Drug design package software

with in silico drug discovery

MolDesk Screening

Ver.1.1.89

Manual

IMSBIO Co., Ltd.

Contents

1. Screening calculations

Operations that can be performed in MolDesk Basic You can run all even MolDesk Screening.

This manual describes the operations that can be performed only in MolDesk Screening.
Please refer to the "MolDesk Basic Manual" for common operations and MolDesk Basic.

Ligand Box is a database that includes 200 million pieces of low-molecular-weight compounds. MolDesk Screening The screening calculation, from the compound LigandBox and user-specified, you can narrow down the drug candidate compounds of several hundred to several thousand compounds.

In addition, with respect to the low-molecular-weight compounds specified by the user, processed and as that becomes possible screening calculation, you can screening calculation.

1.1. Installation and activation

The installation method and activation method of MolDesk Screening are the same as MolDesk Basic. See the MolDesk Basic manual.

However, if you want to perform molecular dynamics calculations in parallel at high speed with MPI or nVIDIA graphic board, you need to install MPI and CUDA operating environment separately.
For the specific method, refer to the chapter "High-speed parallel computing of MD calculation by MPI / GPU".

※ Mac does not support parallel computing with his MPI or his CUDA for molecular dynamics calculations.

1.2. Preparing for LigandBox

LigandBox preparation is required when screening from LigandBox compounds.

* LigandBox is distributed on the download site. If you do not know the information about the download site, please contact IMSBIO Co., Ltd. moldesk@imsbio.co.jp by e-mail. We will inform you of the URL of the download site, account and password.

The LigandBox currently distributed are LigandBox ver.2210, LigandBox ver.2104, LigandBox ver.2004 and LigandBox ver.1906.

LigandBox ver.2210 consists of 8 data compressed files.
The contents are as stated below.

LB_drug_Namiki2204.gz LB_drug_Namiki2204.ligandImage.zip
LB_agri_Namiki2204.gz LB_agri_Namiki2204.ligandImage.gz
LB_drug_Kishida2210.zip LB_drug_Kishida2210.ligandImage.zip
LB_agri_Kishida2210.zip LB_agri_Kishida2210.ligandImage.zip

Each decompression deployment results in four data:
LB_drug_Namiki2204 : 3 million compounds for Namki Shoji / Pharmaceutical
LB_agri_Namiki2204 : 3 million compounds for Namki Shoji / Pesticides
LB_drug_Kishida2210 : 1 million compounds for Kishida Shoji / Pharmaceutical
LB_agri_Kishida2210 : 1 million compounds for Kishida Shoji / Pesticides

* Unzip and unzip LB_drug_Namiki2204.gz and LB_drug_Namiki2204.ligandImage.zip to the same folder (directory).
The same applies to the other 3 data.

If you unzip and unzip it to a folder (directory), it will have the following structure.

LB_drug_Namiki2204
- ligand : mol2 file of compounds
- mts_data : Protein-compound interaction matrix
- protein : 181 proteins
- pro_list (file) : 181 protein list
- pro. list (file) : 181 protein list
- version (file) : The version of sievgene used in DB creation
- ligandImage : Image files for MolDesk Screening

(The other three are the same.)

* For Linux machines, the unzip command is, for example,
  unzip LB_drug_Namiki2204.gz.

* Please make sure that the destination path does not contain spaces. If you unzip to a path that contains spaces If you unzip the file (e.g. C:\Program Files), MolDesk Screening will not work properly.

* When decompressing and deploying, each will be up to about 63 G bytes, so please pay attention to the capacity of the storage medium of the installation destination.

* When using decompression software on Windows, be careful because the file size is large.

Depending on the decompression software, the size may be too large to decompress.

(For example, I was able to decompress with a free decompression software called Explzh (x64). I have confirmed the operation of "Explzh" and "7-Zip". If you get an error, please use these decompression software. .)

If you have a Linux machine, it is easier to unzip it on your Linux machine.

Finally, refer to "8.3.2 Screening" and set LigandBox to MolDesk Screening so that it can be used for screening calculation.

1.3. Preparing ChEMBL sdfs

Preparation of ChEMBL sdfs is necessary for regression analysis of the various properties of compounds.
It is made up of a single data compression file.

The contents are as follows.

chembl_24_sdfs_moldesk.zip

When you extract and expand to a folder (directory), the following configurations are used.

chembl_24_sdfs_moldesk - c000 : compound sdf file
ChEMBL ID is the file name
- chembl_id.lst (file) : Compound list

Finally, refer to "8.3.2 Screening" and set ChEMBL sdfs to MolDesk Screening so that it can be used as a compound to be referenced in regression analysis (Make Regression model and Predict with Regression model).

1.4. Preparing a user-specified screening compound DB

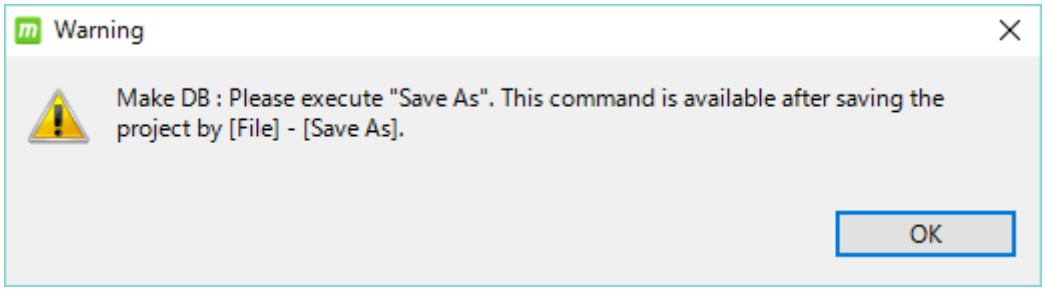
If you are using LigandBox to perform screening calculations, skip this section.

In Preparation [Preparation]- [Make DB for Screening], you can create a database for screening by inputting a user-specified compound file in addition to the distributed LigandBox. This allows you to screen user-specified compounds other than LigandBox. The compound file to be entered can be multiple sdf files.

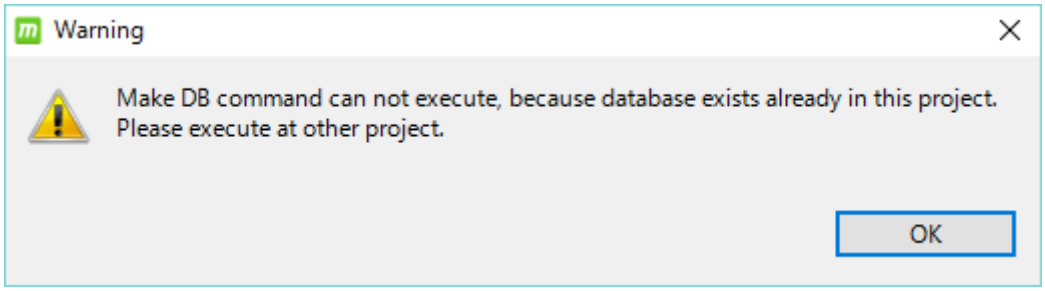
※ The number of compounds included in the database used for screening should be at least several hundred to several thousand. If the number of compounds is small, no hit compounds will be output to the screening result table.
This is because the number of molecules must be at least several hundred to several thousand to determine the parameters of the multiple linear regression equation used in the prediction model for compound screening.

1.4.1. Save project

If you have not saved the project, a warning screen will appear prompting you to save it. Save the project using the [File] - [Save as] menu. The compound database for screening will be created under the saved folder, so save it in a location with sufficient capacity. A capacity of about 6GB is required for every 100,000 compounds.



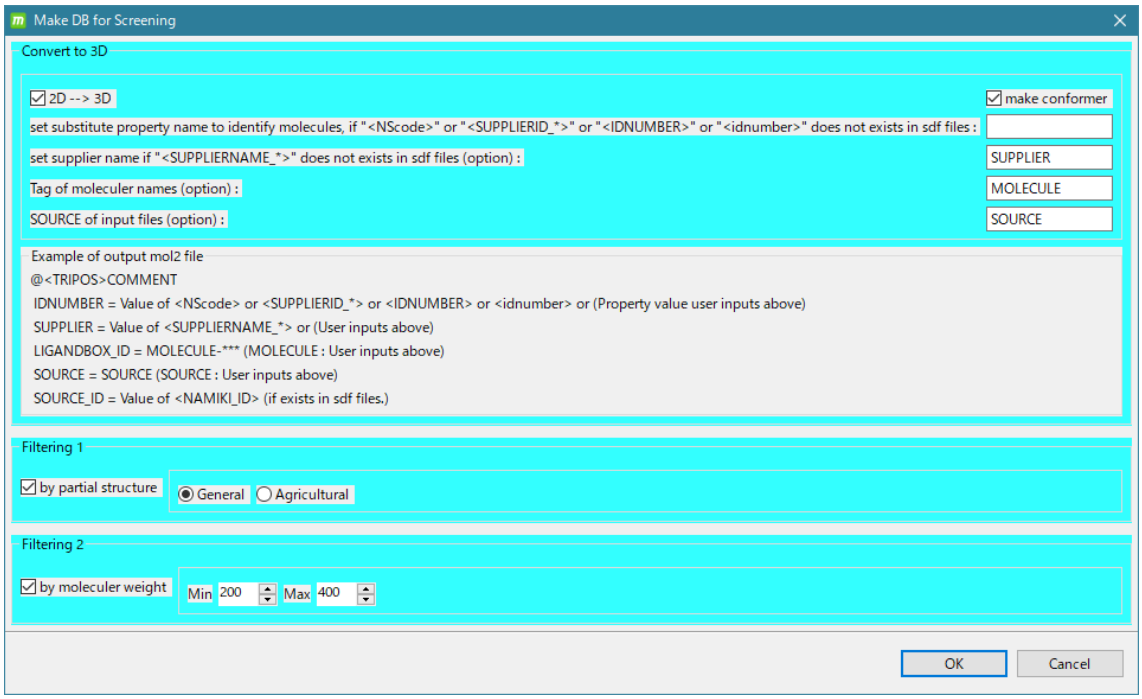
Also, if you have already created a project and opened it with [File] – [Open Project], the following warning screen may appear when you click [Make DB for Screening].



Since you have already tried to create a database in a project that has a work \ database folder that will be the output destination, it will not be possible to execute it, so please execute it in another project that does not have work \ database.

1.4.2. Set up database creation criteria

If the project is saved andthe work \ database folder does not exist, the following database creation conditions setting screen appears.



On the second line,
[set substitute property name to identify molecules, if "<NScode>" or "<SUPPLIERID_*>" or "<IDNUMBER>" or "<idnumber>" does tag does not exists in sdf file :]

It's important to type a string inIn the following Let's take a closer look.

First, check the contents of the input sdf file in a text editor or the like.

※ If you want to open the sdf file on Windows and check the contents, the free software TeraPad is convenient.

Additional information in the sdf file is that ifthe property name is described as NScode or SUPPLIERID_* or IDNUMBER, i.e. idnumber,When

```
> <NScode>
***
```

```
> <SUPPLIERID_*>
***
```

```
> <IDNUMBER>
***
```

```
> <idnumber>
***
```

(However, * is an arbitrary string)

is described, MolDesk describes the character string of the above property value of the sdf file in the ID NUMBER = of the comment line of the automatically generated mol2 file as follows.
This makes it possible to identify the output molecule.

(example mol2 file description)

```
@<TRIPOS>COMMENT
 LIGANDBOX_ID = MOLECULE-00000001-01
 SUPPLIER = SUPPLIER
 SOURCE = SOURCE
 IDNUMBER = NS-000000001-0001
 MOLECULAR_FORMULA = C8H9NO4
 MOLECULAR_WEIGHT = 183.163
 MOLECULAR_CHARGE = 0
 SUM_OF_ATOMNUMBER = 96
 SUM_OF_ATOMNUMBER_MINUS_CHARGE = 96
 NUM_OF_DONOR = 5
 NUM_OF_ACCEPTOR = 4
 HOMO = -9.2167
```

```
@<TRIPOS>MOLECULE
MOLECULE-00000001-01
 22 22 0 0 0
SMALL
USER_CHARGES

@<TRIPOS>ATOM
   1 C1        0.2340    0.2060   -0.1420 C.ar     1  LGD      -0.0357
   2 C2        1.5030   -1.9990    0.1260 C.2      1  LGD       0.3443
   3 C3        1.5630   -0.5300   -0.3070 C.3      1  LGD      -0.1662

. . .
```

Since the generated mol2 file is subject to screening, this value is also described as the ID NUMBER term in the screening calculation result list, and it can be linked with the input sdf molecule.

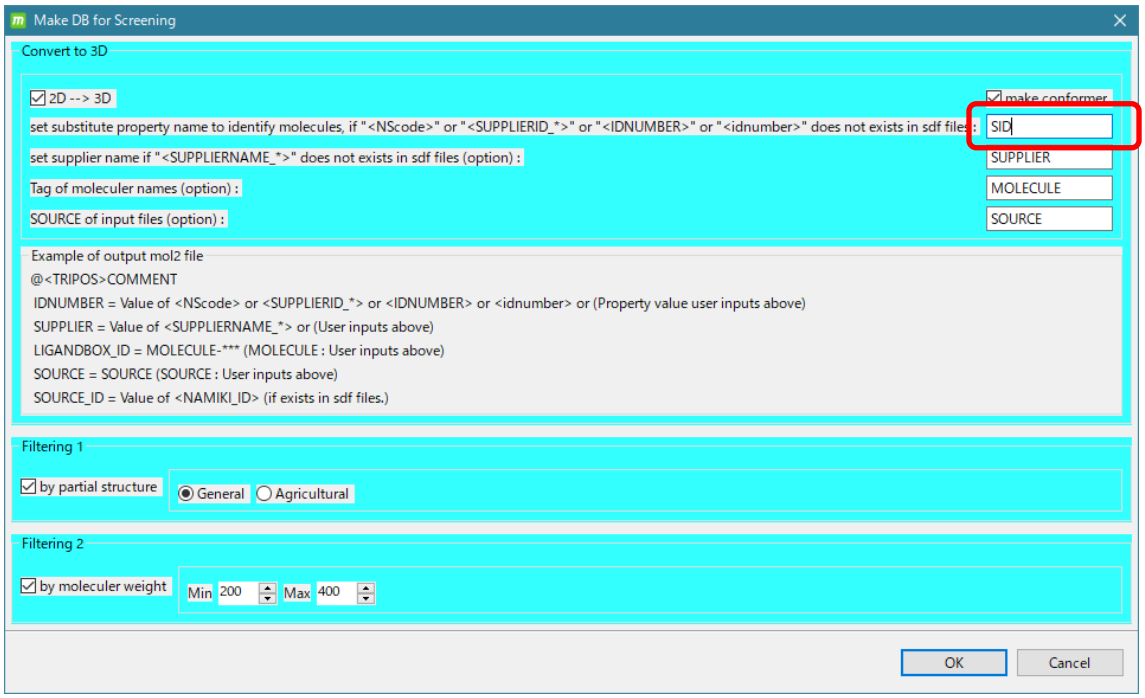
Here, suppose that the numerator description of the input sdf file is as follows, and there is no description of NScode or SUPPLIERID_ * or IDNUMBER or idnumber as the property name as additional information.

```
  Mrv1622910011607582D

 14 13  0  0  0  0            999 V2000
    0.2198    0.0635    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.9343   -1.1740    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.9343   -0.3490    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    0.2198    0.8885    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.4946   -0.3490    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.2091    0.0635    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -0.4946    1.3010    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
   -1.2091    0.8885    0.0000 C   0  0  0  0  0  0  0  0  0  0  0  0
    1.6488   -1.5865    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    1.6488    0.0635    0.0000 N   0  0  0  0  0  0  0  0  0  0  0  0
    0.2198   -1.5865    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -1.9236   -0.3490    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
   -0.4946    2.1260    0.0000 O   0  0  0  0  0  0  0  0  0  0  0  0
    3.6524    0.0000    0.0000 Cl  0  0  0  0  0  0  0  0  0  0  0  0
  2  3  1  0  0  0  0
  3  1  1  0  0  0  0
  4  1  2  0  0  0  0
  5  1  1  0  0  0  0
  6  5  2  0  0  0  0
  7  4  1  0  0  0  0
  8  6  1  0  0  0  0
  9  2  2  0  0  0  0
 10  3  1  0  0  0  0
 11  2  1  0  0  0  0
```

```
 12   6   1   0   0   0   0
 13   7   1   0   0   0   0
  8   7   2   0   0   0   0
M  END
>  <SID>
NS-000000001-0001

$$$$
 . . .
```

In this sdf file, IDNUMBER = cannot be described in the output mol2 file as it is, so the numerator of the input sdf file and the output mol2 file are not linked.
Therefore, instead, the description of the property name SID will be described as ID NUMBER = in the comment line of the mol2 file.

In that case, as shown below, on the second line,
[set substitute property name to identify molecules, if "<NScode>" or "<SUPPLIERID_*>" or "<IDNUMBER>" or "<idnumber>" does tag does not exist in sdf file :]
Describe the SID and the above property name in.



If the property name of the input sdf file is blank, it is not recognizedas a property name.

> <entry name>

molecule.001

In this case, for example, replace the tag name of the sdf file in bulk with an editor, for example, to eliminate white space before using it.

> <entry_name>
molecule.001

The contents of each item are as follows.

[Convert to 3D]

item	substance
2D → 3D	If checked, it will be three-dimensional. Follow the procedure below to make it three-dimensional. AMBER GAFF2 Performs three-dimensional calculation by energy minimization calculation by force field. At that time, H atom is added and electric charge is generated. The addition of H atoms is such that acidic / basic functional groups are dissociated in water, and the charge is generated by MOPAC7 AM1. If unchecked, 3D will not be performed. At this time, the Mol2 file is output by reflecting the original structure as it is without adding H atoms or generating electric charges. If the molecule does not need to be three-dimensionalized because it has already been three-dimensionalized, uncheck it.
make conformer	Check if you want to generate a molecular conformer when making it three-dimensional. Generated for the part of the ring structure with 4 or more member rings. If a chiral center is present in the molecule, an optical isomer is also generated at the same time.

set substitute property name to identify molecules, if "<NScode>" or "<SUPPLIERID_*>" or "<IDNUMBER>" or "<idnumber>" does tag does not exists in sdf file :	When the property name of <NScode> or <SUPPLIERID_*> or <IDNUMBER> or <idnumber> does not exist in the input sdf file, describe the property value of another property name in the output mol2 file as IDNUMBER =. Enter another property name that you want to be recognized as IDNUMBER. If there is no entry, the above three prpperty names are automatically determined and set as ID NUMBER. If the above three property names do not exist and the user-input alias property name does not exist, IDNUMBER = is not added to the output mol2 file (again, it can be three-dimensionalized, but the numerator and output of the input sdf file). Molecules in mol2 files cannot be linked).

[Convert to 3D] The following are options: The specification is not required.

item	substance
Set supplier name if "<SUPPLIERNAME_*>" does not exists in sdf files	When <SUPPLIERNAME_*> does not exist in the input sdf file, the stringentered here can be recorded in the output Mol2 file as SUPPLIER = (cannot be specified per molecule). If there is a <SUPPLIERNAME_*> in the input sdf file, that description takes precedence and is describedasSUPPLIER= in the output Mol2 file. (Supplier part below).
Tag of moleculer name	Specify the tag at the beginning of the molecular name. The molecular name is the string on the next line of @<TRIPOS>MOLECULE in the output Mol2 file. (Part of MOLECULE below). This is the molecular identification ID number that the program generates independently.
SOURCE of input files	Specify the source of the input file. In the C OMMENT line ofthe output Mol2 file,it is listed asSOURCE=. (Source part below).

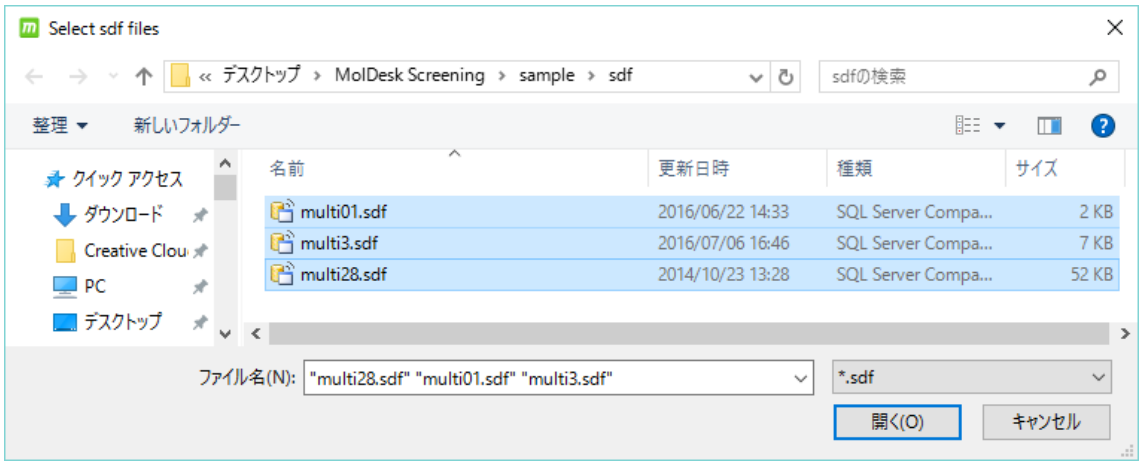
(example mol2 file description)

```
@<TRIPOS>COMMENT
 LIGANDBOX_ID = MOLECULE-00000001-01
 SUPPLIER = SUPPLIER
 SOURCE = SOURCE
 IDNUMBER = NS-000000001-0001
 MOLECULAR_FORMULA = C8H9NO4
 MOLECULAR_WEIGHT = 183.163
 MOLECULAR_CHARGE = 0
 SUM_OF_ATOMNUMBER = 96
 SUM_OF_ATOMNUMBER_MINUS_CHARGE = 96
 NUM_OF_DONOR = 5
 NUM_OF_ACCEPTOR = 4
 HOMO = -9.2167
 LUMO = -0.5693
 NUM_OF_CHIRAL_ATOMS = 1

@<TRIPOS>MOLECULE
MOLECULE-00000001-01
```

[Filtering]

item	substance
[by partial structure] Check when filtering by partial structure	Choose whether to exclude structures that are not suitable for general drugs or structures that are not suitable for pesticides.
[by moleculer weight] Check when filtering by molecular weight	Specify the minimum molecular weight and the maximum molecular weight.

[OK] Click to get a file selector, select the input file, and then click Open.



In this example, we selected three sdf files in the MolDesk Screening folder:

MolDesk Screennng -> sample -> sdf -> multi01.sdf (contains 1 compound)

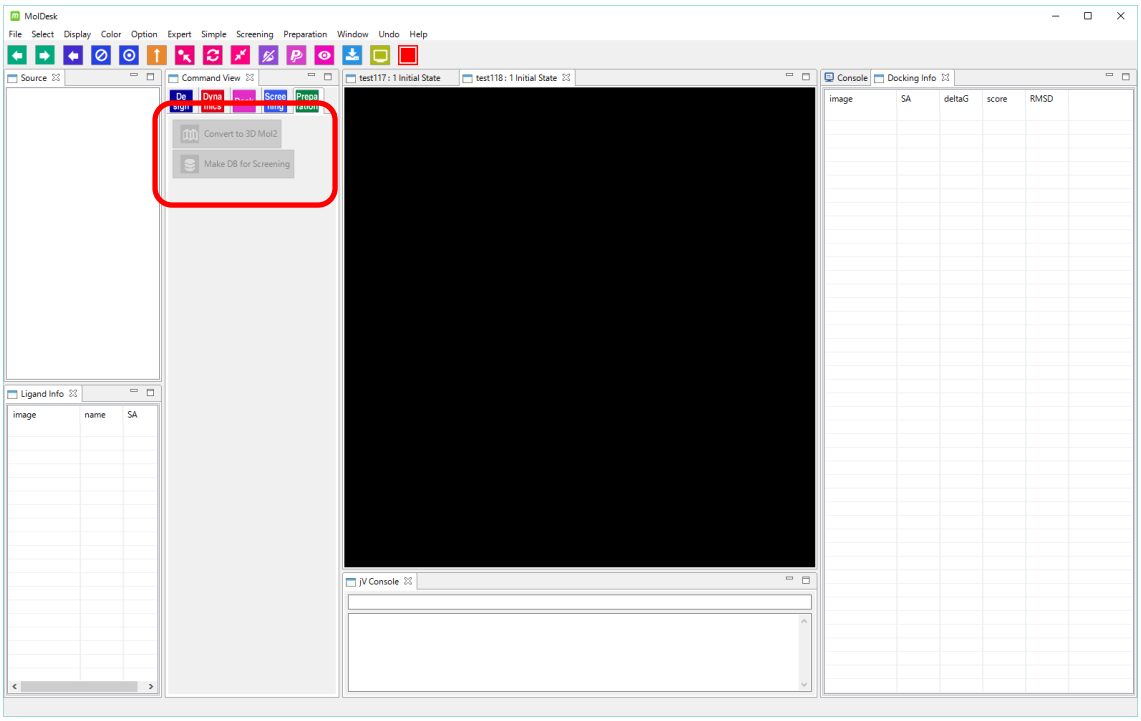
MolDesk Screennng -> sample -> sdf -> multi3.sdf (contains 3 compounds)

MolDesk Screennng -> sample -> sdf -> multi28.sdf (contains 28 compounds)

[Open] to start the calculation.

1.4.3. Database creation calculation

When you start the screening calculation, the command button is grayed out.
Calculations are in place while the command button is gray.



The calculation time for creating a compound database for screening is as shown in the table below.

Calculation method	**Intel Core i7-4790K 4.0GHz / 16GB memory / windows8.1 Run in 8 parallels**	**Xeon E5-2697 v2 @ 2.70GHz x 2 (24 cores 48 processors) / 64GB Memory / Linux CentOS6 Runs in 48 parallels**
Calculation time	641 hours (7 hours for 26 days)	191 hours (23 hours for 7 days)

Example of calculating 259,868 molecules

Multiply the actual calculation time proportionally by the number of compounds for which you want to create a database.

Although it is calculated at high speed by parallel calculation, the amount of memory required when creating a compound database increases as the number of parallels increases.

For example, 8 parallels requires 16GB and 48 parallels requires 32GB.

The number of parallels can be specified by setting the Thread number in [Help]-[Preference]-[Screening]. By default, the maximum number of processors in the machine is set.
The larger the number of parallels, the more memory is consumed, so reduce the Thread number value on a machine with a small amount of memory.

It cannot be calculated with Window 32bit (due to insufficient memory). Please prepare a machine with as good specifications as possible for Windows 64bit or Linux 64bit.

Creating a compound database for screening calculations of 300,000 molecules consumes approximately 4.5GB of storage media.
If it stops in the middle, it needs to be recalculated from the beginning.

1.4.4. Database creation location

When the database creation calculation is finished, the command button changes from gray to be available. The location where the database is created is the folder of the saved project as PROJECT.

[PROJECT] -> work -> database

Database consists of the following folder configurations:

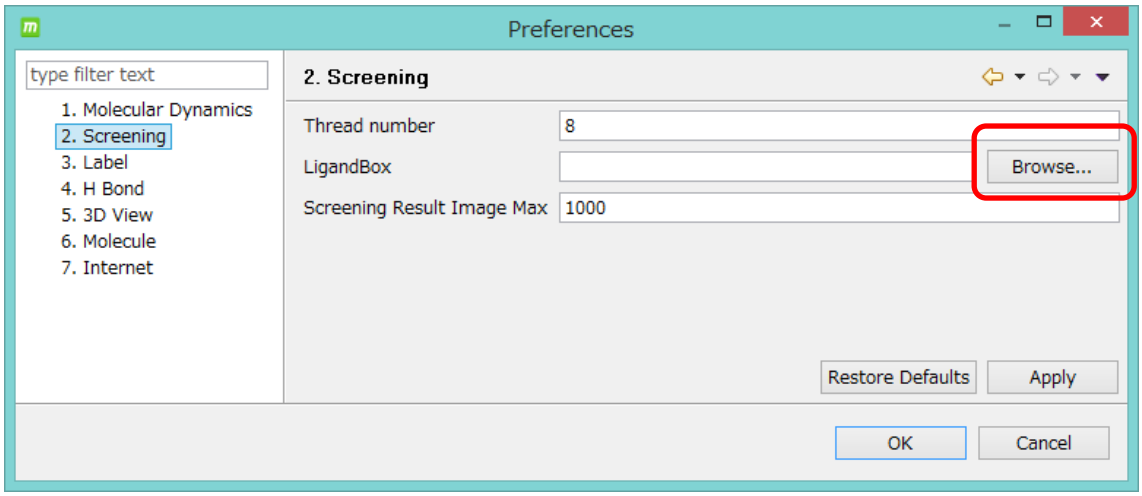
```
[PROJECT] - work - database - ligand
                            - ligandImage
                            - mol2_files
                            - mts_data
                            - protein
                            - all.mol2 (file)
                            - all_exclude.mol2 (file)
                            - error.log (file)
                            - exclude.info (file)
                            - pro_list (file)
                            - version (file)
```

The contents of each are as follows.

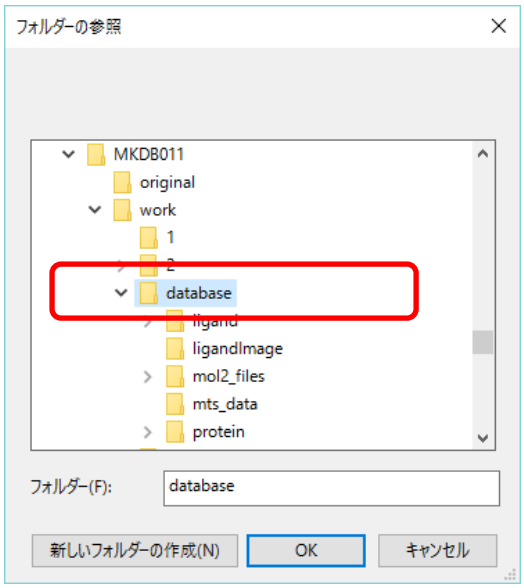
item	substance
ligand	For every 100,000 molecules, create a three-dimensional mol2 file in a folder called c***. After filtering.
ligandimage	Image files for 2D diagrams
mol2_files	Create a three-dimensional mol2 file in a folder called 3d* per input file. Before filtering.
mts_data	Interaction matrix file of compounds and 181 proteins
protein	181 Input files for docking calculations of proteins are stored in folders for each protein
all.mol2	mol2_files multi mol2 file that merges all the following mol2 files: Before filtering.
all_exclude.mol2	all. A mol2 file after filtering it by its partial structure.
error.log	Error log during 3D calculation. We can confirm molecules that could not generate a three-dimensional structure due to errors.
exclude.info	An information file for the presence of a partial structure of each molecule that is used when performing partial structure filtering
pro.list	List of 181 proteins
version	The version file of the database.

If you refer to "8.3.2 Screening" and set the database created here, you will be able to perform screening calculations for the database. Specifically, it is as follows.

Open the [Help]-[Preference] screen, select "2. Screening", and click [Browse].



Select the database folder of the saved project (MKDB011 in the example) as shown below, and click [OK].



Verify that the database folder is specified, and then click [OK].

1.5. Repartitioning of Compound DB for Screening

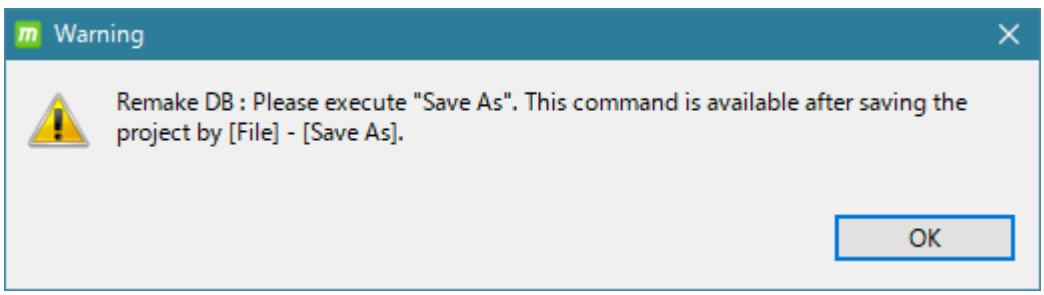
If you use LigandBox for screening calculations, skip this section.

[Preparation]-  [Remake DB for Screening] allows you to subpartition the database for screening.
The purpose of the subdivision is to speed up the screening calculation, as explained below.

The database created by  [Make DB for Screening] in the previous section and LigandBox are internally divided into 10,000 compounds. Since LigandBox has 2 million compounds, it will be divided into 200, and if the number of compounds in the database created by the user with  [Make DB for Screening] is 300,000, it will be divided into 30. Depending on the parallel calculation, there may be a computer that can perform parallel calculation with a larger number of threads in parallel. In this case, subdividing with a larger number of divisions and recreating the database will complete the screening calculation faster.

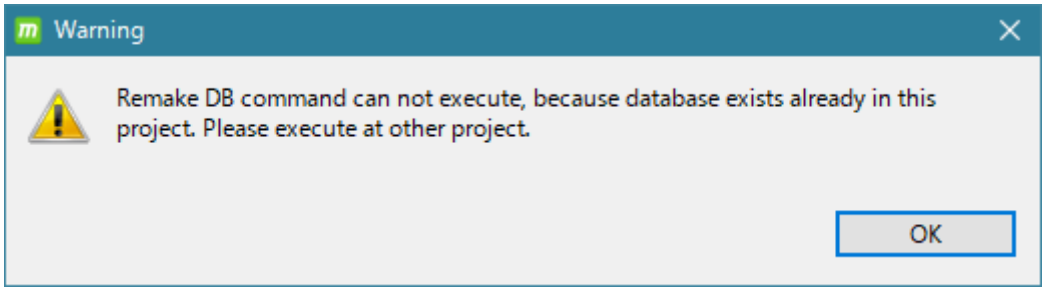
1.5.1. Save project

If you have not saved the project, a warning screen will appear prompting you to save it. Save the project using the [File] – [Save as] menu. The compound database for screening that has been subpartitioned will be created under the folder where it was saved, so save it in a location with sufficient capacity. A capacity of about 6GB is required for every 100,000 compounds.



Also, if you have already created a project and opened it with [File] – [Open Project], the

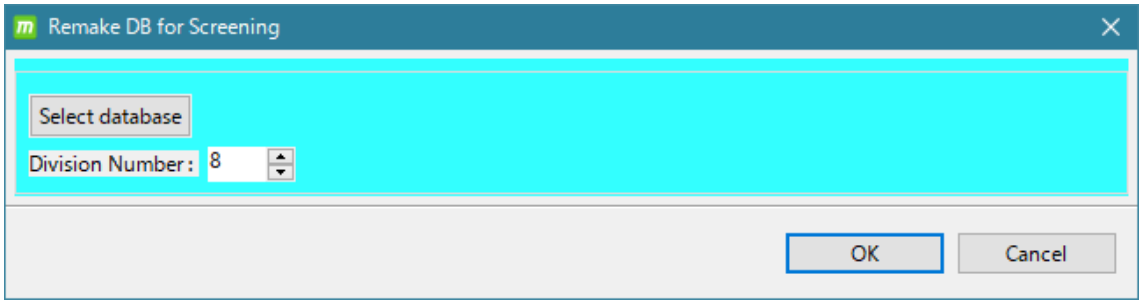
following warning screen may appear when you click  [Remake DB for Screening].



It is not feasible because you tried to create a database in a project that already has a work \database folder to be output to, so run it in another project that does not havea work \database.

1.5.2. Set up database resyn

If the project issaved and the work \database folder does not exist, you will see the following database creation condition settings screen.

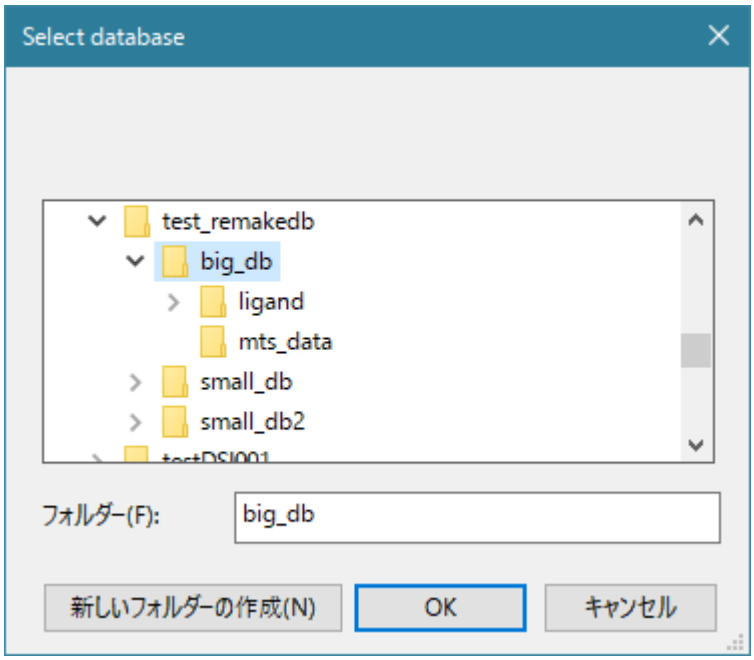


The contents of each item are as follows.

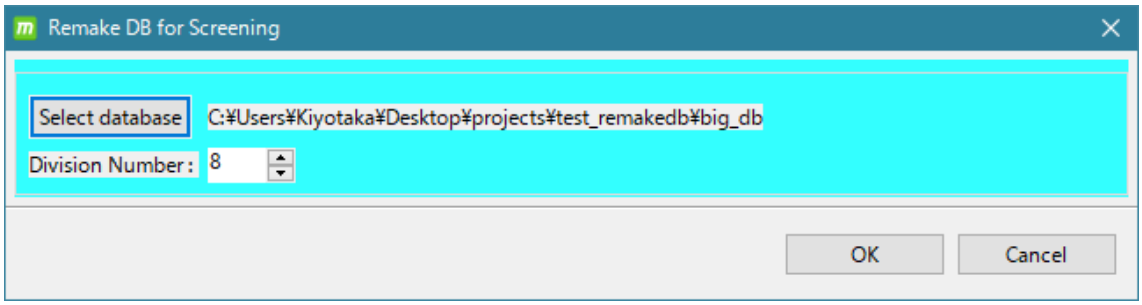
item	substance
Select database	Select the folder (directory) of the database to be re-partitioned.
Division Number	Number of splits when re-splitting

Click [Select database] to display the following folder selection screen. Select the folder directly under the ligand and mts_data folders. The required folders for the subdivision

calculation are ligand and mts_data. Other than that, it is not necessary because it is not used for subdivision.



[OK] Click to enter the selected folder as shown below.



Then select Division Number, which by default shows the maximum number of threads possible on the calculator (8 on the calculator above).

Click OK to start subdividing.

1.5.3. Database creation location

When the database creation calculation is finished, the command button changes from gray to use. The location where the database is created is set to PROJECT, if the folder of thesavedproject is PROJECT .

[PROJECT] -> work -> database

Database consists of the following folder configurations:

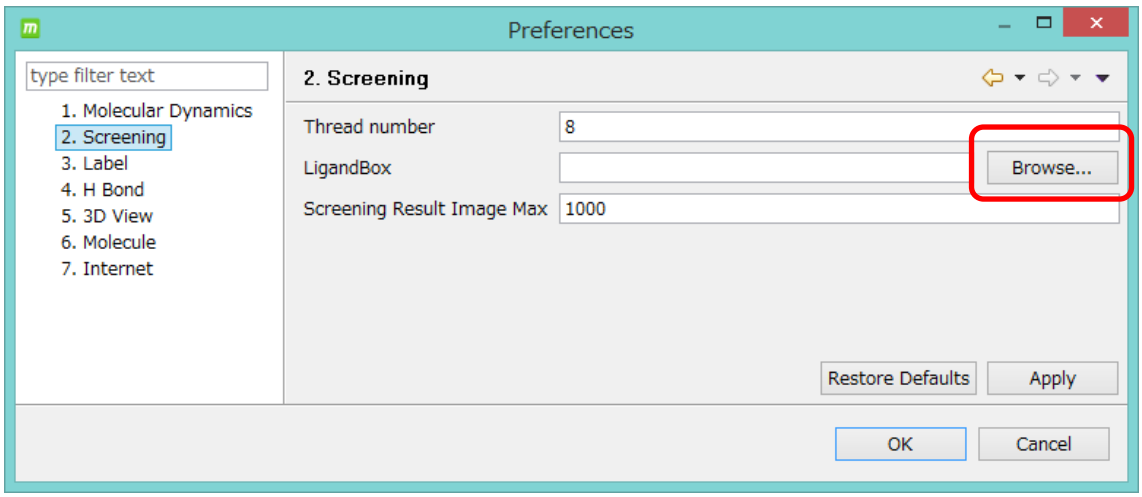
```
[PROJECT] – work – database – ligand
                                  - ligandImage
                                  - mts_data
                                  - protein
                                  - pro_list (file)
                                  - version (file)
```

The contents of each are as follows.

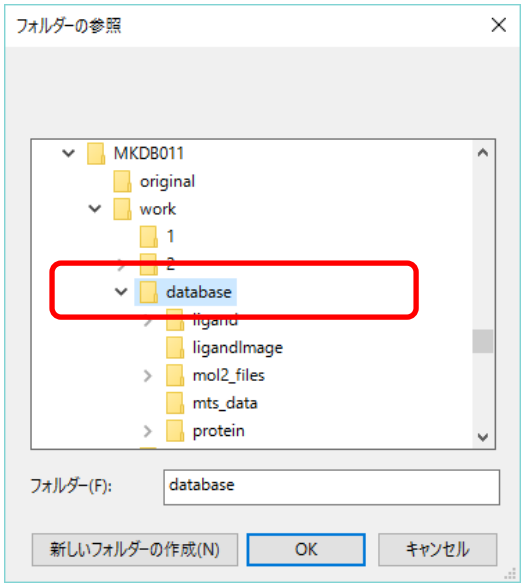
item	substance
ligand	For every 100,000 molecules, create a three-dimensional mol2 file in a folder called c***. After filtering.
ligandimage	Image files of 2D diagrams
mts_data	Interaction matrix file of compounds and 181 proteins
protein	181 Input files for docking calculations of proteins are stored in folders for each protein
pro.list	List of 181 proteins
version	The version file of the database.

If you refer to "8.3.2 Screening" and set the database created here, you will be able to perform screening calculations for the database. Specifically, it is as follows.

Open the [Help]-[Preference] screen, select "2. Screening", and click [Browse].

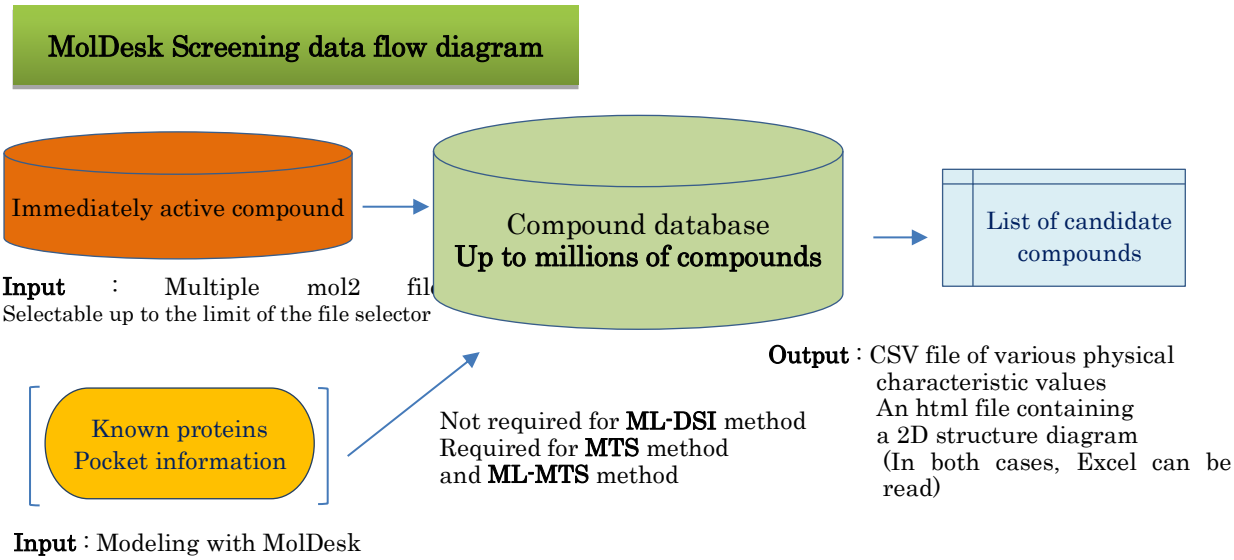


Select the database folder of the saved project (MKDB011 in the example) as shown below, and click [OK].



Verify that the database folder is specified, and then click [OK].

1.6. Screening calculation overview



※ **The compound database can be the user's molecule or the provided LigandBox.**

- In silico screening can be performed using myPresto's following techniques:
  ① Docking score order
  ② MTS Method (MTS)
  ③ Machine Learning MTS (ML-MTS)
  ④ Machine Learning DSI (ML-DSI)

- Both methods are based on docking calculations, so the most accurate active compounds have molecular weights between approximately 200 and 400 Da.
- We recommend that you use at least five active compounds for machine learning.
- Searches include ligandBox's 2 million compounds and user-added compounds.

- The input of each screening calculation method is as follows.

Calculation method	target Proteins (PDB)	Known activity ligand (mol2)	By adding users Compounds to be searched (mol2)
Docking score order	◎		○
MTS	◎		○
ML-MTS	◎	◎	○
ML-DSI		◎	○

◎Required ○ Optional

※ This sectionassumes that ligandBox is screened, but it is also possible to screen from the user's compound in the current version.

1.7. Parallel number, amount of memory, and time for screening calculations

Here is an estimate of the screening calculation time:

Calculation method	Intel Core i7-4790K 4.0GHz / 16GB memory / windows8.1 Run in 8 parallels	Xeon E5-2697 v2 @ 2.70GHz x 2 (24 cores 48 processors) / 64GB Memory / Linux CentOS6 Runs in 48 parallels
docking By score Or MTS	35 hours 31 minutes	10 hours 3 minutes
ML-MTS	45 hours 7 minutes	13 hours 12 minutes
ML-DSI	8 hours 26 minutes	2 hours 49 minutes

Example of calculation ofLigandBox + 174 compounds with 8,928 atoms on the receptor side containing proteins

No special settings are required for parallel computing (thread parallel computing is used).
Immediately after installing and activating MolDesk Screening, you can perform parallel calculations immediately.

