Drug design package software with in silico drug discovery

MolDesk Screening Ver. 1.1.89

Manual



Contents

1.	Scr	eenir	ng calculations	5
	1.1.	Inst	allation and activation	5
	1.2.	Pre	paring for LigandBox	6
	1.3.	Pre	paring ChEMBL sdfs	8
	1.4.	Pre	paring a user-specified screening compound DB	9
	1.4	.1.	Save project	9
	1.4	.2.	Set up database creation criteria	10
	1.4	.3.	Database creation calculation	18
	1.4	.4.	Database creation location	19
	1.5.	Rep	partitioning of Compound DB for Screening	22
	1.5	.1.	Save project	22
	1.5	.2.	Set up database resyn	23
	1.5	.3.	Database creation location	25
	1.6.	Scre	eening calculation overview	27
	1.7.	Par	allel number, amount of memory, and time for screening calculations	28
	1.8.	Scre	eening calculation procedure	30
	1.9.	Pre	pare mol2 files	32
	1.10.	S	creening by MTS method or docking score order	33
	1.1	0.1.	Create a project	33
	1.1	0.2.	Data id for screening calculations	37
	1.1	0.3.	Adding compounds with mol2 files	40
	1.1	0.4.	Enter ligandBox compounds to search for	42
	1.1	0.5.	Enter the size of the screening results list	43
	1.1	0.6.	Enter docking calculation method	44
	1.1	0.7.	Start screening calculations	45
	1.1	0.8.	Review screening calculation results	46
	1.1	0.9.	Check docking pose	50
	1.1	0.10.	File output of screening calculation results	50
	1.11.	Ν	Il-MTS Calculation Procedure	51
	1.1	1.1.	Create a project	51
	1.1	1.2.	Enter data for screening calculations	51
	1.1	1.3.	Addingcompounds with m ol2 files	52
	1.1	1.4.	Enter ligandBox compounds to search for	57
	1.1	1.5.	Confirmation of screening calculation results, and others	57
	1.12.	Ν	II-DSI Calculation Procedure	58

1.1	2.1.	Create a project	58
1.1	2.2.	Enter data for screening calculations	58
1.1	2.3.	Addingcompounds with m ol2 files	59
1.1	2.4.	Enter ligandBox compounds to search for	65
1.1	2.5.	Review screening calculation results	65
1.13.	Repe	eat screening calculation 1	66
1.1	3.1.	Selection of receptor molecules	67
1.1	3.1.	Data entry for screening calculations	68
1.1	3.1.	Addingcompounds with m ol2 files	70
1.1	3.2.	Review screening calculation results	72
1.14.	Repe	eat screening calculation 2	73
1.1	4.1.	Selection of receptor molecules	73
1.1	4.2.	Enter data for screening calculations	74
1.1	4.3.	Review screening calculation results	76
2. Poo	eket sear	rch and docking calculation by Molsite	77
2.1.	Read F	PDB file and add hydrogen atoms and charges	77
2.2.	Pocket	Search with Molsite	78
2.3.	Paralle	el number and time of pocket search by Molsite	80
2.4.	Dockin	ng calculations	80
3. Do	cking Sc	ore QSAR (Predict Activity)	84
3.1.	Acquis	itionof CH EMBL experimental value data	86
3.2.	Calcul	ating regression parameters	90
3.2	.1. 01	verview	90
3.2	.2. Cr	reate a project	90
3.2	.3. Ca	alculate regression parameter	91
3.2	.4. Co	onfirmation of the calculation results of regression parameters by g	raphs
	94		
3.2	.5. Re	eview regression parameter files	95
3.3.	Predict	tion calculation of active values	98
3.3	.1. Pe	erform active value prediction calculations	98
3.3	.2. Co	onfirmation of the results of the active value prediction calculation.	101
4. Pre	ediction	of characteristic values of compounds by regression analysis (Predict	t with
Regress	sion mod	del)	103
4.1.	Acquis	itionof CHEMBL experimental value data	104
4.2.	Create	input experiment data file	108
4.2	.1. Cr	reate a project	108

