interaction matrix list file

The interaction matrix list is a file that indicates the names of the interaction matrix data files used in screening. One interaction matrix data file name is entered on each line.

```
matrix1.dat
matrix2.dat
```

Fig. 0.8. Format of interaction matrix list file

In the case of this sample, use the following file without editing:
```
{HOME}/screening1/dsi/merge_list
```

This file contains entries similar to the following:
```
../../matrix/pro.list_lig0.dat
../matrix/pro.list_lig1.dat
```

S4. Preparation of screening program settings file (DSI method)

The screening program settings file contains the DSI screening program settings. The protein list file name, hit list file name, number of orders output, and interaction matrix list file name are indicated.

```
protein.list           (Protein list file name)
../../sample/hit_lists/0cox1.lst (Hit list file name)
1 2
30
300                (Number of orders output)
merge.lst            (Interaction matrix list file name)
```

Parameter for specification of the axis in the DSI method. Use this without changing the value.

Fig. 0.9. Format of DSI screening program input file
n
In the case of this sample
Use the following file without editing:

{HOME}/screening1/dsi/dsi.inp

This file contains the following entries:

pro.list
hit_list
1 2
30
100
merge_list

n
In the case of this sample
Execute the following command in the
{HOME}/screening1/dsi/ directory:

> ../bin/selectDSI<dsi.inp

> ../bin/selectDSI<DSI screening program settings file name

n
In the case of this sample

> ../bin/selectDSI<dsi.inp

 Execute the command below in the
 {HOME}/screening1/dsi/ directory to execute the
 DSI screening program.

> ../bin/selectDSI<DSI screening program settings file name
Verification of results (DSI method)

The results of DSI screening are output to the {HOME}/screening1/dsi/ directory using the file name comp_list_PCA. The ranking, internal number of the low molecule, low molecule name, and DSI score are entered on each line in that order. An example of comp_list_PCA is shown in Fig. 5.10.

```
1     5325 0016245-01    0.5061
2     9132 0036475-01    0.5435
3     5394 0016451-01    0.5955
4     4504 0014570-01    0.6858
5     5824 0017979-01    0.6957
```

Fig. 0.10. DSI screening program output data (comp_list_PCA)

In addition, the following files are output:
- final.db (Distribution of low molecules specified in the hit list file)
- out_log_PCA (Interaction matrix data file of low molecules remaining after screening)

In the case of this sample, the result of screening is output to the {HOME}/screening1/dsi/ directory using the file name comp_list_PCA. An output example of comp_list_PCA is attached in Appendix 2.

```
1     5325 0016245-01    0.5061
2     9132 0036475-01    0.5435
3     5394 0016451-01    0.5955
4     4504 0014570-01    0.6858
5     5824 0017979-01    0.6957
```

Note that the output order of the data can be specified.
5.5.2. Screening by the MTS method

The MTS screening program uses the following four files:

- Protein list file
- Hit list file
- Interaction matrix list file
- Screening program settings file (for MTS)

After the above files have been prepared, execute the program to begin screening.

S1. Preparation of protein list file

The protein list file for the MTS method is the same as for the DSI method. The file format is indicated in Fig. 5.6. 

```
{HOME}/screening1/MTS/pro.list
```

is a sample protein list file that lists 182 proteins. To add proteins, add the protein names to the protein list file.

In the case of this sample

Use the following file without editing:

```
{HOME}/screening1/MTS/pro.list
```

S2. Preparation of hit list file

The hit list file used for the MTS method is the same as for the DSI method. The file format is indicated in Fig. 5.7. In the MTS method, a setup that does not use a hit list file is also possible (see "S4. Preparation of screening program settings file (MTS)").

In the case of this sample

A hit list file is not used.

S3. Preparation of the interaction matrix list file

The interaction matrix list file used for the MTS method is the same as for the DSI method. The file format is indicated in Fig. 5.8.