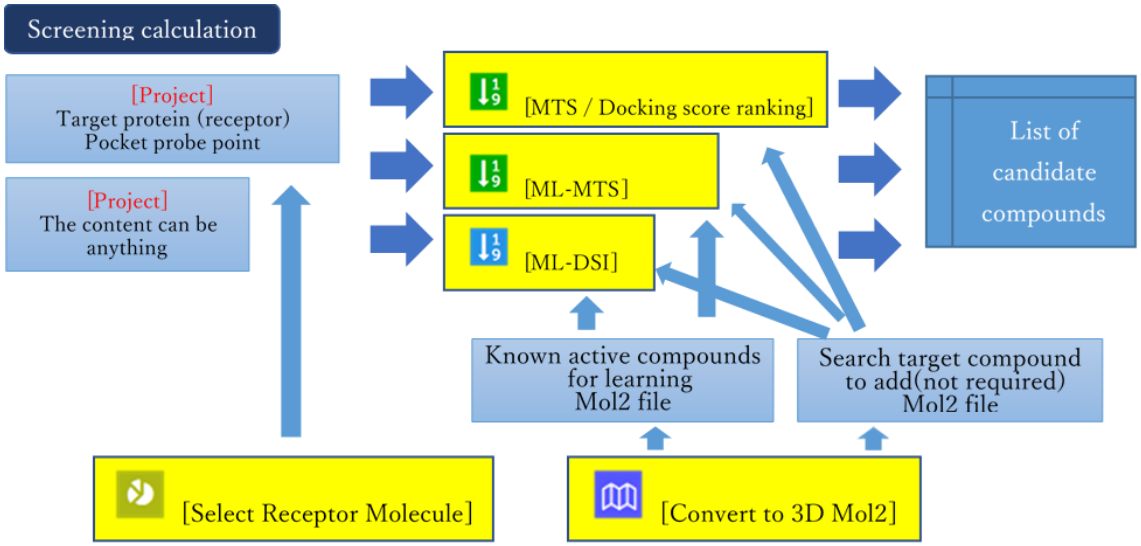
The amount of memory required for screening calculations increases as the number of parallels increases.
For example, 8 parallels requires 16GB and 48 parallels requires 32GB.

The number of parallels can be specified by setting the Thread number in [Help]-[Preference]-[Screening]. By default, the maximum number of processors in the machine is set.
The larger the number of parallels, the more memory is consumed, so reduce the Thread number value on a machine with a small amount of memory.

In case of Window 32bit, screening calculation cannot be executed due to insufficient memory.
When performing screening calculations, please prepare a machine with the best specifications of Windows 64bit or Linux 64bit.

Each screening calculation consumes about 5GB of storage media.

1.8. Screening calculation procedure



1. Prepare mol2 files

   The ML-MTS and ML-DSI methods require a mol2 file of known active ligands for learning (not required for the MTS / Docking score ranking method).

   Also, if there are compounds to be searched for that you want to add, you will also need their mol2 files.

   For the procedure for preparing the mol2 file, refer to "1.9 Preparing the mol2 file".

2. Preparing target proteins and pockets

   The MTS / Docking score ranking method and the ML-MTS method require modeling of the target protein and pocket (the ML-DSI method is a ligand-based calculation method, so modeling of the target protein and pocket is not required).

   There are two ways to create a project in which the target protein and pocket are modeled.

   A) Create a project that models the target protein and pocket probe points. Refer to the MolDesk Basic manual for how to create a pocket probe point.

   B) Load the PDB file of the target protein that has already been modeled and the PDB file of the probe point of the pocket into the project. Refer to the MolDesk Basic manual for how to read the PDB file.

After modeling the target protein and the pocket, and then select the receptor molecules in the  [Select Receptor Molecule]. At this time, the space of the pocket of the receptor, please be chosen to vacate.

3. Click any of  [MTS / Docking score ranking],  [ML-MTS], or  [ML-DSI] to display the screen entry dialog.
   Enter the mol2 file created above and the name of the target protein.

4. [OK] click to start the screening calculation.

5. The input methods for target proteins, known active regands, and compounds to be searched are as follows:

What to enter	How to enter
Target protein	Create a project, create a target protein and pocket, or enter a PDB file for the target protein you have already created and a PDB file for the pocket.
Known active rigand And Compounds to be searched	Enter one mol2 file per compound from the file selection dialog.

This documentdescribes the steps to perform screening calculations using the sample data contained in the MolDesk Screening folder created on the desktop when MolDesk Screening was installed.

1.9. Prepare mol2 files

If there are known active compounds for learning, or if there are compounds to be searched that the user wants to add, it is necessary to prepare mol2 files (three-dimensional structured ones) of them in advance.

The mol2 file can be entered in both a multi-format format in which multiple molecules are described in one file and a single format in which one molecule is described in one file.

However, since the three-dimensionalization and charge addition of compound molecules are not performed in the screening calculation, it is necessary to perform three-dimensionalization and charge addition in advance for compound molecules that do not have three-dimensionalization and charge addition. there is.

Execute  [Convert to 3D Mol2] to generate a mol2 file in which the compound molecule is three-dimensionalized and the charge is added.

For details on  [Convert to 3D Mol2], refer to the MolDesk Basic manual.

1.10. Screening by MTS method or docking score order

In the MTS method or docking score order screening calculation, the MTS method and docking score order are calculated simultaneously in a one calculation.

1.10.1. Create a project

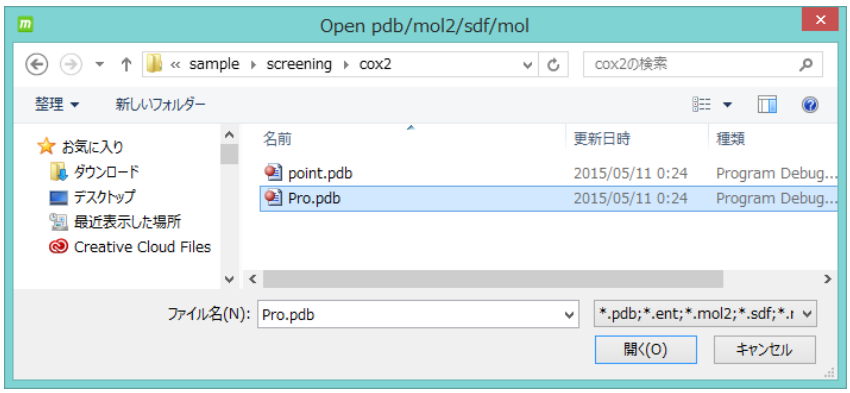
Create a project, model a target protein, and create a pocket.

This example reads protein and pocket information from a file.

You can also create a pocket with  [Make Pocket] or  [Find Pocket]. See the MolDesk Basic manual for how to create a pocket.

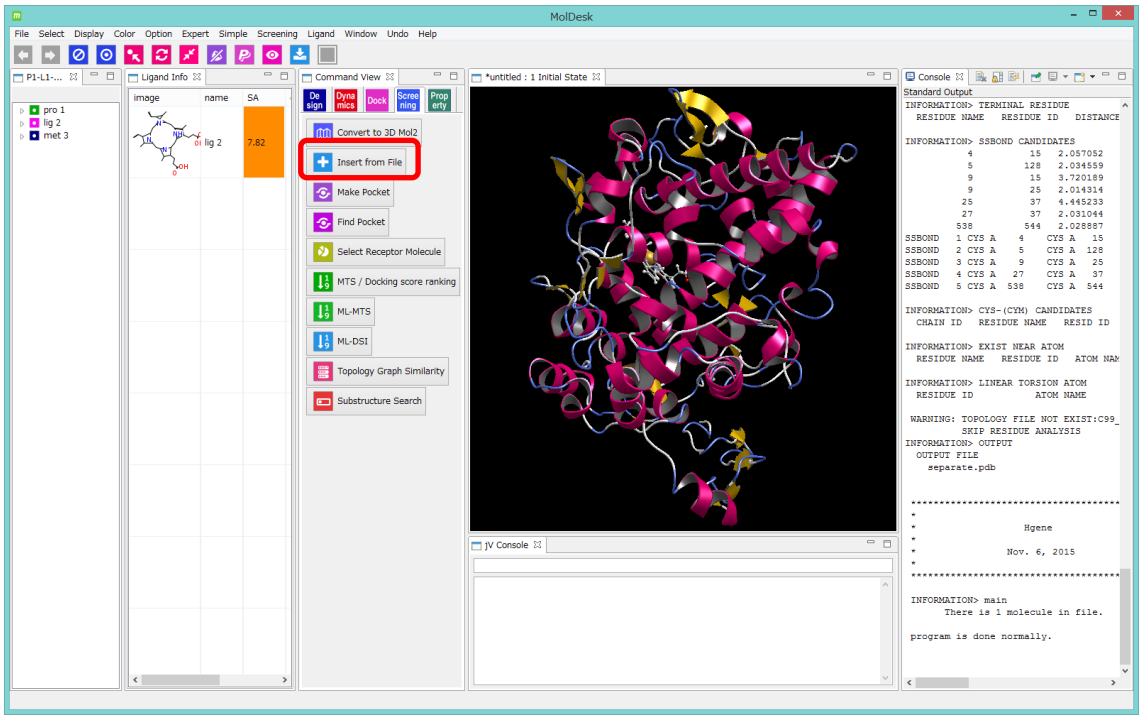
Create a project with the File-Open Molecular File menu.
In this example, select the following PDB file included in the MolDesk Screening folder to create a new project.

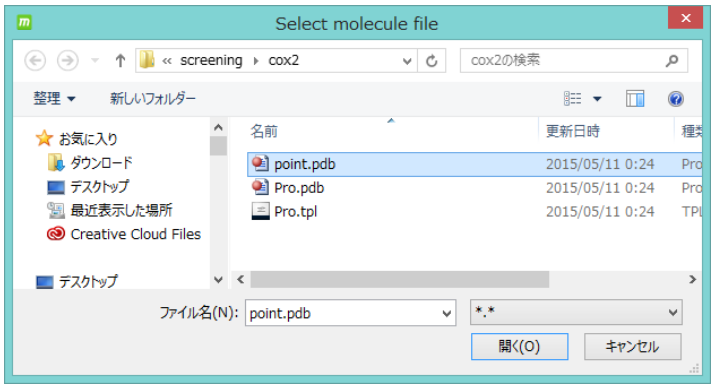
MolDesk Screenng -> sample -> screening -> cox2 -> Pro.pdb

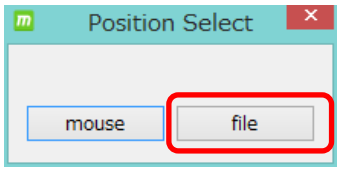


[Insert from File] click , and then select the pocket file below.

MolDesk Screenng -> sample -> screening -> cox2 -> point.pdb



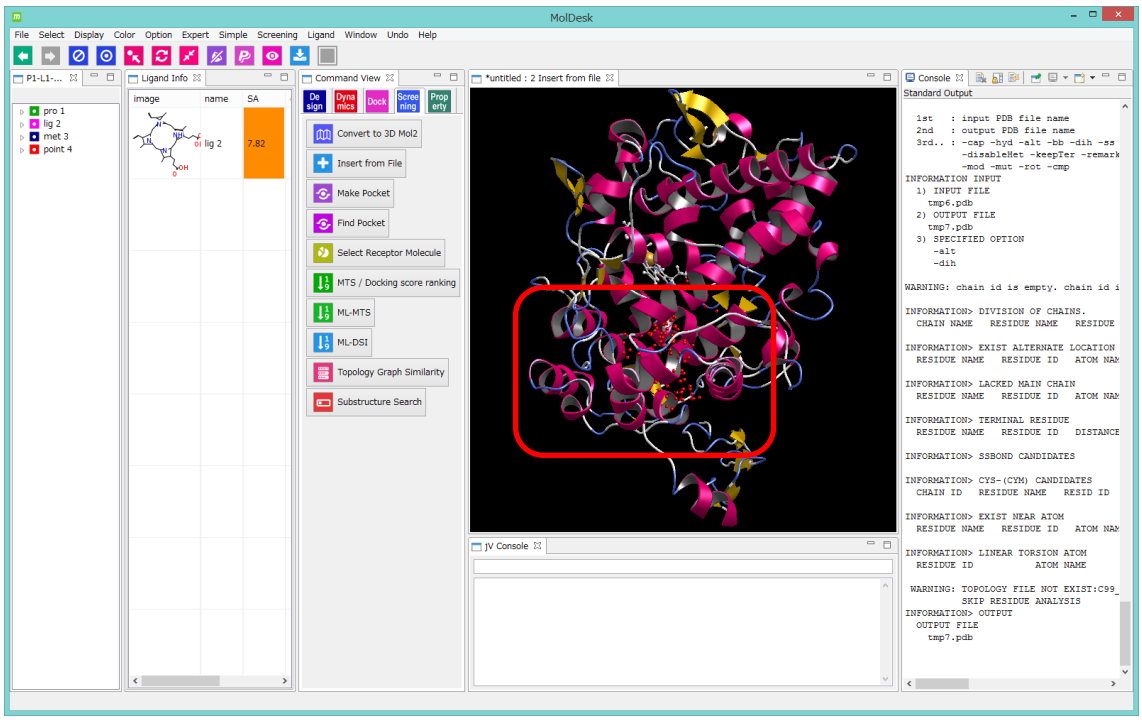




Select [file].

If you select [mouse], the pocket file will be entered at the coordinates where the user clicks the mouse, and accurate calculation will not be possible.

The pocket will be entered at the position specified in the file.

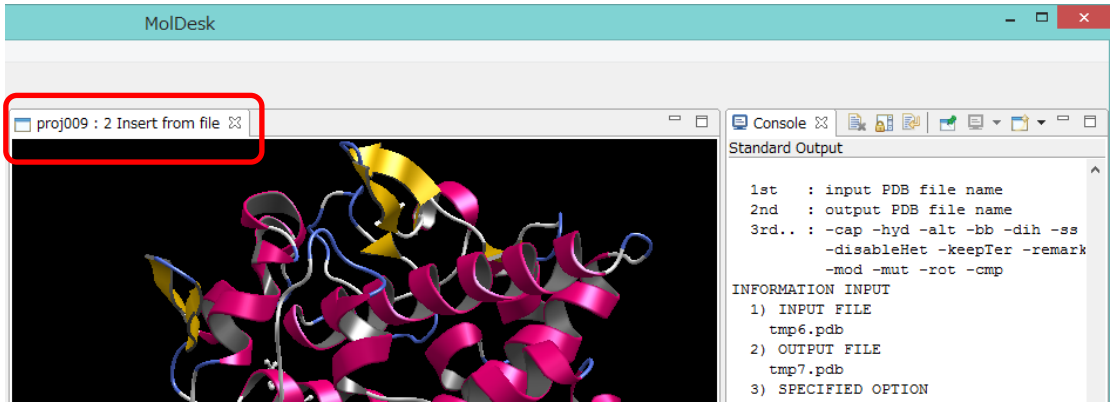


Save this project under the name "proj009".

See the MolDesk Basic documentation for instructions on how to save the project.

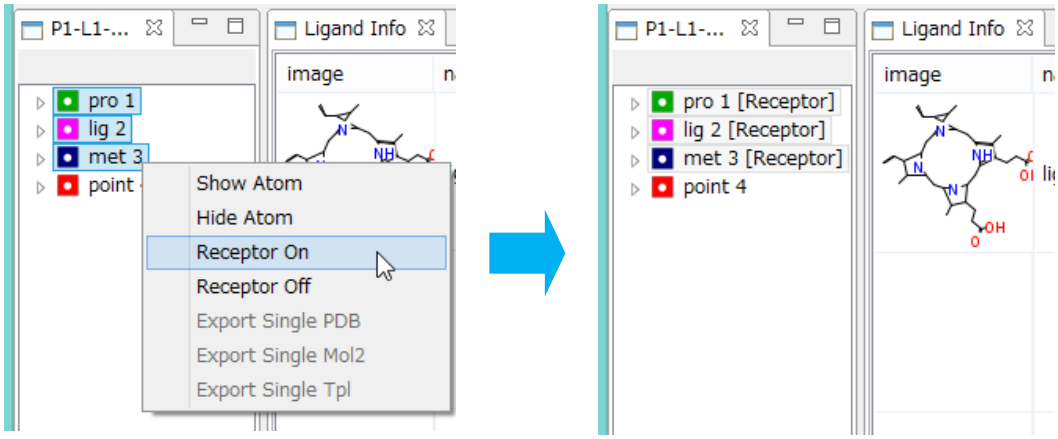
Because screening calculations generate a large amount of data in a calculation (~a few GB),you must save the project before calculatingand confirm the folder where the data will be stored.

When you save a project, the tab name on the 3D screen changes to the project name.



Specifies the receptor for the docking calculation.

Here, on the tree display screen, select pro1, lig 2, and meta 3 by Ctrl + click, and select [Receptor On] from the right click. (Since lig2 and met3 are in a place that has nothing to do with your pocket, you can select only pro 1). Receptors choose to open pocket space.



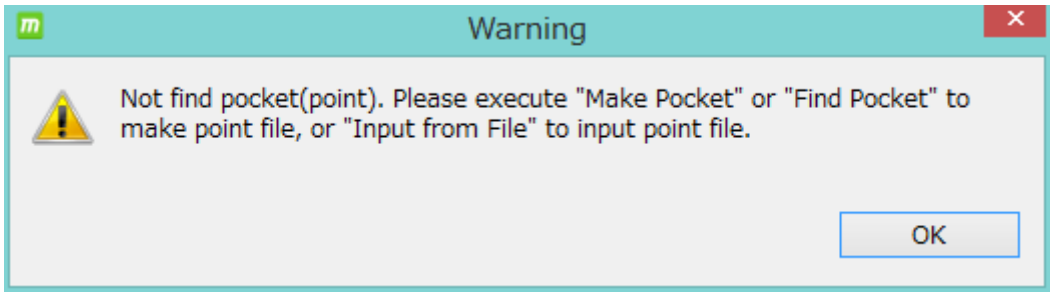
See the MolDesk Basic manual for more information on how to select receptors.

1.10.2. Data id for screening calculations

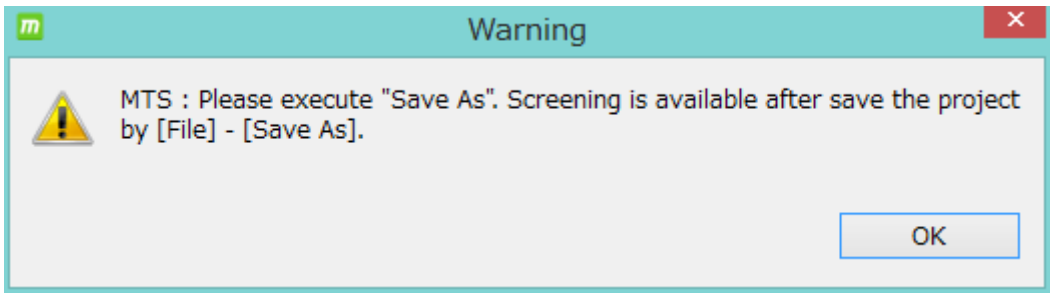
Click  [MTS / Docking score ranking].

If the required work has not been performed, the following warning dialog will be displayed. After performing the required work, click  [MTS / Docking score ranking] again.

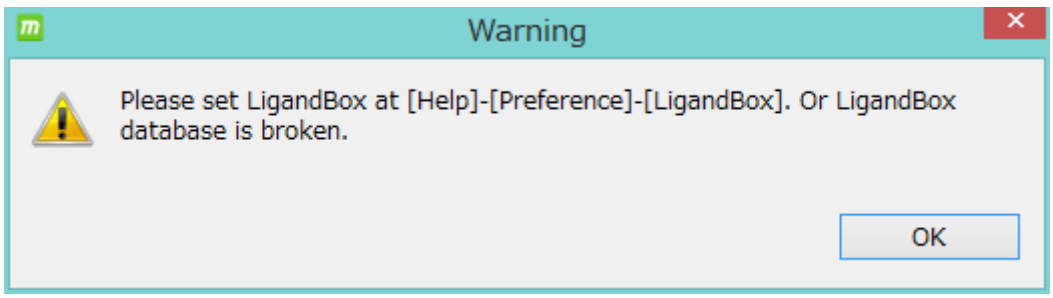
- If you have not created a pocket, the following warning dialog will be displayed, so please create a pocket.



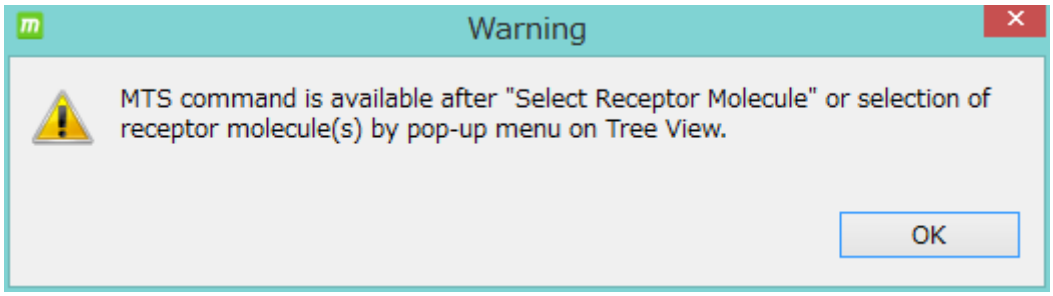
- If the project is not saved, the following warning dialog will be displayed, so please save the project.



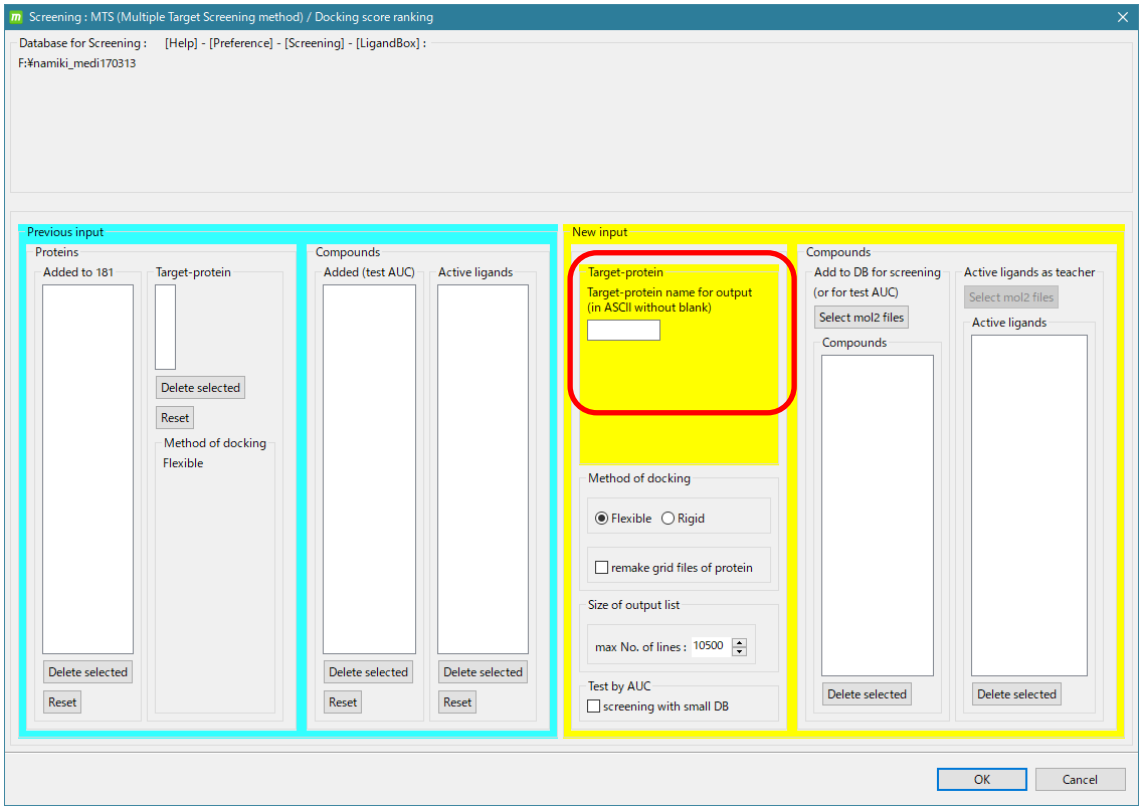
If LigandBox is not set, the following warning dialog will be displayed. Refer to "1.2 Preparing LigandBox" and set LigandBox.



- If no receptor molecules are selected, the following warning dialog will be displayed, so please select the receptor.



If all the necessary work has been done, the data entry dialog for screening calculations is displayed.

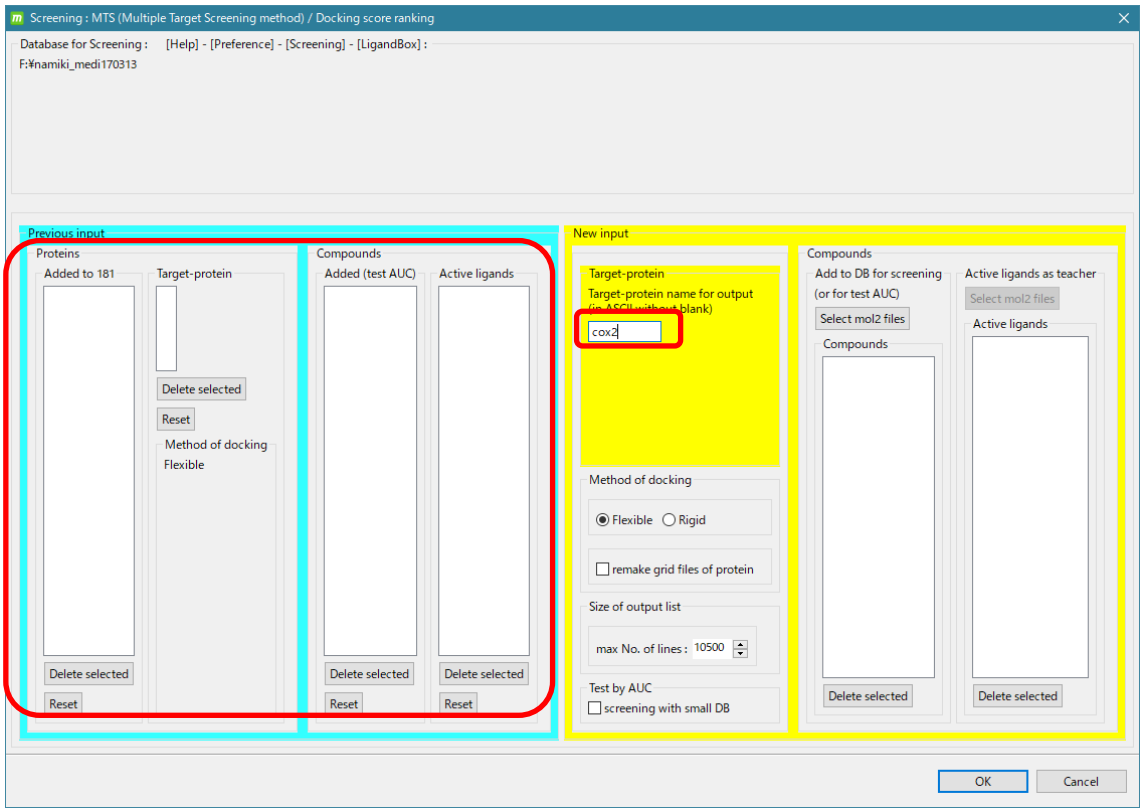


For screening calculations by the MTS method / docking score order, you must enter the name of the target protein ([Target protein]).
For the name of the target protein, enter any name in alphanumeric characters without spaces.

Others are not required.

Known active ligands ([Active ligands as teacher]) cannot be entered because they are not used in the screening calculation by the MTS method / docking score order.
In this example, enter "cox2" as the name of the target protein.

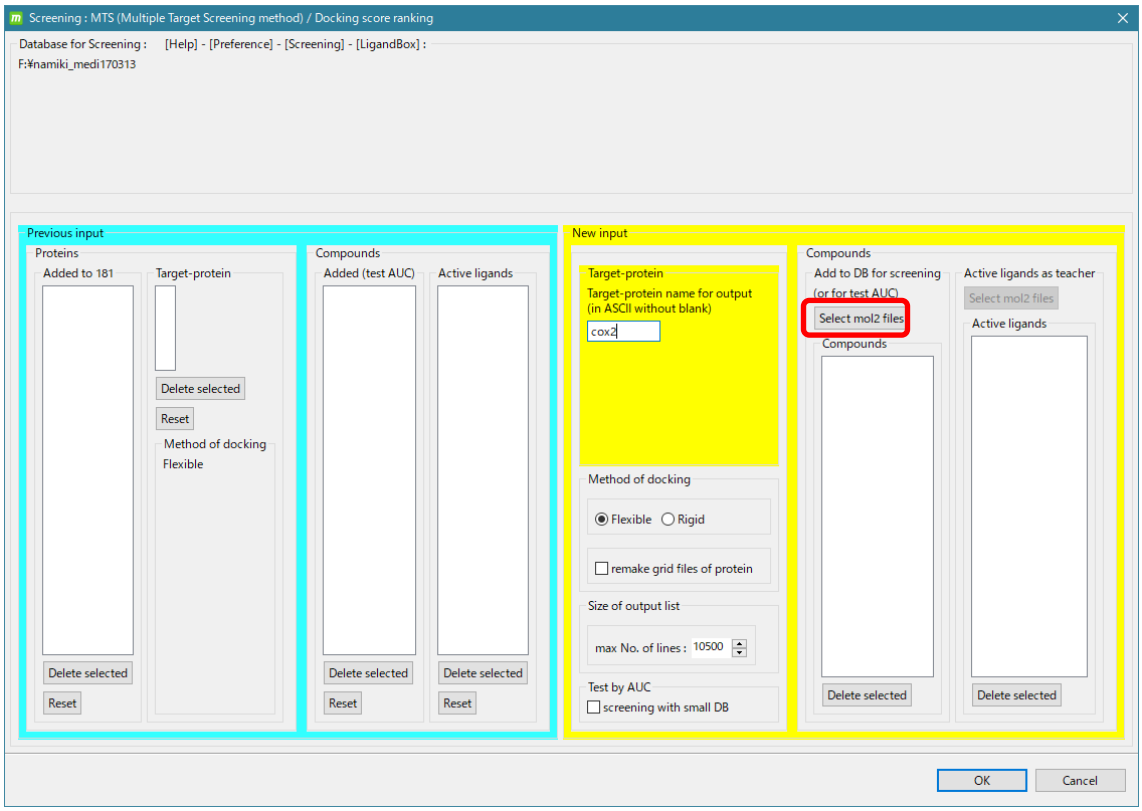
The blue background area on the left side of the dialog shows the input contents of the previous calculation. In this example, it is blank because it is the first time.



1.10.3. Adding compounds with mol2 files

In addition to LigandBox, enter the compound to be screened in a mol2 file.

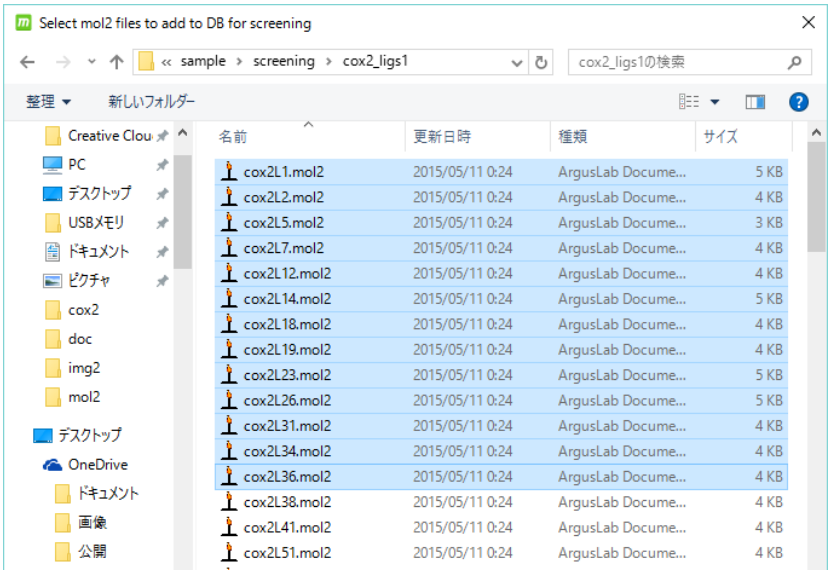
In this example, we arbitrarily enter a known active ligand and see how high the known active compounds are listed. After the screening calculation is completed, the database enrichment curve is displayed to verify the calculation accuracy by AUC (Area under the curve).

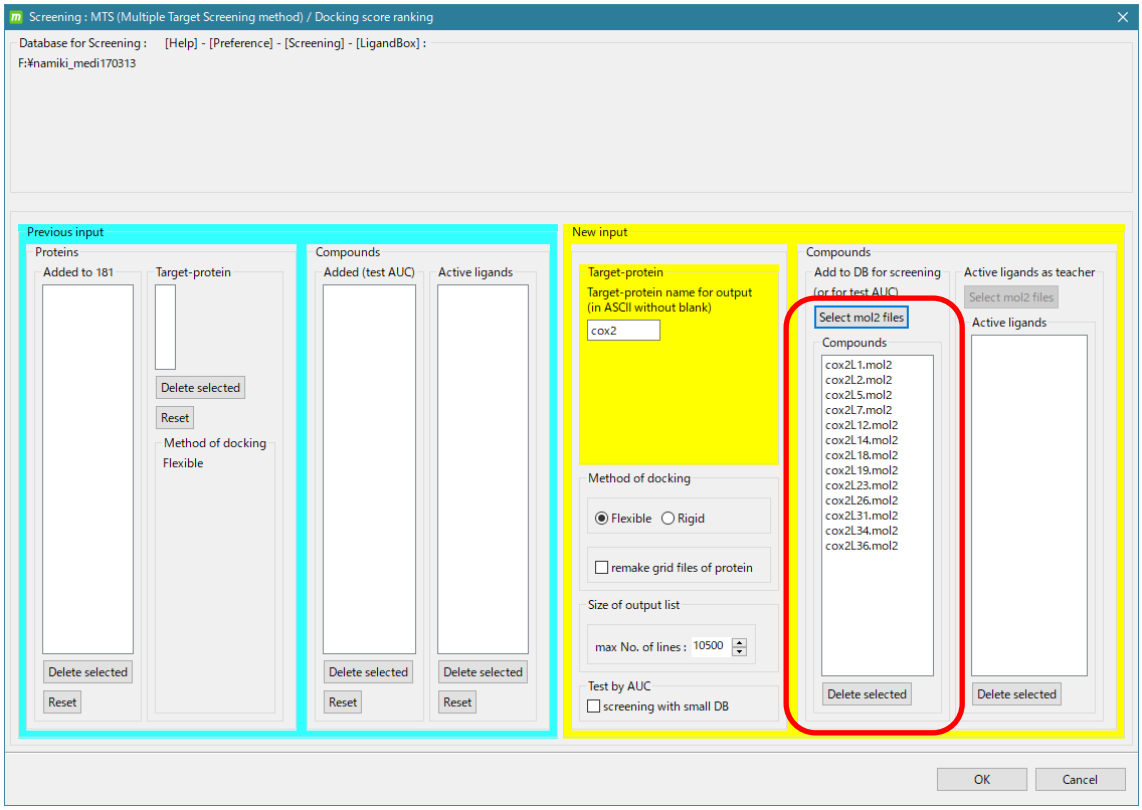


Select [Select mol2 files] of [Add to DB for screening], select the following 13 files, and click [Open].

MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L1.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L2.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L5.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L7.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L12.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L14.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L18.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L19.mol2

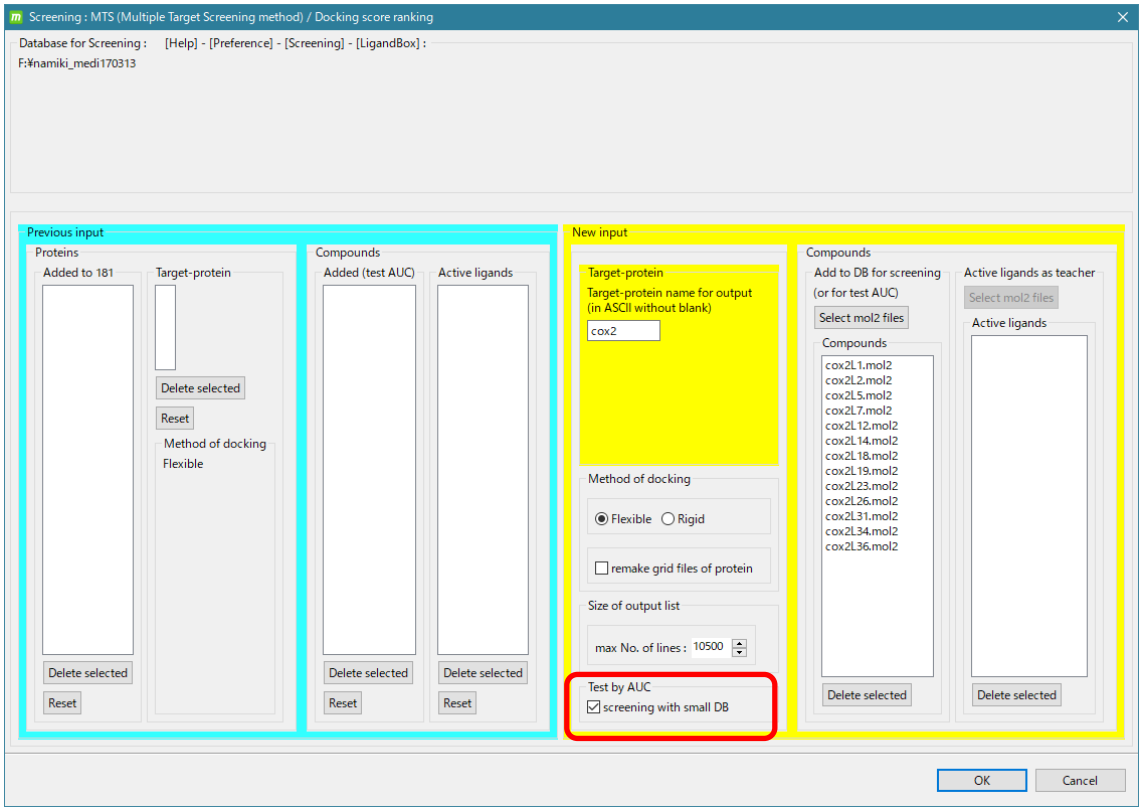
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L23.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L26.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L31.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L34.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L36.mol2





1.10.4. Enter ligandBox compounds to search for

If you check [screening with small DB], you can limit the search target compounds of LigandBox to a small group (20,000 compounds or less) for test calculation.



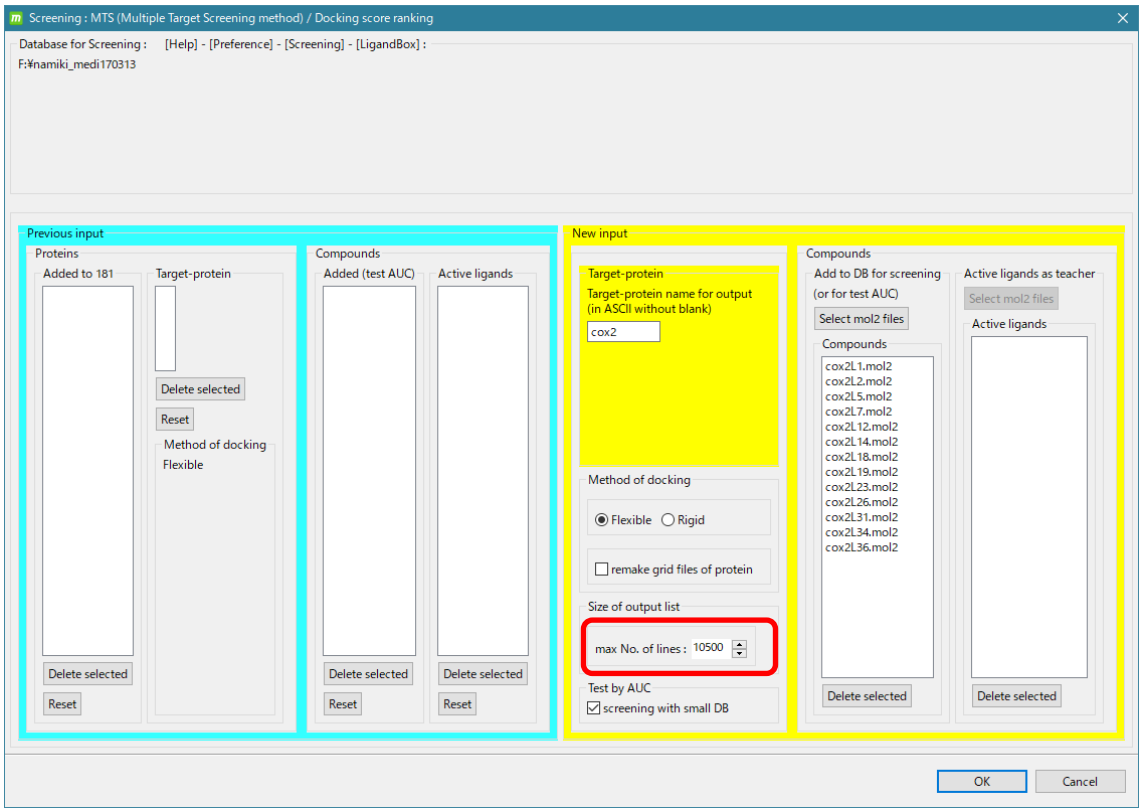
If unchecked, screening calculations will be performed for all compounds.

In this example, check [screening with small DB] to save test time.

1.10.5. Enter the size of the screening results list

[max No. of lines] is the number of results displayed after the screening calculation. The hit compounds are displayed by this number from the one with the highest score.

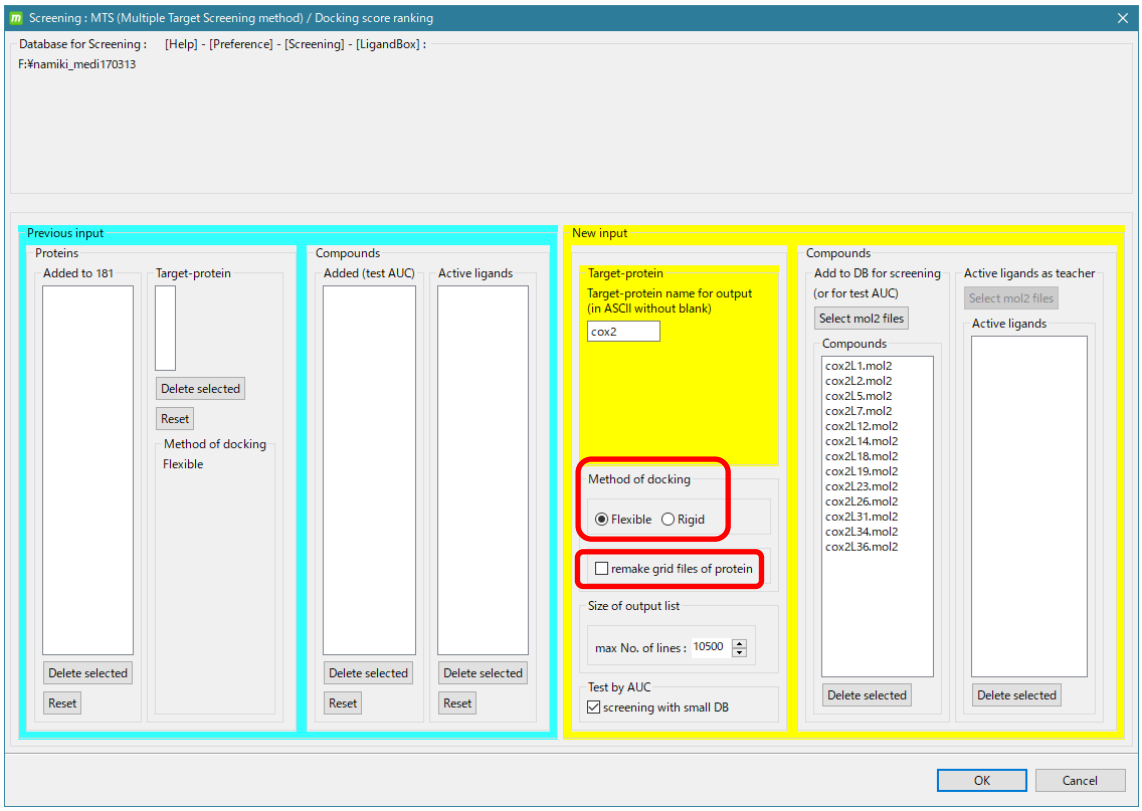
This time, the default value of 10,500 is used.



1.10.6. Enter docking calculation method

In [Method of docking], when performing docking calculation of an additionally input compound, the structure of the compound is calculated by flexible (generating a large number of candidate structures) or rigid (rigid body with the input structure as it is). Select. Normally, select flexible. Select rigid if you want to enter the representative structure of the compound and dock the ligand as a rigid body.

In [remake grid files of protein], select whether to use the grid file created in the previous calculation or recalculate and recreate the grid file for the reference protein used in the screening calculation of the MTS method. To do. Normally, leave the default and proceed without checking.