4.2	2.2. How to create from ChEMBL data file	
4.2	2.3. Example of needing to edit the data type of an experiment	tal data file113
4.2	2.4. Example of no need to edit the data type in the experimen	ntal data file114
4.2	2.5. A method for editing all experimental data files by the use	er115
4.3.	Calculating regression parameters	116
4.3	3.1. Input items for regression parameter calculation	116
4.3	3.2. Performing regression parameter calculations	
4.3	3.3. Confirmation of the calculation results of regression para	meters by graphs
	123	
4.3	3.4. Review regression parameter files	
4.4.	Predicting characteristic values	
4.4	4.1. Perform characteristic value prediction calculations	
4.4	4.2. Characteristic value Check the result of the forecast calcu	lation129
5. MV	VO Screening	131
5.1.	Query molecular selection	131
5.2.	Select molecules to search for	132
5.3.	View results	133
6. Sin	milar structure search	135
6.1.	Select query molecules	135
6.2.	Select, calculate, and display results of search target molecule	s136
7. Par	rtial structure search	138
7.1.	Query molecular selection	138
7.2.	Selecting, calculating, and displaying results for substructu	re search target
mole	ecules	139
8. Fas	ast parallel computing of MD calculations using MPI/GPU	140
8.1.	How to set up an MPI operating environment	142
8.1	1.1. Windows 64bit	142
8.1	1.2. Linux 64bit	142
8.2.	How to set up a CUDA operating environment	143
8.2	2.1. Windows 64bit	143
8.2	2.2. Linux 64bit	143
8.3.	Preference settings	144
8.3	3.1. Molecular Dynamics	144
8.3	3.2. Screening	148
8.3	3.3. Molecule	153
8.4.	MD calculation by psygene / psygene -G	

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

 $LB_drug_Kishida 2210$: 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.

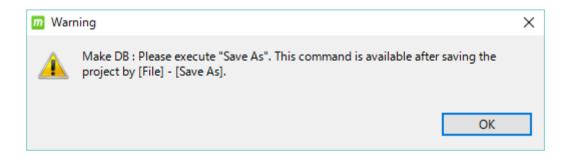
In Preparation]- [Make DB for Screening], you can create a database for screening by inputting a user-specified compound file in addition to the distributed LigandBox. This allows you to screen user-specified compounds other than LigandBox. The compound file to be entered can be multiple sdf files.

* 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.

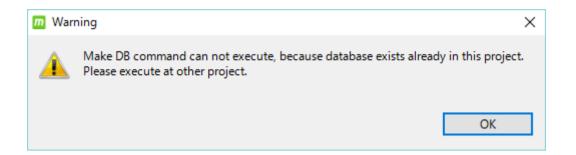
1.4.1. Save project

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 and the work \land database folder does not exist, the following database creation conditions setting screen appears.

m Make DB for Screening	×
Convert to 3D	
2D> 3D set substitute property name to identify molecules, if " <nscode>" or "<supplierid_">" or "<idnumber>" or "<idnumber>" does not exists in sdf files</idnumber></idnumber></supplierid_"></nscode>	make conformer
set supplier name if " <suppliername_*>" does not exists in sdf files (option) : Tag of moleculer names (option) : SOURCE of input files (option) :</suppliername_*>	SUPPLIER MOLECULE SOURCE
Example of output mol2 file @ <tripos>COMMENT IDNUMBER = Value of <nscode> or <supplierid_*> or <idnumber> or <idnumber> or (Property value user inputs above) SUPPLIER = Value of <suppliername_*> or (User inputs above) LIGANDBOX_ID = MOLECULE-*** (MOLECULE : User inputs above) SOURCE = SOURCE (SOURCE : User inputs above) SOURCE_ID = Value of <namiki_id> (if exists in sdf files.)</namiki_id></suppliername_*></idnumber></idnumber></supplierid_*></nscode></tripos>	
Filtering 1 Dy partial structure General Agricultural Filtering 2	
by moleculer weight Min 200 🗭 Max 400 🖨	
ОК	Cancel

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 if the 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-00000001-0001
MOLECULAR_FORMULA = C8H9NO4
MOLECULAR_FORMULA = C8H9NO4
MOLECULAR_WEIGHT = 183.163
MOLECULAR_CHARGE = 0
SUM_OF_ATOMNUMBER = 96
SUM_OF_ATOMNUMBER = 96
SUM_OF_ATOMNUMBER_MINUS_CHARGE = 96
NUM_OF_ATOMNUMBER_MINUS_CHARGE = 96
NUM_OF_ACCEPTOR = 4
HOMO = -9.2167
```