In the case of this sample

Use the following file without editing:

```
{HOME}/screening1/mts/merge_list
```

The file format is indicated in Fig. 5.8.

```
{HOME}/screening1/mts/merge_list
```

The MTS screening program uses the following four files:

- Protein list file
- Hit list file
- Interaction matrix list file
- Screening program settings file (for MTS)

After the above files have been prepared, execute the program to begin screening.

S1. Preparation of protein list file

The protein list file for the MTS method is the same as for the DSI method. The file format is indicated in Fig. 5.6.

```
{HOME}/screening1/MTS/pro.list
```

is a sample protein list file that lists 182 proteins. To add proteins, add the protein names to the protein list file.

In the case of this sample

Use the following file without editing:

```
{HOME}/screening1/MTS/pro.list
```

S2. Preparation of hit list file

The hit list file used for the MTS method is the same as for the DSI method. The file format is indicated in Fig. 5.7. In the MTS method, a setup that does not use a hit list file is also possible (see "S4. Preparation of screening program settings file (MTS)").

In the case of this sample

A hit list file is not used.

S3. Preparation of the interaction matrix list file

The interaction matrix list file used for the MTS method is the same as for the DSI method. The file format is indicated in Fig. 5.8.

In the case of this sample

Use the following file without editing:

```
{HOME}/screening1/mts/merge_list
```

The file format is indicated in Fig. 5.8.
Preparation of screening program settings file (MTS)

The MTS screening program settings file contains settings for the MTS screening program. The protein list file name, hit list file name, number of orders output, and interaction matrix are indicated. If a hit list file is not used, enter "n" on the hit list file line. The format of the screening program settings file is shown in Fig. 5.11.

```
protein.lst            (Protein list file name)
proteinA              (Protein name)
../../sample/hit_lists/0cox1.lst  (Hit list file name or n)
100                (Number of orders output)
merge_list              (Interaction matrix file list file name)
```

In the case of this sample
Use the following file without editing:
```
{HOME}/screening1/mts/mts.inp
```

Execution of screening program (MTS)
Enter the following command to execute the MTS screening program:
```
>  ../bin/selectMTS < {MTS screening program settings file name}
```

In the case of this sample
Execute the following command in the {HOME}/screening1/dsi/ directory:
```
> ../bin/selectMTS < mts.inp
```
Verification of results (MTS)

The results of MTS screening are output to the `{HOME}/screening1/mts/` directory with the file name `comp_list_UN`. Each line shows the ranking, low molecule internal number, low molecule name, and DSI score in that order. An example of `comp_list_UN` is shown in Fig. 5.12.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Internal Number</th>
<th>Molecule Name</th>
<th>DSI Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3338</td>
<td>0012157-01</td>
<td>0.3889</td>
</tr>
<tr>
<td>2</td>
<td>5937</td>
<td>0018199-01</td>
<td>0.5634</td>
</tr>
<tr>
<td>3</td>
<td>9256</td>
<td>0039111-01</td>
<td>0.3794</td>
</tr>
<tr>
<td>4</td>
<td>6703</td>
<td>0020485-01</td>
<td>0.5293</td>
</tr>
<tr>
<td>5</td>
<td>5805</td>
<td>0017939-01</td>
<td>0.3860</td>
</tr>
</tbody>
</table>

Fig. 5.12 MTS screening program output data (comp_list_UN)

The following files are also output:

- `final.db` (distribution of low molecules specified in hit list)
- `comp_list_MTS`, `comp_list_RS` (intermediate results of screening)
- `out_log_UN`, `out_log_MTS`, `out_log_RS` (interaction matrix data file using high-ranking low molecules in the results and intermediate results)
- `plot_1.data`, `pca_plot.data` (data for plotting the screening results)