Reference protein:

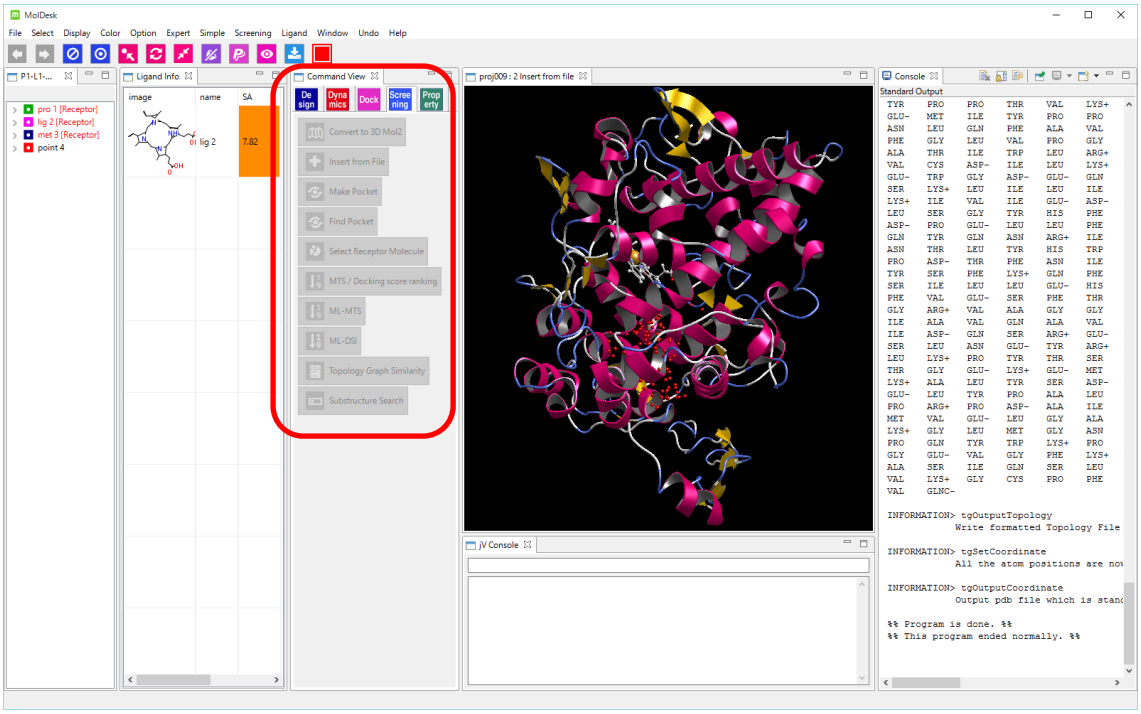
Refers to the 181 reference proteins used in the docking calculations of all screening calculations. See the myPresto manual for details.

1.10.7. Start screening calculations

Click [OK] to start the screening calculation.

In this example, the search target is limited to about 20,000 molecules, so the calculation can be completed in a few hours even on a normal PC.

The command button is grayed out when the screening calculation is started. Calculation is in progress while the command button is grayed out.



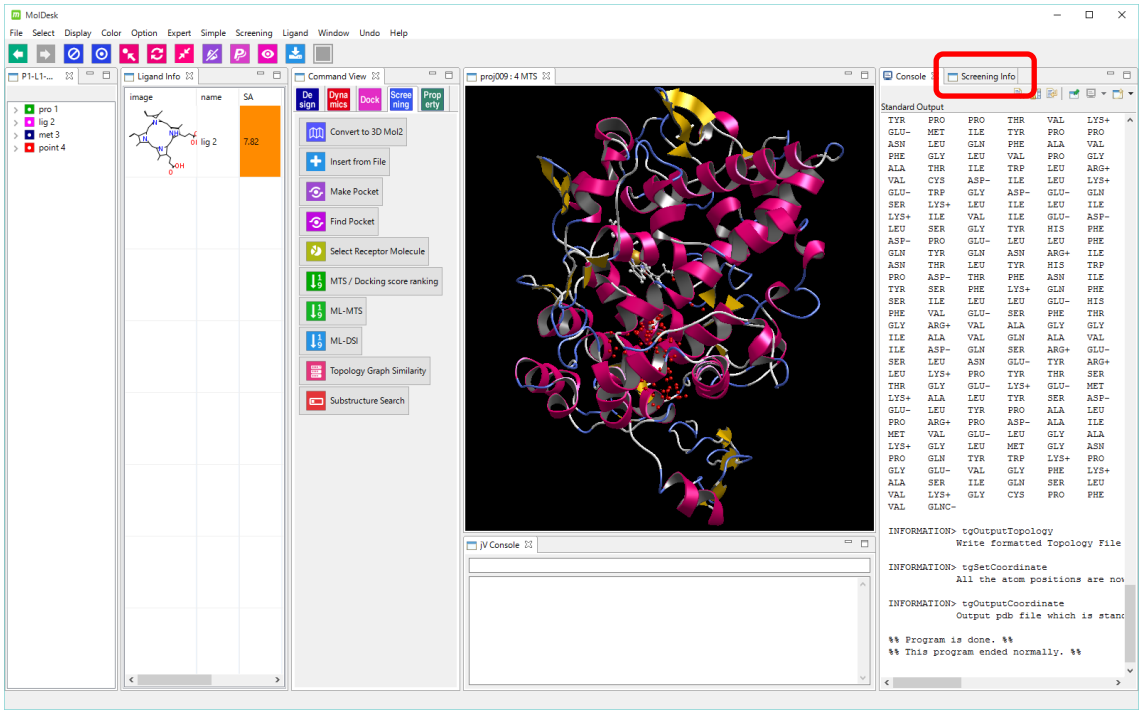
If you have another project open in advance, you can operate that project even during the screening calculation. However, depending on the processor occupancy, the operation may become extremely slow.

The number of parallel calculations for parallel calculation in the screening calculation can be set by Thread number of [Help]-[Preference] – [Screening]. For details, refer to "1.7 Number of Parallel Screening Calculations, Memory Amount, and Time".

1.10.8. Review screening calculation results

When the screening calculation is completed, the command button returns to the original display.

In addition, the screening results are listed on the [Screening Info] tab.



- If you don't see the Screening Info tab, click the Windows-Screening Info menu to display it.

If the list is difficult to see, expand the MolDesk window or drag the Screening Info tab with the mouse to make it appear alone outside the window.

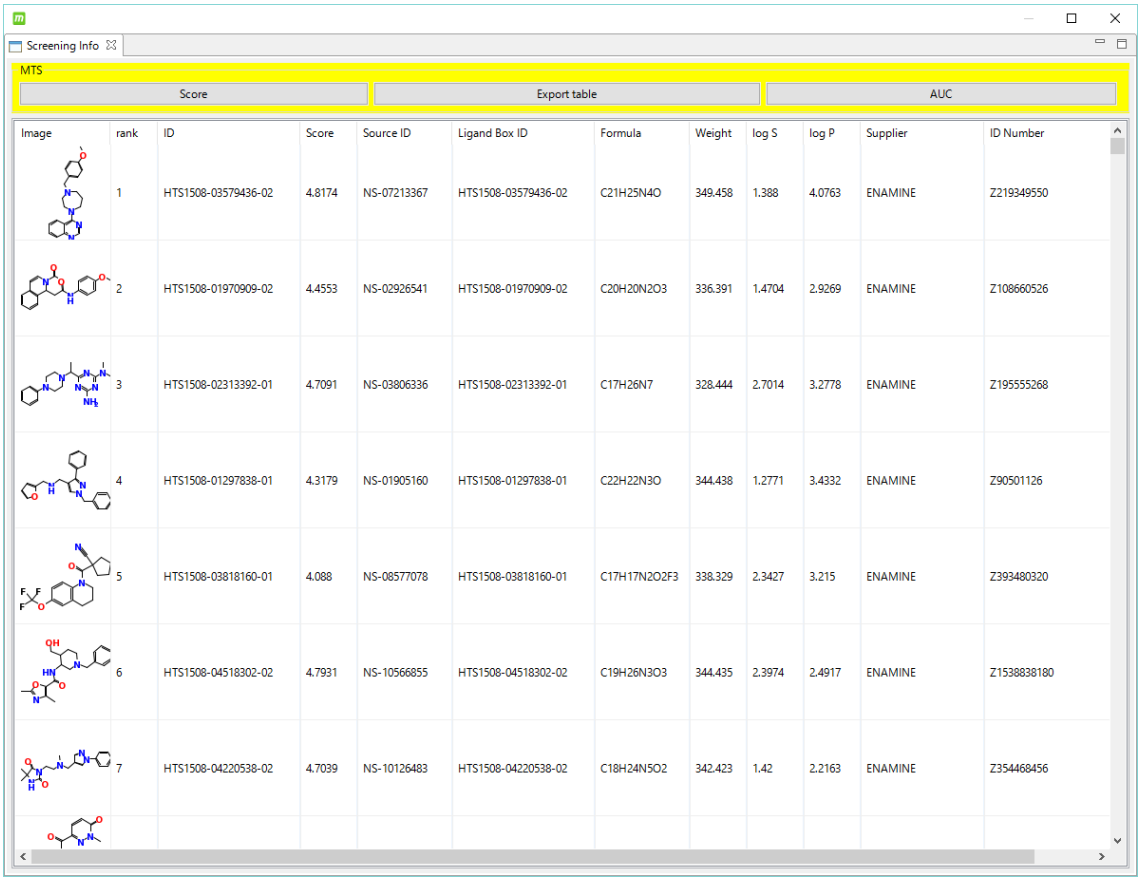
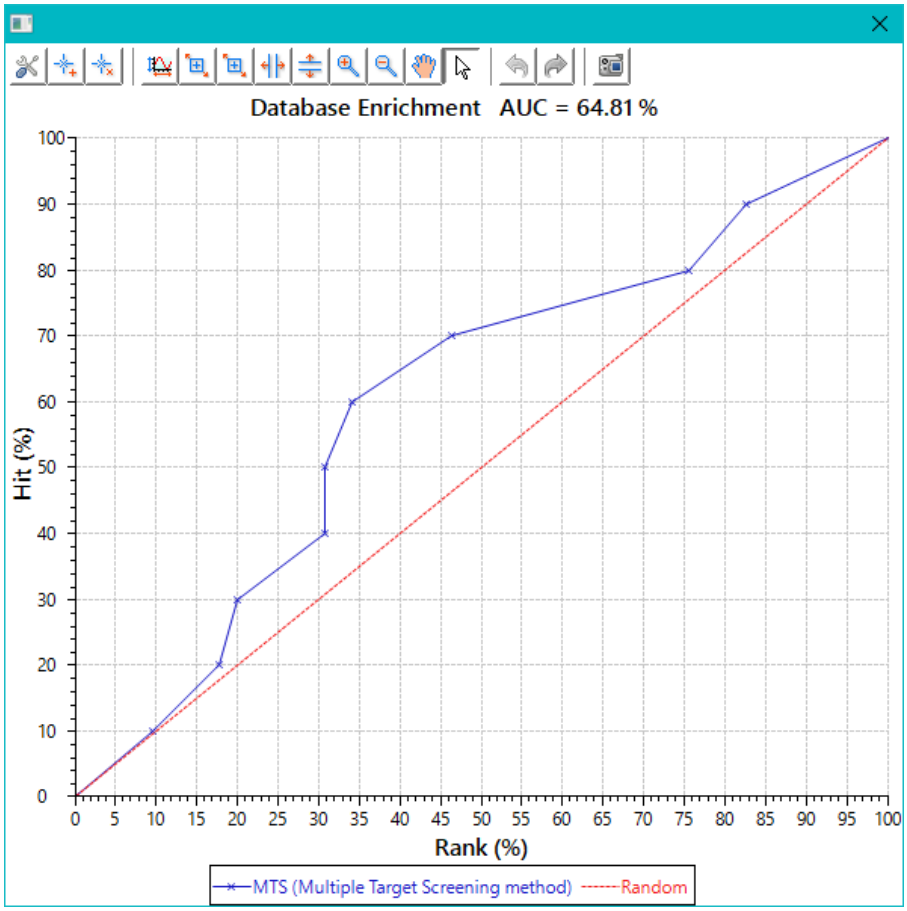


Image	rank	ID	Score	Source ID	Ligand Box ID	Formula	Weight	log S	log P	Supplier	ID Number
	1	HTS1508-03579436-02	4.8174	NS-07213367	HTS1508-03579436-02	C21H25N4O	349.458	1.388	4.0763	ENAMINE	Z219349550
	2	HTS1508-01970909-02	4.4553	NS-02926541	HTS1508-01970909-02	C20H20N2O3	336.391	1.4704	2.9269	ENAMINE	Z108660526
	3	HTS1508-02313392-01	4.7091	NS-03806336	HTS1508-02313392-01	C17H26N7	328.444	2.7014	3.2778	ENAMINE	Z195555268
	4	HTS1508-01297838-01	4.3179	NS-01905160	HTS1508-01297838-01	C22H22N3O	344.438	1.2771	3.4332	ENAMINE	Z90501126
	5	HTS1508-03818160-01	4.088	NS-08577078	HTS1508-03818160-01	C17H17N2O2F3	338.329	2.3427	3.215	ENAMINE	Z393480320
	6	HTS1508-04518302-02	4.7931	NS-10566855	HTS1508-04518302-02	C19H26N3O3	344.435	2.3974	2.4917	ENAMINE	Z1538838180
	7	HTS1508-04220538-02	4.7039	NS-10126483	HTS1508-04220538-02	C18H24N5O2	342.423	1.42	2.2163	ENAMINE	Z354468456

By default, candidate compounds are displayed in ranking order. You can sort by item by clicking each item at the top of the list.

Click the [AUC] button at the top of the list to see a graph of the database enrichment curve.

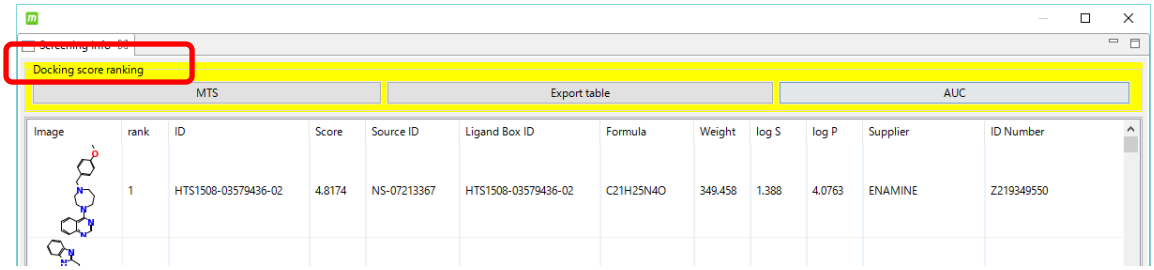


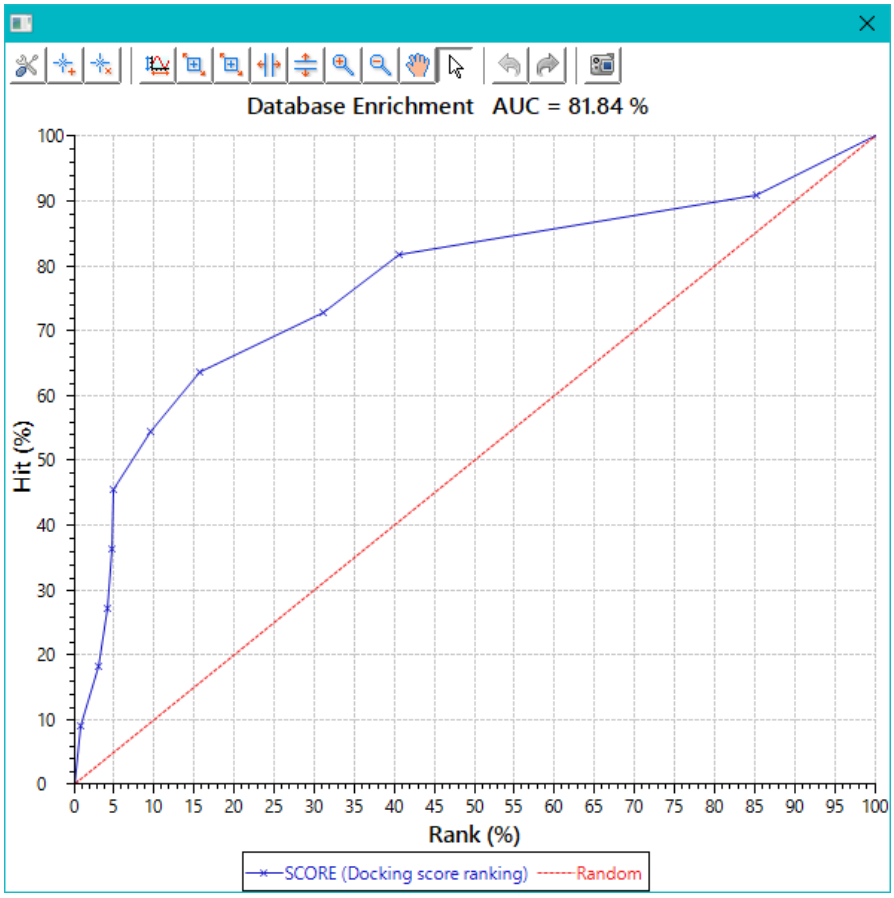
This graph can be displayed when a known active ligand is input and is used to confirm the accuracy of the method.

In this example, it was confirmed that AUC (Area under the curve) is 64.81%.
* The value will change slightly depending on the execution.

Click the Score button at the top of the Screening Info list to see the screening calculation results in order of docking score.

The title of the list changes from "MTS" to "Docking score ranking".



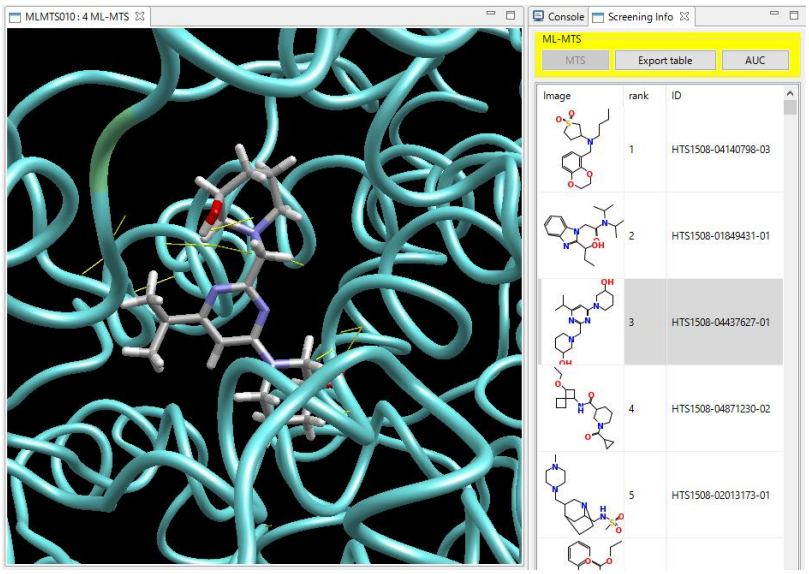


Click the [AUC] button to display the AUC of the screening calculation in order of docking score. In this example it was 81.84%.
* The value will change slightly depending on the execution.

In general, docking score-ordered calculations tend to be more accurate than MTS calculations.

1.10.9. Check docking pose

When you select a compound in the list, you can see its docking pose on the3D screen.



The selection of compounds istoggled by ↑ ↓ key, and the display of the 3D screen is also switched in conjunction.

1.10.10. File output of screening calculation results

You can output the screening results to a file.

By clicking the [Export table] button at the top of the list, you can output the data of the displayed list as a csv file (separated by commas) or an HTML file.

When outputting as an HTML file, a folder called user-specified character string .html_image is generated, and all image files are output with the file name id.png in that folder.
The output order is the default ranking order. This HTML file with image data can be read into Excel.

1.11. MI-MTS Calculation Procedure

The screening calculation for the ML-MTS method is exactly the same as the screening calculation for the MTS method / docking score order up to the point where the target protein and pocket are prepared.

1.11.1. Create a project

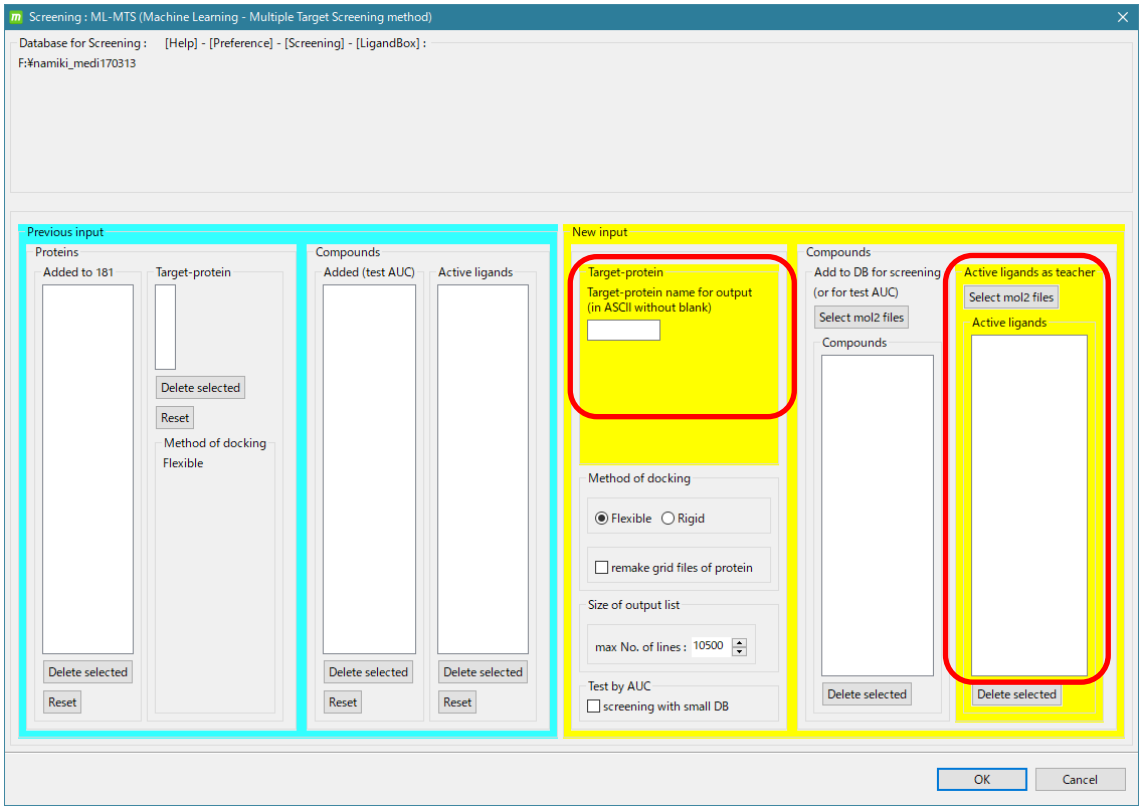
Similar to MTS　method / docking score order,1.10.1Create a project Project".

1.11.2. Enter data for screening calculations

Click  [ML-MTS].

If a warning dialog is displayed, refer to "1.10.2 Entering data for screening calculation" and perform the necessary work.

If all the required work has been done, the Screening Calculation Data Entry dialog is displayed.



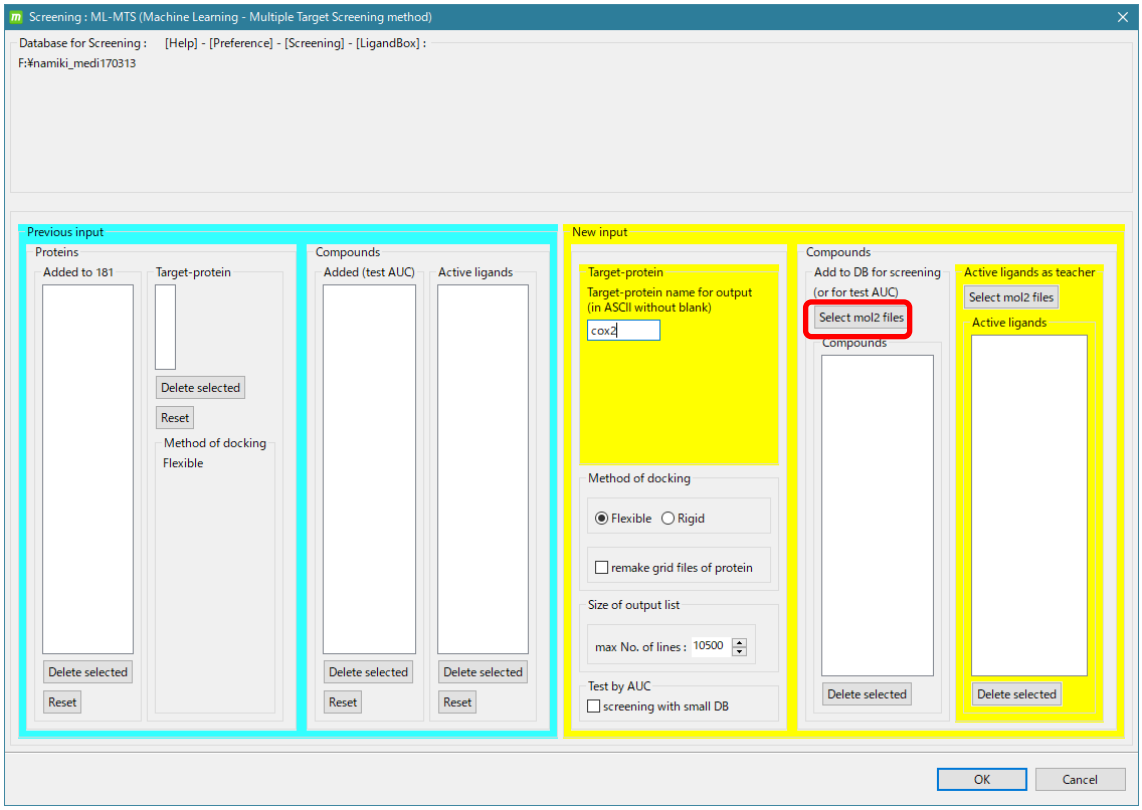
Screening calculations using the ML-MTS method require the name of the target protein ([Target protein]) and the known active ligands ([Active ligands as teacher]) to be entered.
For the name of the target protein, enter any name in alphanumeric characters without spaces.

Others are not required.

In this example, enter "cox2" as the name of the target protein.

1.11.3. Addingcompounds with m ol2 files

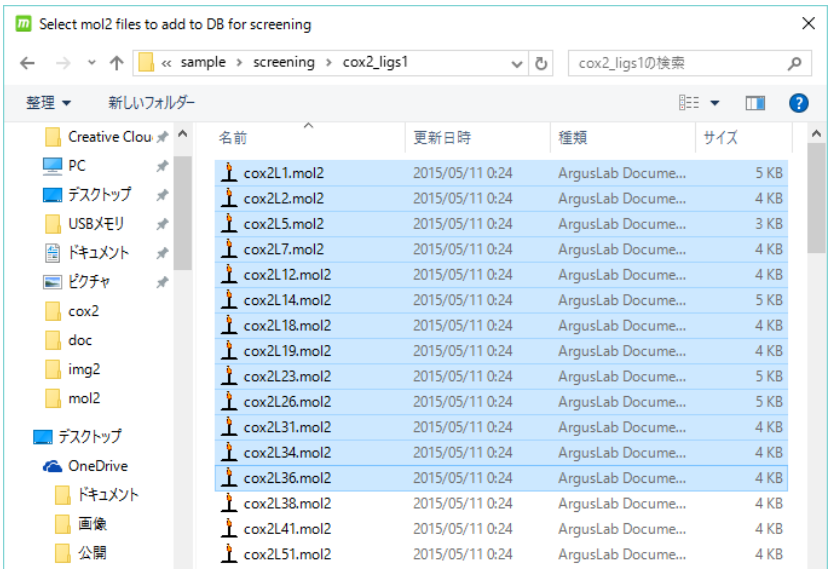
In addition to LigandBox, enter the compound to be screened in a mol2 file.

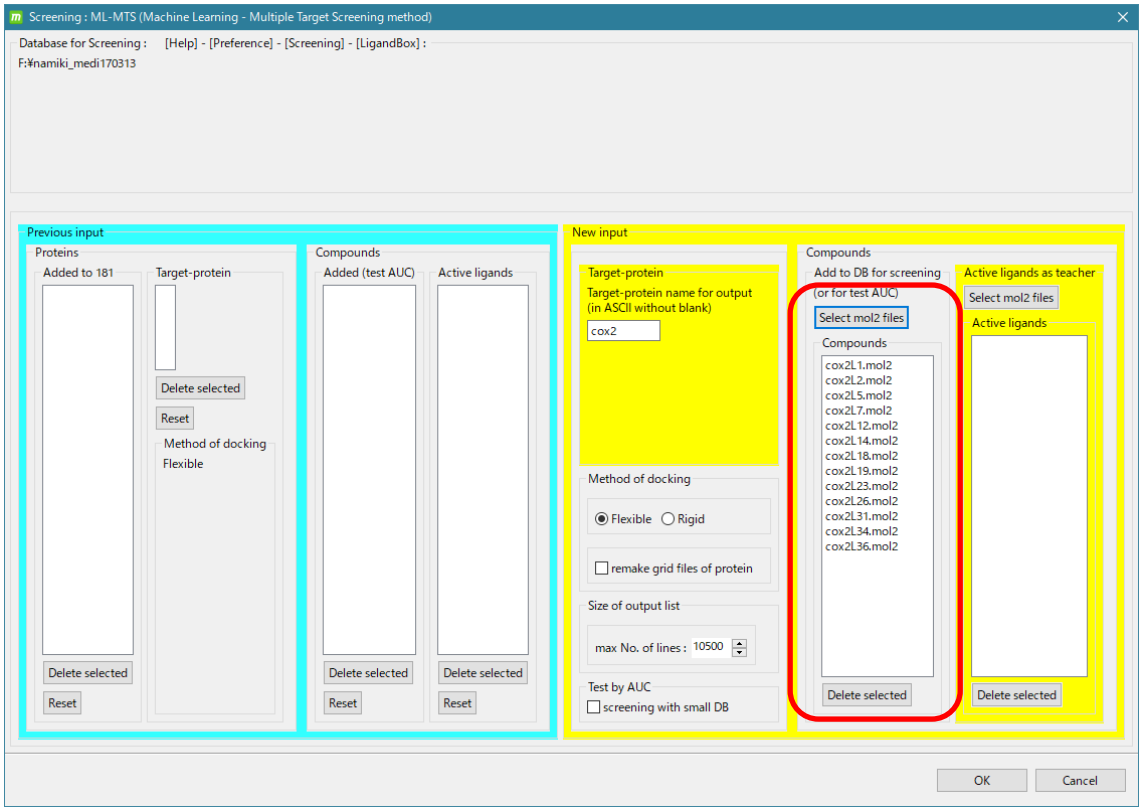


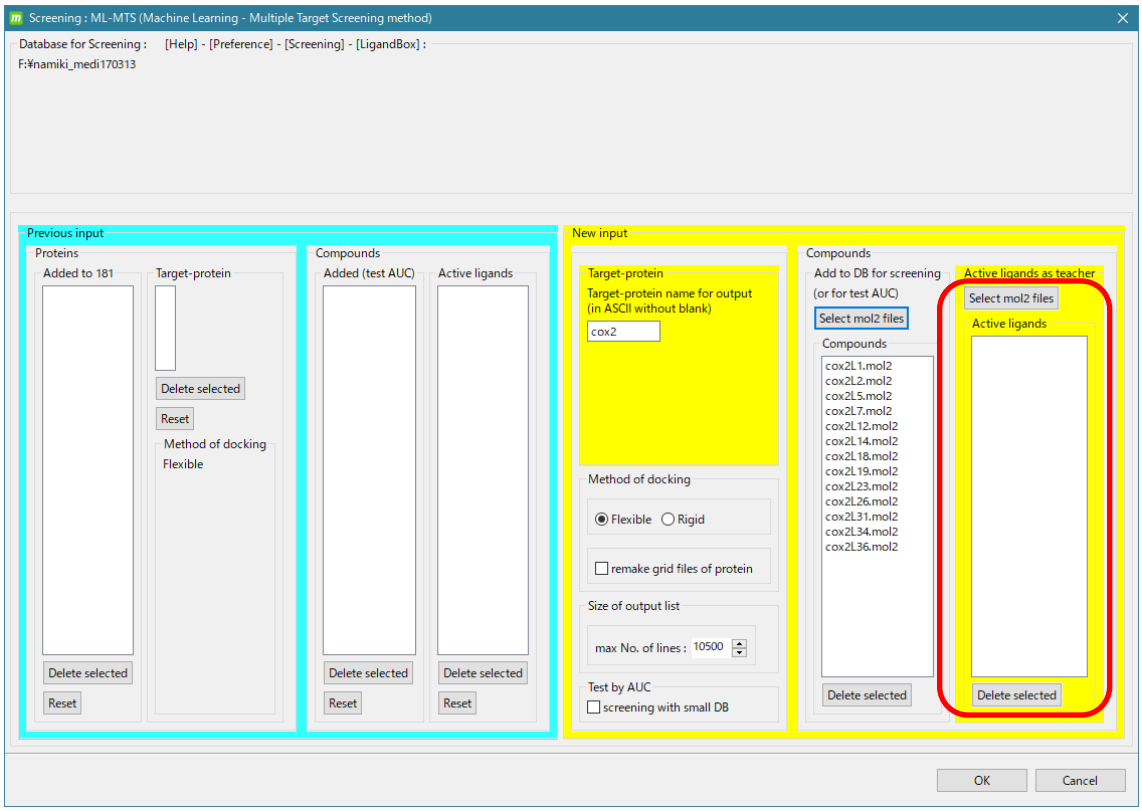
Select [Select mol2 files] of [Add to DB for screening], select the following 13 files, and click [Open].

MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L1.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L2.mol2

MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L5.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L7.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L12.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L14.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L18.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L19.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L23.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L26.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L31.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L34.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L36.mol2



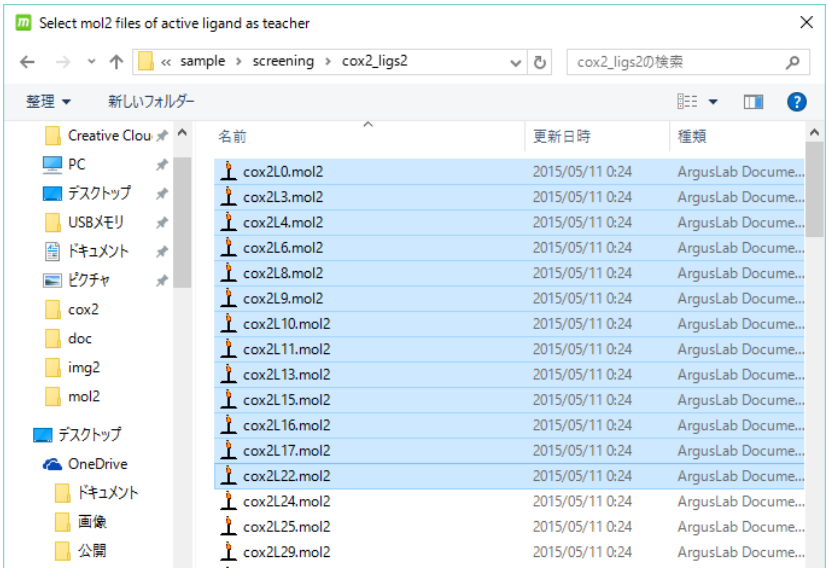


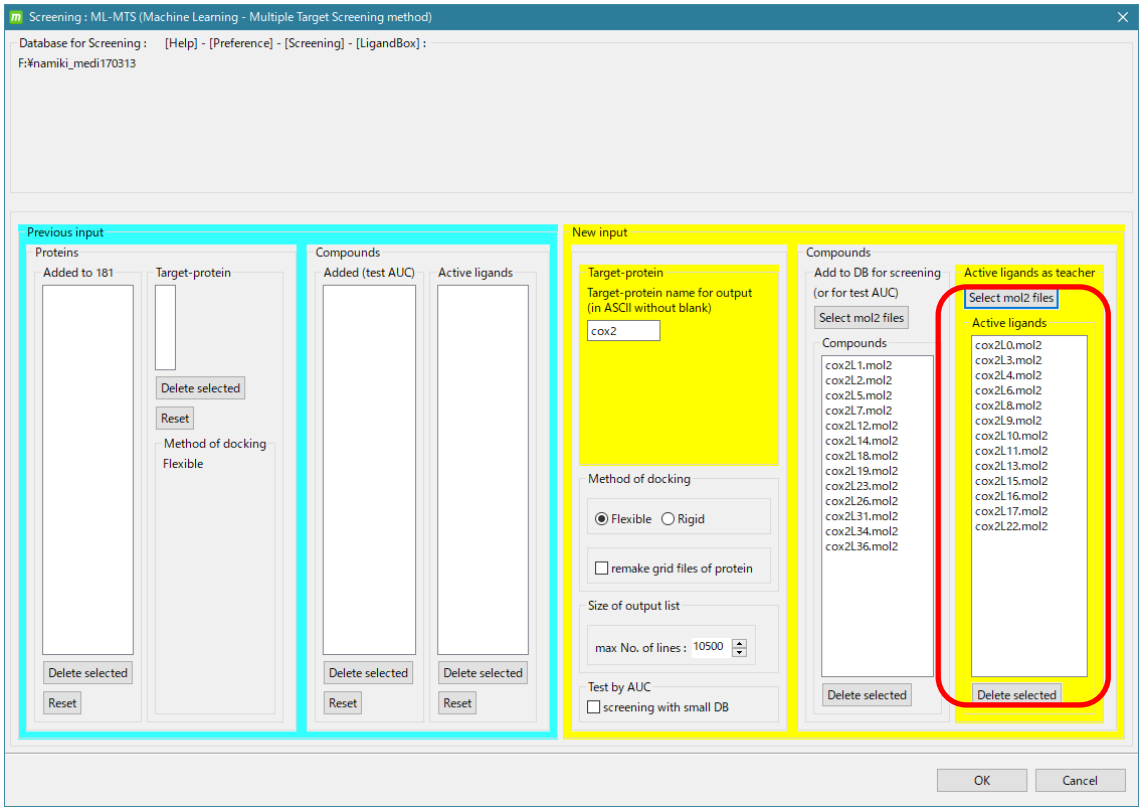


Select [Select mol2 files] under [Active ligands as teacher], select the following 13 files,

and click [Open].

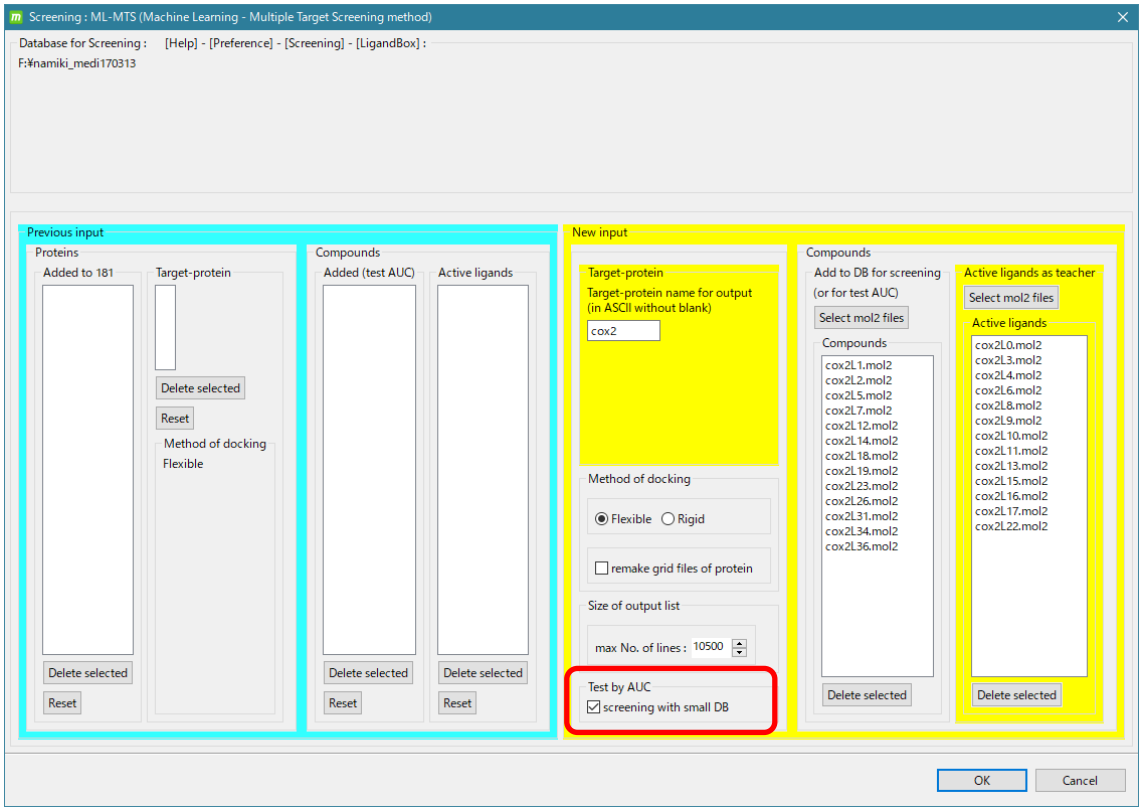
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L0.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L3.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L4.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L6.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L8.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L9.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L10.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L11.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L13.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L15.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L16.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L17.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L22.mol2




Screening : ML-MTS (Machine Learning - Multiple Target Screening method)
Database for Screening : [Help] - [Preference] - [Screening] - [LigandBox] :
F:¥namiki_medi170313
Previous input
Proteins
Added to 181
Target-protein
Delete selected
Reset
Method of docking
Flexible
Delete selected
Reset
Compounds
Added (test AUC)
Active ligands
Delete selected
Reset
Delete selected
Reset
New input
Target-protein
Target-protein name for output (in ASCII without blank)
cox2
Method of docking
Flexible
Rigid
remake grid files of protein
Size of output list
max No. of lines : 10500
Test by AUC
screening with small DB
Compounds
Add to DB for screening (or for test AUC)
Select mol2 files
Compounds
cox2L1.mol2
cox2L2.mol2
cox2L5.mol2
cox2L7.mol2
cox2L12.mol2
cox2L14.mol2
cox2L18.mol2
cox2L19.mol2
cox2L23.mol2
cox2L26.mol2
cox2L31.mol2
cox2L34.mol2
cox2L36.mol2
Delete selected
Active ligands as teacher
Select mol2 files
Active ligands
cox2L0.mol2
cox2L3.mol2
cox2L4.mol2
cox2L6.mol2
cox2L8.mol2
cox2L9.mol2
cox2L10.mol2
cox2L11.mol2
cox2L13.mol2
cox2L15.mol2
cox2L16.mol2
cox2L17.mol2
cox2L22.mol2
Delete selected
OK
Cancel

1.11.4. Enter ligandBox compounds to search for

If you check [screening with small DB], you can limit the search target compounds of LigandBox to a small group (20,000 compounds or less) for test calculation.



If unchecked, screening calculations will be performed for all compounds.

In this example, check [screening with small DB] to save test time.

Click OK to start the screening calculation.

1.11.5. Confirmation of screening calculation results, and others

Other setting items, screening calculation result confirmation method, file output method, etc. are almost the same as in the case of screening calculation by MTS method / docking score order.
Refer to "1.10.5 Entering the size of the screening result list" to "1.10.10 Outputting the screening calculation result to a file".

1.12. MI-DSI Calculation Procedure

The ML-DSI method is a regand-based screening technique that allows the target protein to
It is not necessary.

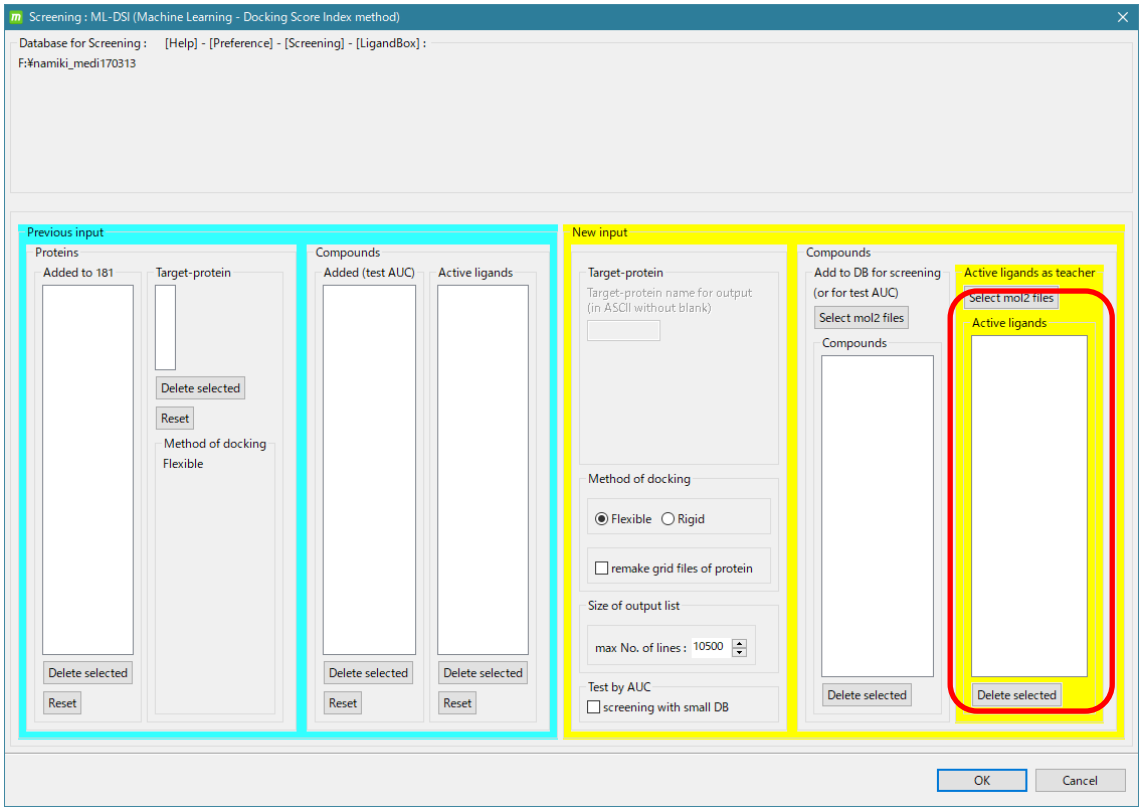
1.12.1. Create a project

In the File-New Project menu, create an empty project and save it.
Refer to the MolDesk Basic manual for how to save the project.

1.12.2. Enter data for screening calculations

Click  [ML-DSI].
If a warning dialog is displayed, click "1.10.2 Screening calculation data entry".
Please refer to it and perform the necessary work.

If all the required work has been done, the Screening Calculation Data Entry dialog is displayed.