```
@<TRIPOS>MOLECULE
MOLECULE-00000001-01
22\ 22\ 0\ 0\ 0
SMALL
USER_CHARGES
@<TRIPOS>ATOM
                   0.2060 -0.1420 C.ar
                                        1 LGD
                                                   -0.0357
  1 C1
           0.2340
  2 C2
                            0.1260 C.2
           1.5030 -1.9990
                                        1 LGD
                                                   0.3443
  3 \text{ C}3
           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	V200	00										
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 0	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 -1.9236 -0.3490	0.0000 0 0.0000 0	0 0	0	0 0	0	0	0	0	0	0 0	0 0	0 0	0
-0.4946 2.1260	0.0000 0	0	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	0.0000 C1	0	0	0	0	0	0	0	0	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													
6520000													
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
     7
        2
  8
           0
              0
                  0
                     0
  END
М
>
  <SID>
NS-00000001-0001
$$$$
. . .
```

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.

m Make DB for Screening					
Convert to 3D					
☑ 2D> 3D	make conformer				
set substitute property name to identify molecules, if " <nscode>" or "<supplierid_*>" or "<idnumber>" or "<idnumber>" does not exists in sdf file:</idnumber></idnumber></supplierid_*></nscode>	SID				
set supplier name if " <suppliername_*>" does not exists in sdf files (option) :</suppliername_*>	SUPPLIER				
Tag of moleculer names (option) :	MOLECULE				
SOURCE of input files (option) :	SOURCE				
Example of output mol2 file @ <tripos>COMMENT IDNUMBER = Value of <nscode> or <supplierid_*> or <idnumber> or <idnumber> or (Property value user inputs above) SUPPLIER = Value of <suppliername_*> or (User inputs above) LIGANDBOX_ID = MOLECULE-**** (MOLECULE : User inputs above) SOURCE = SOURCE (SOURCE : User inputs above) SOURCE_ID = Value of <namiki_id> (if exists in sdf files.)</namiki_id></suppliername_*></idnumber></idnumber></supplierid_*></nscode></tripos>	IDNUMBER = Value of <nscode> or <supplierid_*> or <idnumber> or <idnumber> or (Property value user inputs above) SUPPLIER = Value of <suppliername_*> or (User inputs above) LIGANDBOX_ID = MOLECULE +*** (MOLECULE : User inputs above) SOURCE = SOURCE (SOURCE : User inputs above)</suppliername_*></idnumber></idnumber></supplierid_*></nscode>				
Filtering 1 by partial structure General Agricultural					
Filtering 2					
by moleculer weight Min 200 R Max 400 R					
OK	Cancel				

If the property name of the input sdf file is blank, it is not recognized as 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 :</idnumber></idnumber></supplierid_*></nscode>	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 property 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).</idnumber></idnumber></supplierid_ </nscode>

item	substance
Set supplier name if " <suppliername_*>" does not exists in sdf files</suppliername_*>	 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).</suppliername_*> If there is a <suppliername_*> in the input sdf file, that description takes precedence and is describedasSUPPLIER= in the output Mol2 file.</suppliername_*> (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.</tripos> (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 of the output Mol2 file, it is listed asSOURCE=. (Source part below).

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

(example mol2 file description)

```
@<TRIPOS>COMMENT
 LIGANDBOX_ID = MOLECULE-00000001-01
 SUPPLIER = SUPPLIER
 SOURCE = SOURCE
 IDNUMBER = NS-00000001-0001
 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.

Select sdf files				×
← → · ↑	スクトップ > MolDesk Screening > sample > sdf	~ ⊽	sdfの検索	Q
整理 ▼ 新しいフォルダー				. 🔟 🚯
🖈 クイック アクセス 🔷	へ 名前	更新日時	種類	サイズ
🖊 ダウンロード 🛛 🖈	🖺 multi01.sdf	2016/06/22 14:33	SQL Server Compa	2 KB
Creative Clou 🖈	😭 multi3.sdf	2016/07/06 16:46	SQL Server Compa	7 KB
PC 🖈	😭 multi28.sdf	2014/10/23 13:28	SQL Server Compa	52 KB
 デスクトップ 🖈 🗸	٢			>
วราม	名(N): "multi28.sdf" "multi01.sdf" "multi3.sdf"	~	*.sdf	~
			開<(O)	キャンセル

In this example, we selected three sdf files in the MolDesk Screening folder: MolDesk Screenng -> sample -> sdf -> multi01.sdf (contains 1 compound) MolDesk Screenng -> sample -> sdf -> multi3.sdf (contains 3 compounds) MolDesk Screenng -> 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.