In the case of this sample, the screening results are output to the `{HOME}/screening1/mts/` directory with the file name `comp_list_UN`. An output example of `comp_list_UN` is shown in Appendix 3.
<table>
<thead>
<tr>
<th>Directory</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>screening1/</td>
<td>(Sample screening work directory)</td>
</tr>
<tr>
<td>“¥ bin/</td>
<td>(Execution program directory)</td>
</tr>
<tr>
<td>“¥ make_docking_score.csh</td>
<td>(Docking score generation shell)</td>
</tr>
<tr>
<td>“¥ make_score_data.csh</td>
<td>(Interaction matrix generation shell)</td>
</tr>
<tr>
<td>“¥ make_grid.csh</td>
<td>(Grid generation shell: Used to add proteins)</td>
</tr>
<tr>
<td>“¥ selectDSI</td>
<td>(DSI screening program)</td>
</tr>
<tr>
<td>“¥ selectMTS</td>
<td>(MTS screening program)</td>
</tr>
<tr>
<td>“input/</td>
<td>(Interaction matrix preparation settings directory)</td>
</tr>
<tr>
<td>“¥ s0.inp</td>
<td>(Docking score generation settings file)</td>
</tr>
<tr>
<td>“¥ s0grid.inp</td>
<td>(Grid generation settings file: Used to add proteins)</td>
</tr>
<tr>
<td>“¥ list/</td>
<td>(List directory)</td>
</tr>
<tr>
<td>“¥ pro.list</td>
<td>(Protein list: For preparation of interaction matrix)</td>
</tr>
<tr>
<td>“¥ lig0</td>
<td>(Low molecule list)</td>
</tr>
<tr>
<td>“¥ matrix/</td>
<td>(Interaction matrix directory)</td>
</tr>
<tr>
<td>“¥ pro.list_lig1.dat</td>
<td>(Created interaction matrix)</td>
</tr>
<tr>
<td>“¥ result/</td>
<td>(Docking score output directory)</td>
</tr>
<tr>
<td>“¥ work/</td>
<td>(Docking log output directory)</td>
</tr>
<tr>
<td>“¥ grid/</td>
<td>(Grid directory)</td>
</tr>
<tr>
<td>“¥ dsi/</td>
<td>(DSI screening directory)</td>
</tr>
<tr>
<td>“¥ pro.list</td>
<td>(Protein list: For DSI screening)</td>
</tr>
<tr>
<td>“¥ hit_list</td>
<td>(Hit list file)</td>
</tr>
<tr>
<td>“¥ matrix_list</td>
<td>(Interaction matrix list file)</td>
</tr>
<tr>
<td>“¥ dsi.inp</td>
<td>(DSI screening settings file)</td>
</tr>
<tr>
<td>“¥ comp_list_PCA</td>
<td>(DSI screening results)</td>
</tr>
<tr>
<td>“¥ mts/</td>
<td>(MTS screening directory)</td>
</tr>
<tr>
<td>“¥ pro.list</td>
<td>(Protein list: For MTS screening)</td>
</tr>
<tr>
<td>“¥ hit_list</td>
<td>(Hit list file: Not used in some cases)</td>
</tr>
<tr>
<td>“¥ matrix_list</td>
<td>(Interaction matrix list file)</td>
</tr>
<tr>
<td>“¥ mts.inp</td>
<td>(MTS screening settings file)</td>
</tr>
<tr>
<td>“¥ comp_list_UN</td>
<td>(MTS screening results)</td>
</tr>
</tbody>
</table>

*Note: myPresto 4.2 is a software tool used for analyzing protein-ligand interactions.*
### Appendix 'Q Example of Output Results of DSI Screening Program