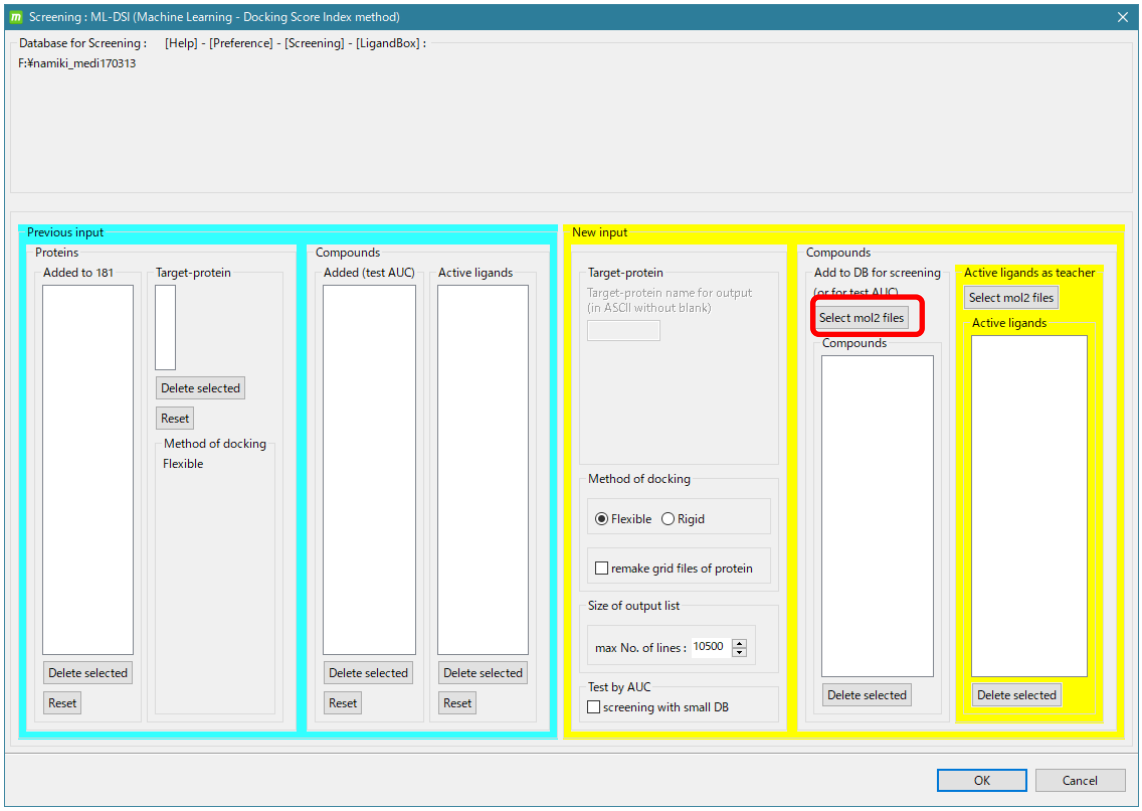
For screening calculations by the ML-DSI method, it is necessary to enter known active ligands ([Active ligands as teacher]).

Others are not required.

The target protein ([Target protein]) is calculated by the ML-DSI method.
Since it is not used, it cannot be entered.

1.12.3. Addingcompounds with m ol2 files

In addition to LigandBox, enter the compound to be screened in a mol2 file.



Select [Select mol2 files] of [Add to DB for screening], select the following 13 files, and click [Open].

MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L1.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L2.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L5.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L7.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L12.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L14.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L18.mol2

(Next page)

MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L19.mol2

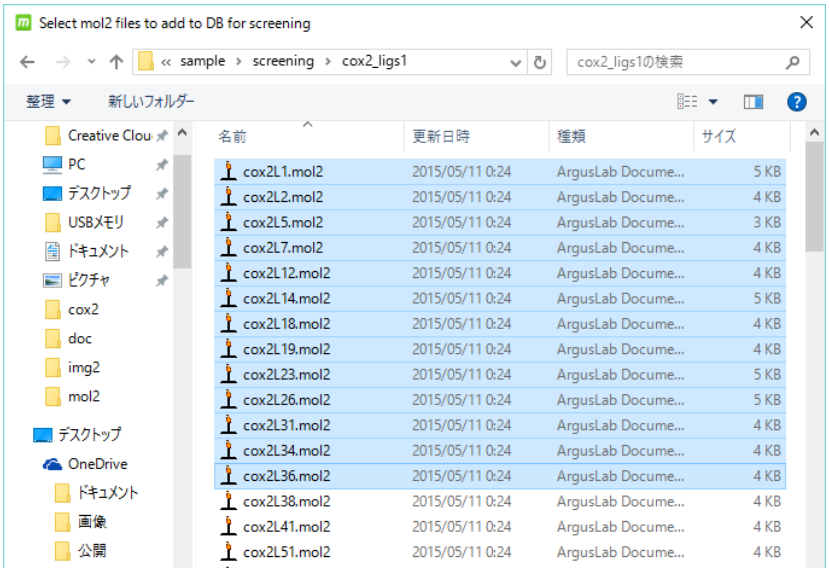
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L23.mol2

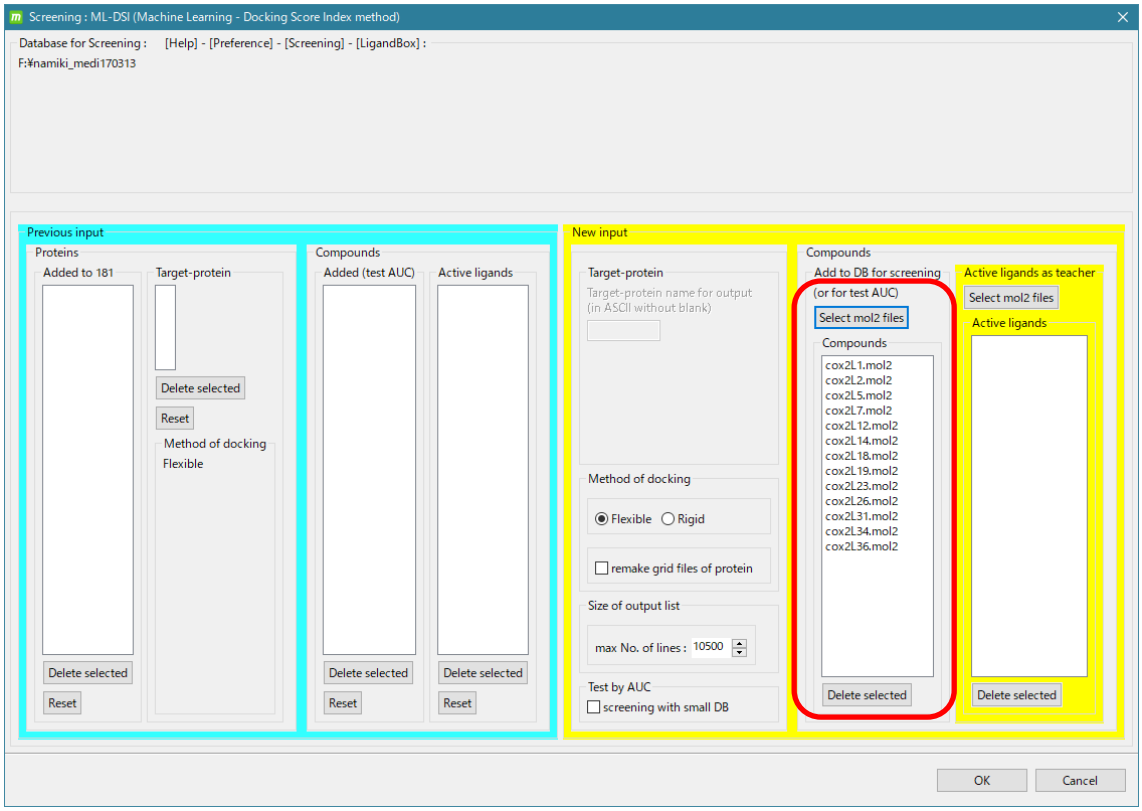
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L26.mol2

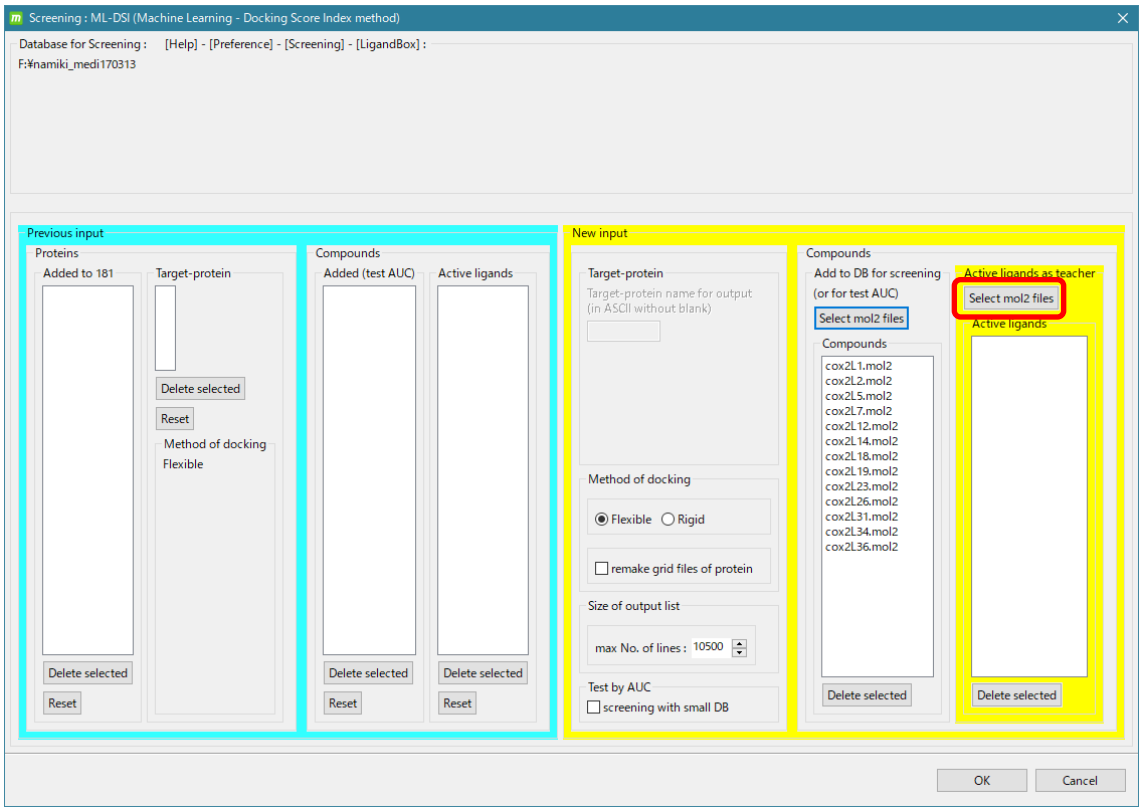
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L31.mol2

MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L34.mol2

MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L36.mol2



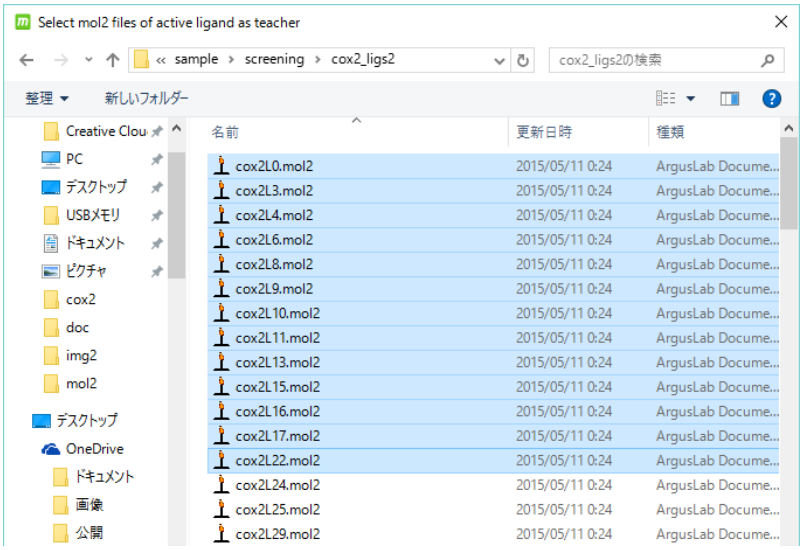


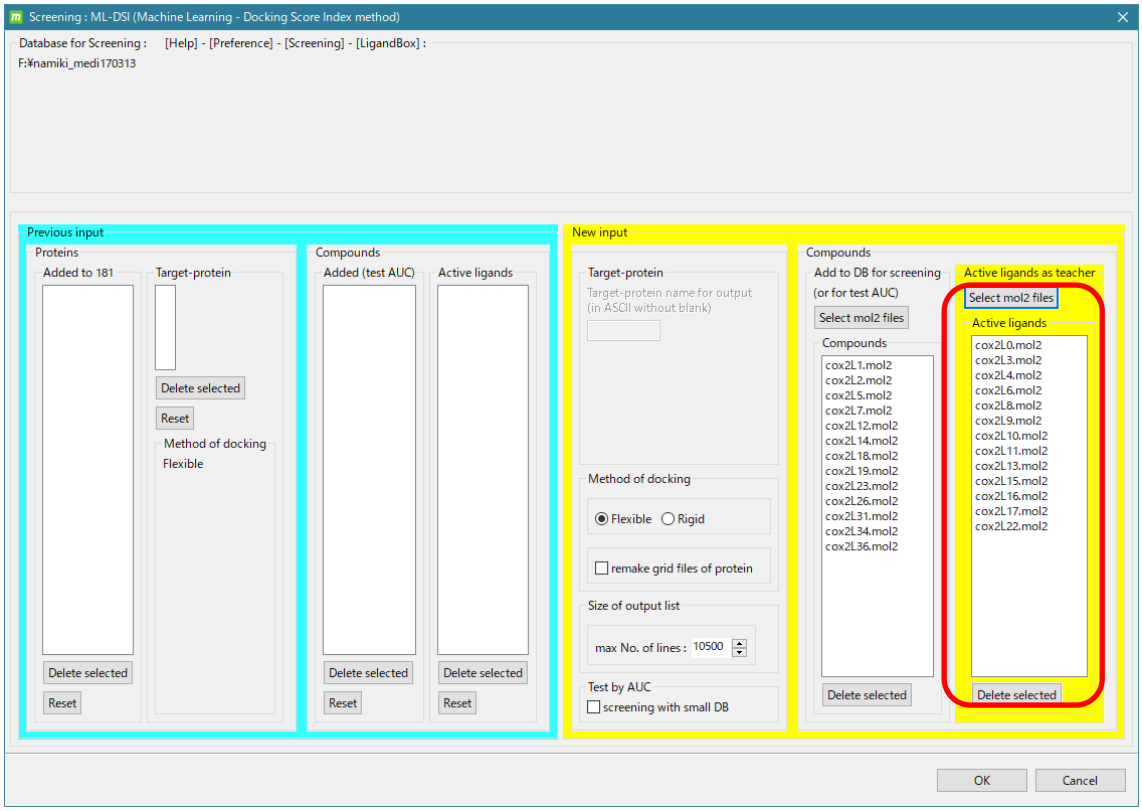


Select [Select mol2 files] under [Active ligands as teacher], select the following 13 files,

and click [Open].

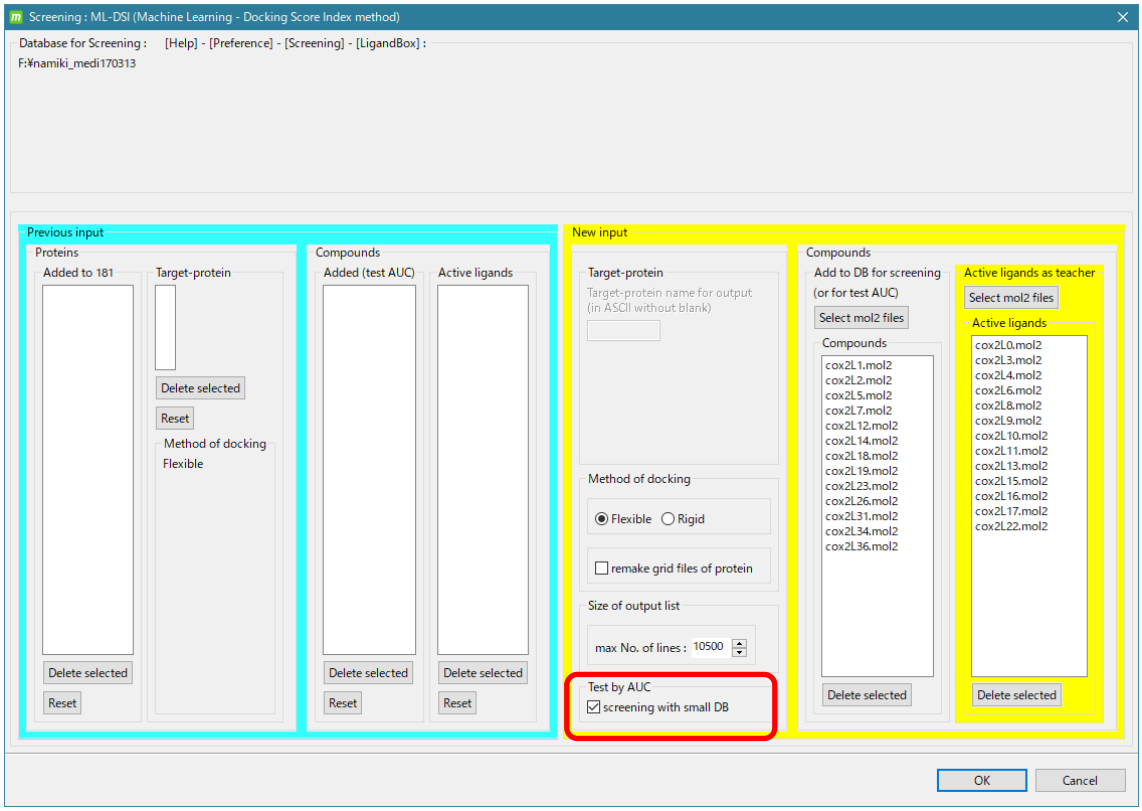
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L0.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L3.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L4.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L6.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L8.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L9.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L10.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L11.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L13.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L15.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L16.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L17.mol2
MolDesk Screening -> sample -> screening -> cox2_lig2 -> cox2L22.mol2




Screening : ML-DSI (Machine Learning - Docking Score Index method)
Database for Screening : [Help] - [Preference] - [Screening] - [LigandBox] :
F:¥namiki_medi170313
Previous input
Proteins
Added to 181
Target-protein
Delete selected
Reset
Method of docking
Flexible
Delete selected
Reset
Compounds
Added (test AUC)
Active ligands
Delete selected
Reset
Delete selected
Reset
New input
Target-protein
Target-protein name for output
(in ASCII without blank)
Method of docking
Flexible
Rigid
remake grid files of protein
Size of output list
max No. of lines : 10500
Test by AUC
screening with small DB
Compounds
Add to DB for screening
(or for test AUC)
Select mol2 files
Compounds
cox2L1.mol2
cox2L2.mol2
cox2L5.mol2
cox2L7.mol2
cox2L12.mol2
cox2L14.mol2
cox2L18.mol2
cox2L19.mol2
cox2L23.mol2
cox2L26.mol2
cox2L31.mol2
cox2L34.mol2
cox2L36.mol2
Delete selected
Active ligands as teacher
Select mol2 files
Active ligands
cox2L0.mol2
cox2L3.mol2
cox2L4.mol2
cox2L6.mol2
cox2L8.mol2
cox2L9.mol2
cox2L10.mol2
cox2L11.mol2
cox2L13.mol2
cox2L15.mol2
cox2L16.mol2
cox2L17.mol2
cox2L22.mol2
Delete selected
OK
Cancel

1.12.4. Enter ligandBox compounds to search for

If you check [screening with small DB], you can limit the search target compounds of LigandBox to a small group (20,000 compounds or less) for test calculation.



If unchecked, screening calculations will be performed for all compounds.

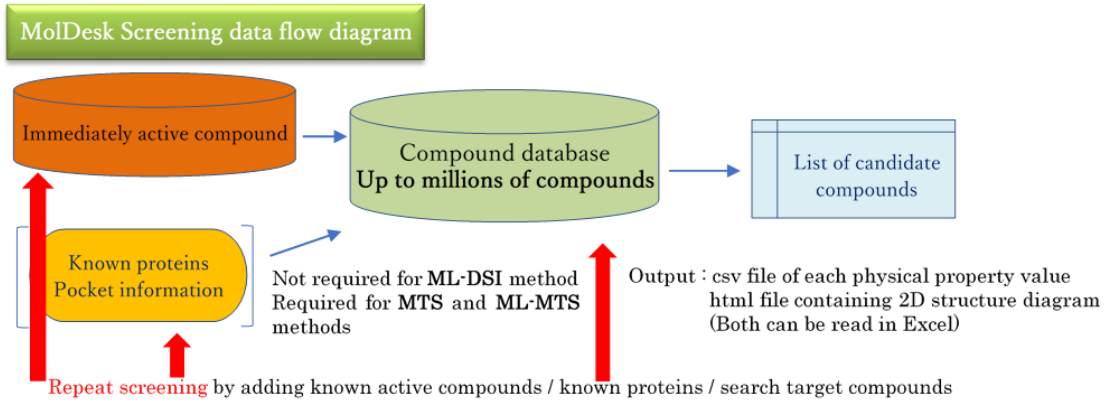
In this example, check [screening with small DB] to save test time.

Click OK to start the screening calculation.

1.12.5. Review screening calculation results

Other setting items, screening calculation result confirmation method, file output method, etc. are almost the same as in the case of screening calculation by MTS method / docking score order.
Refer to "1.10.5 Entering the size of the screening result list" to "1.10.10 Outputting the screening calculation result to a file".

1.13. Repeat screening calculation 1



The screening calculation can be repeated. The calculation method can be different.

In this example, assuming a case where the active ligand of the receptor is initially unknown and later found, the procedure for performing the screening calculation by the MTS method / docking score order and then the screening calculation by the ML-MTS method will be described. increase.

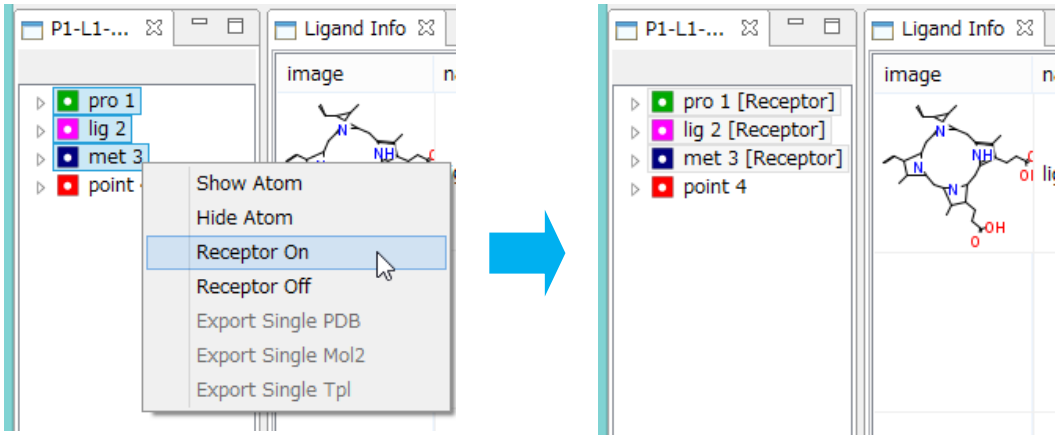
In addition, since the target protein was found later, it can be assumed that the screening calculation by the ML-DSI method is executed and then the screening calculation by the ML-MTS method is executed.

1.13.1. Selection of receptor molecules

Use the project (proj009) created in "1.10.1 Creating a project".

Specifies the receptor for docking calculations.

On the tree view screen, Ctrl + click to select pro 1, lig 2, and met 3, right-click and select Receptor On. (Since lig2 and met3 are in a place that has nothing to do with your pocket, you can select only pro 1). Receptors choose to open pocket space.



See the MolDesk Basic manual for more information on how to select receptors.

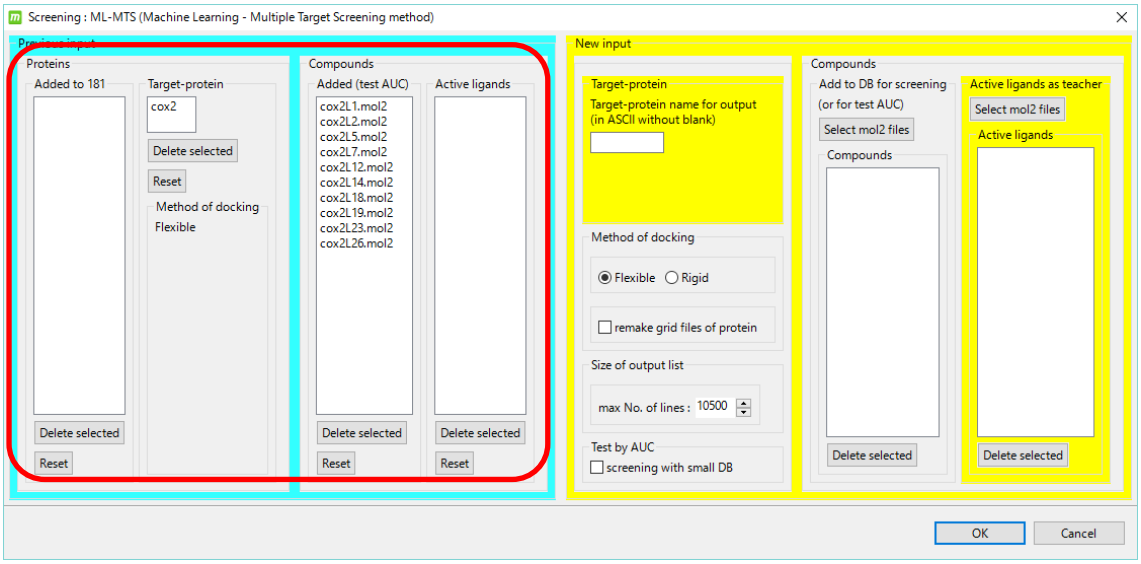
- The receptor setting is removed each time you perform a screening calculation, so set it every time.

1.13.1. Data entry for screening calculations

Click  [ML-MTS].

If a warning dialog is displayed, refer to "1.10.2 Entering data for screening calculation" and perform the necessary work.

If all the required work has been done, the Screening Calculation Data Entry dialog is displayed.



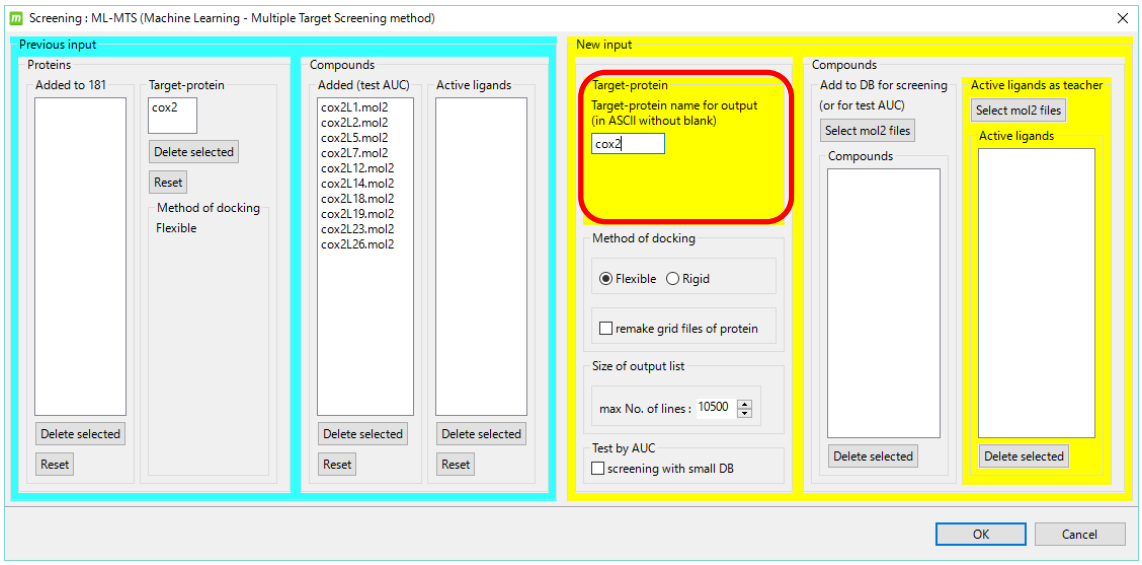
The blue background area on the left side of the dialog shows the input contents of the previous calculation.

You can confirm that the previously entered [Target-protein], the previously entered mol2 file, and the ligand structure were calculated with Flexible.

Add the mol2 file here, add the compound to the previous screening calculation, and repeat the screening calculation.

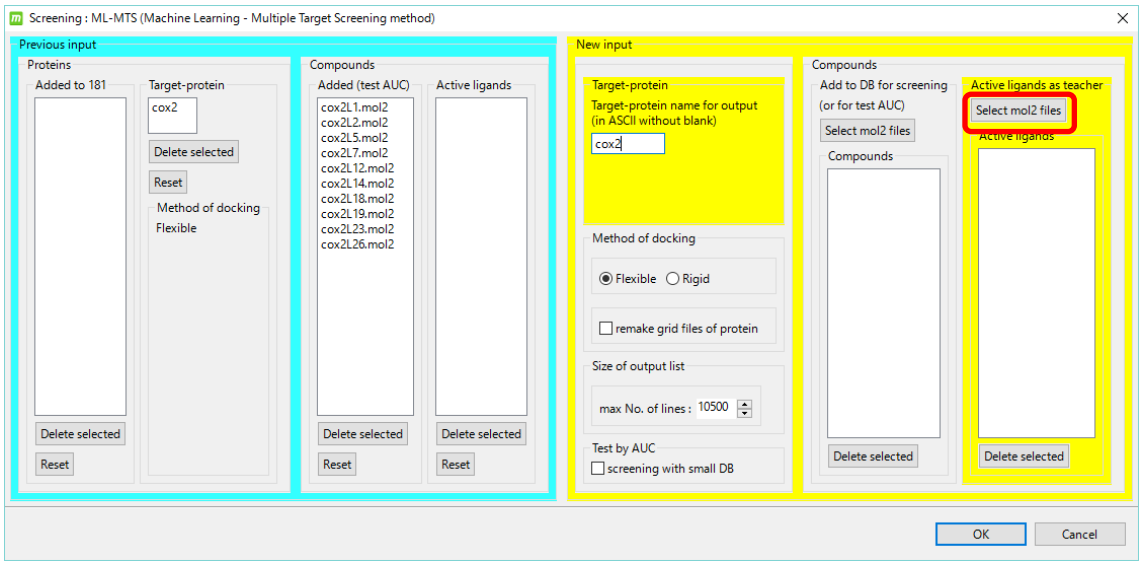
In Target-protein, enter the name of the target protein.

Here, enter cox2 as before.



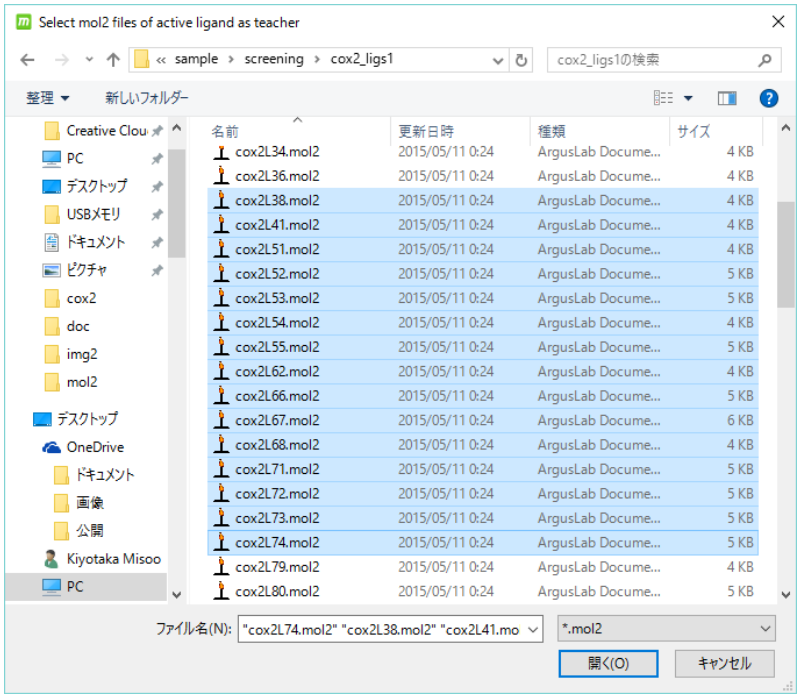
It is also possible to add another target protein. Refer to "1.14 Repeated Screening Calculation 2" for an example of execution when adding another target protein.

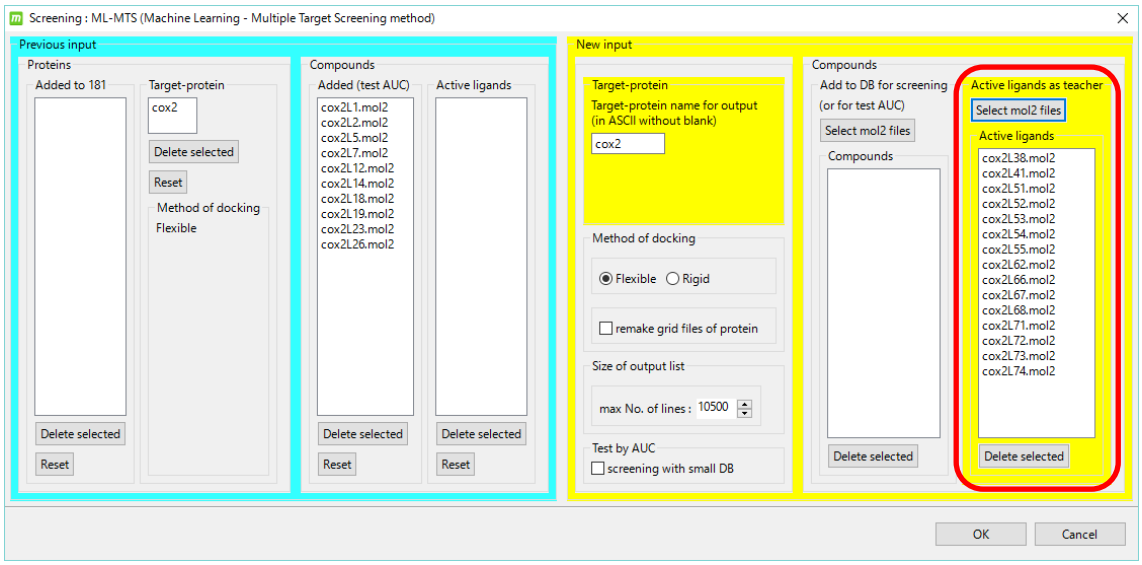
1.13.1. Addingcompounds with m ol2 files



Select [Select mol2 files] under [Active ligands as teacher], select the following 15 files, and click [Open].

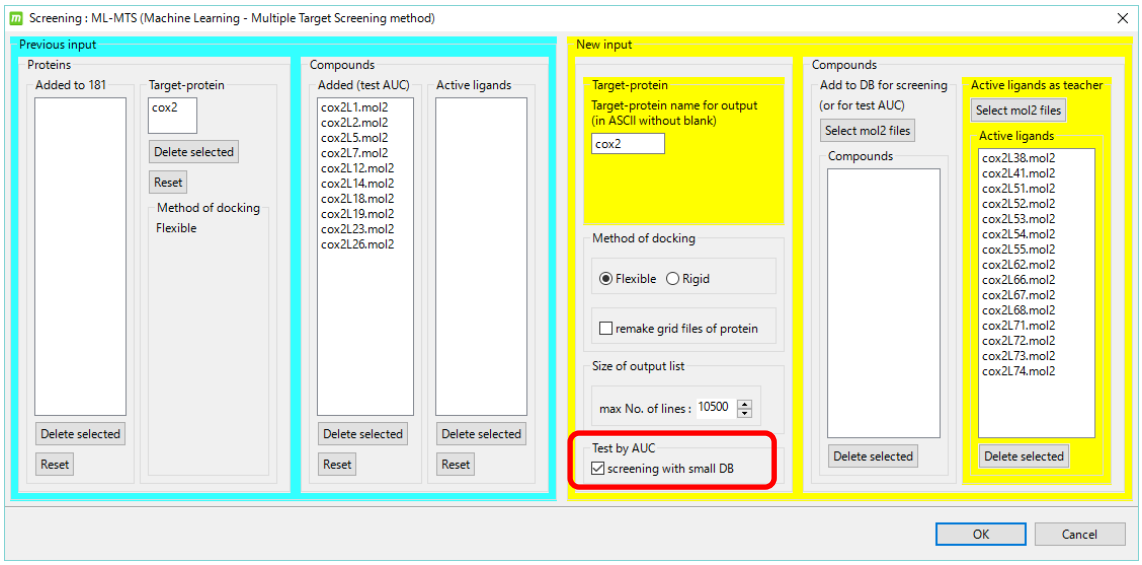
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L38.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L41.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L51.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L52.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L53.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L54.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L55.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L62.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L66.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L67.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L68.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L71.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L72.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L73.mol2
MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L74.mol2


Select mol2 files of active ligand as teacher
« sample › screening › cox2_ligs1
cox2_ligs1の検索
整理
新しいフォルダー
名前
更新日時
種類
サイズ
cox2L34.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L36.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L38.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L41.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L51.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L52.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
cox2L53.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
cox2L54.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L55.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
cox2L62.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L66.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
cox2L67.mol2 2015/05/11 0:24 ArgusLab Docume... 6 KB
cox2L68.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L71.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
cox2L72.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
cox2L73.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
cox2L74.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
cox2L79.mol2 2015/05/11 0:24 ArgusLab Docume... 4 KB
cox2L80.mol2 2015/05/11 0:24 ArgusLab Docume... 5 KB
Creative Clou
PC
デスクトップ
USBメモリ
ドキュメント
ピクチャ
cox2
doc
img2
mol2
デスクトップ
OneDrive
ドキュメント
画像
公開
Kiyotaka Misoo
PC
ファイル名(N): "cox2L74.mol2" "cox2L38.mol2" "cox2L41.mo
*.mol2
開く(O)
キャンセル


Screening : ML-MTS (Machine Learning - Multiple Target Screening method)
Previous input
Proteins
Added to 181
Target-protein
cox2
Delete selected
Reset
Method of docking
Flexible
Delete selected
Reset
Compounds
Added (test AUC)
cox2L1.mol2
cox2L2.mol2
cox2L5.mol2
cox2L7.mol2
cox2L12.mol2
cox2L14.mol2
cox2L18.mol2
cox2L19.mol2
cox2L23.mol2
cox2L26.mol2
Delete selected
Reset
Active ligands
Delete selected
Reset
New input
Target-protein
Target-protein name for output (in ASCII without blank)
cox2
Method of docking
Flexible
Rigid
remake grid files of protein
Size of output list
max No. of lines : 10500
Test by AUC
screening with small DB
Compounds
Add to DB for screening (or for test AUC)
Select mol2 files
Compounds
Delete selected
Active ligands as teacher
Select mol2 files
Active ligands
cox2L38.mol2
cox2L41.mol2
cox2L51.mol2
cox2L52.mol2
cox2L53.mol2
cox2L54.mol2
cox2L55.mol2
cox2L62.mol2
cox2L66.mol2
cox2L67.mol2
cox2L68.mol2
cox2L71.mol2
cox2L72.mol2
cox2L73.mol2
cox2L74.mol2
Delete selected
OK
Cancel

In this example, check [screening with small DB] to save test time.
For details, refer to "1.12.4 Input of compounds to be searched in LigandBox".

Click [OK] to start the screening calculation by the ML-MTS method.



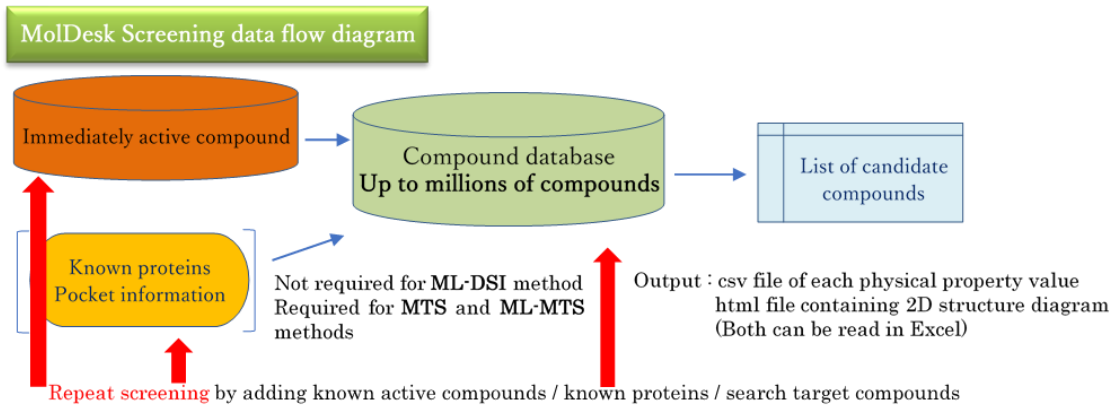
The first calculation is reused without recalculating, so the calculation is completed in less time than the previous one.

1.13.2. Review screening calculation results

Other setting items, screening calculation result confirmation method, file output method, etc. are almost the same as in the case of screening calculation by MTS method / docking score order.
Refer to "1.10.5 Entering the size of the screening result list" to "1.10.10 Outputting the screening calculation result to a file".

In the screening calculation by the ML-MTS method this time, known active compounds were input and machine learning was performed, so the AUC was significantly higher than the screening calculation by the previous MTS method (AUC = 64.81%) and docking score order (AUC = 81.84%). It will get better.

1.14. Repeat screening calculation 2



Add more target proteins1.13Repeat screening calculation 1toperform screening calculations usingml-MTS method.

- We have found a more precise structure for the target protein calculated in the past, and we expect a case to recalculate.

1.14.1. Selection of receptor molecules

Use the project of "1.13 Repeated Screening Calculation 1" to select the receptor molecule according to the procedure of "1.13.1 Receptor molecule selection".

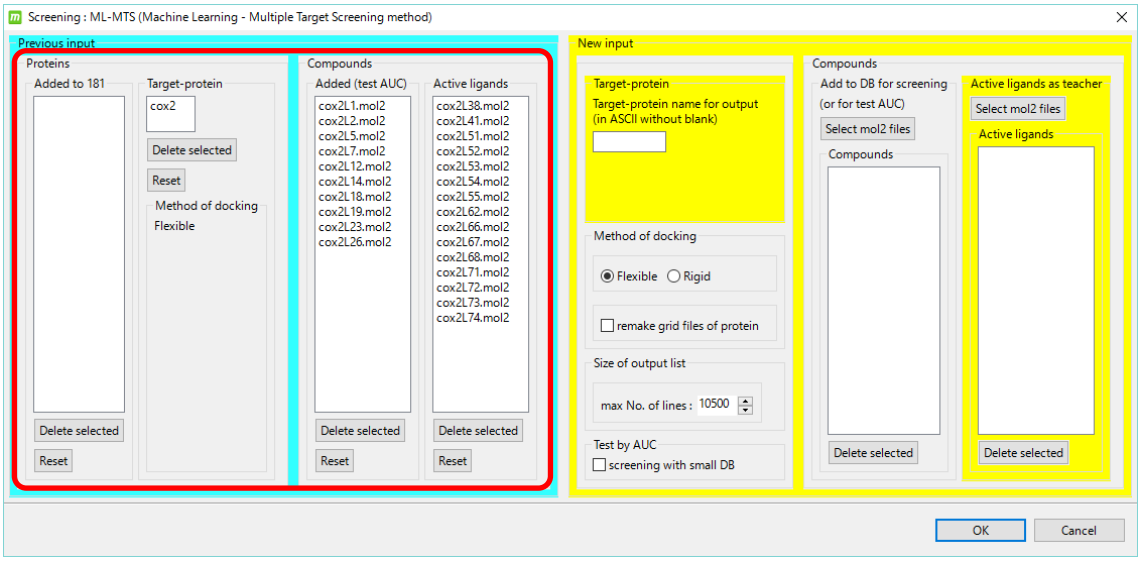
- The receptor setting is canceled every time you perform a screening calculation, so please set it every time.

1.14.2. Enter data for screening calculations

Click [ML-MTS].

If a warning dialog is displayed, refer to "1.10.2 Entering data for screening calculation" and perform the necessary work.

If all the required work has been done, the Screening Calculation Data Entry dialog is displayed.

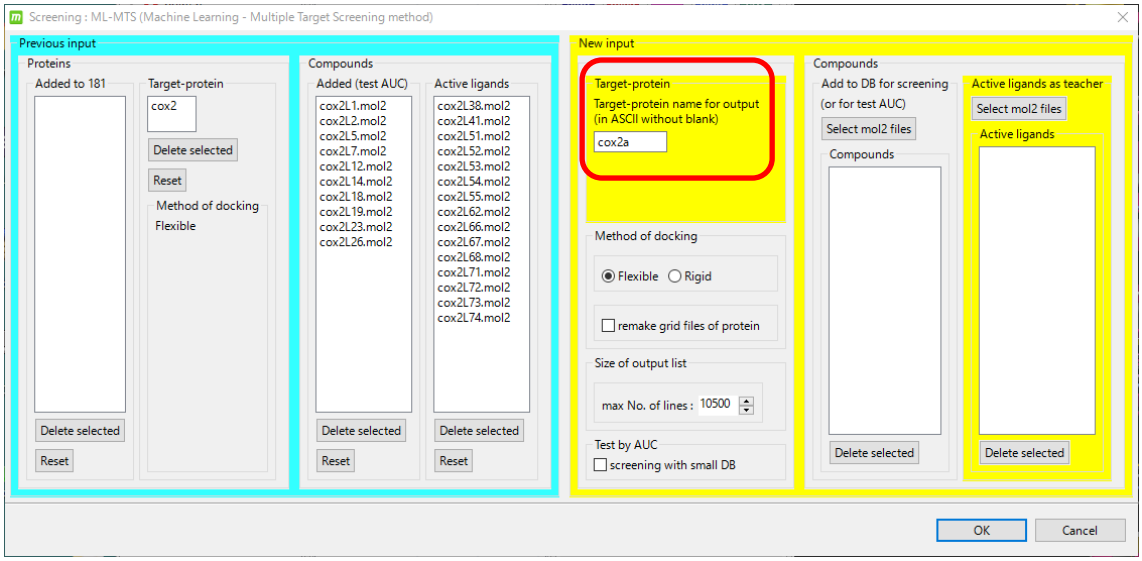


The blue background area on the left side of the dialog shows the input contents of the previous calculation.

You can check the mol2 file entered in "1.13 Repeated Screening Calculation 1" in [Active ligands].

[Target-protein] In , enter the name of the target protein.

In this case, I will specify a different name from the previous one and enter cox2a.



- I have omitted the explanation here, but if you want to add another target protein, you need to model the project. Specifically, the target protein is input, new pockets are generated, and receptors are selected.