MolDesk													-	
e Select [Display Co	olor Optio	n Expert Simple	Screening	Preparation	Window Undo Help								
	• 0	\odot	1 K C	* %	PO	🛃 🔲 🔳								
Source 🛙							💼 test118 : 1 Initial State 🕴		Console	Docking In	fo 23			-
			De Dyna	Scree	Prepa				image		deltaG		RMSD	
			De Dyna sign mics	ming	ration				inage	34	Genad	score	NWI3D	
			fff Conve	rt to 3D Mol2										
			😑 Make	DB for Screenin	9									
					- I									
gand Info														
	52													
		SA												
						j/ Console 🕸								
						j/ Console ⊠		• •						
						j/Console ∞								
						j/ Console 🕾								
						j// Console ⊠								
igand into						Di N Console 32								
						□ N Censole 38								

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	641 hours	191 hours
time	(7 hours for 26 days)	(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:

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].

m	Prefe	erences		- 🗆 🗙
type filter text	2. Screening			(
1. Molecular Dynamics 2. Screening	Thread number	8		
3. Label 4. H Bond	LigandBox			Browse
5. 3D View	Screening Result Image Max	1000		
6. Molecule 7. Internet				
			Destaus Defaults	Annha
			Restore Defaults	Apply
			ОК	Cancel

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

フォルダーの参照	×
✓ MKDB011	^
📙 original	
v 🔒 work	
1	
✓ database	
igand	
ligandImage	
> mol2_files	
mts_data	
> protein	~
	_
フォルダー(F): database	
新しいフォルダーの作成(N) OK キャンセル	,

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.

ration [Preparation]- [Remake DB for Screening] allows you to subpartition the database for screening.

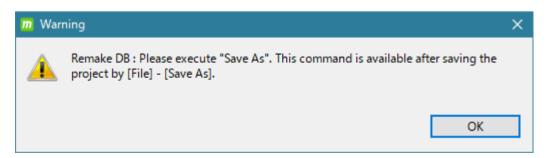
The purpose of the subdivision is to speed up the screening calculation, as explained below.

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 **Solution** [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.

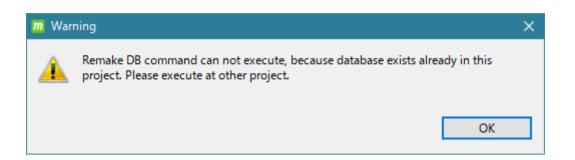
1.5.1. Save project

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 have a 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.

m Remake DB for Screening	×
Select database Division Number: 8	
OK Cancel	

The contents of each item are as follows.

item	substance
Select database	Select the folder (directory) of the database to be repartitioned.
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.

Select database				×
✓ test_ren	makedb			^
🗸 🔤 big	_db			
>	ligand			
	mts_data			
> 🔒 sma	all_db			
> 🔤 sma	all_db2			
	001		 	~
フォルダー(F):	big_db			
新しいフォルダーの	作成(N)	OK	キャンセル	

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

m Remake DB for Screening	×
Select database C:¥Users¥Kiyotaka¥Desktop¥projects¥test_remakedb¥big_db Division Number: 8	
OK Cancel	

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:

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.

m	Prefe	erences		- • ×
type filter text	2. Screening			← ▼ ⇒ ▼
1. Molecular Dynamics 2. Screening 3. Label 4. H Bond	Thread number LigandBox	8		Browse
5. 3D View 6. Molecule 7. Internet	Screening Result Image Max	1000		
			Restore Defaults	Apply
			ОК	Cancel

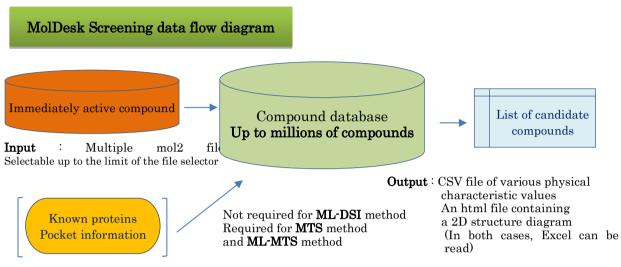
Open the [Help]-[Preference] screen, select "2. Screening", and click [Browse].