<table>
<thead>
<tr>
<th>ID</th>
<th>Code</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>548</td>
<td>1975467-01</td>
<td>1.9974</td>
</tr>
<tr>
<td>265</td>
<td>0754933-01</td>
<td>2.0652</td>
</tr>
<tr>
<td>1002</td>
<td>6coxz</td>
<td>2.0918</td>
</tr>
<tr>
<td>422</td>
<td>1371845-01</td>
<td>2.1246</td>
</tr>
<tr>
<td>278</td>
<td>1044490-01</td>
<td>2.1746</td>
</tr>
<tr>
<td>1001</td>
<td>4coxz</td>
<td>2.1773</td>
</tr>
<tr>
<td>683</td>
<td>2335847-01</td>
<td>2.1931</td>
</tr>
<tr>
<td>743</td>
<td>2674121-01</td>
<td>2.1934</td>
</tr>
<tr>
<td>597</td>
<td>1978627-02</td>
<td>2.2097</td>
</tr>
<tr>
<td>1000</td>
<td>1pxxz</td>
<td>2.2158</td>
</tr>
<tr>
<td>296</td>
<td>1045155-01</td>
<td>2.2242</td>
</tr>
<tr>
<td>706</td>
<td>2336969-01</td>
<td>2.2473</td>
</tr>
<tr>
<td>753</td>
<td>2674468-01</td>
<td>2.2689</td>
</tr>
<tr>
<td>786</td>
<td>2675806-02</td>
<td>2.2945</td>
</tr>
<tr>
<td>276</td>
<td>1044420-01</td>
<td>2.3025</td>
</tr>
<tr>
<td>282</td>
<td>1044648-01</td>
<td>2.3078</td>
</tr>
<tr>
<td>421</td>
<td>1371797-01</td>
<td>2.3203</td>
</tr>
<tr>
<td>104</td>
<td>0445234-01</td>
<td>2.3403</td>
</tr>
<tr>
<td>613</td>
<td>1979062-01</td>
<td>2.3411</td>
</tr>
<tr>
<td>723</td>
<td>2337709-01</td>
<td>2.3480</td>
</tr>
<tr>
<td>858</td>
<td>3058128-02</td>
<td>2.8097</td>
</tr>
</tbody>
</table>

### Appendix 'R Example of Output Results of MTS Screening Program

<table>
<thead>
<tr>
<th>ID</th>
<th>Code</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0001568-01</td>
<td>0.5666</td>
</tr>
<tr>
<td>987</td>
<td>3421297-01</td>
<td>0.7832</td>
</tr>
<tr>
<td>81</td>
<td>0007711-01</td>
<td>0.4554</td>
</tr>
<tr>
<td>63</td>
<td>0006064-01</td>
<td>0.7809</td>
</tr>
<tr>
<td>868</td>
<td>3058890-01</td>
<td>0.4945</td>
</tr>
<tr>
<td>959</td>
<td>3419883-01</td>
<td>0.7065</td>
</tr>
<tr>
<td>418</td>
<td>1371633-01</td>
<td>0.4573</td>
</tr>
<tr>
<td>675</td>
<td>2335192-01</td>
<td>0.6835</td>
</tr>
<tr>
<td>941</td>
<td>3419426-01</td>
<td>0.4531</td>
</tr>
<tr>
<td>658</td>
<td>2333939-01</td>
<td>0.6782</td>
</tr>
<tr>
<td>26</td>
<td>0002156-01</td>
<td>0.4591</td>
</tr>
<tr>
<td>873</td>
<td>3059336-01</td>
<td>0.6701</td>
</tr>
<tr>
<td>840</td>
<td>3056621-01</td>
<td>0.4339</td>
</tr>
<tr>
<td>503</td>
<td>1654750-01</td>
<td>0.6657</td>
</tr>
<tr>
<td>820</td>
<td>3055224-01</td>
<td>0.4896</td>
</tr>
<tr>
<td>496</td>
<td>1654082-01</td>
<td>0.6646</td>
</tr>
<tr>
<td>499</td>
<td>1654416-01</td>
<td>0.4956</td>
</tr>
<tr>
<td>183</td>
<td>0751704-01</td>
<td>0.6586</td>
</tr>
<tr>
<td>637</td>
<td>2331096-01</td>
<td>0.4412</td>
</tr>
<tr>
<td>925</td>
<td>3418675-01</td>
<td>0.6586</td>
</tr>
<tr>
<td>282</td>
<td>1044601-01</td>
<td>0.5837</td>
</tr>
</tbody>
</table>
myPresto 4.2

- in silico screening -