- The iterative screening calculation can be performed any number of times. You can accumulate and calculate data on compounds and proteins.

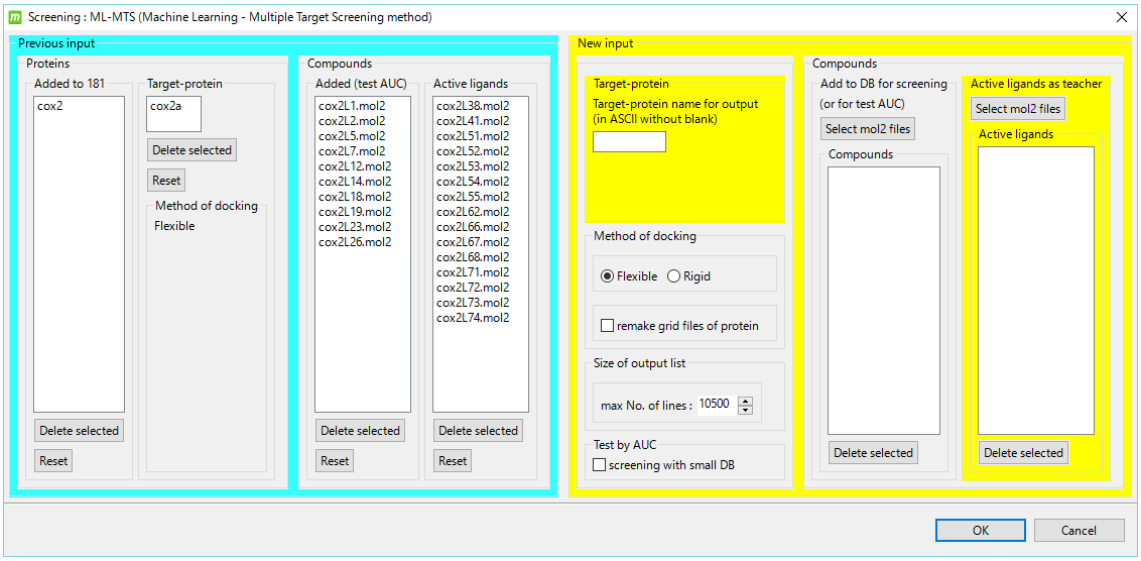
- Click OK to start the iterative screening calculation by the ML-MTS method.

Click [OK] to start the iterative screening calculation by the ML-MTS method.

1.14.3. Review screening calculation results

Other setting items, screening calculation result confirmation method, file output method, etc. are almost the same as in the case of screening calculation by MTS method / docking score order.
Refer to "1.10.5 Entering the size of the screening result list" to "1.10.10 Outputting the screening calculation result to a file".

If you click  [ML-MTS] again after this operation, the target protein cox2 from the previous time is displayed in [Added to 181], and the previous target protein cox2a is displayed in [Target-protein].



In this way, each time the target protein is added and the screening calculation is repeated, the protein is also added to [Added to 181].
Proteins added to [Added to 181] are treated the same as reference proteins.

Reference protein:
Refers to the 181 reference proteins used in the docking calculations of all screening calculations. See the myPresto manual for details.

2. Pocket search and docking calculation by Molsite

MolDesk Screening allows Molsite to perform accurate pocket searches.

2.1. Read PDB file and add hydrogen atoms and charges

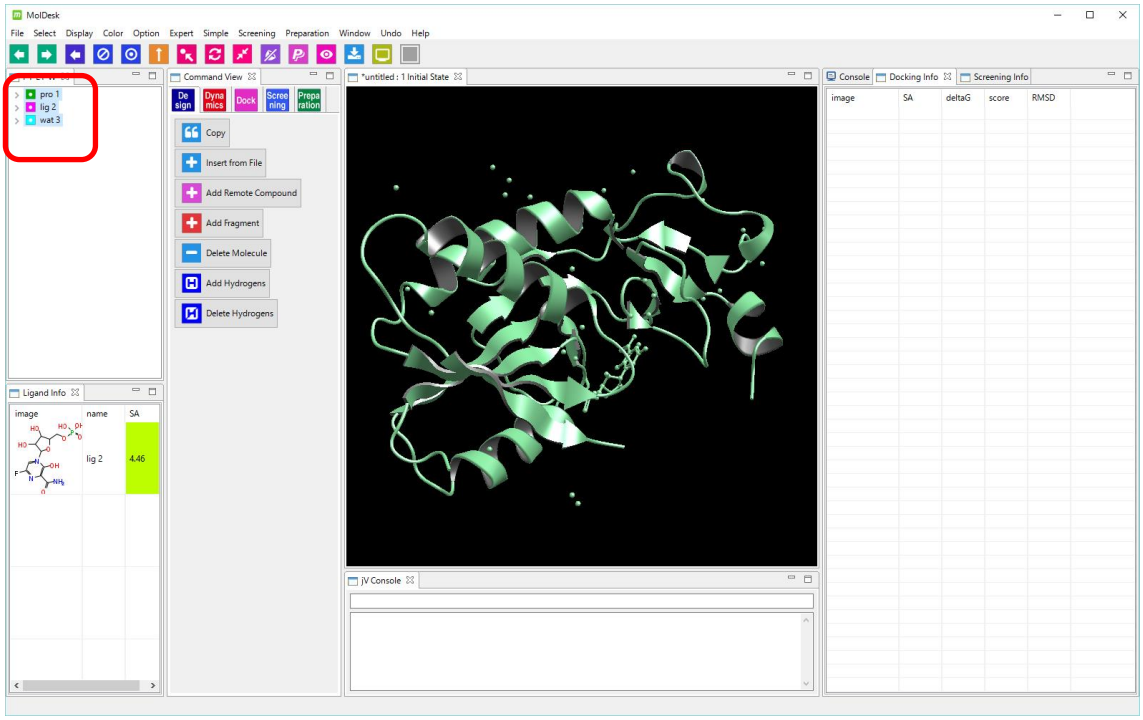
[File] – [Open Molecular File] loads the following files in the MolDesk Screening folder created on the desktop when MolDesk was installed.

MolDesk Screening-> sample-> pdb-> 4KN6.pdb

For [Open Molecular File], refer to the MolDesk Basic manual.

"4kn6" used in the example is a compound in which ribose-5'-1 phosphate is bound to HGPRT (hypoxanthine-guanine phosphoribosyl transferase), which is one of the purine-metabolizing enzymes, and favipiravir, a candidate for the treatment of Ebola fever. It is included.

On the tree display screen, select protein, compound, and water of crystallization with the mouse (shift + click to select multiple).



With [Add Hydrogens], add all the missing hydrogen atoms.

At this time, an electric charge is also added to the protein.

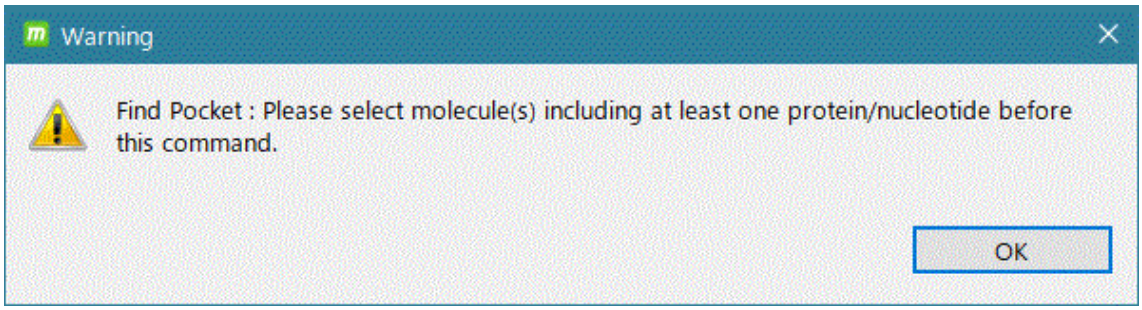
※ For hydrogen atom addition to a compound, there are three options to choose from: -p / -h / -m. Please refer to the "MolDesk Basic Manual" for the details of the contents. Here, the default -p option is selected.

Next, execute [Partial Charge] with lig2 selected, and perform charge calculation with MOPAC7 AM1 to add charge to the compound.

2.2. Pocket Search with Molsite

Select pro 1 on the tree view screen and click [Find Pocket].

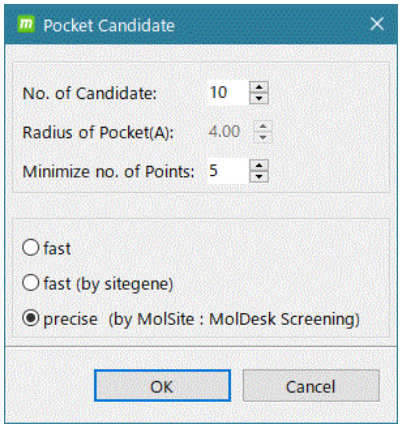
If you execute [Find Pocket] without selecting a molecule, the following warning dialog will be displayed. In that case, select the molecule and try again.



Here, select at least one protein molecule. If you select at least one protein molecule, you can include multiple non-protein molecules. Performs a pocket search on the surface of the selected molecule. For this reason, select the molecule so that there is space in the pocket.

* Molsite supports pocket searches for protein molecules and does not target nucleic acid molecules.

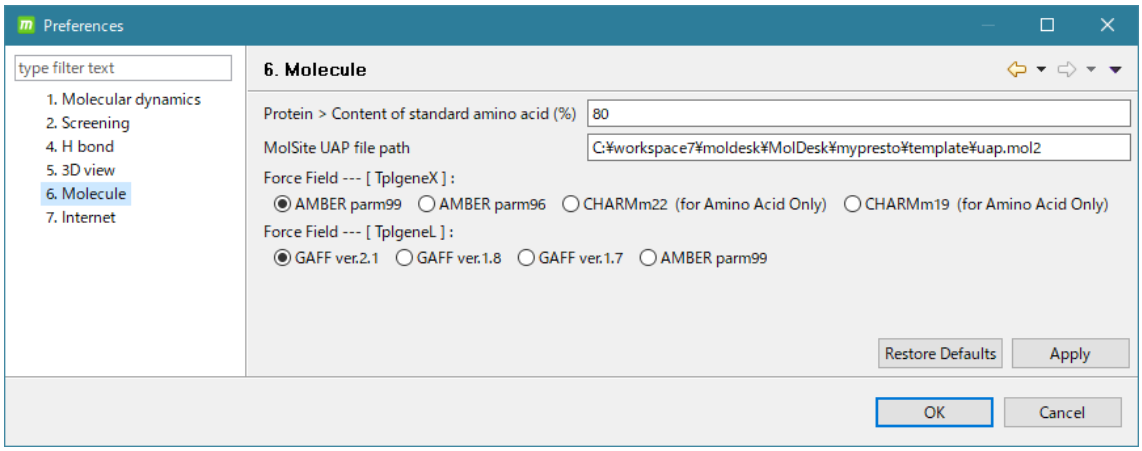
A dialog appears to set pocket search conditions.



In this example, select precise (by Molsite: MolDesk Screening).

Refer to the MolDesk Basic manual for the explanation of each setting item.

※ * In MolSite, a large number of ligand candidate molecules called UAP molecules are docked to pocket candidates one after another, but UAP molecules are displayed by selecting [Help]-[Preferene]-[Molecule] on the screen below, which is displayed by default. The user can set any molecule other than the set molecule.
If you want to change the UAP molecule, enter the mol2 file path of the molecule you want to change in the [MolSite UAP file path] below, and click [Apply] or [OK] to confirm and then calculate.

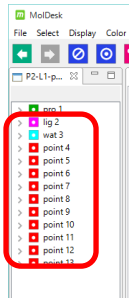


2.3. Parallel number and time of pocket search by Molsite

The guideline for the calculation time of pocket search by Molsite is as follows.
No special settings are required for parallel computing (thread parallel computing is used).

	Intel Core i7-4790K 4.0GHz / 16GB memory / windows8.1 Run in 8 parallels
PDB 4kn6 (1555 atoms)	11 minute
PDB 1m17 (4744 atoms)	15 minutes

In this example, 10 pocket candidates are displayed in score order after the calculation is completed.

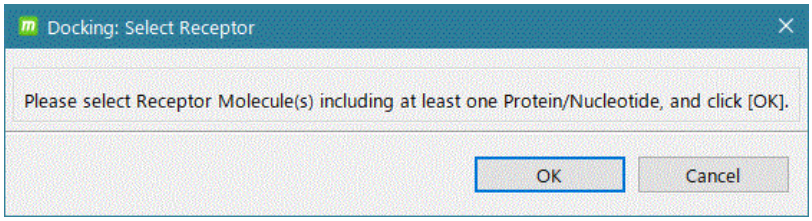


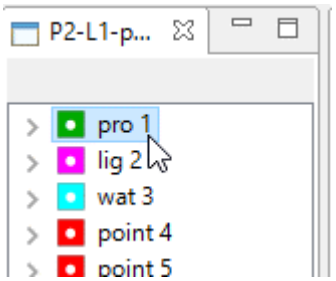
2.4. Docking calculations

Next, perform the docking calculation.

Click  [Docking].

A message dialog is displayed prompting you to select a receptor.

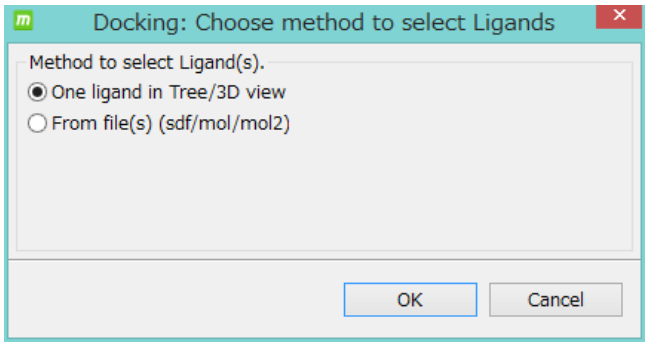




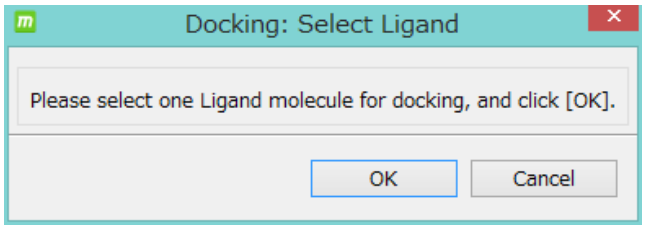
In this example, select pro1 and click [OK].

Receptor molecules must contain at least one protein molecule. It may contain compounds or metals.

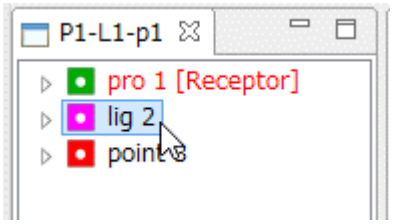
A dialog appears that specifies how to select a regand.



In this example, leave the default One Ligand in Tree / 3D view and click [OK].

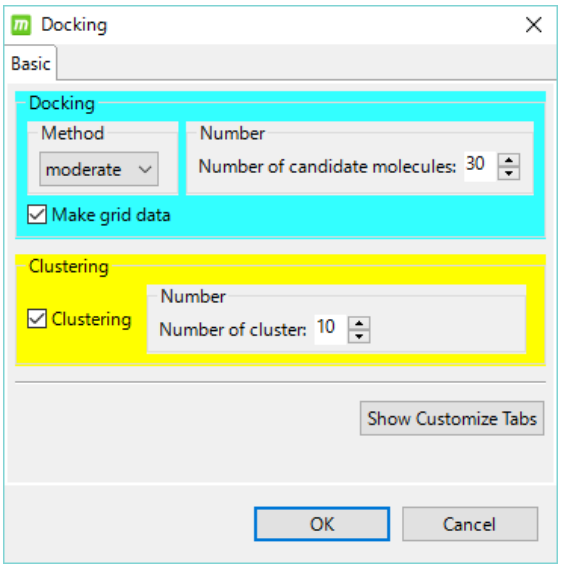


A dialog will appear prompting you to select a ligand.
Select one compound to use as a ligand from the tree display screen or 3D screen.



In this example, select lig 2, and then click OK.

A dialog appears to enter docking calculation conditions.



Enter the accuracy of the docking calculation, the presence or absence of structural clustering, and so on.

In this example, it is calculated by default.
Click [OK] as it is.

For the meaning of the setting items, refer to the MolDesk Basic manual.

Docking calculations are performed using the pocket with the highest score **(the one at the top of the tree display screen in the pocket).**

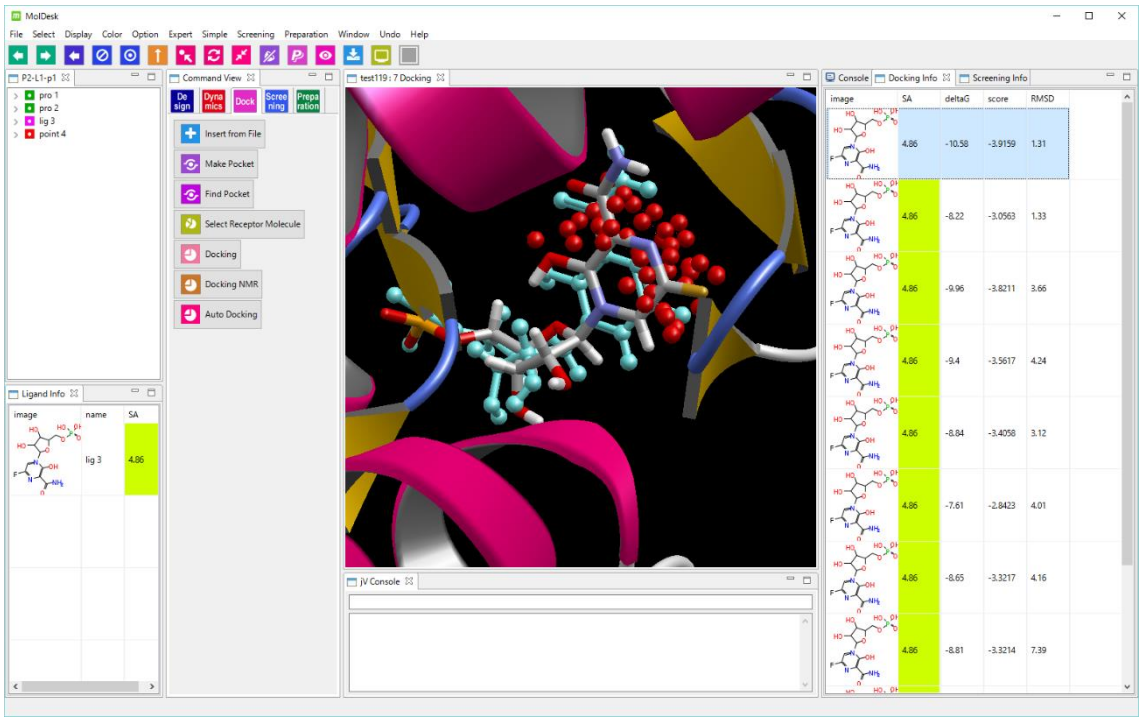
All other pockets will be deleted automatically.

If you want to use a different pocket for docking calculations

Delete all pockets with better scores with [Delete Molecule]
Perform the docking calculation.

For checking and saving the docking result, refer to the MolDesk Basic manual.

In this example, RMSD = 1.31 Å as a result of comparing the predicted structure with the best score and the correct structure (displayed in light blue with lig3, [Color] – [Atom] – [Any]).



3. Docking Score QSAR (Predict Activity)

Docking Score Predicts the activity of compound molecules using the QSAR method. Predicts the activity value of a specified compound for a specific protein.

Docking Score QSAR is for creating regression parameters for specific proteins.

[Preparation]-[Make DB to predict Activity] 

And, using the regression parameters created above, calculate the activity values of multiple compounds for a particular protein at once.

[Screening]-[Predict Activity] 

There are two buttons.

In order to predict the activity value with [Predict Activity], it is necessary to create a data file created by learning the regression parameters with [Make DB to predict Activity].
Both [Make DB to predict Activity] and [Predict Activity] require a long calculation time because the docking calculation of the input compound and about 600 proteins inside the program is performed in a round-robin manner. (500 compounds, about 3 hours on a normal 8-thread parallel CPU machine).

Docking Score QSAR :
A method of estimating the binding free energy with a weighted average of docking scores for a large number of proteins. The pharmacophore is represented by 600 kinds of proteins, the 600 docking score of a certain compound is analyzed as a principal component, and the experimental data ΔG is subjected to regression analysis by the least squares method. The estimation model is calculated by descriptor-based weighted PCR using ridge regression, and outliers are excluded using robust estimation (M estimation). Affinity data and structural data used for regression are obtained from ChEMBL and PDB (public database). All affinity data (IC50 value,% inhibition value, activity value, etc.) obtained from ChEMBL are converted into binding free energy ΔG. However, since ChEMBL lacked the experimental information required for conversion, some

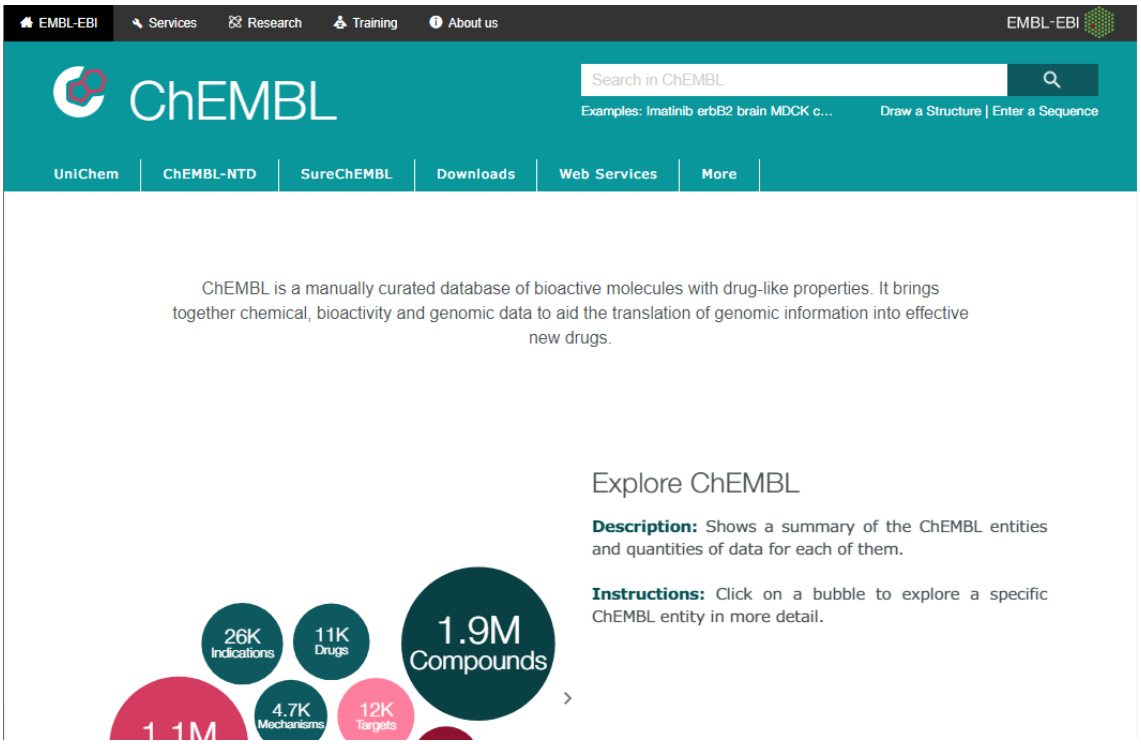
assumptions are made (Kd = Ki, etc.).

3.1. Acquisitionof CH EMBL experimental value data

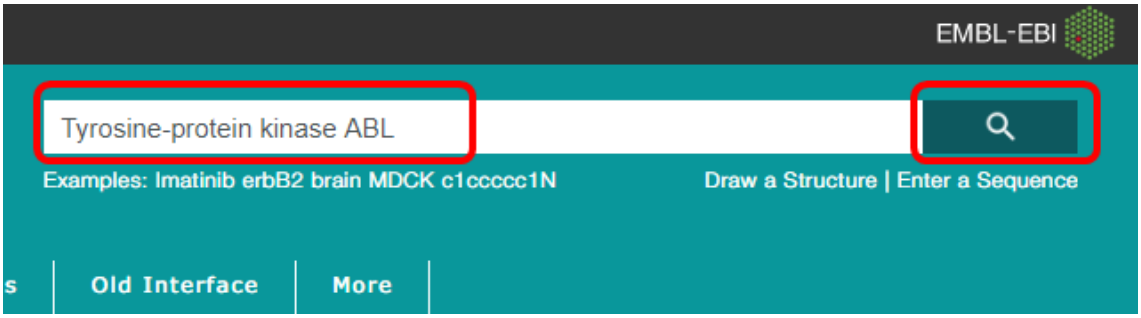
The target protein can be freely downloaded and selected by the user from ChEMBL.

First, obtain the affinity data (IC50 value,% inhibition value, activity value, etc.) of various compounds for a specific protein from ChEMBL as a file by the following procedure.

https://www.ebi.ac.uk/chembl/ (Access the ChEMBLE top page.

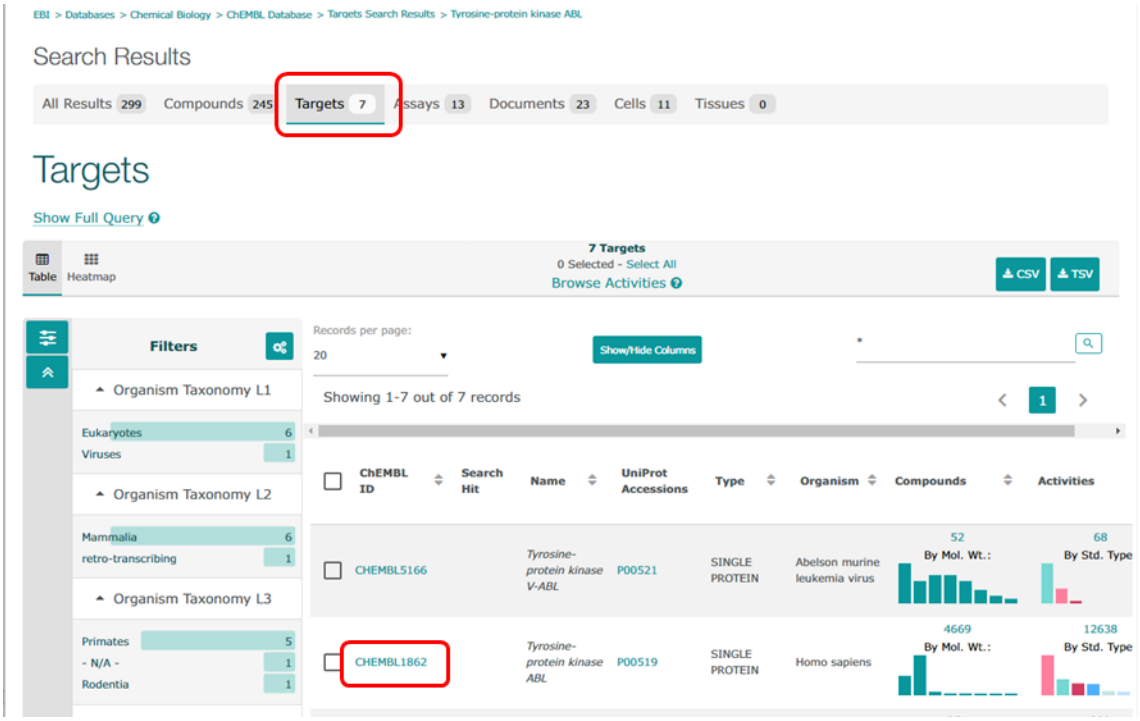


Enter the protein you want to predict and select (click) the search button.
In this example, enter Tyrosine-protein kinase ABL.

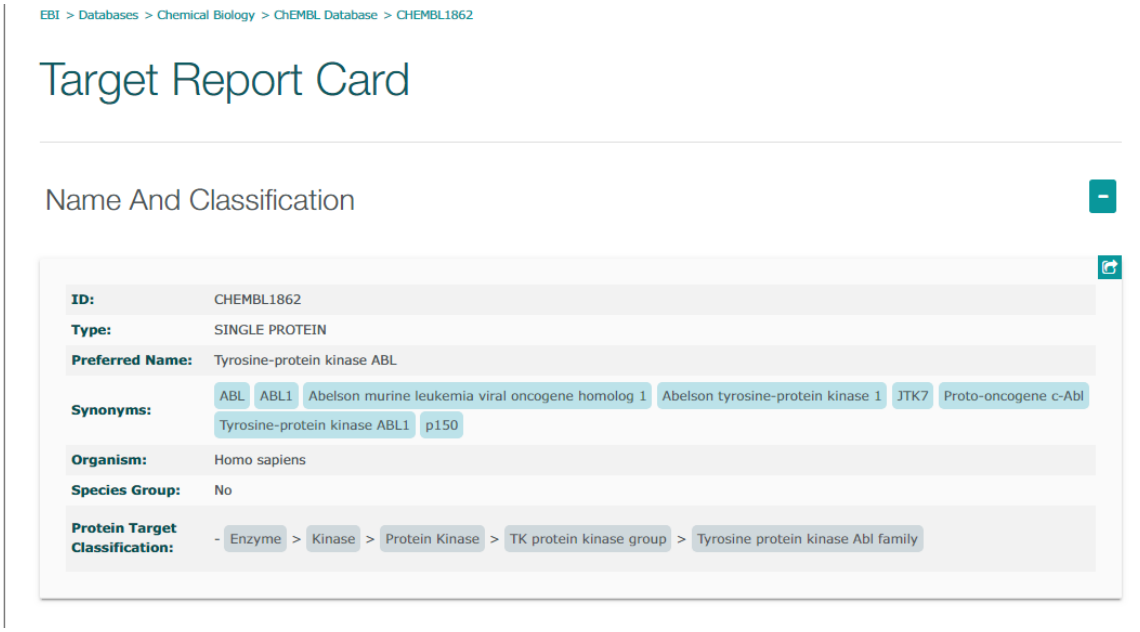


Select (click) the "Targets" group from the displayed search results to display it, and select the appropriate protein.

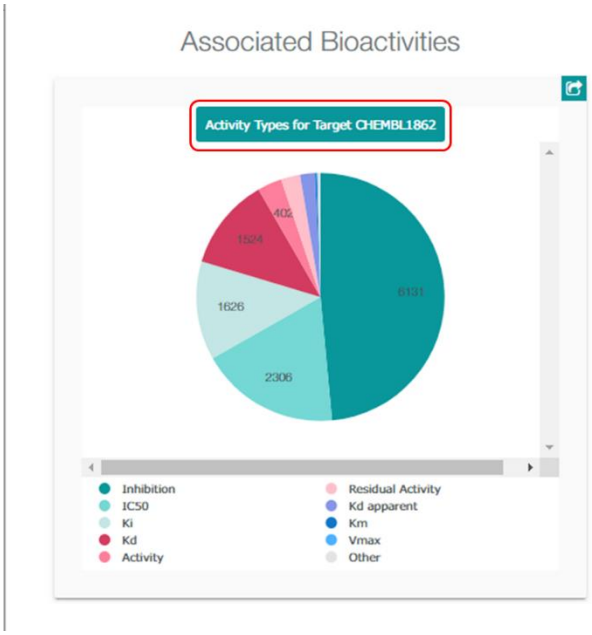
In this example, select CHEMBL1862.



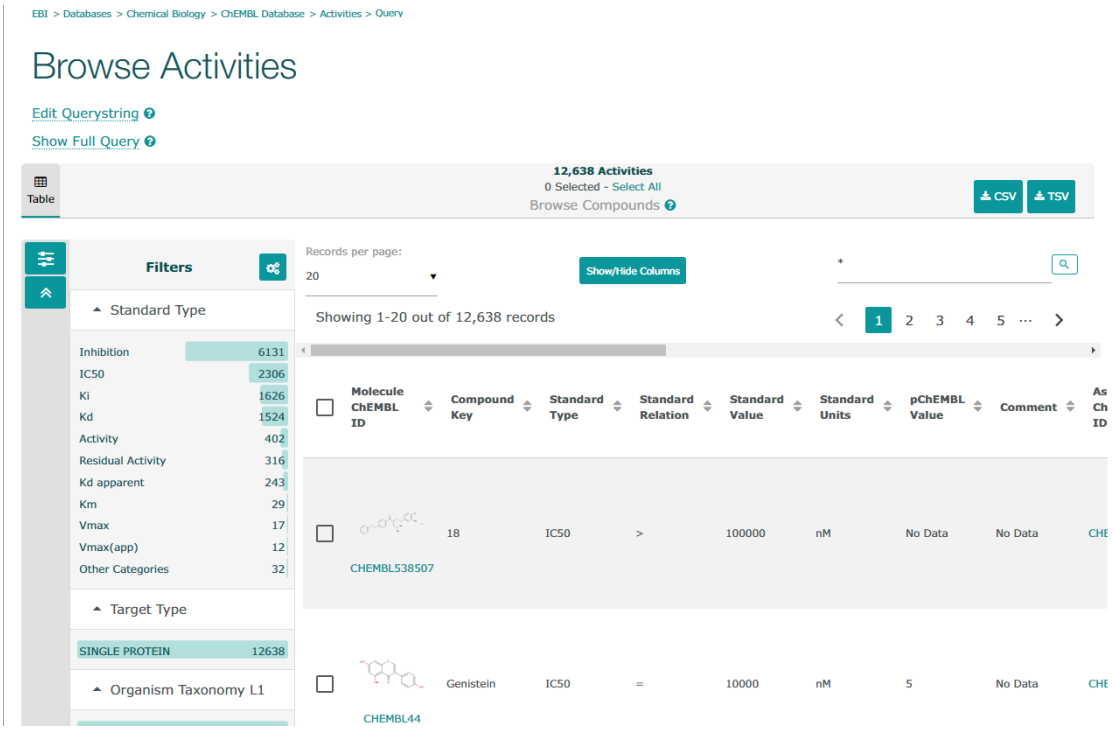
The following pages appear.



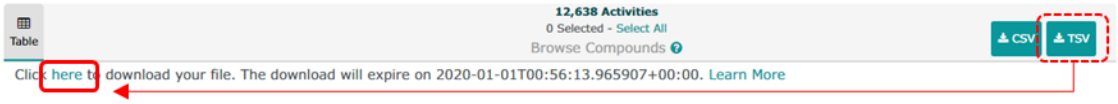
In the pie chart belowonthe page, select (click Activity Types for Target CHEMBL1862).



The following pages appear:



When you select (click) [TSV] at the top of the page, a download link for the tab-delimited text file will be generated. Select (click) the generated [here].

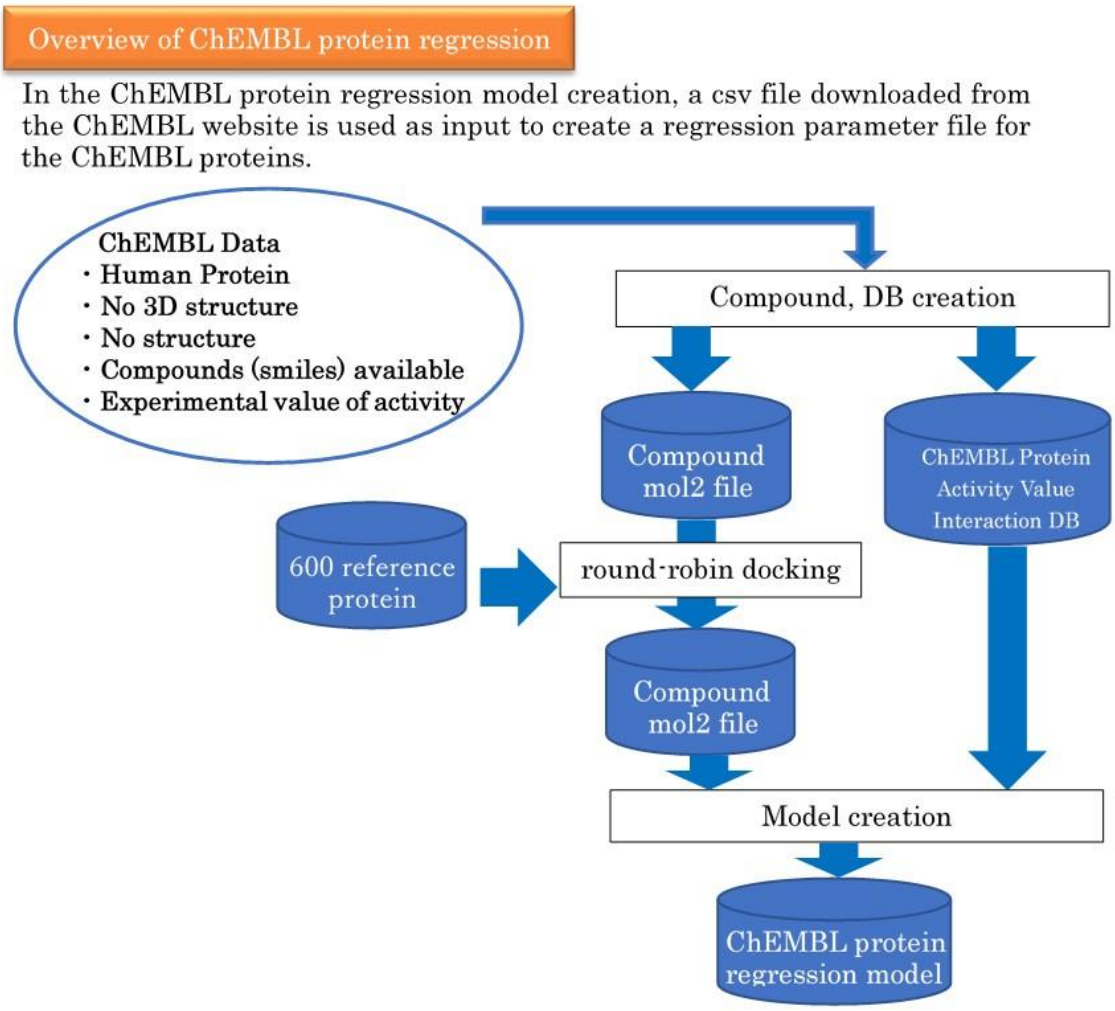


A file called CHEMBL25-chembl_activity-XXXX.tsv.gz will be downloaded. (* XXXX is a long random alphanumeric symbol)

Since the file is compressed in gz format, decompress it into a tab-delimited text file in tsv format with an appropriate decompression software.
CHEMBL25-chembl_activity-XXXX.tsv.gz
↓
CHEMBL25-chembl_activity-XXXX.tsv

3.2. Calculating regression parameters

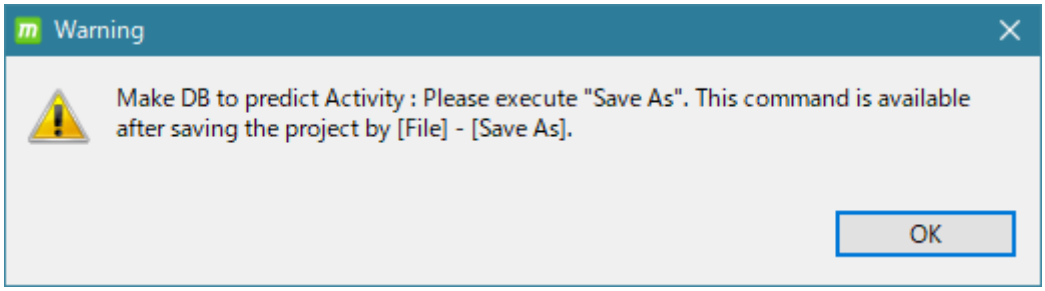
3.2.1. Overview



3.2.2. Create a project

Now let's get back to working with MolDesk Screening.
In the File-New Project menu, create an empty project and save it.
Refer to the MolDesk Basic manual for how to save the project.

- If the project is not saved, the following warning dialog will be displayed, so please save the project.



3.2.3. Calculate regression parameter

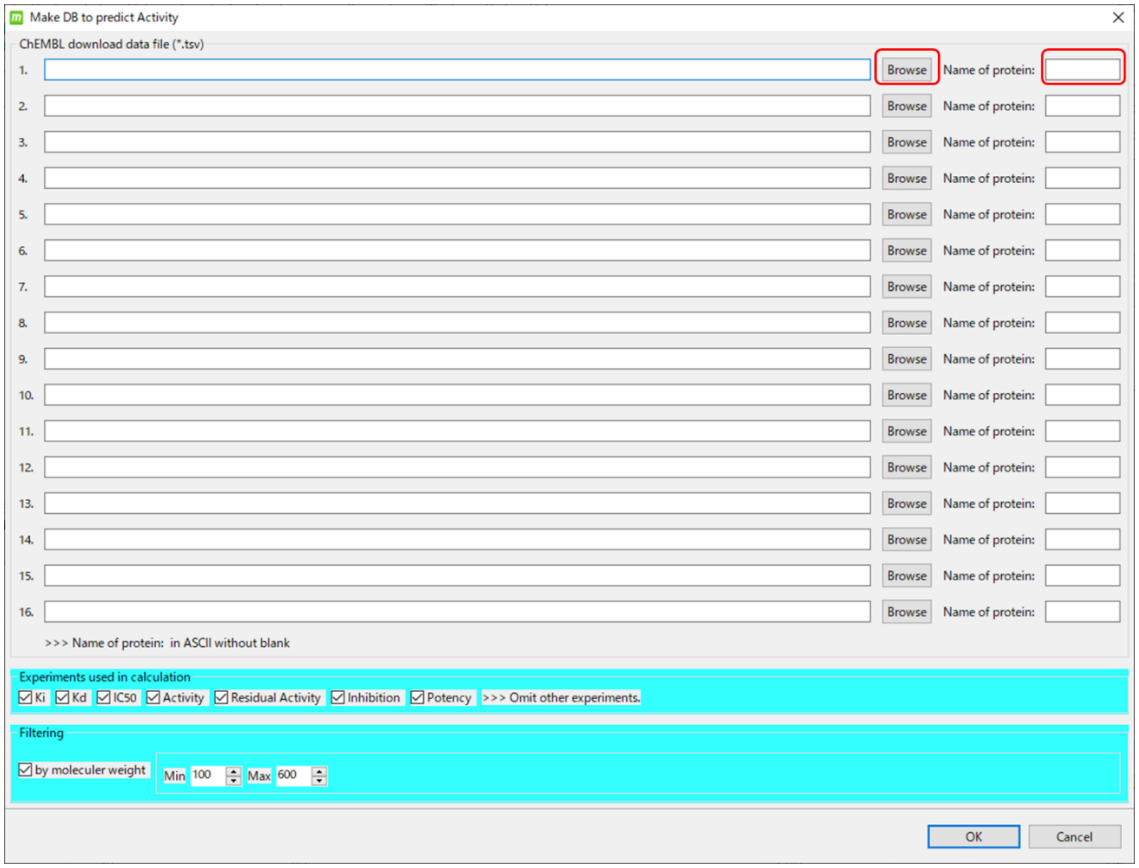
To create regression parameters for a particular protein

Click [Preparation] - [Make DB to predict Activity] .

Then the following screen will be displayed.

Click the [Browse] button and use the file selector to select the experimental data file for the activity value of a particular protein that you downloaded from ChEMBL in the previous section.
When the selection is complete, the file path will be displayed as shown below.

In addition, enter the name of the specific protein in the [Name of protein]: field. This input name is used for calculations such as the name of the output file, so enter it in alphanumeric characters without spaces. In this example, enter ABL.



On this page, you can enter up to 16 experimental data files downloaded from ChEMBL. That is, the regression parameters for 16 specific proteins can be calculated in parallel at one time. However, it generally takes a long time to calculate (48 parallel machines, a few days per machine).

[Experiments used in calculation]
In the data file downloaded from ChEMBL, select the experimental data to be used for regression model creation. Experimental data other than the checked items will be ignored.

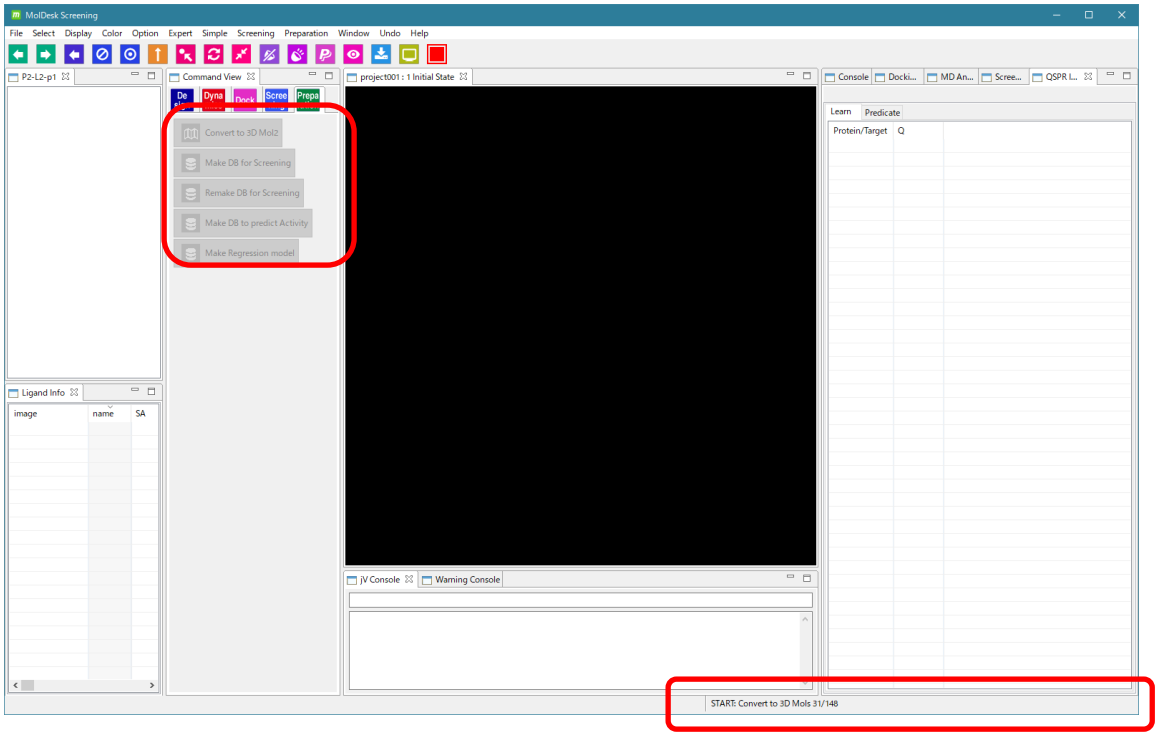
[Filtering]

You can select whether or not to perform filtering by the molecular weight of the compound with [Filtering] – [by molecular weight].