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

フォルダーの参照	×
 MKDB011 original work 1 	
database database ingand ligandImage mol2_files mts_data	
> protein フォルダー(F):	✓
新しいフォルダーの作成(N) OK キャンセル	

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

1.6. Screening calculation overview



Input : Modeling with MolDesk

X 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	\bigcirc		0
MTS	\bigcirc		0
ML-MTS	\bigcirc	\bigcirc	0
ML-DSI		O	0

 \bigcirc Required \circ 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 of LigandBox + 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.

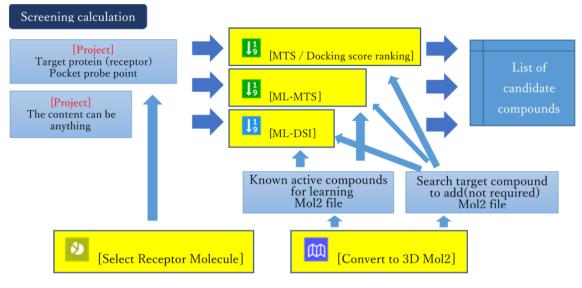
The larger the number of parallels, the more memory is consumed, so reduce the Thread number value on a machine with a small amount of memory.

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	Enter one mol2 file per compound from the file selection
And	dialog.
Compounds to be	
searched	

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 (threedimensional 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 2 [Make Pocket] or 2 [Find Pocket]. See the MolDesk Basic manual for how to create a pocket.

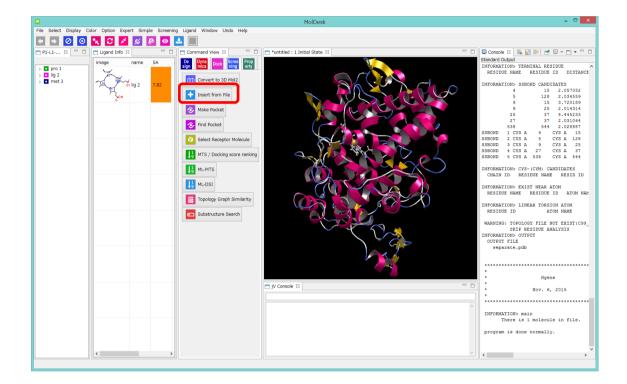
Create a project with the File-Open Molecular File menu.

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

m	Open pdb/mol	2/sdf/mol	×
🛞 🌛 🔻 ↑ 퉬 « sample	e → screening → cox2	✓ C cox2の検索	Q
整理 ▼ 新しいフォルダー			
☆ お気に入り	名前	更新日時 種類	
🚺 ダウンロード	🐏 point.pdb	2015/05/11 0:24 Program	m Debug
📰 デスクトップ	🐏 Pro.pdb	2015/05/11 0:24 Program	m Debug
 2010 2011 <			
¥	<		>
ファイル名(M	I): Pro.pdb	 ✓ *.pdb;*.ent;*.mol2;*.sc< 間√(O) 年で2 	

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



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

m	Select m	nolecule file			x
🔄 🏵 🗉 🕇 🌆	≪ screening → cox2	✓ C cox2の検索		,	P
整理 ▼ 新しいフォ	カレダー				?
쑭 お気に入り	^ 名前	^	更新日時		種
📜 ダウンロード	🐏 point.pdb		2015/05/11	0:24	Pi
📰 デスクトップ	🐏 Pro.pdb		2015/05/11	0:24	Pi
温 最近表示した場所	Pf = Pro.tpl		2015/05/11	0:24	Т
Oreative Cloud	d Files				
📰 デスクトップ	✓ <				3
:	ファイル名(N): point.pdb	✓ *.*) + p	ッンセル	*

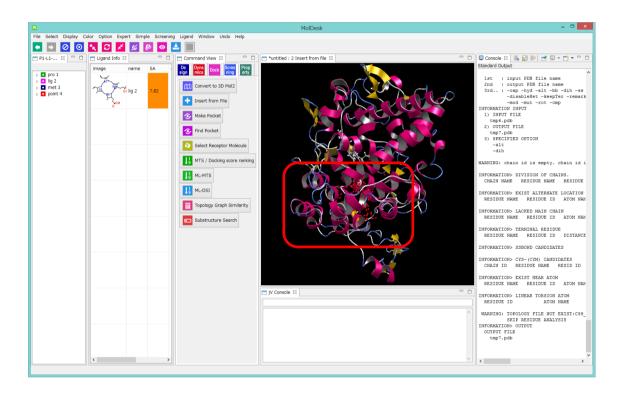
m	Position Select		
	mouse	file	

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.