MolDesk Screening uses myPresto's sievgene for docking calculations, which reduces the accuracy of docking calculations with a molecular weight of 600 or more, so it is set to 100 – 600 by default. Normally, the default is used for calculation.

Click OK to start the (parallel) calculation.

The command button will be grayed out when the calculation starts. Calculation is in progress while the command button is grayed out. In addition, a simple calculation status during calculation is displayed in the red frame at the bottom right.



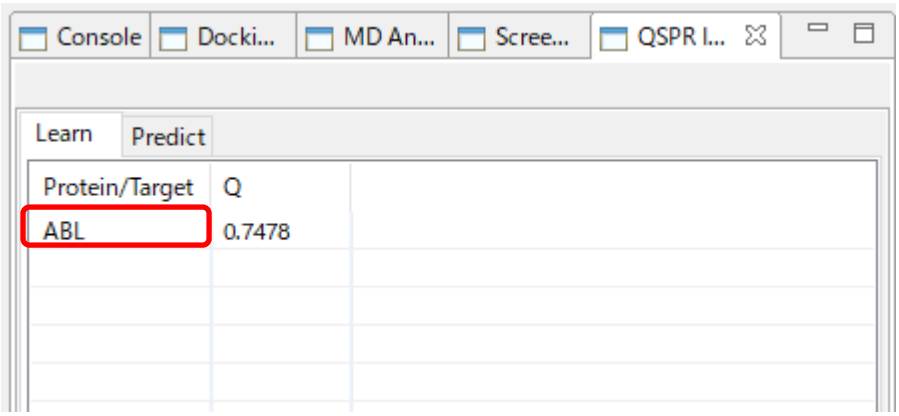
You can work with other projects during the calculation, but be aware that it may be extremely slow depending on the processor occupancy.

The number of parallels when calculating in parallel can be set by Thread number of [Help]-[Preference] – [Screening]. For details, refer to "1.7 Number of Parallel Screening Calculations, Memory Amount, and Time".

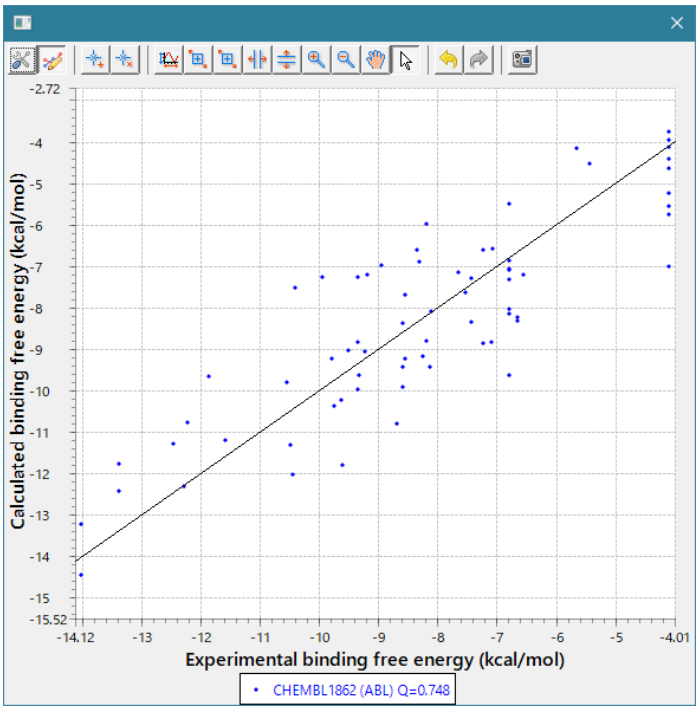
3.2.4. Confirmation of the calculation results of regression parameters by graphs

When the calculation of the regression parameter is completed, the command button changes from gray to available. Also, [END: Make DB to predict Activity] is displayed at the bottom right of the screen.

In addition, on the right side of the screen, display the Learn tab of the QSPR Info screen as shown below.
Lists the protein name entered when the regression model was created and the correlation coefficient (Q value) of the regression model created.



Now, double-click the protein name in the [Protein / Target] column in the red frame above.
Then, you can check the reliability of learning by displaying the experimental data ΔG derived from ChEMBL used when creating the regression parameter and the graph of the calculated value of ΔG when calculating the regression parameter as follows. ..



3.2.5. Review regression parameter files

The location where the regression parameter file of the calculation result is created is

[PROJECT]-> work-> database_qsar-> 09.param

, assuming that the saved project folder is [PROJECT].
The file name will be

[PROTEIN] .param

if the name of the specific protein entered in the [Name of protein:] field in the previous section is [PROTEIN]. This file is important because it will be used in the calculation of the activity value prediction described in the next section.

The following folders and files are created in database_qsar.
Users don't have to worry about these contents, but the contents are as follows.

```
[PROJECT] - work - database_qsar - 00.log
                                 - 01.download
                                 - 02.mol2
                                 - 03.topology
                                 - 04.optimize
                                 - 05.db
                                 - 06.lignad
                                 - 07.work
                                 - 08.score
                                 - 09.param
                                 - protein_qsar
                                 - ChEMBL.list (file)
                                 - pro_list  (file)
```

item	substance
00.log	[PROTEIN] Eachfileis printedin the PROTEIN.log / error folder. Each [PROTEIN] calculation produces an error that occurred during the calculation.
01.download	The experimental data file downloadedfrom ChEMBL entered in the previous section is saved.
02.mol2	For each [PROTEIN] folder, output a mol2 file of the compounds described in the experimental data below. In the experimental data file, it is recorded by SMILES,but it is convertedto mol2, and it also performs three-dimensionalization and charge generation.
03.topology	For each [PROTEIN], create each compound folder below the[PROTEIN] folder to generate a topology file for myPresto.
04.optimize	Energy-to-system calculation results of the three-dimensional structure of each of the above compounds
05.db	ChEMBL Experimental Data File Machining Data File
06.logand	ChEMBL Experimental Data File Compound mol2 File
07.work	Folder × docking calculations for 600 proteins and mol2 files

08.score	600 proteins× mol2 file docking calculation full score save file
09.param	Regression Parameter Data File
protein_qsar	600 Grid file creation folder for protein docking calculations
ChEMBL.list	Experimental data file name list downloaded from ChEMBL
pro_list	PDB ID list of 600 proteins

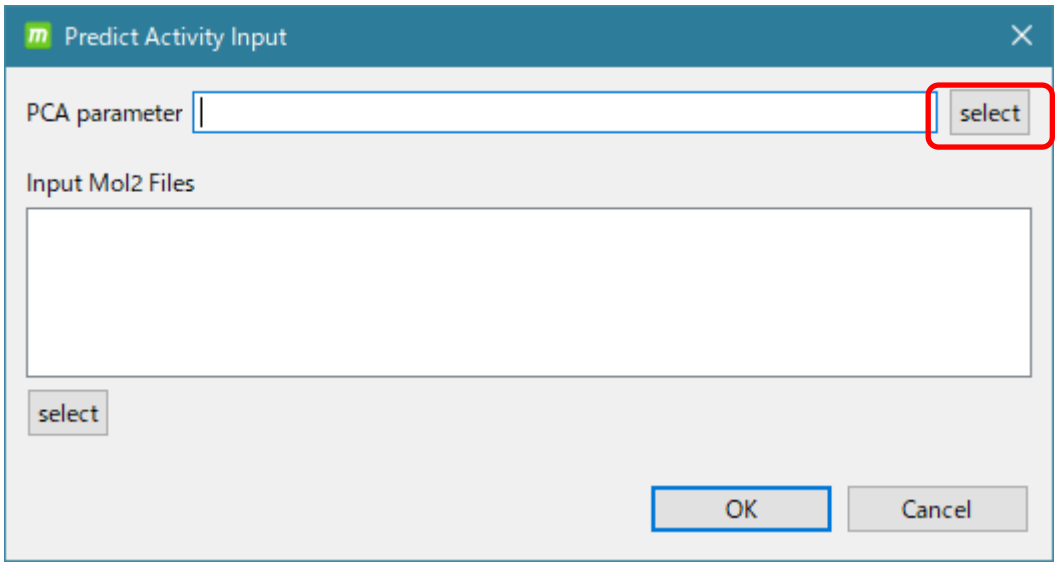
3.3. Prediction calculation of active values

Using the regression parameter file created in the previous section, the activity values of multiple compounds for a specific protein are calculated at once (parallel).

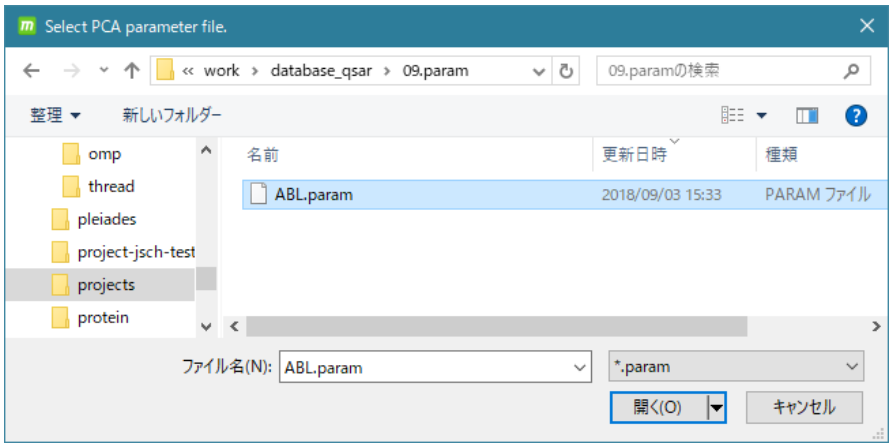
3.3.1. Perform active value prediction calculations

[Screening] - [Predict Activity]

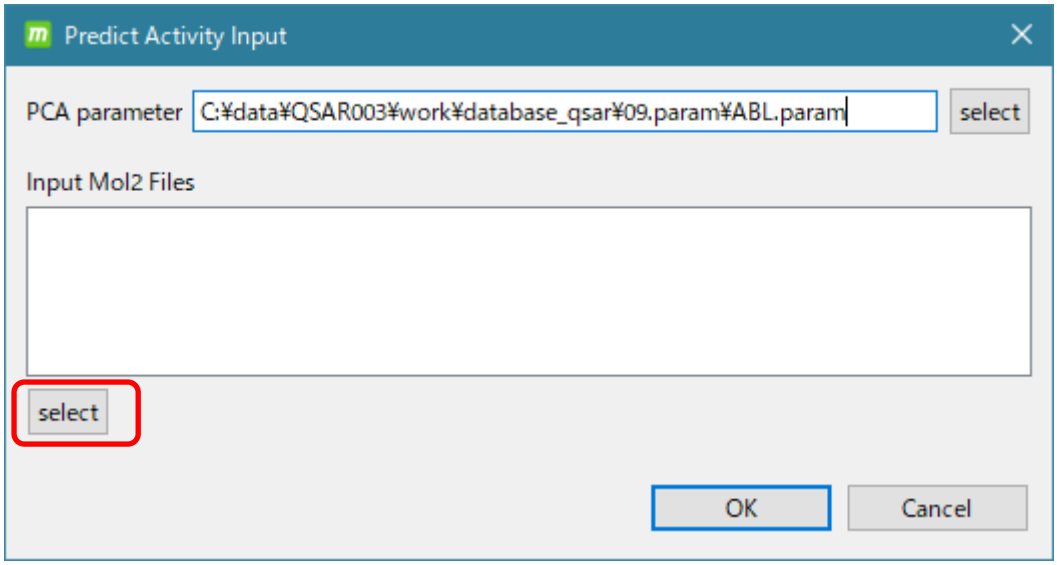
Click. The following input screen will appear.



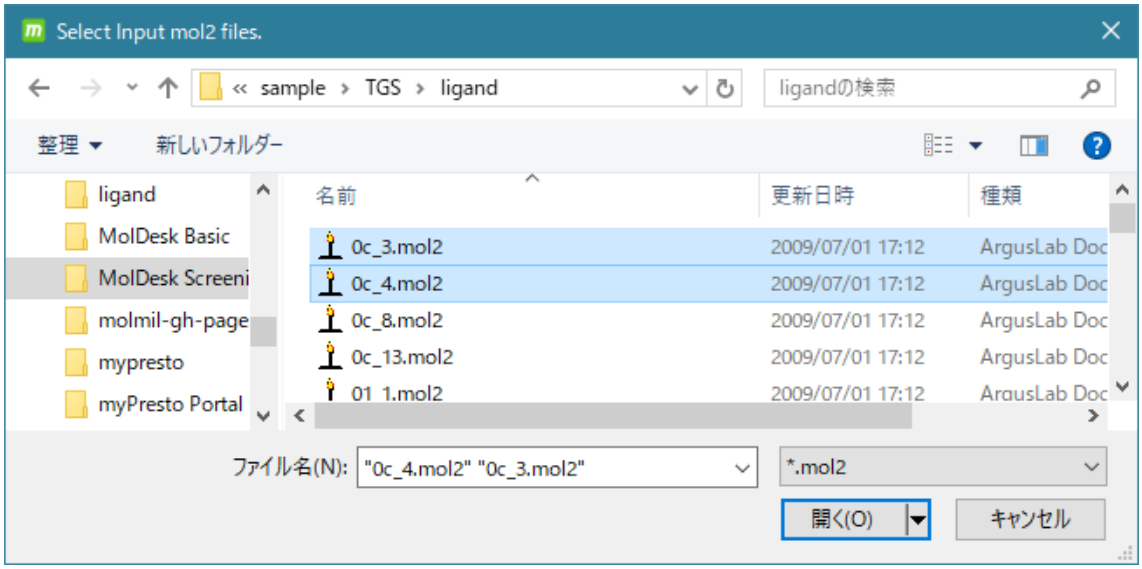
In the image above, click selectinthe red frame to select the regression parameter file calculated in the previous section in the figure below.



For example,selectPROJECT -> work -> database_qsar -> 09.param -> ABL.param.



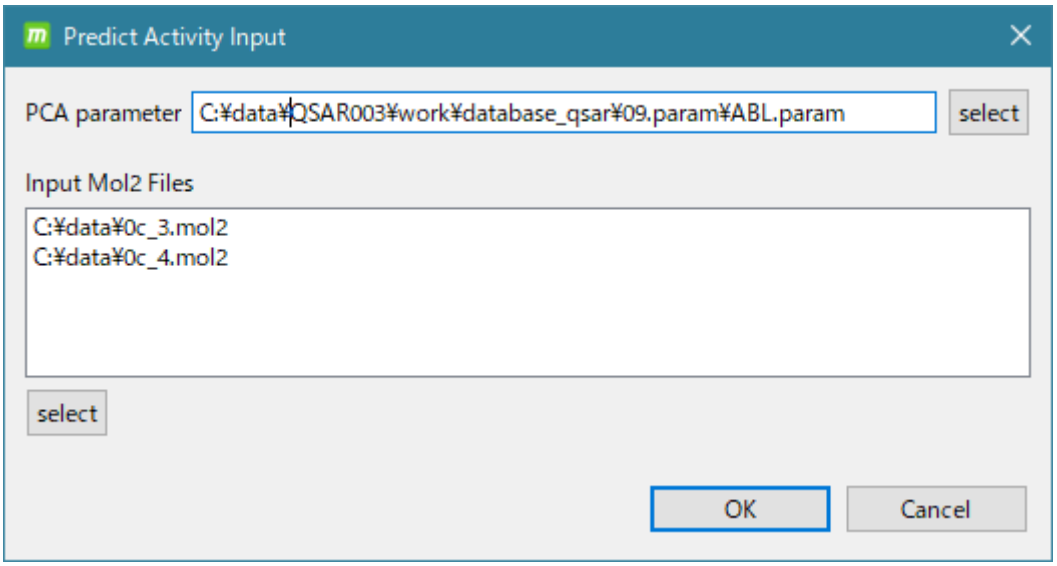
Next, click [select] in the red frame in the above figure, and select the mol2 file of the compound whose activity value you want to predict in the figure below.



Here, as an example, I selected the above two.

※ Make sure that the mol2 file of the compound contains only one molecule. You can use the Mol2 file created by [Preparation] - [Convert to 3D Mol2] of MolDesk Screening as input.

※ We plan to add a function in the near future to enable batch input of multiple moleculesin the sdf file.



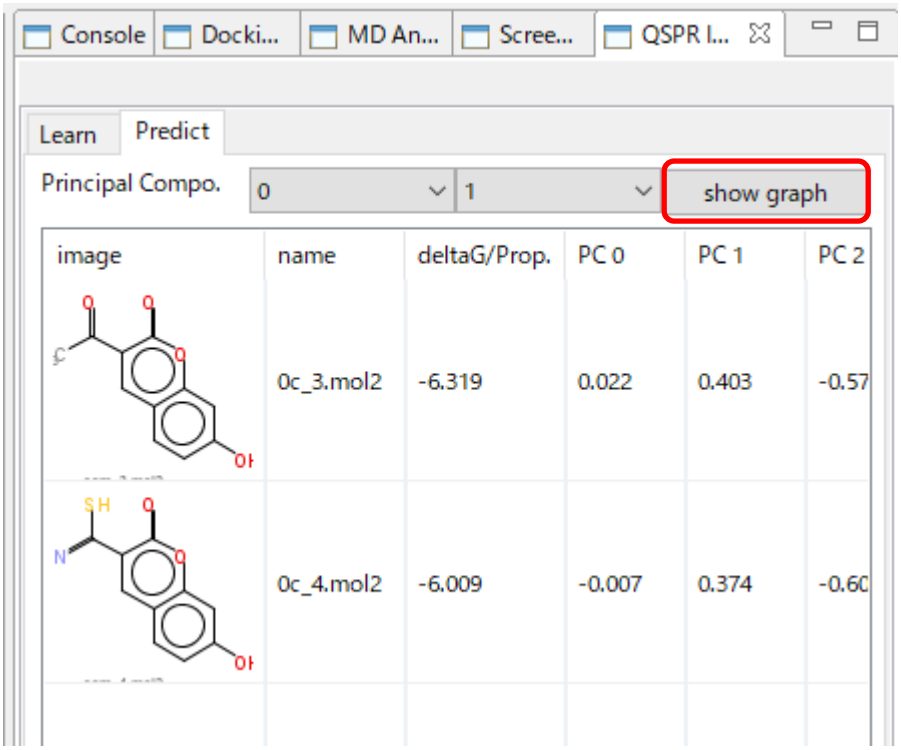
When the input is completed, it will be as shown in the above figure. Click [OK] to start the prediction calculation.

During the forecast calculation, the calculation status is displayed at the bottom right of the screen.

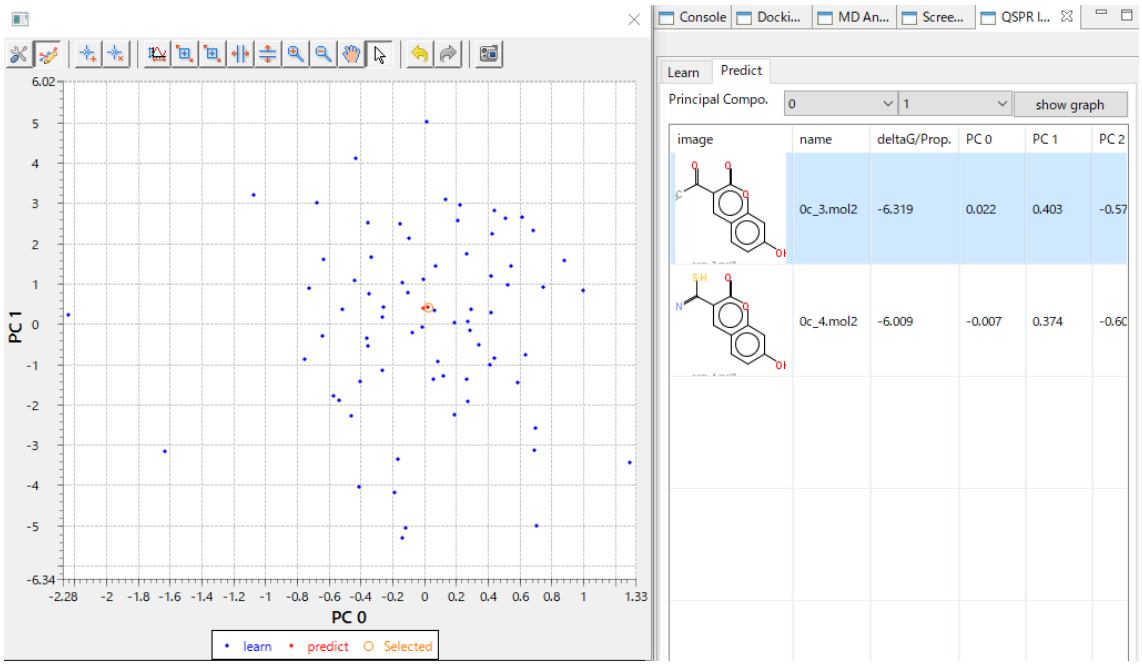
3.3.2. Confirmation of the results of the active value prediction calculation

When the calculation of the activity value prediction is completed, the command button changes from gray to available. Also, [END: Predict Activity] is displayed at the bottom right of the screen.

In addition, on the right side of the screen, display the Predict tab on the QSPR Info screen as shown below.



Now, click the [show graph] button in the red frame above to display the PCA graph. The example figure displays a 0-axis and 1-axis PCA graph. The axis can be selected arbitrarily from 0 to 9.



If you click on a compound in the list, you can see the predicted compound position in the PCA graph with a red circle.

Conversely, clicking the red circle in the graph will focus on the compounds in the list.

The blue dots are the data derived from ChEMBL used for learning. You can check whether the predicted compound is far from the compound derived from ChEMBL, so you can evaluate the reliability of the calculation.

The predicted activity value (logarithmic conversion) is displayed in the deltaG / Prop. (Kcal / mol) column of the table.

4. Prediction of characteristic values of compounds by regression analysis (Predict with Regression model)

Predicts various characteristic values of a compound by regression analysis.

For creating regression parameters from experimental data files of various characteristic values of compounds

[Preparation]-[Make Regression model] 

And, using the regression parameters created above, calculate various property values of multiple compounds at once.

[Screening]-[Predict with Regression model] 

There are two buttons.

In order to predict various characteristic values with [Predict with Regression model], it is necessary to create a data file created by learning regression parameters with [Make Regression model] in advance.

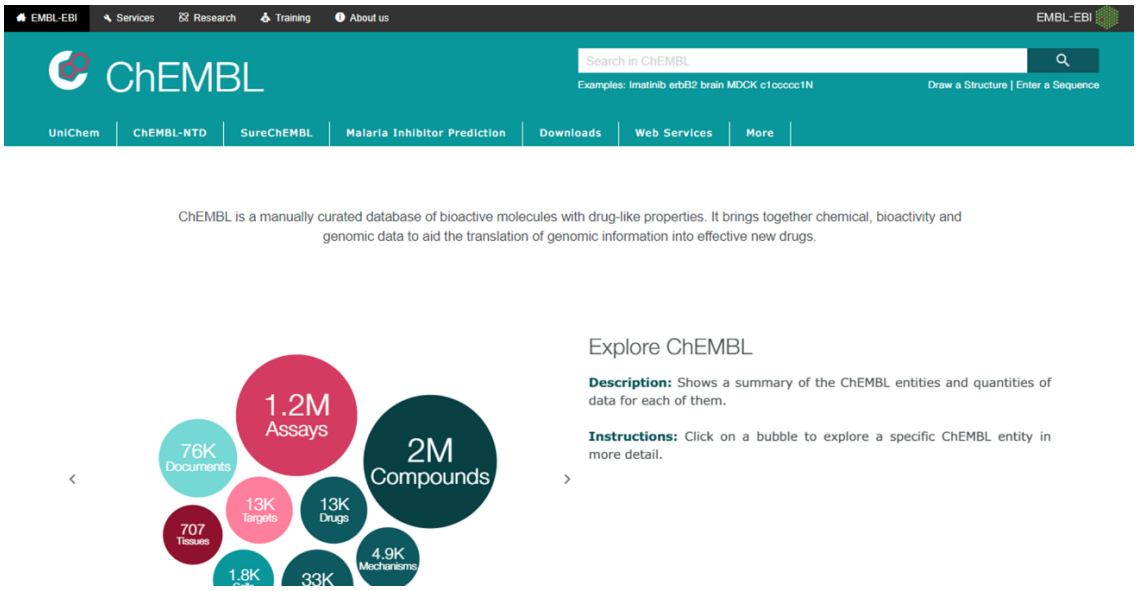
Both [Make Regression model] and [Predict with Regression model] require a relatively long calculation time to create a descriptor of the input compound.

4.1. Acquisitionof CHEMBL experimental value data

Describes the procedure for obtaining experimental value data for various characteristic values from CHEMBL.

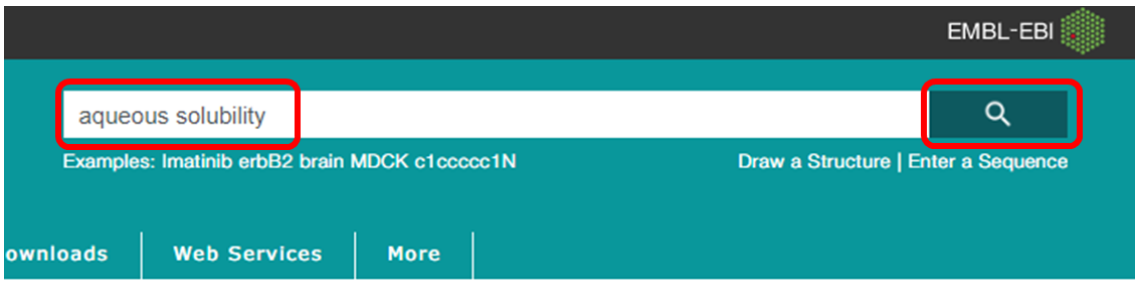
https://www.ebi.ac.uk/chembl/ (ChEMBLE Top Page)

Access the .

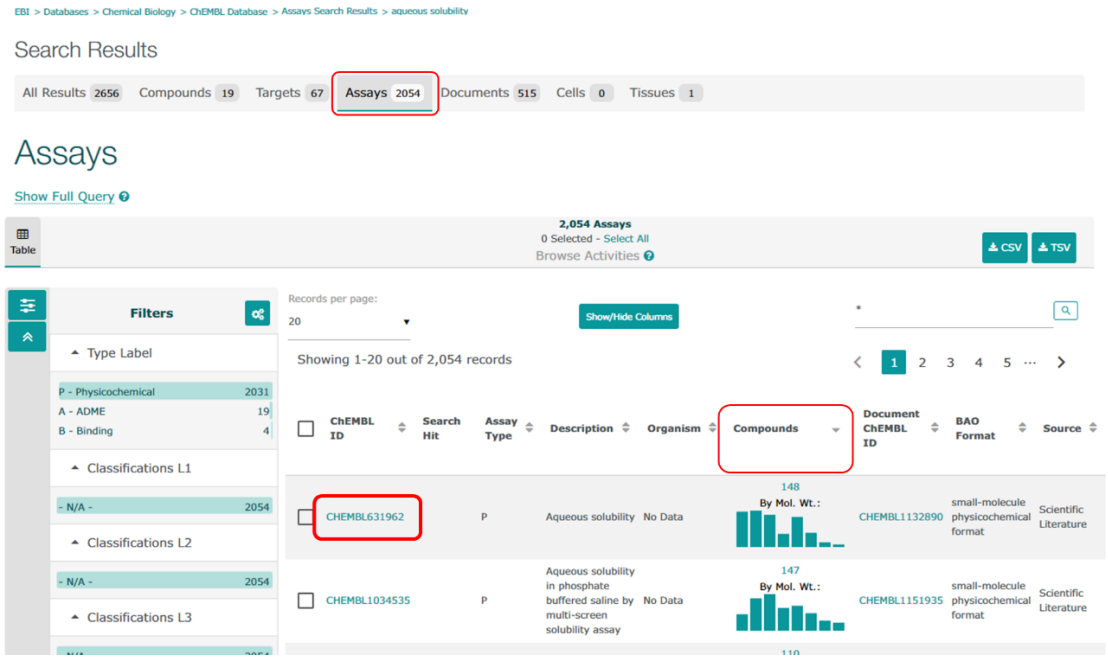


Enter the name of the physical property you are interested in, such as "aqueous solubility" or "permeability", in the "Search in ChEMBL" field as a keyword when searching, and select (click) the search button.

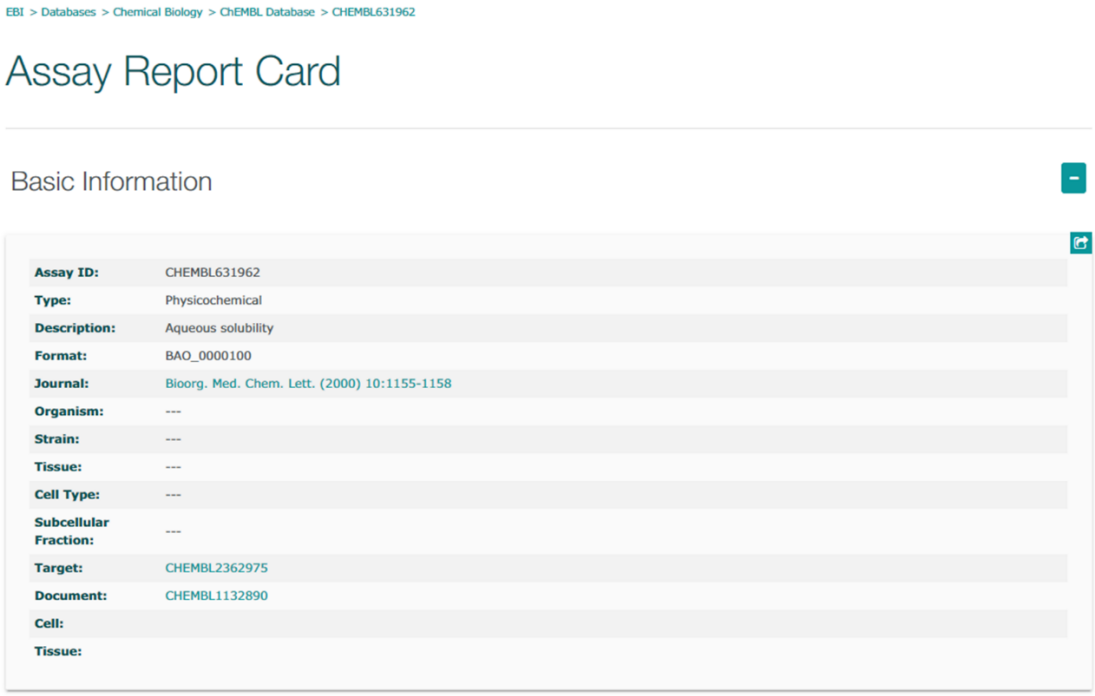
In this example, enter aqueous solubility.



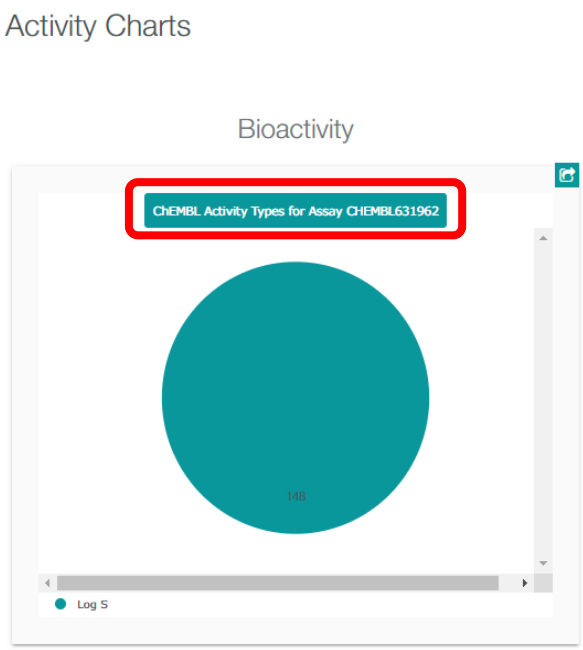
Select (click) the "Assays" group from the displayed search results to display it, and sort the target list by "Compounds" in descending order of the number of data.
Select the appropriate protein from the list.
In this example, select "CHEMBL631962".



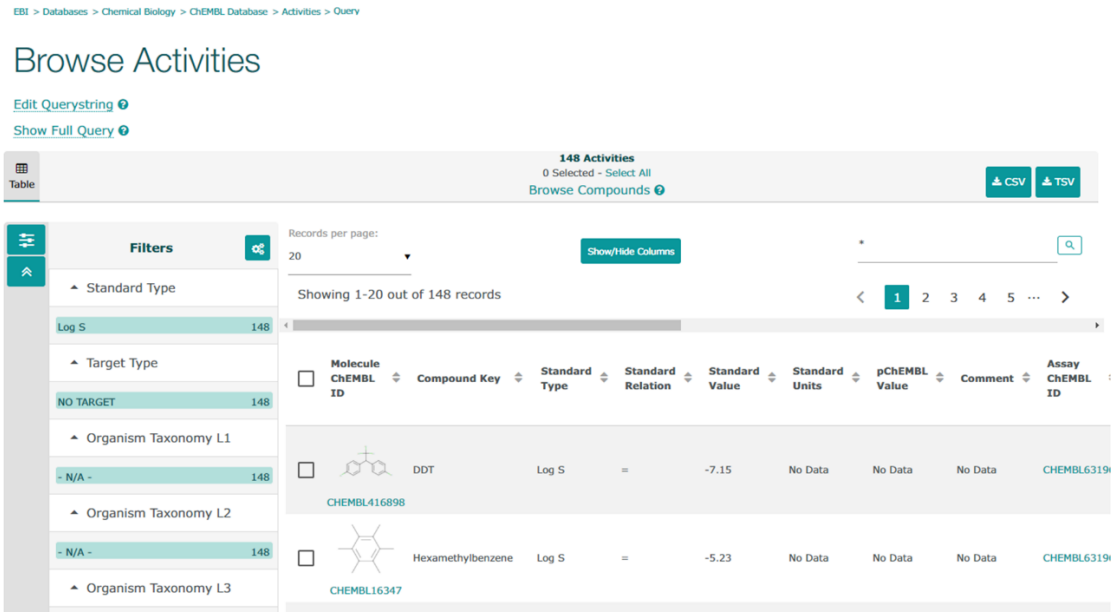
The following screen will be displayed. (The "Assay Report Card" page is displayed)



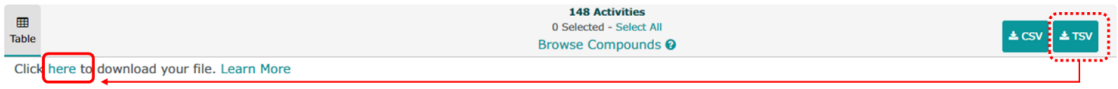
Select (click) Activity Types for Target CHEMBL631962 in the pie chart below on the page.



The following pages appear:



Select (click) [TSV] at the top of the page to generate a download link for the tab-delimited text file. Select (click) the generated [here].



A file called DOWNLOAD-XXXX.tsv.gz will be downloaded. (* XXXX is a long random alphanumeric symbol)

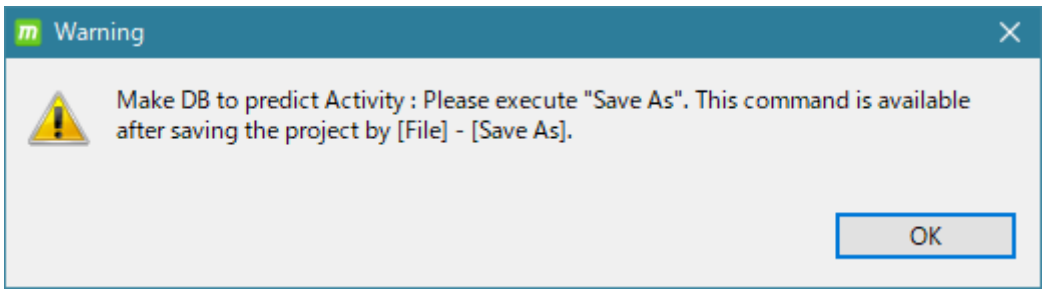
Since the file is compressed in gz format, decompress it into a tab-delimited text file in tsv format with an appropriate decompression software.
DOWNLOAD-XXXX.tsv.gz
↓
DOWNLOAD-XXXX.tsv

4.2. Create input experiment data file

4.2.1. Create a project

Now let's get back to working with MolDesk Screening.
In the File-New Project menu, create an empty project and save it.
Refer to the MolDesk Basic manual for how to save the project.

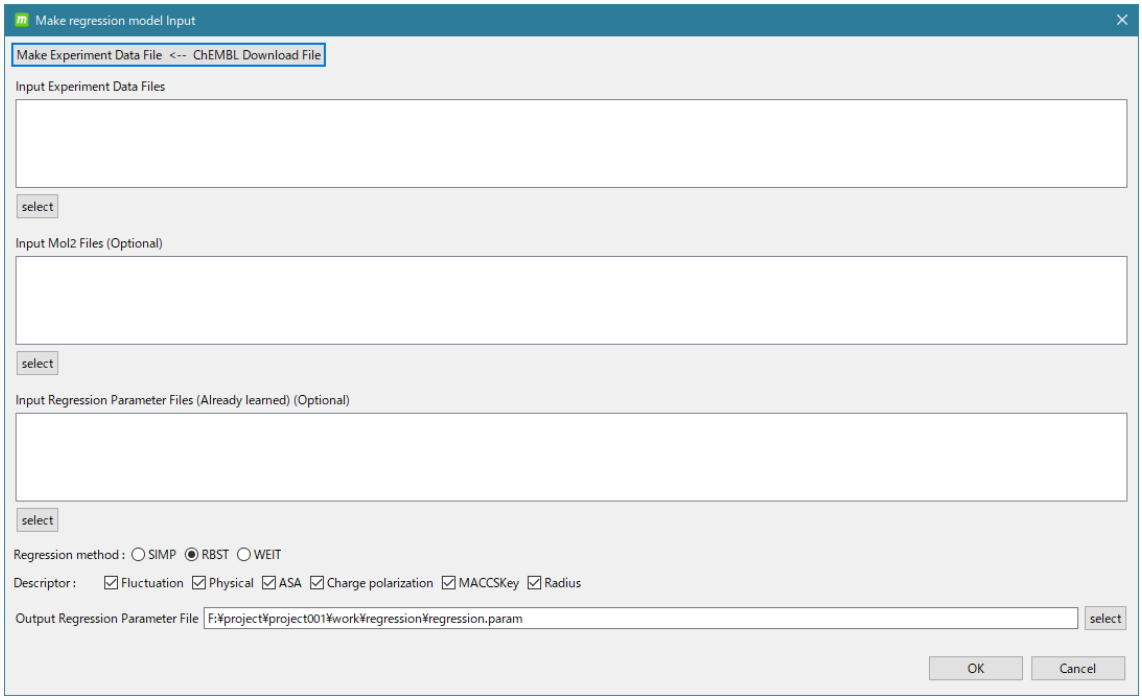
- If the project has not been saved, the following warning dialog will be displayed. Save the project.



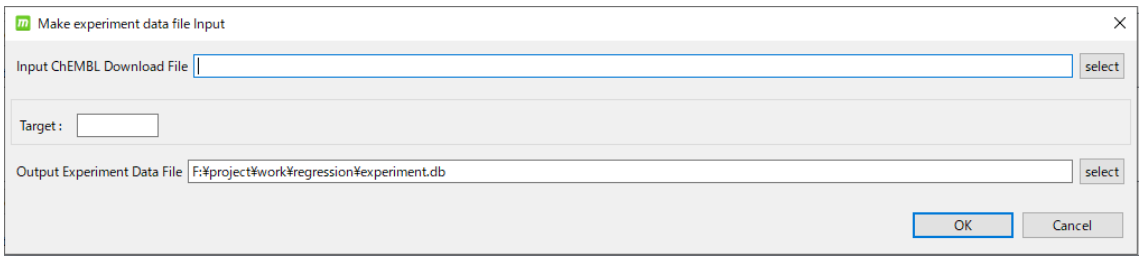
4.2.2. How to create from ChEMBL data file

[Preparation] - [Make Regression model] 

Click. Then the following screen will be displayed.



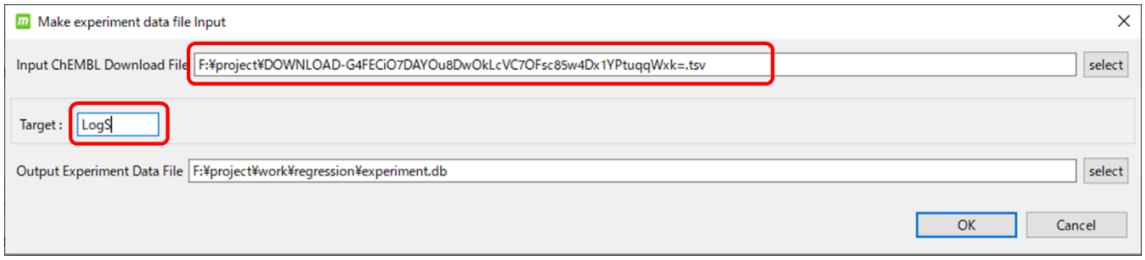
Now, to create an experimental data file, click the top button
[Make Experiment Data File ← ChEMBL Download File] ClickI will.
The following screen will appear.



The input items are as follows.

item	substance
Input ChEMBL Download File	Data file downloaded with ChEMBL (Input Required)
Target	Strings about the type of experiment and the target protein (Input Required) Examples : LogS, LogP, CHEMBL1785, etc.
Output Experiment Data File	Path of experimental data file to output (Input Required)

In this example, enter the following and click OK.
(For the input of [Input ChEMBL Download File], select the downloaded file on the file selection screen that appears when you click [Select].)



At this time, a text file with the following contents is output as an experiment.db file.
The "LogS" entered above is output in red below. Each column is space separated.

```
TARGET LogS VER CHEMBL631962 COMP CHEMBL15844 Log_S -1.21 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL15891 Log_S -4.6 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL14687 Log_S 0.62 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL275626 Log_S 0.58 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL278489 Log_S -3.05 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL279816 Log_S -1.96 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL103 Log_S -4.42 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL211456 Log_S -0.47 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL504760 Log_S -1.96 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL72 Log_S -3.66 - 1.0
・・・
```

The meaning of each column (space delimiter) is as follows.

column	substance	Use in calculations
1	String "TARGET" (fixed)	×
2	Text entered in Target on the input screen	×
3	String"VER"(fixed)	×
4	ASSAY_ID	×
5	String"COMP"(fixed)	×
6	CMPD_CHEMBLID(compound ID)	○
7	STANDARD_TYPE(data type)	○
8	STANDARD_VALUE(data value)	○
9	STANDARD_UNITS (units)	×
10	Weights (default 1.0)	○

※ Here, the character string in the 7th column corresponding to the data type may need to be edited by the user. Please edit with Excel etc.

This is because the calculation program needs to convert the non-logarithmic data value to logarithm in order to perform the regression calculation in logarithm.

The data type of the non-logarithmic data value must be edited to the data type in the table below. (It is OK without editing logarithmic data as in the above example (Log_S).)

Data type	Conversion expression
S	sc = log(sc)/log(10.0)
P	
D	

Pa	if(sc .lt. 0.001) sc = 0.001
	sc = log(sc)/log(10.0) -6.0
	if (sc .le. -30.0) sc =-30.0
	if (sc . ge. 30.0) sc = 30.0

Papp	
Pe	
Peff	

For data values that are not logarithmic,
change to one of the data types S, P, or D for physical property values such as solubility and fat solubility.

For data values related to membrane permeation, change to one of the following data types: Pa, Papp, Pe, Peff.

4.2.3. Example of needing to edit the data type of an experimental data file

An example in which the user needs to edit the data type of the experimental data file created by the method in the previous section is explained below.

```
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL572342  permeability  770.0 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL550752  permeability  1060.0 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL550761  permeability  380.0 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL550905  permeability  1.0 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL551385  permeability  30.0 10'-6_cm/s 1.0
・・・
```

This experimental data file is an example of creating an experimental data file from the ChEMBL download file downloaded by entering "permeability" in the "Search ChEMBL" field as a keyword when searching with ChEMBL, but the data type is "permeability". And the data value is not logarithmicized. (Pa is a string entered by the user in Target.)

In order to calculate accurately with a calculation program, it is necessary to edit to one of the data types Pa, Papp, Pe, Peff with an editing program such as Excel as shown below.

```
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL572342  Pa  770.0 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL550752  Pa  1060.0 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL550761  Pa  380.0 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL550905  Pa  1.0 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL1034536 COMP CHEMBL551385  Pa  30.0 10'-6_cm/s 1.0
・・・
```

In principle, regression prediction of characteristic values other than solubility / fat solubility and membrane permeation can be calculated.
In that case, if the characteristic value is not logarithmically converted, edit it to one of the data types S, P, D or Pa, Papp, Pe, Peff. In the case of a characteristic value that has been logarithmically converted, it can be calculated as it is without editing.

4.2.4. Example of no need to edit the data type in the experimental data file

The following is an example of an experimental data file that you created using the method in the previous section that you do not need to edit.

Example 1

```
TARGET Pa  VER CHEMBL3430218 COMP CHEMBL1294  Papp  28.65 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL3430218 COMP CHEMBL421362  Papp  0.93 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL3430218 COMP CHEMBL3425620  logPapp  0.98 - 1.0
TARGET Pa  VER CHEMBL3430218 COMP CHEMBL3425623  logPapp  1.24 - 1.0
TARGET Pa  VER CHEMBL3430218 COMP CHEMBL3425624  Papp  5.95 10'-6_cm/s 1.0
TARGET Pa  VER CHEMBL3430218 COMP CHEMBL3425629  logPapp  -1.05 - 1.0
・・・
```

In the case of this experimental data file, “Papp” and “logPapp” are mixed as data types, but the data value of “logPapp” is already logarithmic, and the data value of “Papp” is in the calculation program. It is logarithmized with, so there is no need to edit it. (Pa is a string entered by the user in Target.)

Example 2

```
TARGET LogD  VER CHEMBL3301363 COMP CHEMBL638  LogD7.4  1.7 - 1.0
TARGET LogD  VER CHEMBL3301363 COMP CHEMBL639  LogD7.4  2.46 - 1.0
TARGET LogD  VER CHEMBL3301363 COMP CHEMBL642  LogD7.4  -0.27 - 1.0
TARGET LogD  VER CHEMBL3301363 COMP CHEMBL645  LogD7.4  0.1 - 1.0
TARGET LogD  VER CHEMBL3301363 COMP CHEMBL652  LogD7.4  1.13 - 1.0
・・・
```

In the case of this experimental data file, the data type is “LogD7.4”, but the data values are already logarithmic, so there is no need to edit them. (LogD is a string entered by the user in Target.)

4.2.5. A method for editing all experimental data files by the user

Instead of relying on the ChEMBL download file, users can also create experimental data files using their own experimental data that they can edit.

In that case, edit the experimental data file explained in the previous section from scratch.

The meaning of each column (separated by spaces) is as follows.

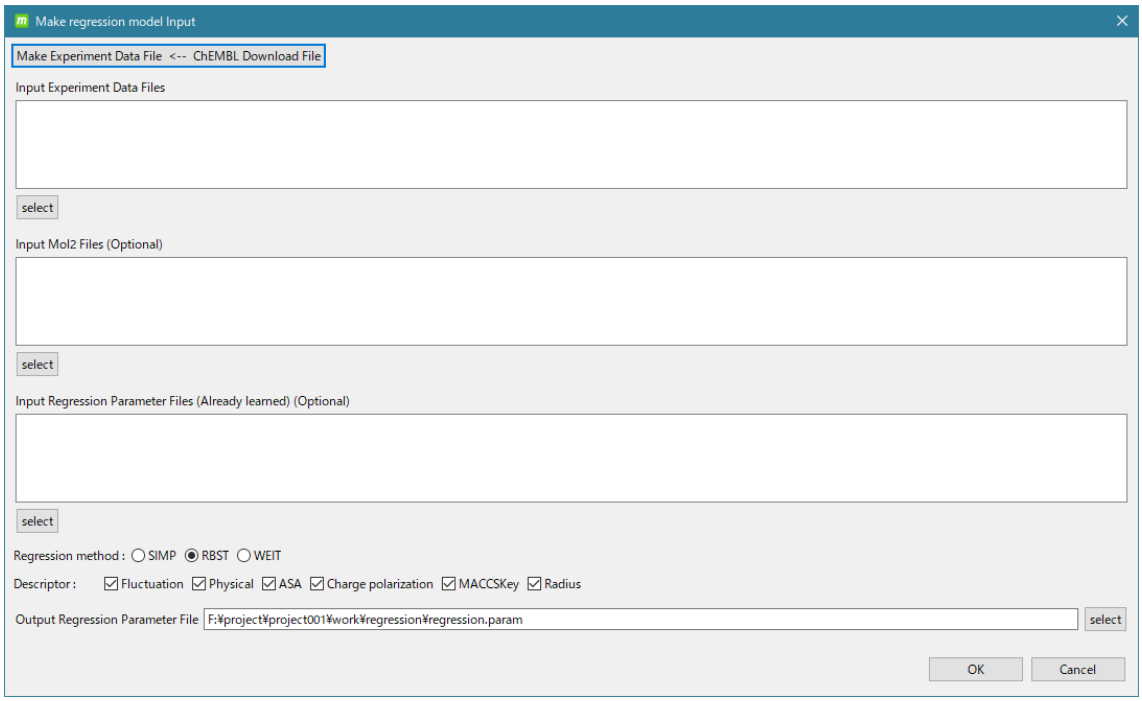
column	substance	Use in calculations
1	String "TARGET" (fixed)	×
2	Text entered in Target on the input screen	×
3	String"VER"(fixed)	×
4	ASSAY_ID	×
5	String"COMP"(fixed)	×
6	CMPD_CHEMBLID(compound ID)	○
7	STANDARD_TYPE(data type)	○
8	STANDARD_VALUE(data value)	○
9	STANDARD_UNITS (units)	×
10	Weights (default 1.0)	○

In this case, you want to focus on the 6th, 7th, 8th, and 10th columns.

4.3. Calculating regression parameters

4.3.1. Input items for regression parameter calculation

When you have finished creating the experiment data file, you will be returned to the screen below when you click [Preparation]-[Make Regression model]  .



The contents of each input item are as follows.

item	substance
[Make Experiment Data File ←ChEMBL Download File] Button	Open the screen for creating an experimentaldatafileto be entered in Input ExperimentData Files from a data file downloaded in ChEMBL (described in the previous section).
Input Experiment Data Files	Experimental data file to be entered (multiple selections allowed)(Input Required)
Input Mol2 Files (Option)	The mol2 file to enter (multiple selections allowed). If no input is available, compute using ChEMBL sdfs. When entering, enter the mol2 file of the compound described in the experimental data file. (Input is not required)
Input Regression Parameter Files (Already learned) (Option)	Regression parameter files already calculated in the past (Multiple selections allowed)(Input is not required)
Regression method	How to calculate regression ・SIMP : All data have the same weight ・RBST: Auto-adjust weights with robust estimation (default) ・WEIT : Use the weights listed in the experimental data file
Descriptor	Type of desyn code for calculation (all defaults on) ・Fluctuation: Fluctuations and dispersion of physical quantities ・Physical: Physical ・ASA: ASA ・Charge polarization: The number of hydrogen bonds and the charge polarity of atoms that can be hydrogen bonded MACCSKey: MACCSKey Radius :Molecular radius (average radius,Rgyr at 3 poles)
Output Regression Parameter File	Output regression paramator file (Input Required)

On this screen, the file input item is

[Input Experiment Data Files]
[Input Mol2 Files (Option)]
[Input Regression Parameter Files (Already learned) (Option)]
[Output Regression Parameter File]

There are four fields that require file entry:

[Input Experiment Data Files]
[Output Regression Paramater File]

There are only two.

In [Input Experiment Data Files], enter the experiment data file created by the method in the previous section. You can enter more than one.

In [Input Mol2 Files (Option)], multiple mol2 files of the compound created by the user can be input and used for the calculation of compound descriptor creation.
If there is no input here, the compound descriptor will be calculated using her CheMBL sdfs set in [Help]-[Preference]-[2.Screening].
The mol2 file entered here has the following restrictions.
1) The file name iscompound ID .mol2 in column 6of theexperimental data file.
2) One molecule, one file.

In [Input Regression Parameter Files (Already learned) (Option)], you can input the regression parameter files already created in the past by this function. You can enter more than one.

In [Output Regression Parameter File], set the path of the regression parameter file to be output. By default
[PROJECT]-> work-> regression-> regression.param
Is set, but if you want to change the path, such as when you want to change the file name, edit it.

In [Regression method], select the method of regression calculation. When WEIT (use the weight described in the experiment data file) is selected, the weight in the 10th

column of the experiment data file is used for the calculation, so the user should edit the weight of the experiment data file as appropriate (the weight of the experiment data file is used). The red part in the example below).

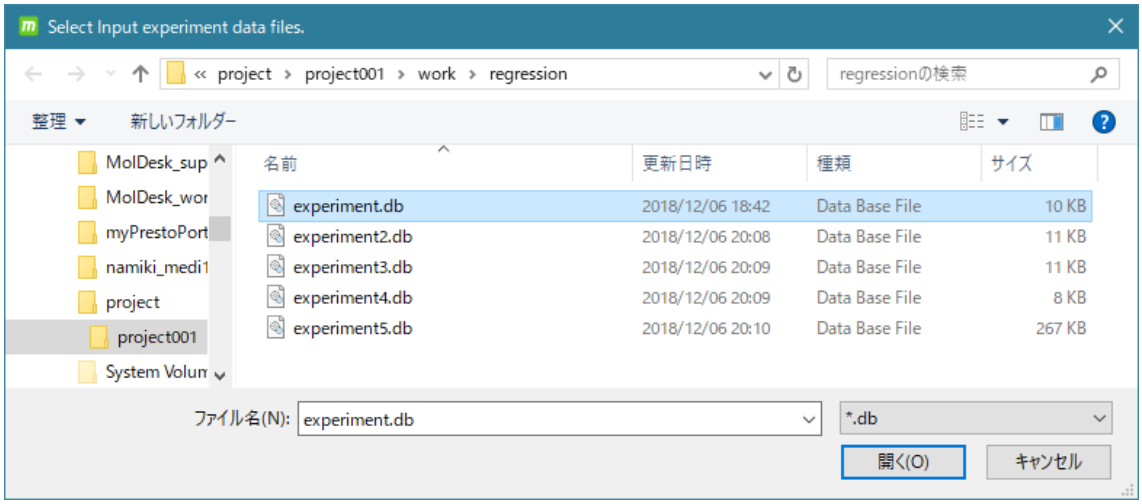
```
TARGET Pa VER CHEMBL3430218 COMP CHEMBL1294 Papp 28.65 10'-6_cm/s 1.0
TARGET Pa VER CHEMBL3430218 COMP CHEMBL421362 Papp 0.93 10'-6_cm/s 1.0
TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425620 logPapp 0.98 - 2.0
TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425623 logPapp 1.24 - 2.0
TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425624 Papp 5.95 10'-6_cm/s 1.0
TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425629 logPapp -1.05 - 2.0
・・・
```

In [Despriptor], the type used in the calculation of the compound descriptor is specified by ON / OFF. By default, all are used (ON).

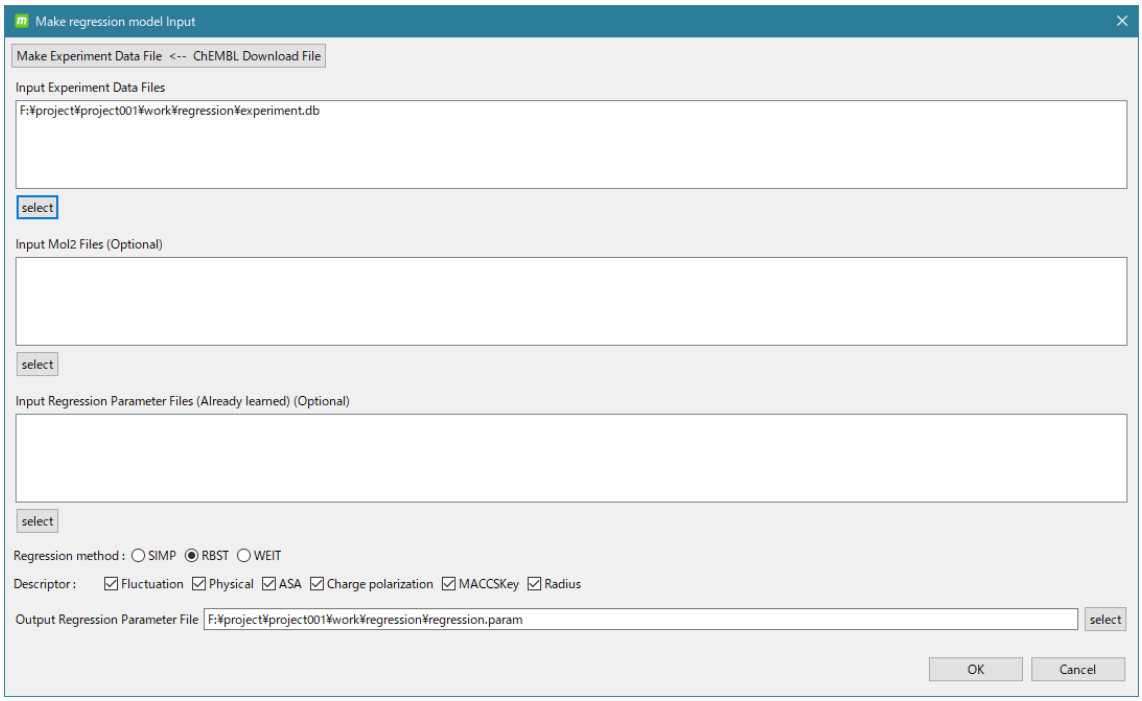
4.3.2. Performing regression parameter calculations

This example describes the case where you enter one experiment data file in Input Experiment Data Files and perform the calculation with the default values for the others.

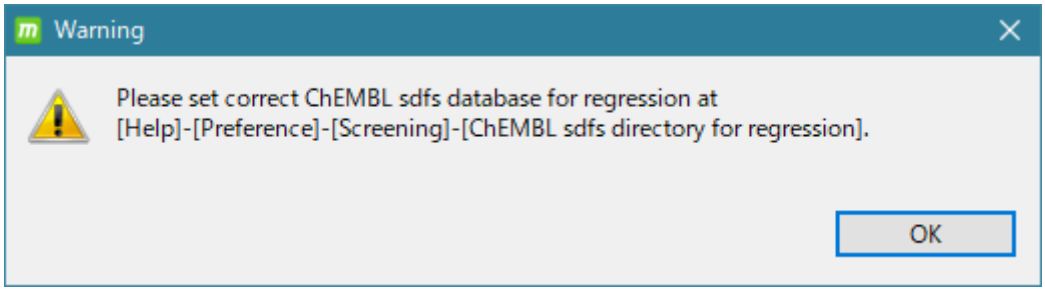
Click [Select] of [Input Experiment Data Files] to display the following file selection screen. Select the experiment data file already created by the method in the previous section.
(Multiple selections are possible, but only one is selected here.)



The selected experimental data file is then taken in as follows.



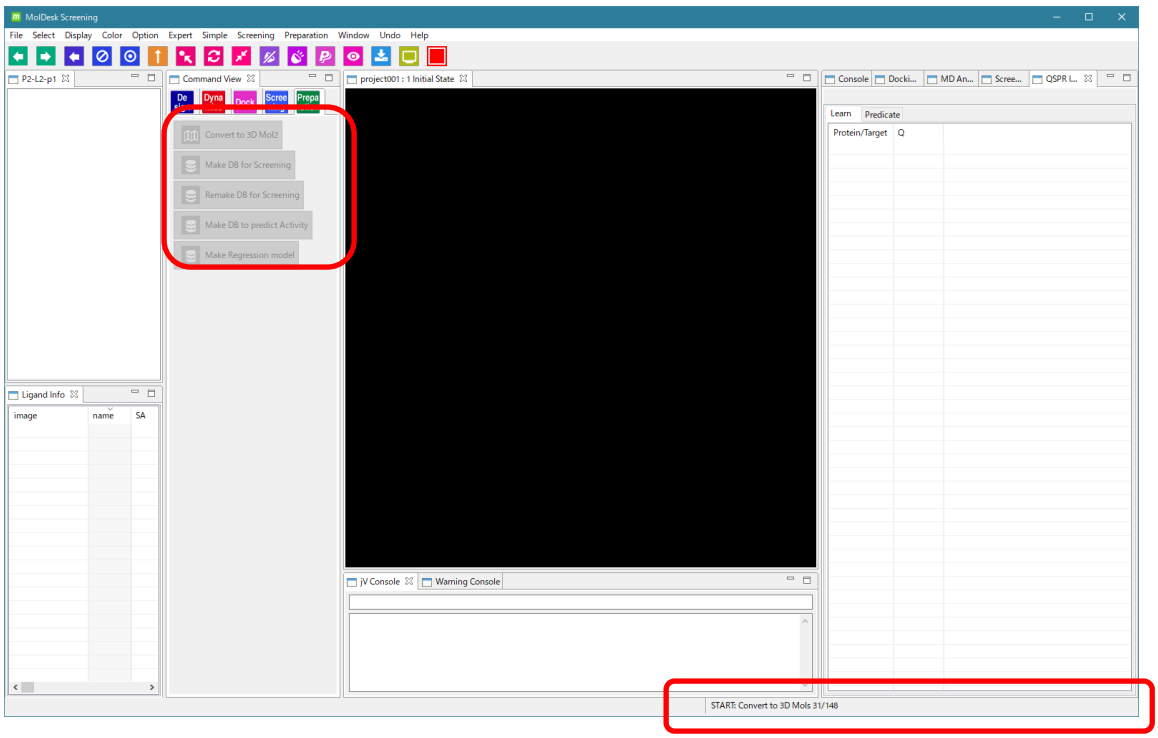
If you click [OK], the following warning screen will appear if ChEMBL sdfs is not set in Preference.



Since you have not entered Input Mol2 Files (Option), the compound descriptor is calculated using ChEMBL sdfs, so the setting of ChEMBL sdfs is mandatory. Refer to the setting method in " 8.3.2 Screening" and set in Preference.

If ChEMBL sdfs is set in Preferecne, the (parallel) calculation will start after clicking OK.

The command button will be grayed out when the calculation starts. Calculation is in progress while the command button is grayed out. In addition, a simple calculation status during calculation is displayed in the red frame at the bottom right.



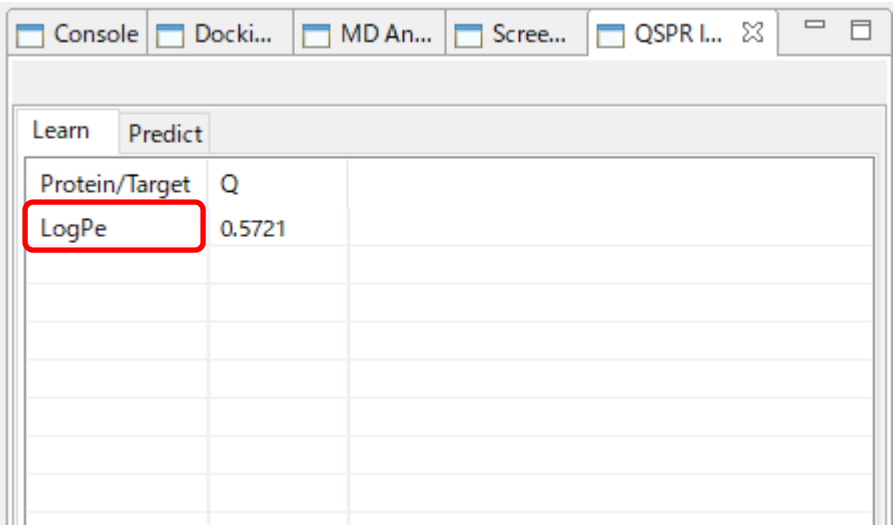
You can work with other projects during the calculation, but be aware that it may be extremely slow depending on the processor occupancy.

The number of parallels when calculating in parallel can be set by Thread number of [Help]-[Preference] – [Screening]. For details, refer to "1.7 Number of Parallel Screening Calculations, Memory Amount, and Time".

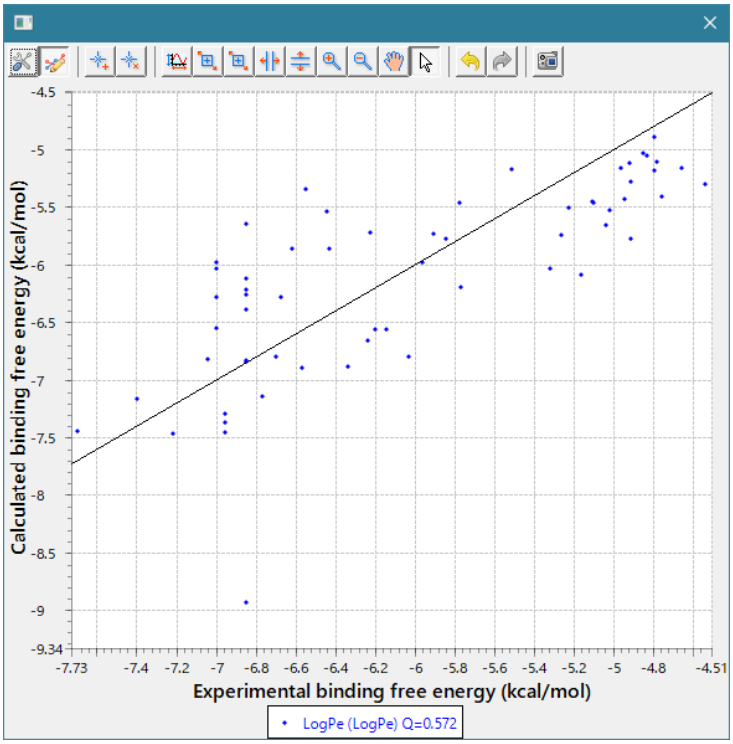
4.3.3. Confirmation of the calculation results of regression parameters by graphs

When the calculation of the regression parameter is completed, the command button changes from gray to available. Also, [END: Make Regression model] is displayed at the bottom right of the screen.

In addition, on the right side of the screen, display the Learn tab of the QSPR Info screen as shown below.
The data type of the experimental data file input when the regression model was created and the correlation coefficient (Q value) of the created regression model are displayed in a list.



Now, double-click the data type name in the [Protein / Target] column in the red frame above.
Then, you can check the reliability of learning by displaying the experimental data values used when creating the regression parameters and the graph of the calculated values when calculating the regression parameters as follows.



4.3.4. Review regression parameter files

The location where the regression parameter file of the calculation result is created is the path specified in [Output Regression Parameter File]. The default is

[PROJECT]-> work-> regression-> regression.param.

This file is important because it will be used in the calculation of the characteristic value prediction described in the next section.

The following folders and files are created in regression.
Users don't have to worry about these contents, but the contents are as follows.

```
[PROJECT] - work - regression - input
                              - mol2
                              - mol2list_*** (file)
                              - error_MakeRegressionModel.log (file)
                              - learn.inp (file)
                              - learn.out (file)
                              - regression.param (file)
```

item	substance
input	The input experimental data file (required) and the regression paramator file (only if there is input) are saved.
mol2	The mol2 and descriptor files of the compound are saved. descriptor input file during calculation (*. inp) and the standard output file (*. (stdout) is also saved.
mol2list_***	List of compound file names (no extensions)
error_MakeRegressionModel.log	Error output during regression parameter calculation (if there is an error)
learn.inp	Input file during learning calculation
learn.out	Standard output file during learning calculation
regression.param	Regression parameter files generated by learning calculations

4.4. Predicting characteristic values

Use the regression parameter data file created in the previous section to calculate the characteristic values of multiple compounds at once (in parallel).

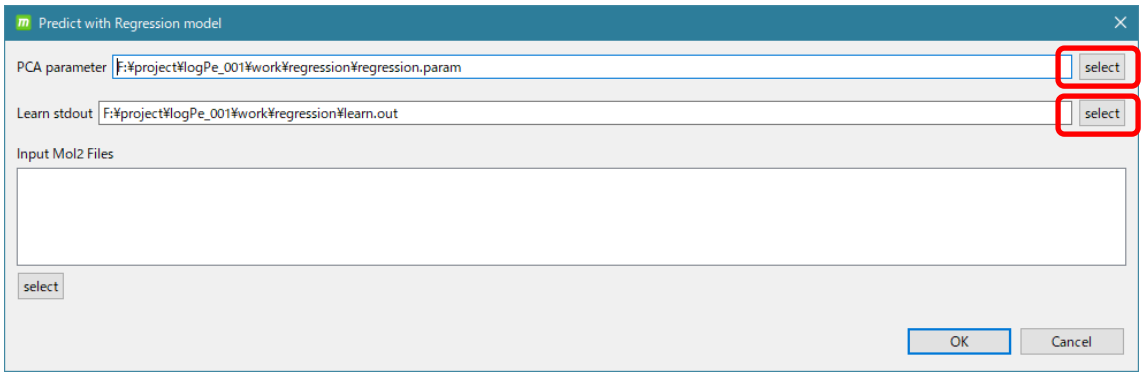
4.4.1. Perform characteristic value prediction calculations

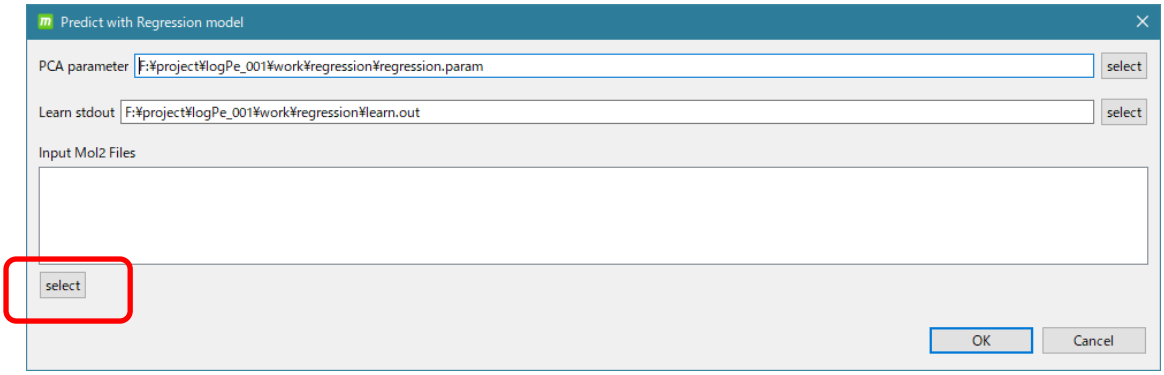
[Screening] - [Predict with Regression model] 

Click. The following input screen will appear.Enter

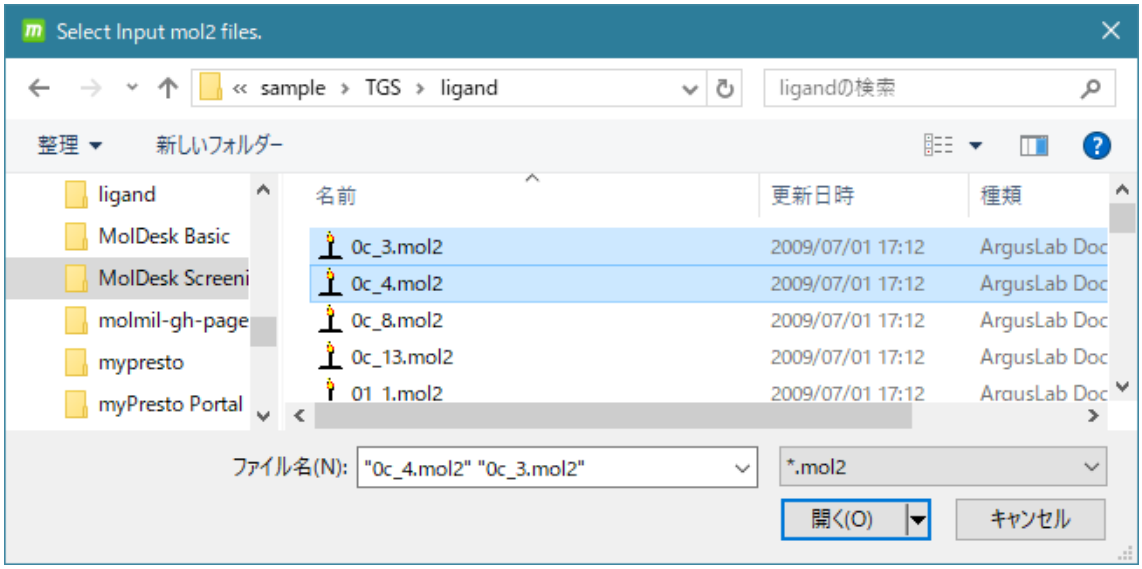
Regression parameter file (created using the method in the previous section)
Standard output file for learning calculations (created using the methods in the previous section)
mol2 file of compounds you want to predict

and click [OK].
The default path is already entered for the regression parameter file and the standard output file for training calculation, but if they are different, click [select] in the red frame and select the correct file on the file selection screen that appears. Please change.





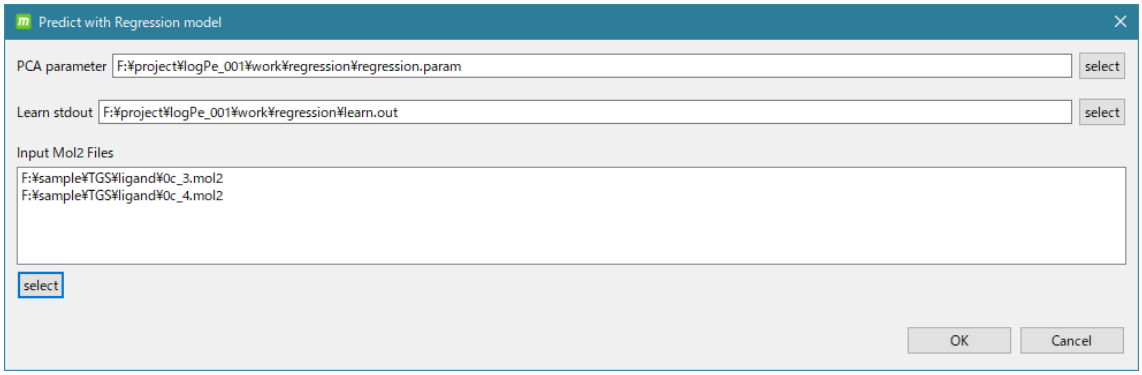
Next, click [select] in the red frame in the above figure, and select the mol2 file of the compound whose activity value you want to predict in the figure below.



Here, as an example, I selected the above two.

※ Make sure that the mol2 file of the compound contains only one molecule. You can use the Mol2 file created by [Preparation] - [Convert to 3D Mol2] of MolDesk Screening as input.

※ We plan to add a function in the near future to enable batch input of multiple molecules in the sdf file.



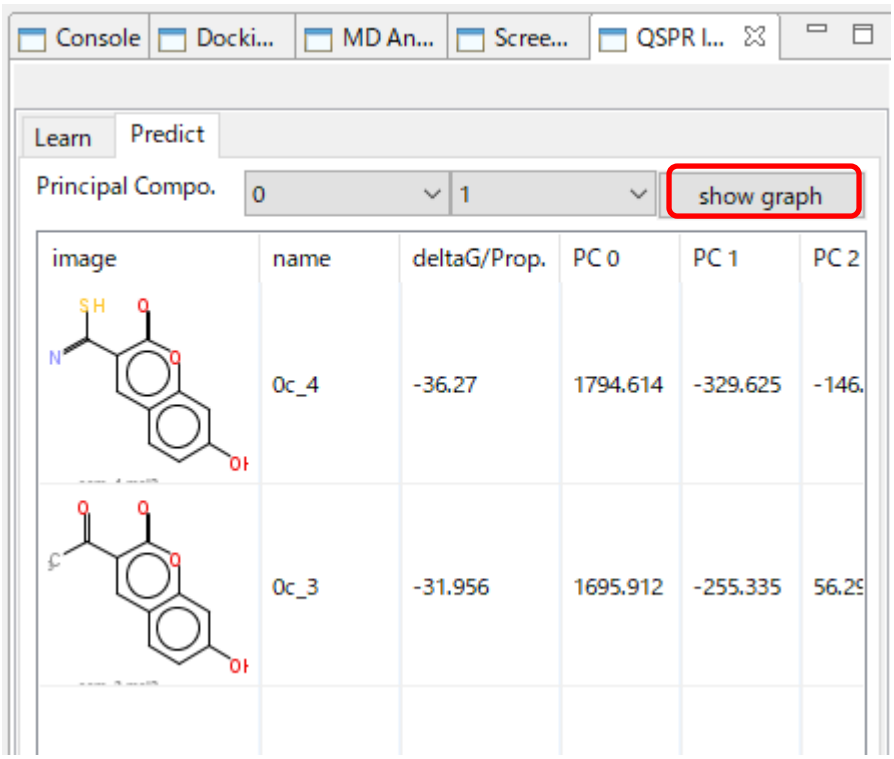
When the input is completed, it will be as shown in the above figure. Click [OK] to start the prediction calculation.

During the forecast calculation, the calculation status is displayed at the bottom right of the screen.

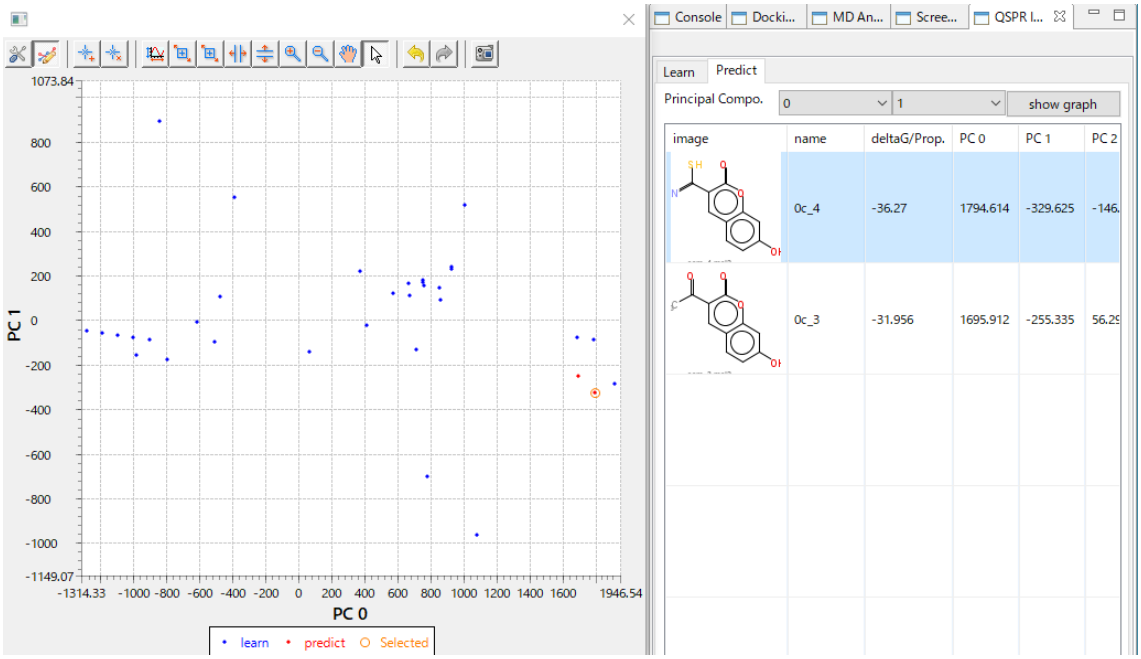
4.4.2. Characteristic value Check the result of the forecast calculation

When the calculation of the characteristic value prediction is completed, the command button changes from gray to available. Also, [END: Predict with Regression model] is displayed at the bottom right of the screen.

In addition, on the right side of the screen, display the Predict tab on the QSPR Info screen as shown below.



Now, click the [show graph] button in the red frame above to display the PCA graph. The example figure displays a 0-axis and 1-axis PCA graph. The axis can be selected arbitrarily from 0 to 9.



If you click on a compound in the list, you can see the predicted compound position in the PCA graph with a red circle.
Conversely, clicking the red circle in the graph will focus on the compounds in the list.
You can check whether it is far from the compound derived from the experimental data file, so you can evaluate the reliability of the calculation.

The predicted characteristic values (logarial conversion) are displayed in the deltaG / Prop. Column of the table.

5. MVO Screening

We search for similar compounds of compound molecules using the MVO Screening method.
Searches for compounds similar to the specified compounds from the compounds entered in the mol2 file.

MVO Screening: (Oldname: MDMVO, also known **asMIN-MVO)**

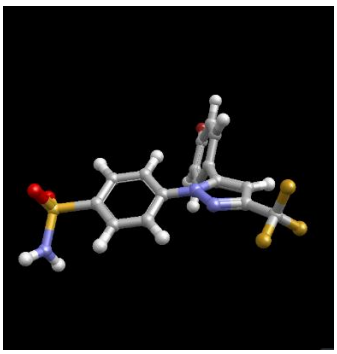
Due to the three-dimensional superposition of two molecules, the one with a large overlap is considered to have high similarity. In superposition, the generation of molecular conformation and the similarity of atomic charges are taken into consideration, and superposition is performed using energy minimization, and the score is the value of% of volume overlap.

5.1. Query molecular selection

Select one compound as the query. This numerator becomes the search query.

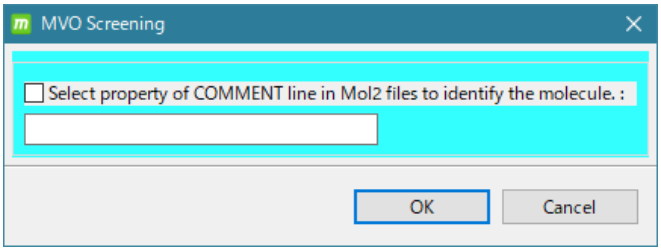
[File] – [Open Molecular File] loads the following files in the MolDesk Screening folder created on the desktop when MolDesk was installed.

MolDesk Screening -> sample -> MVO_screening -> query -> 1cx2_1.mol2



5.2. Select molecules to search for

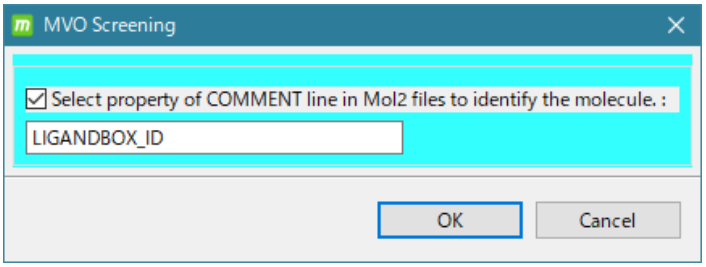
Now, click [MVOScreening] to display the following screen:



Here, enter the property of the COMMENT line (if any) of the mol2 file. Since this property value becomes the ID of the calculation result list, it will be possible to link the molecules, so if there is a property value you want to use for the molecule classification, check it and enter the property value Click [OK].

If the COMMENT line does not exist, just click OK to continue. In that case, by default, the molecule name of the line following the @ <TRIPOS> MOLECULE line becomes the ID.

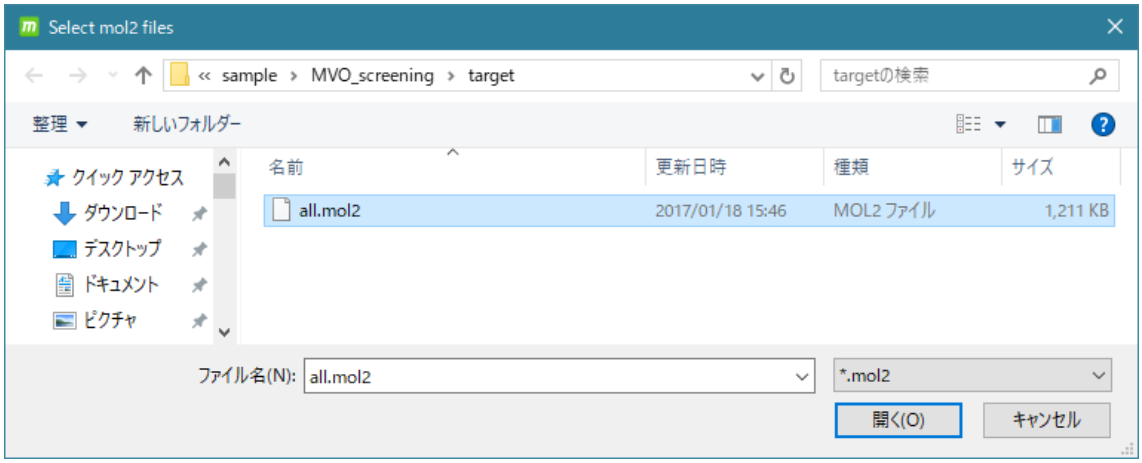
※ Create the mol2 file from the sdf file using the [Convert to 3D Mol2] command.



An example of entry is shown in the figure above. (Enter LIGANDBOX_ID as the property you want to use as the ID, and check the check box.)

Click [OK] to display the file selection screen shown below.
Here, select the mol2 file in the following folder. Click "Open" to start MVO Screening.

MolDesk Screening -> sample -> MVO_screening -> target



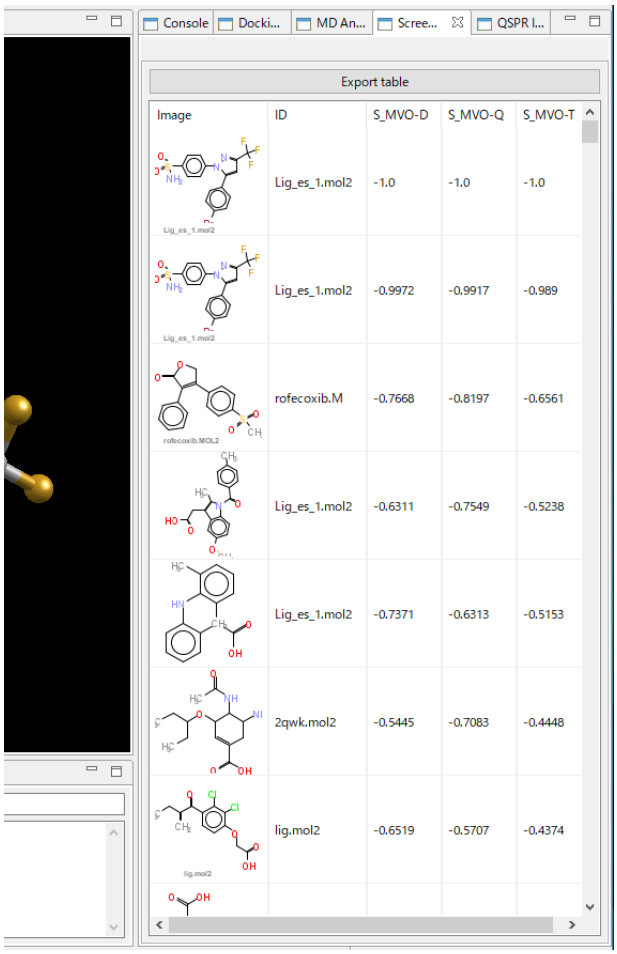
※ In theexample, it is one file, but you can also select multiple files.

5.3. View results

When the calculation is complete, the results are displayed on the Screening Info screen. There are 3 types of scores, SMVO-Q、SMVO-D、and SMVO-T, each of which can be sorted. -1.0 is an exact match, the higher the value (smaller absolute value), the higher the similarity.

The characteristics of each score are as follows.

SMVO-Q:	Small molecules in the database are selected
SMVO-D:	Large molecules in the database are selected
SMVO-T:	Database molecules close to query molecules are selected



In addition, the following molecular properties are also displayed.
Formula, Weight, Charge, Donor, Acceptor, Chiral atoms

You canalsoclick the [Export table] button to save theresultsto a c svfile, anhtml file.

6. Similar structure search

We will search for similar structures of compound molecules using the TGS (Topology Graph Similarity) method.
The compound whose structure is similar to the specified compound is searched from the compounds entered in the mol2 file.

Topology Graph Similarity :

This is a method to search for compound similarity using the molecular graph with the covalent bond of the molecule as the edge and the matrix eigenvalue as an index.
Converts the molecular structure information into a real-valued vector and calculates the similarity from the vector distance.
It is very fast, but it is indistinguishable between optical isomers and conformations.

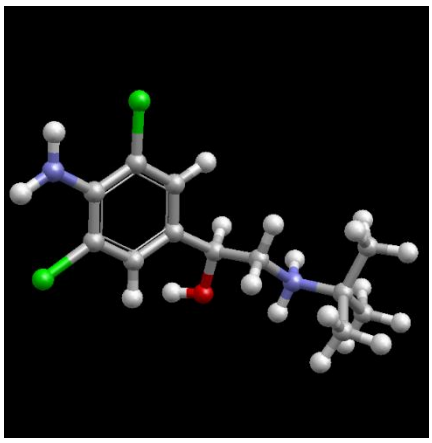
6.1. Select query molecules

Select one compound as the query. This numerator becomes the search query.

Select [File] – [New Project] to load the following files in the MolDesk Screening folder created on the desktop when MolDesk was installed.

MolDesk Screening-> sample-> TGS-> query-> query.mol2

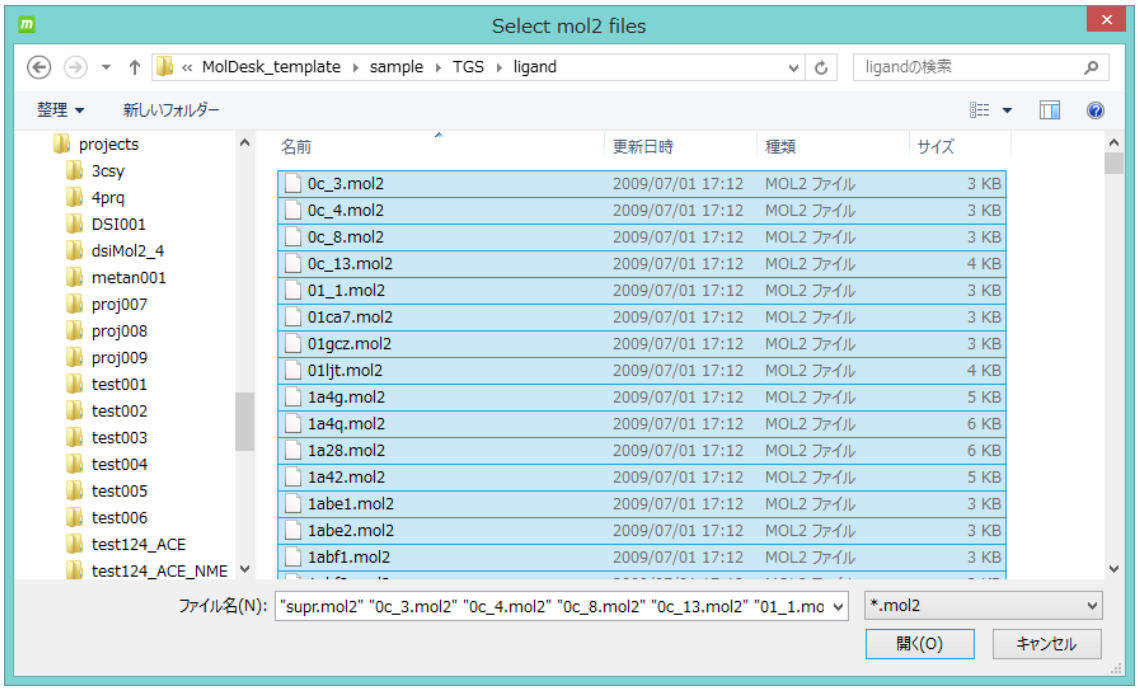
For [New Project], refer to the MolDesk Basic manual.