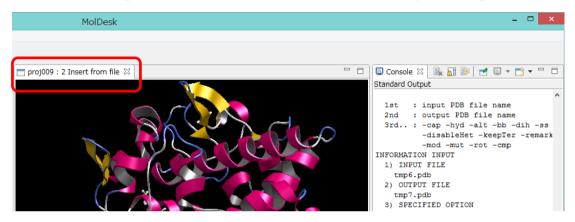


Save this project under the name "proj009".

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

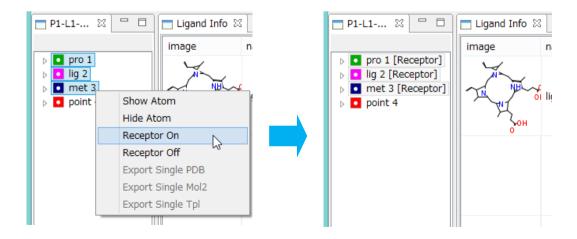
Because screening calculations generate a large amount of data in a calculation (~a few GB),you must save the project before calculating and confirm the folder where the data will be stored.

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 **1** [MTS / Docking score ranking] again.

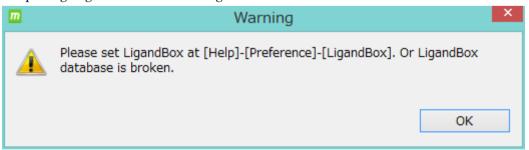
• If you have not created a pocket, the following warning dialog will be displayed, so please create a pocket.

<u>m</u>	Warning	×
	Not find pocket(point). Please execute "Make Pocket" or "Find Pocket" to make point file, or "Input from File" to input point file.	
	ОК	

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

m	Warning
	MTS : Please execute "Save As". Screening is available after save the project by [File] - [Save As].
	ОК

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.

m Screening : MTS (Multiple Target Screening metho	od) / Docking score ranking		×
– Database for Screening : [Help] - [Preference] - [F:¥namiki_medi170313	Screening] - [LigandBox] :		
Previous input		New input	
Proteins	Compounds		Compounds
Added to 181 Target-protein	Added (test AUC) Active ligands	-Target-protein Target-protein name for output (in ASCII without blank)	Add to DB for screening (or for test AUC) Select mol2 files Compounds
Delete selected Reset Method of docking Flexible		Method of docking Flexible Rigid	
		□ remake grid files of protein Size of output list max No. of lines : 10500 🌩	
Delete selected	Delete selected Delete selected	max No. or lines : 10500	
Reset	Reset Reset	Test by AUC	Delete selected
			OK Cancel

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.

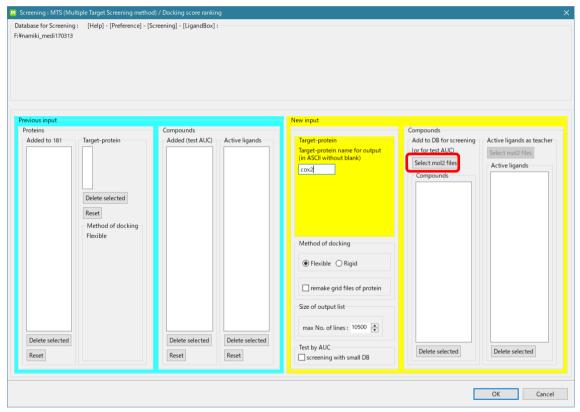
The blue background area on the left side of the dialog shows the input contents of the previous calculation. In this example, it is blank because it is the first time.

	tiple Target Screening metho			×
Database for Screening F:¥namiki_medi170313	: [Help] - [Preference] - [S	creening] - [LigandBox] :		
Previous input			New input	
Proteins Added to 181	Target-protein	Compounds Added (test AUC) Active ligands	Target-protein Target-protein name for output (in ACCI without blank) [cox2]	Compounds Add to DB for screening (or for test AUC) Select mol2 files Compounds
	Delete selected Reset Method of docking Flexible		Method of docking	
Delete selected		Delete selected Delete selected	☐ remake grid files of protein Size of output list max No. of lines : 10500 ♀	
Reset		Reset	Test by AUC	Delete selected Delete selected
				OK Cancel

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 -> cox2L12.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L14.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L14.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L18.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 mol2 files to a	dd to DB for screening			>
← → × ↑ <mark> </mark> <	sample > screening > cox2_ligs	:1 ∨ Ö	cox2_ligs1の検索	م
整理 ▼ 新しいフォ.	Vダ-			- 🔳 🕐
Creative Clou 🖈	^ 名前 [^]	更新日時	種類	サイズ
💻 РС 🛛 🖈	上 cox2L1.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
📃 デスクトップ 🛛 🖈	上 cox2L2.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
📙 USBメモリ 🚽	上 cox2L5.mol2	2015/05/11 0:24	ArgusLab Docume	3 KB
🚆 Kŧュメント 🛛 🖈	cox2L7.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
📰 ピクチャ 🛛 🖈	上 cox2L12.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
cox2	上 cox2L14.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
doc	上 cox2L18.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
	cox2L19.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
img2	cox2L23.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
mol2	cox2L26.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
🔜 デスクトップ	cox2L31.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
ConeDrive	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
画像	cox2L41.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
公開	上 cox2L51.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB

m Screening : MTS (Multiple Target Screenin	ng method) / Docking score ranking		×
– Database for Screening : [Help] - [Prefer F:¥namiki_medi170313	ence] - [Screening] - [LigandBox] :		
Previous input		- New input	
Proteins Added to 181 Target-protein Delete selected Delete selected Delete selected		Target-protein Target-protein name for output (in ASCII without blank) cox2 Method of docking Image: Plexible Plexible Rigid remake grid files of protein Size of output list max No. of lines : 10500 Tremate	Add to DB for screening Active ligands as teacher for for test ALIC Select mol2 files Compounds Cox2L1.mol2 cox2L2.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2
Reset	Reset	Test by AUC	Delete selected Delete selected
			OK Cancel

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.

vious input		- New input	
Added to 181 Delete selected Reset Method of docking Flexible	Compounds Added (test AUC) Active ligands	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 	Add to D8 for screening (or for test AUC) Active ligands as teac Select mol2 files Select mol2 files Compounds Cox2L1.mol2 cox2L3.mol2 cox2L3.mol2 cox2L14.mol2 cox2L18.mol2 cox2L18.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 Active ligands as teac

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.

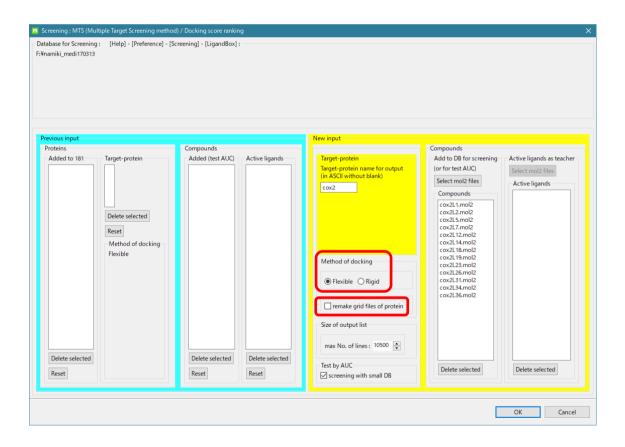
-	tiple Target Screening metho : [Help] - [Preference] - [S		-			
Previous input				- New input		
Proteins Added to 181	Target-protein	Compounds Added (test AUC)	Active ligands	- Target-protein	Compounds Add to DB for screening	Active ligands as teacher
	Delete selected Reset Method of docking Flexible			Method of docking	Cord of test AUC) Select mol2 files Compounds Cox21.1mol2 cox21.2mol2 cox21.2mol2 cox21.2mol2 cox21.12mol2 cox21.12mol2 cox21.12mol2 cox21.12mol2 cox21.3mol2 cox21.3mol2 cox21.3mol2 cox21.34.mol2 co	Active ligands
Delete selected Reset		Delete selected Reset	Delete selected Reset	max No. of lines : 10500 🖨 Test by AUC 🖂 screening with small DB	Delete selected	Delete selected
					Delete selected	Delete selected

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:

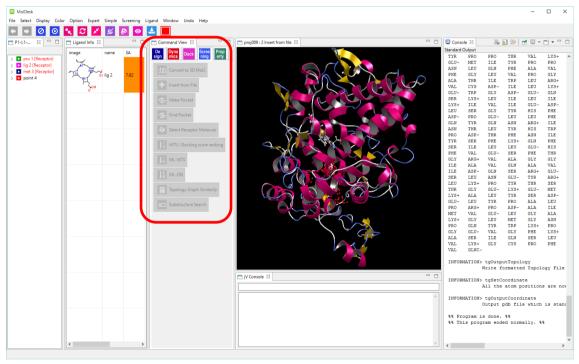
Refers to the 181 reference proteins used in the docking calculations of all screening calculations. See the myPresto manual for details.

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.



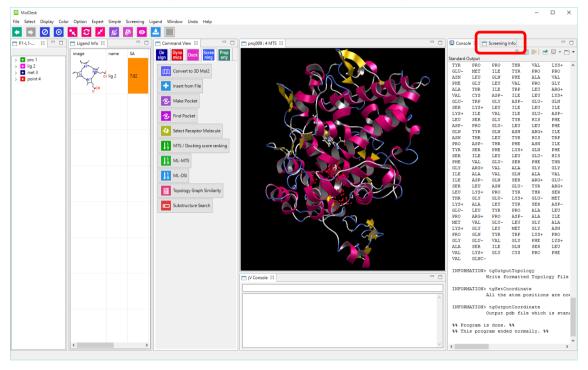
If you have another project open in advance, you can operate that project even during the screening calculation. However, depending on the processor occupancy, the operation may become extremely slow.

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.



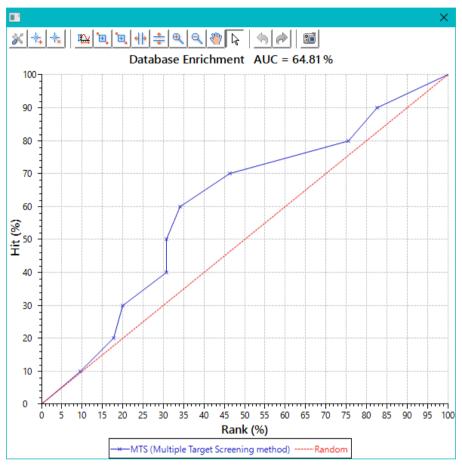
• If you don't see the Screening Info tab, click the Windows-Screening Info menu to display it.

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.

ſS		Score			Export tab	ble				AUC	
		Score			Export de	ле				AUC	
age	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
ri o	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
	54	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
ᡁ ᠉᠆ᡬ᠕ᢆᠰ ᡐ᠆ᠺ᠕ᢆᠰ	⊃ 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.



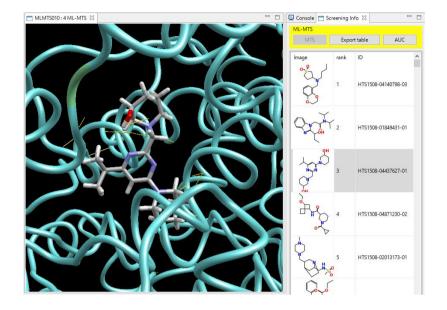
The title of the list changes from "MTS" to "Docking score ranking".

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 $\uparrow \downarrow$ 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, 1.10.1 Create a project Project".

1.11.2. Enter data for screening calculations



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.

-	s (Machine Learning - Multiple g : [Help] - [Preference] - [S			
revious input Proteins Added to 181	Target-protein	Compounds Added (test AUC) Active ligands	New input Target-protein Target-protein name for output (in ASCII without blank)	Compounds Add to D8 for screening (or for test AUC) Select mol2 files Active ligands as teacher Select mol2 files
	Delete selected Reset Method of docking Flexible		Method of docking () Flexible () Rigid	Compounds
Delete selected		Delete selected Reset Reset	□ remake grid files of protein Size of output list max No. of lines : 10500 • Test by AUC □ screening with small DB	Delete selected

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.

Database for Screening :	lachine Learning - Multiple [Help] - [Preference] - [S				×
F:