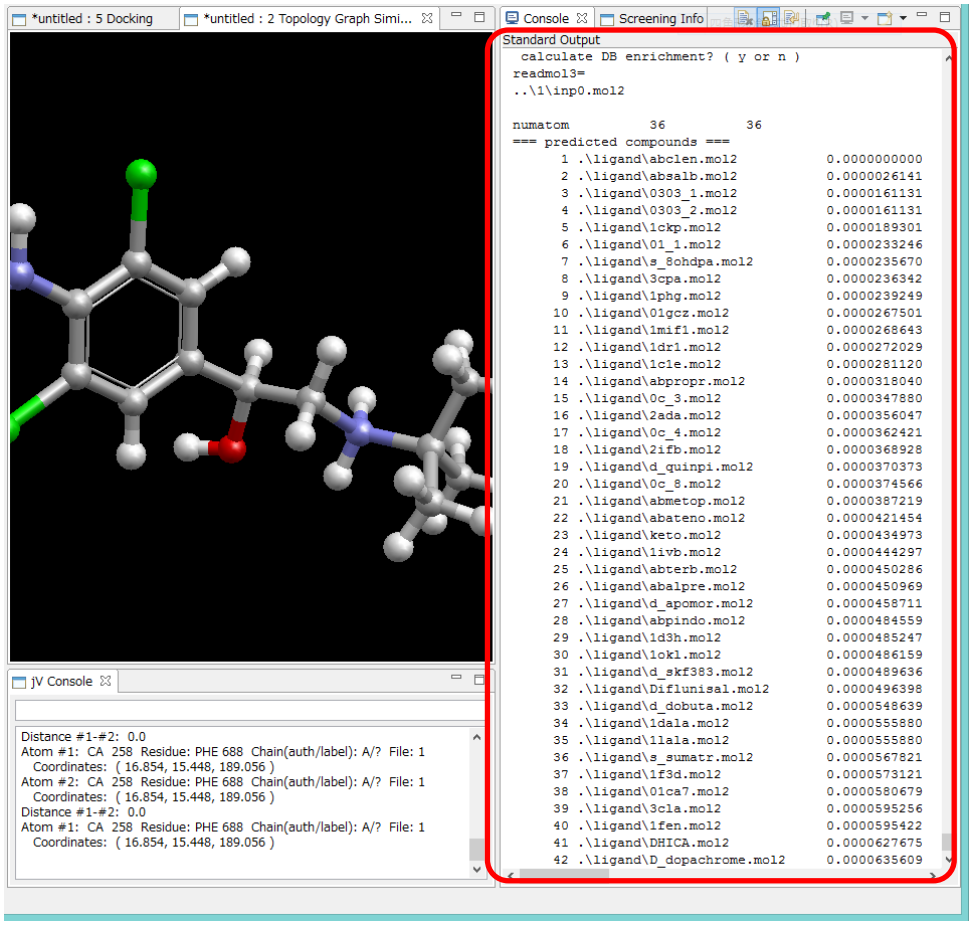
6.2. Select, calculate, and display results of search target molecules

[Topology Graph Similarity] Click , and then select all the mol2 files in the following folders: Click "Open" to start searching for similar structures.

MolDesk Screening -> sample -> TGS -> ligand



When the calculation is complete, the result will be displayed on the Console screen. From the left, the ranking, compound name, and score (0.0 is an exact match, the smaller the value, the higher the similarity) are displayed.



7. Partial structure search

Substructure Search searches for compounds whose partial structure is similar to the specified compound from the compounds entered in the mol2 file.

The molecule is transformed into an edge matrix with chemical bonds as edges, and the partial structures are compared according to Ulmann's theorem. It does not consider molecular conformations or optical isomers.

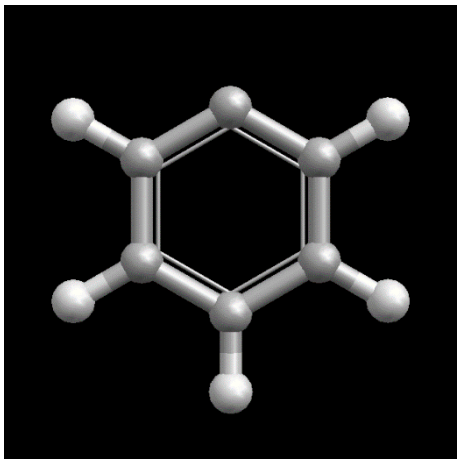
7.1. Query molecular selection

Select one compound as the query. This numerator becomes the search query.

Select [File] – [New Project] to load the following files in the MolDesk Screening folder created on the desktop when MolDesk was installed.

MolDesk Screening-> sample-> substructure_search-> query-> lig1.mol2

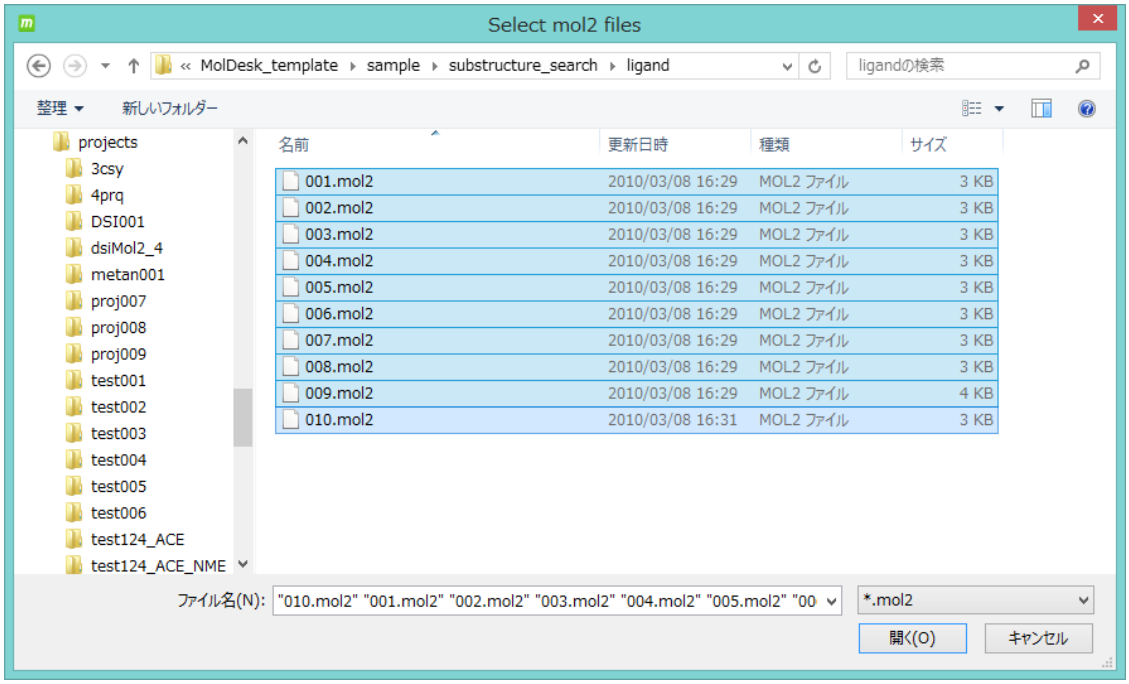
For [New Project], refer to the MolDesk Basic manual.



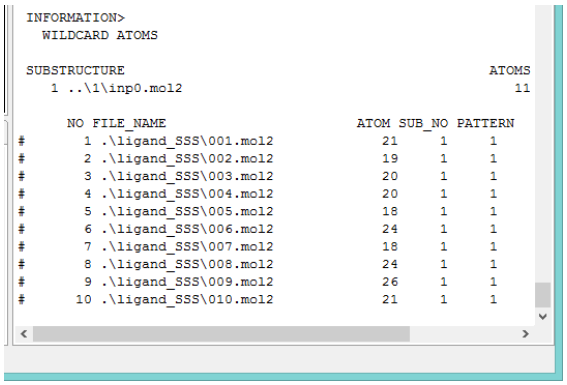
7.2. Selecting, calculating, and displaying results for substructure search target molecules

Click Substructure Search and select all mol2 files in the following folders: Click "Open" to start the similar structure search.

MolDesk Screening-> sample-> substructure_search-> ligand



When the calculation is complete, the result will be displayed on the Console screen. From the left, the internal number, file name, number of atoms, number of searched substructures, and number of substructures found are displayed.



```
 INFORMATION>
   WILDCARD ATOMS

 SUBSTRUCTURE                                          ATOMS
    1 ..\1\inp0.mol2                                      11

      NO FILE_NAME                          ATOM SUB_NO PATTERN
#      1 .\ligand_SSS\001.mol2               21     1      1
#      2 .\ligand_SSS\002.mol2               19     1      1
#      3 .\ligand_SSS\003.mol2               20     1      1
#      4 .\ligand_SSS\004.mol2               20     1      1
#      5 .\ligand_SSS\005.mol2               18     1      1
#      6 .\ligand_SSS\006.mol2               24     1      1
#      7 .\ligand_SSS\007.mol2               18     1      1
#      8 .\ligand_SSS\008.mol2               24     1      1
#      9 .\ligand_SSS\009.mol2               26     1      1
#     10 .\ligand_SSS\010.mol2               21     1      1
```

8. Fast parallel computing of MD calculations using MPI/GPU

You can run four molecular dynamics calculators of myPresto and GROMACS:

※ Mac does not support parallel computing by MPI or CUDA for molecular dynamics calculations of myPresto.

MD Program	Required operating environment	MD calculation capabilities
cosgene	- No particular	All MD calculations are possible
cosgene_MPI	MPI 64bit Limited	All MD calculations are possible
psygene	MPI 64bit Limited	・ Calculations other than periodic boundary conditions (*) are not allowed. ・ The size of one side of the aerodic solution cannot be less than 54Å
psygene-G	MPI CUDA 64bit Limited	・ Calculations other than periodic boundary conditions (*) are not allowed. ・ The size of one side of the aerodic solution cannot be less than 54Å
GROMACS	Windows64bit Linux / MAC depends on user's installation and execution environment	・ Calculations other than periodic boundary conditions (*) are not allowed.

※Cube produces aero solvent

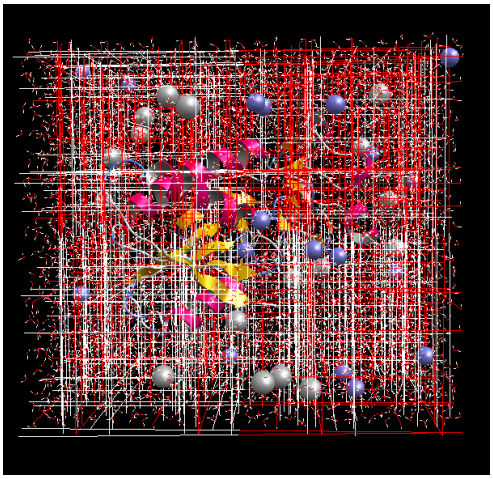
- cosgene_MPI and psygene, MPI must be set.
- psygene - Grequires MPI and CUDA to be set .
- Currently, MPI is not multi-node compatible. Works multicore on one node.

- psygene -G supports up to four multi-GPUs per node.
- psygene -G requires an NVIDIA graphics board.
  - GF100 generation or higher (GTX460 or higher, Compute Capability 2.0 or

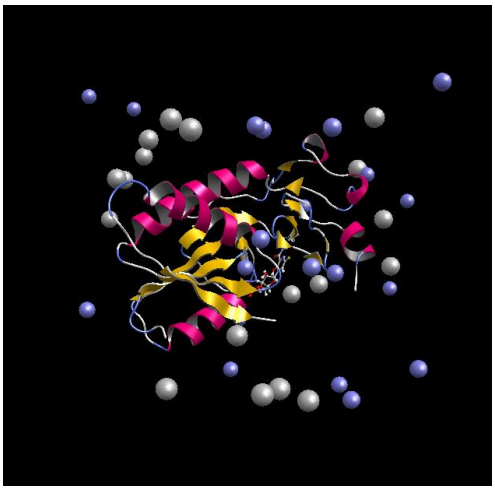
higher) is required.

- The more video memory you have, the more large problems you can calculate.

- Inpsygene -G, without SHAKE, the water molecules at the periodic boundary are displayed in a line like the following, but it is not an anomaly in the calculation.



In this case, right-click and select the water molecule on the tree view screen to hide it in the Hide Atom menu. Other molecules are displayed neatly as shown below.



8.1. How to set up an MPI operating environment

8.1.1. Windows 64bit

Install Microsoft MS-MPI.

https://www.microsoft.com/en-us/download/details.aspx?id=100593

From, click Download to download msmpisetup.exe. Double-click the downloaded msmpisetup.exe to complete the installation. Environment variables are set at the same time.

8.1.2. Linux 64bit

Open MPI(https://www.open-mpi.org/) or MPICH(https://www.mpich.org/)
Install. Please refer to each manual for instructions.

The open MPI installation command is as follows:

ForD ebian 64bit Linux
 $ sudo apt-get install openmpi-bin libopenmpi-dev

For Redha t-based 64-bit Linux
 $ yum install openmpi openmpi-devel

In the case of Ubuntsu, the environment setting is completed at the same time, but in the case of CentOS, it is necessary to set the path such as export PATH = $ PATH: / usr / lib64 / openmpi / bin / in the ~ / .bash_profile file.

8.2. How to set up a CUDA operating environment

The PC must have an NVIDIA graphics board.
For graphics boards, GF100 generation or later (GTX460 or higher, Compute Capability 2.0 or higher) is required. The more video memory you have, the larger problems you can calculate.

8.2.1. Windows 64bit

It works if the latest version of nvidia's graphics drivers is installed.
Download the graphics drivers below.
https://www.nvidia.co.jp/Download/index.aspx?lang=jp

8.2.2. Linux 64bit

It works if the latest version of nvidia's graphics drivers is installed.
Download the graphics drivers below.
https://www.nvidia.co.jp/Download/index.aspx?lang=jp

8.3. Preference settings

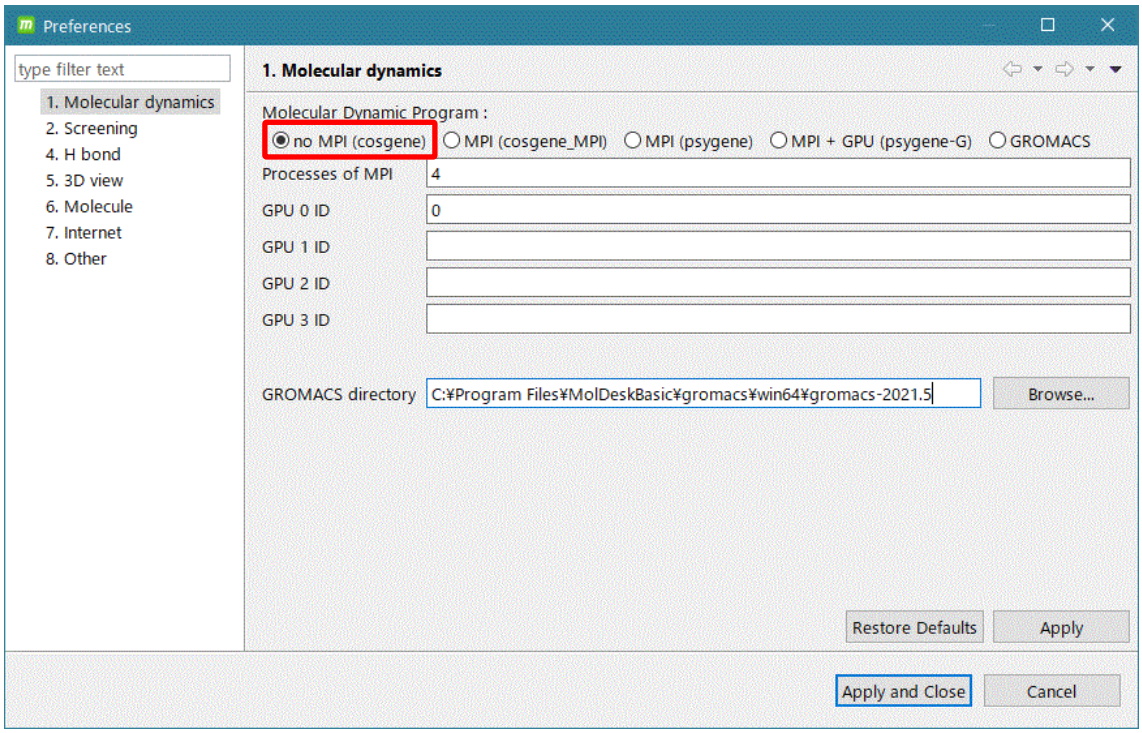
You can set various Preference values with [Help] – [Preference].

This section describes only the "Molecular Dynamics" and "Screening" items that need to be set only in MolDesk Screening. For other items, refer to the MolDesk Basic manual.

8.3.1. Molecular Dynamics

Set up md calculations. Here's a description of each option:

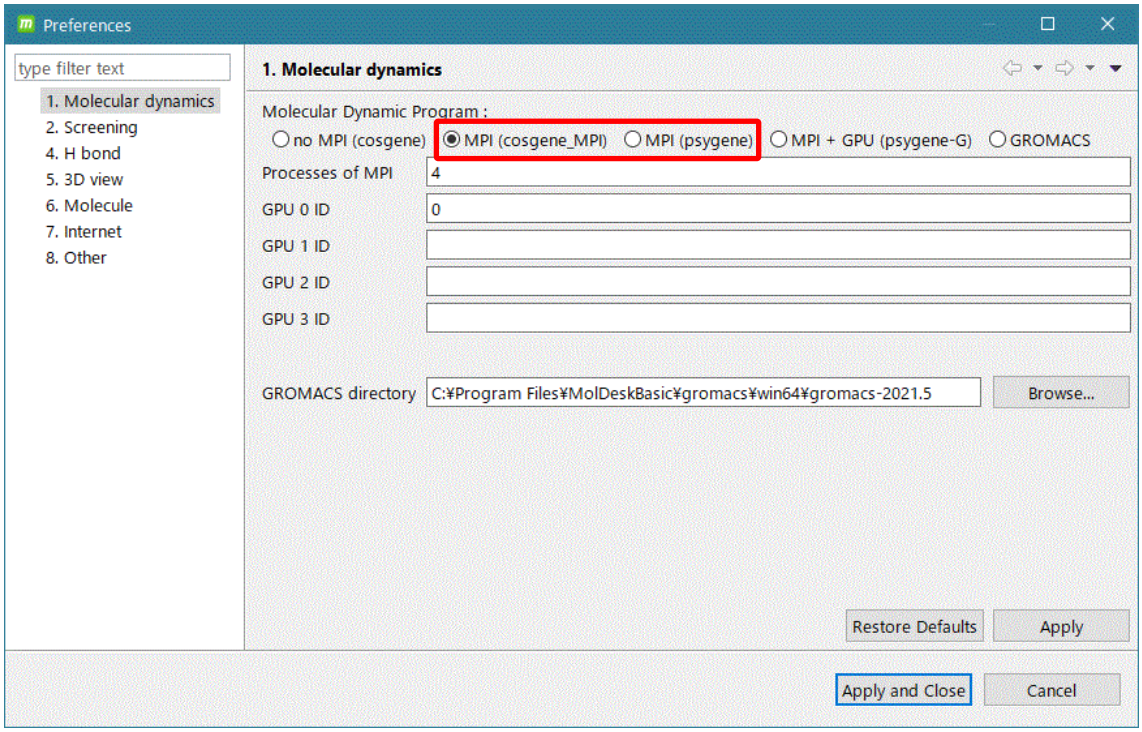
- not MPI (cosgene)



When [not MPI (cosgene)] is selected, parallel calculation is not performed. No environment settings such as MPI are required.

The settings of [Processes of MPI],
[GPU 0 ID], [GPU 1 ID], [GPU 2 ID], and [GPU 3 ID]
are ignored.

- MPI (cosgene_MPI) or MPI (psygene)



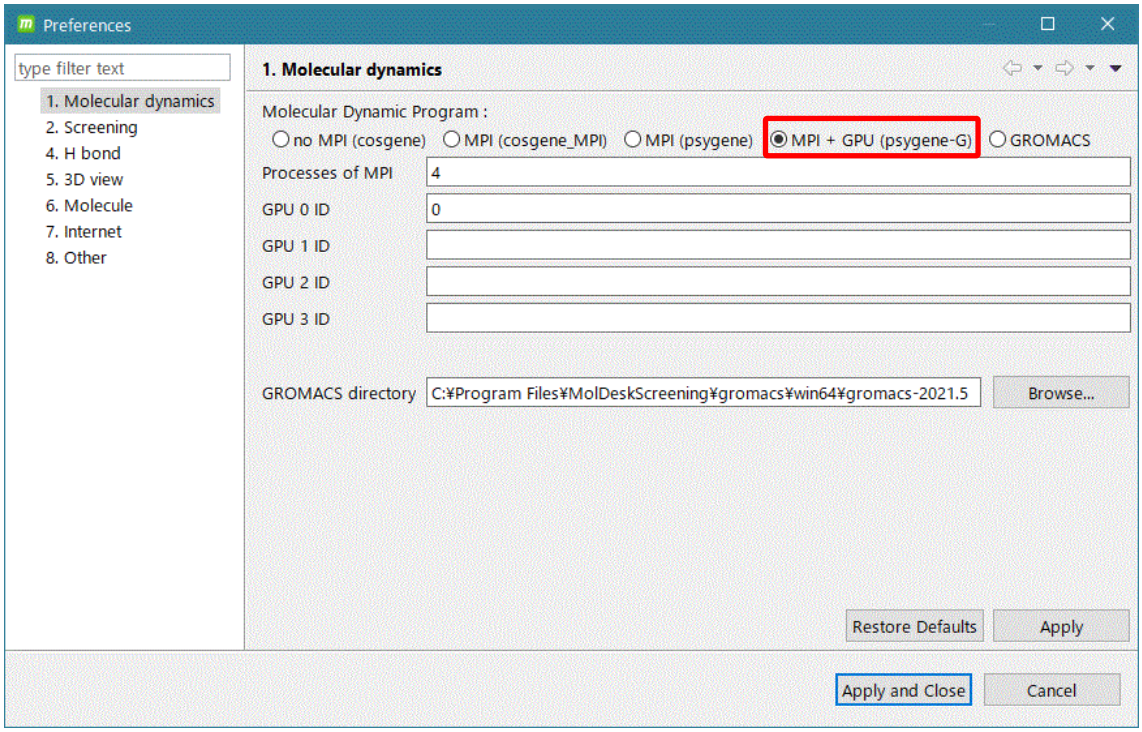
When MPI (cosgene_MPI) or MPI (psygene) is selected, parallel calculation by MPI is performed.

Set the number of MPI parallels in Processes of MPI. The default value is the maximum number of physical processors on the installed system. Normally you do not need to change this value.

[GPU 0 ID] [GPU 1 ID] [GPU 2 ID] [GPU 3 ID] settings are ignored.

※ Mac or Windows 32bit does not support parallel calculation of MD calculation by MPI and CUDA, so this setting screen is not displayed.

- MPI + GPU (sievgene-G)

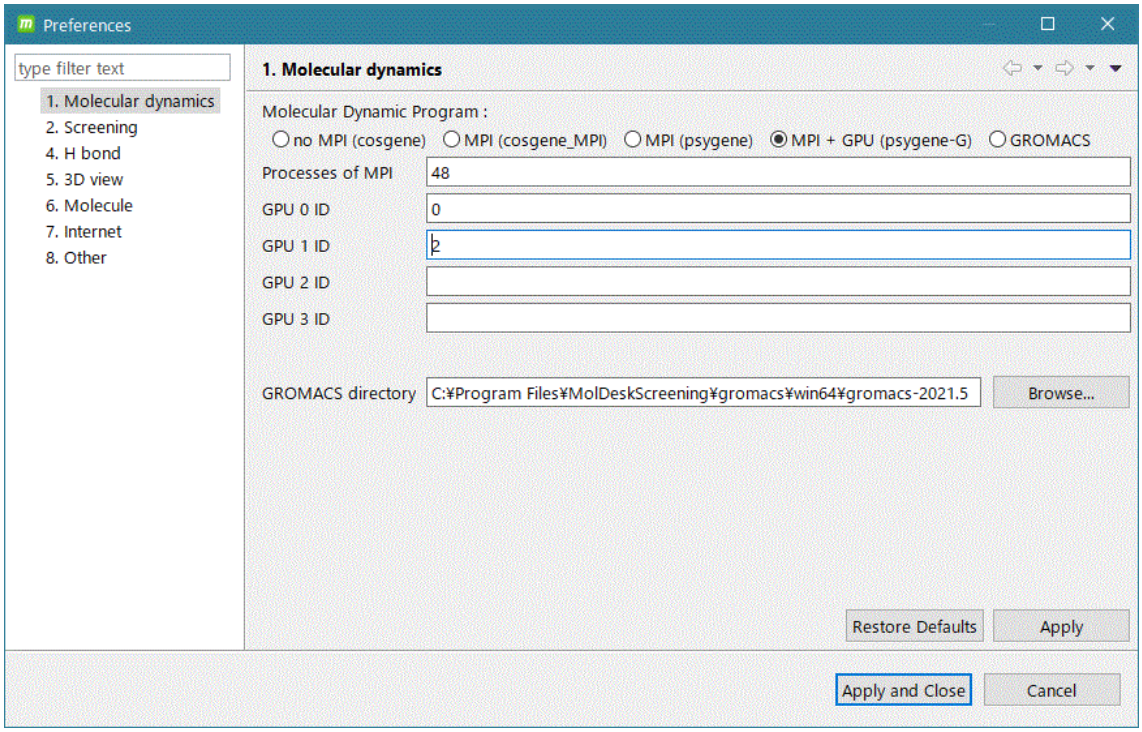


When [MPI + GPU (sievgene-G)] is selected, parallel calculation by MPI + GPU is performed.

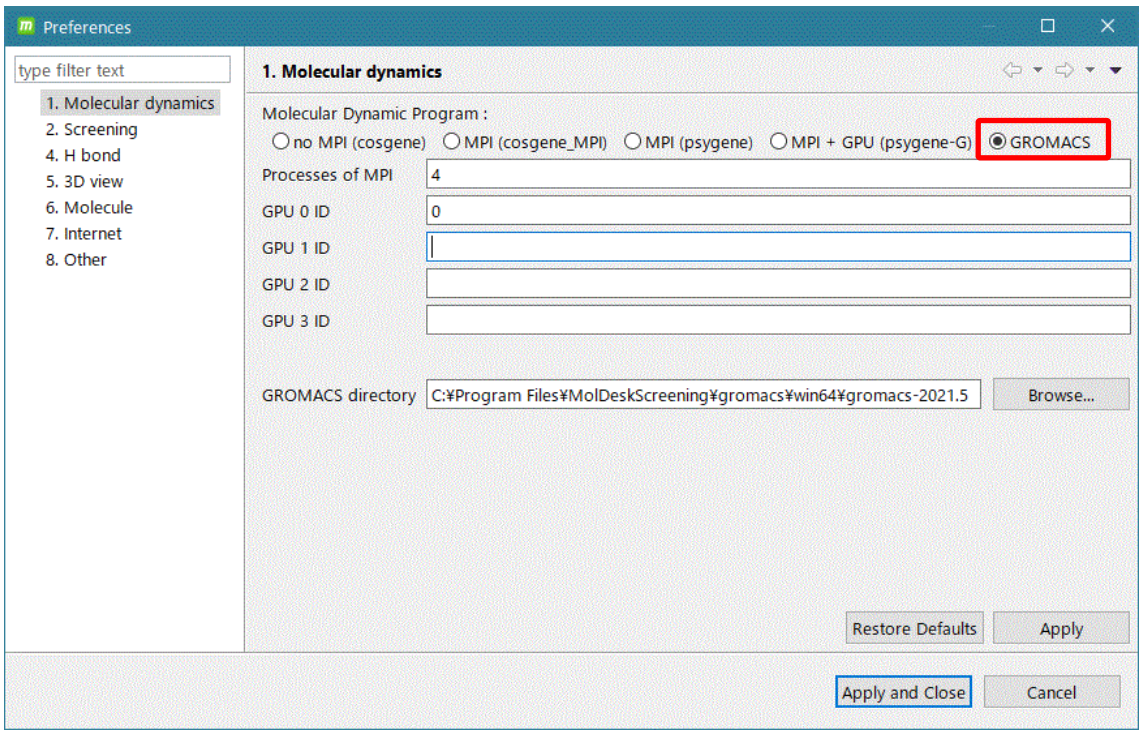
Set the number of MPI parallels in Processes of MPI. The default value is the maximum number of physical processors on the installed system. Normally you do not need to change this value.

[GPU 0 ID] [GPU 1 ID] [GPU 2 ID] [GPU 3 ID]
Set the Device ID of the GPU used in. Up to 4 Device IDs of the GPU to be used can be set, and up to 4 multi-GPU calculations are possible.

This is an example of using two GPU boards with Device ID 0 and Device ID 2 (MPI parallel number is 48).



- GROMACS

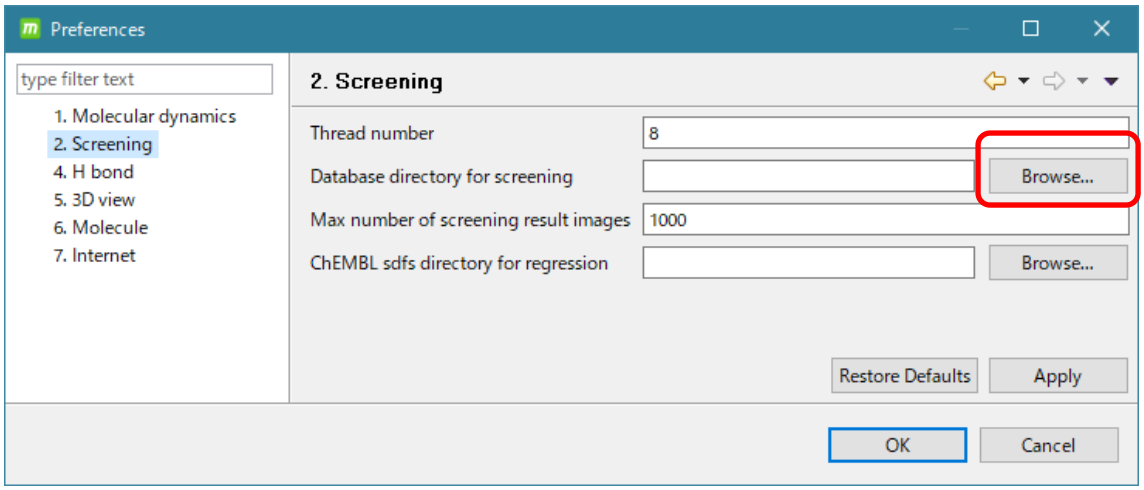


When [GROMACS] is selected, MD calculations with GROMACS are performed.
See the MolDesk Basic Manual for details on usage.

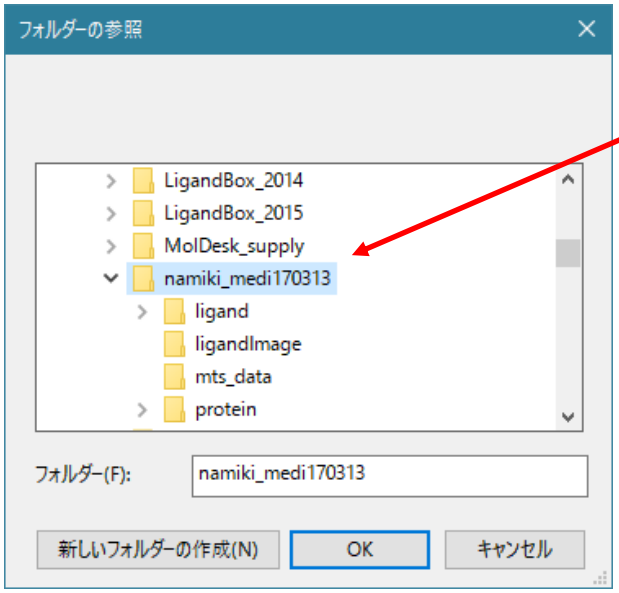
8.3.2. Screening

Set up the LigandBox or user-created database used for screening calculations.

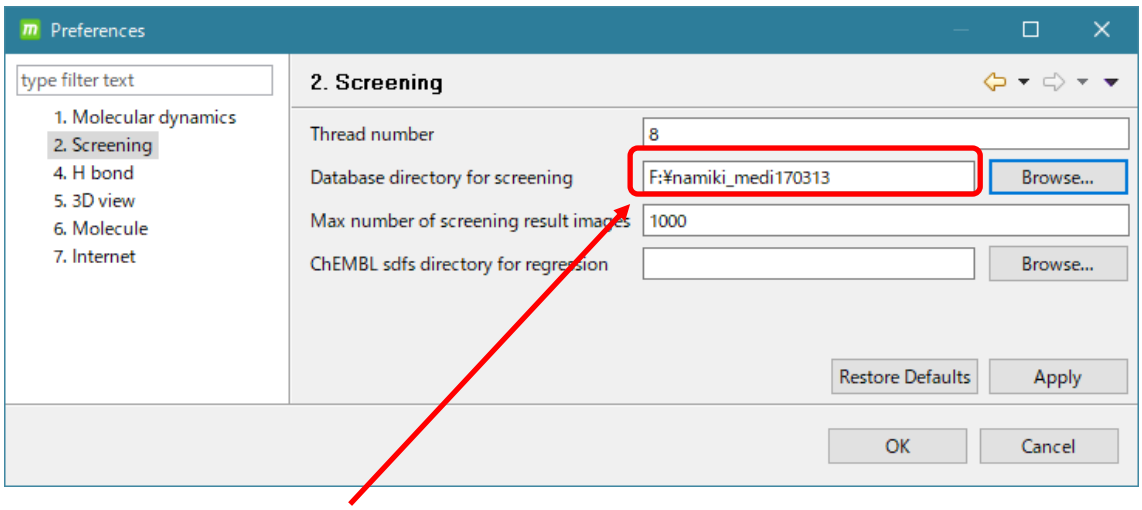
Open the [Help]-[Preference] screen, select "2. Screening", and click [Browse].



Ligand
ligandImage
mts_data
protein
of LigandBox (namiki_medi170313 in the example below) unzipped in "1.2 Preparing LigandBox"

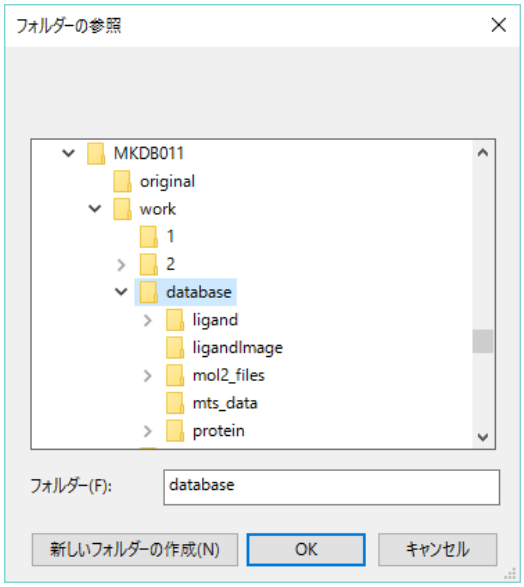


Select the folder (directory) immediately above the folder (directory) and click [OK]..



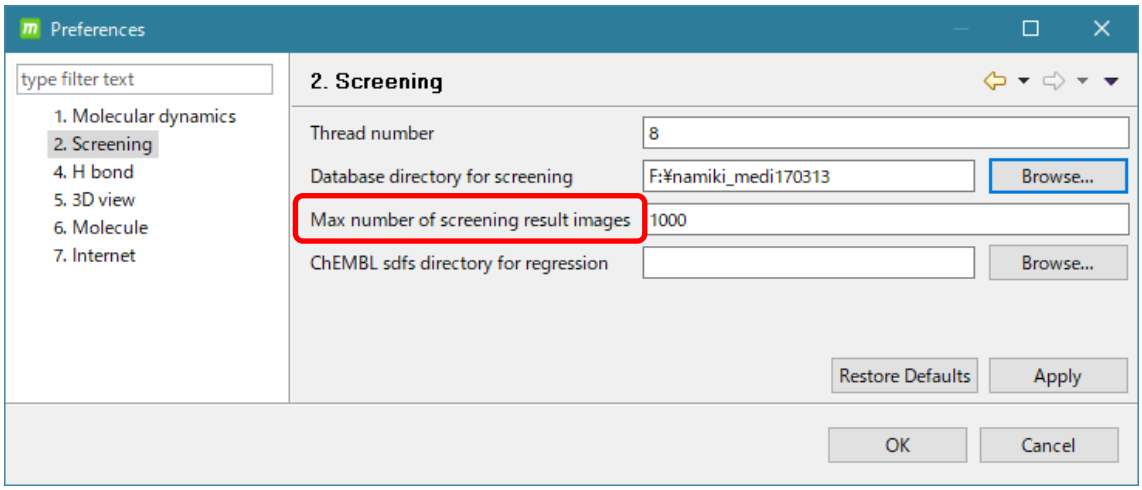
Then, LigandBox will be set as shown above. Click [OK].
Now you can perform the screening calculation.

If you want to use the database created in "1.3 Preparing the compound DB for screening specified by the user" instead of LigandBox, select the database folder of the saved project as shown in the figure below and click [OK].



Make sure the database folder is specified and click OK.

You can change the number of 2D chemical structure diagrams in [Max number of screening result images] on the [Help]-[Preference] screen.



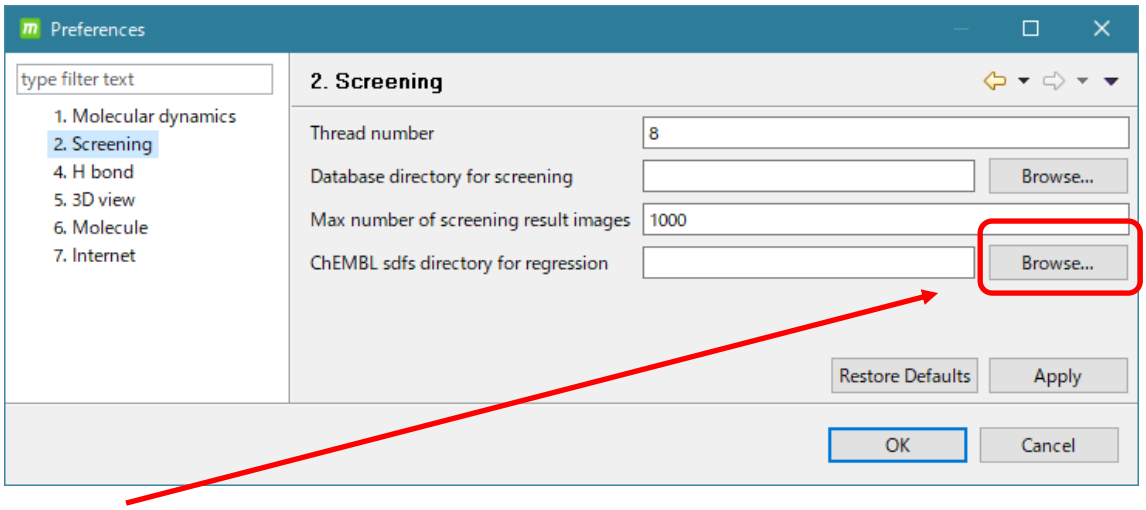
[Screening Info] You can change the number of screening result 2D chemical structure diagrams displayed on the screen.

For example, if this value is set to 1000, even if the screening result is 1000 or more, the 2D chemical structure diagram of the 1001st and subsequent results will not be displayed.

If there is no need to change it, the default (1000) can be used.
You can reduce the resources used by the computer by making this value less than 1000. The larger the size, the longer it will take to generate the 2D chemical structure diagram when the project is opened.

If you change the value, close the project once and then reopen it. The number of 2D diagrams in the screening result list is changed and displayed.

Select the [Help]-[Preference]-[2.Screening] screen and configure the ChEMBL sdfs settings.
Settings for performing regression analysis predictions ([Make Regression model] and [Predict with Regression model]) of various properties of a compound.



Click [Browse].

ChEMBL sdfs unzipped in "1.3 Preparing ChEMBL sdfs"

(In the example below, chembl_24_sdfs_moldesk)

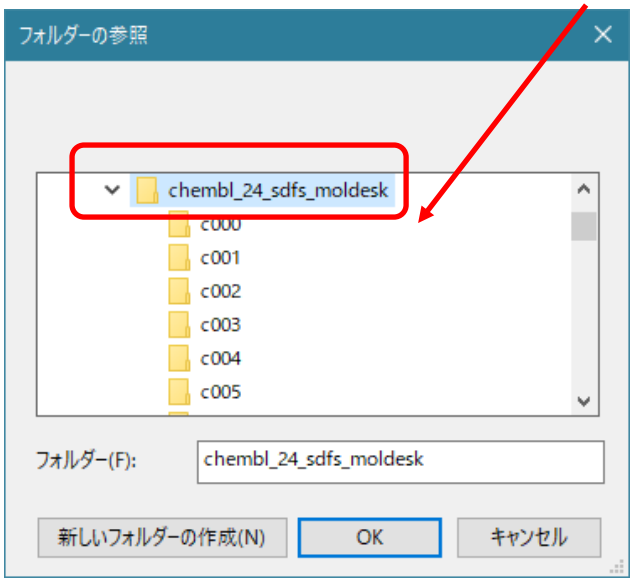
c000

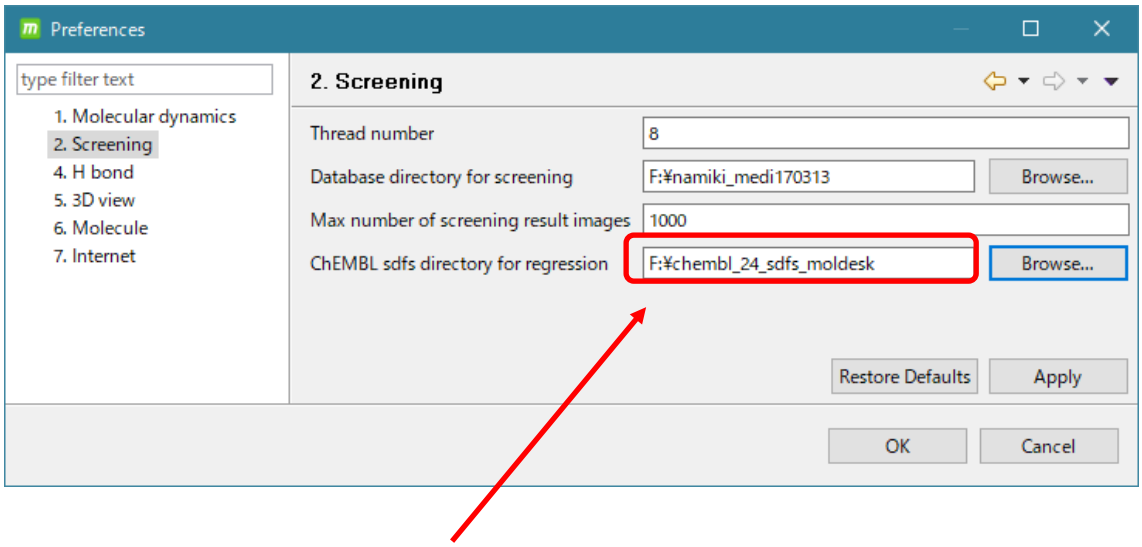
c001

c002

・・・

Select the folder (directory) immediately above the folder (directory) and click [OK].





Then, LigandBox will be set as shown above. Click [OK].

Now you can perform regression analysis predictions ([Make Regression model] and [Predict with Regression model]) of various properties of the compound.

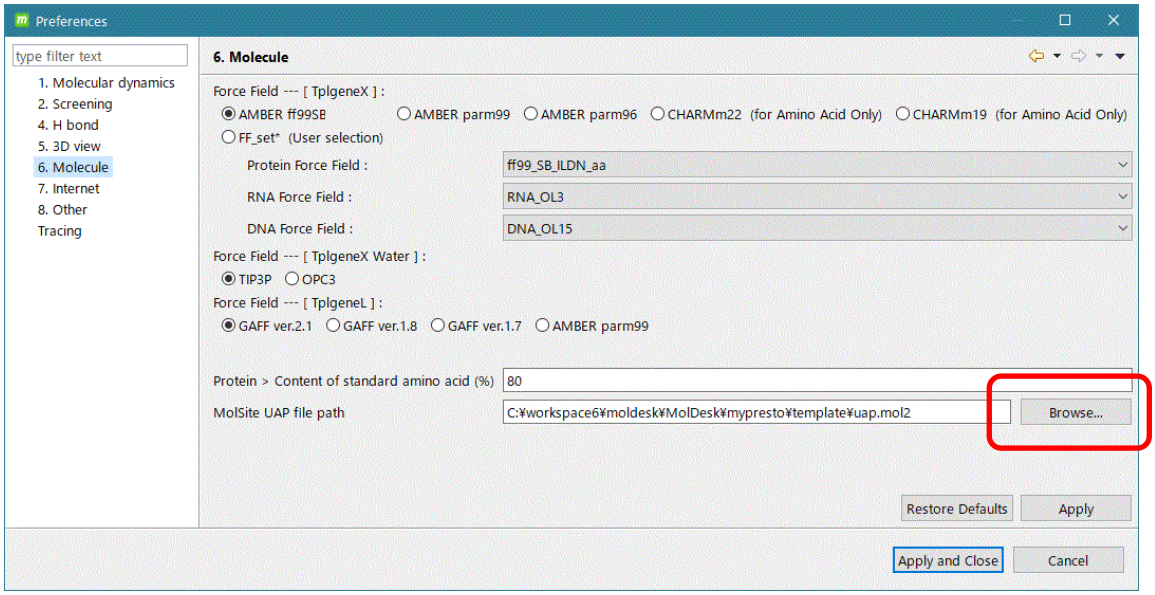
8.3.3. Molecule

Set when you want to change the candidate ligand used for pocket search by MolSite from the default.

The candidate ligand provided by default is already set, but the user can change it when he / she wants to change the candidate ligand. If changed, the candidate ligand will search for a pocket that is easy to bind.

Be sure to set the mol2 file here.

Open the [Help]-[Preference] screen, select "6. Molecule", and click [Browse] under [MolSite UAP file path:].



8.4. MD calculation by psygene / psygene -G

[Help] - [Preference]-"Molecular Dynamics"
When [MPI (sievgene]] or [MPI + GPU (sievgene-G)] is selected in, parallel calculation of molecular dynamics by psygene or psygene-G is executed.

The calculation method is basically the same as for cosgene / cosgene_MPI, so please refer to the MolDesk Basic manual for details on the calculation method.
However, there are the following functional differences between the psygene and cosgene MD calculation programs.

- For psygene / psygene-G, the shape of the water solvent cannot be calculated as Cap (spherical). Only Cube (cuboid). Both cosgene / cosgene_MPI can be calculated.
- Both calculations of solutes in vacuum are possible.
  (It is now possible to calculate with the psygene system without forming Cube water.)
- Generalized Born method calculations can be done with cosgene / cosgene_MPI, but not with psygene / psygene-G.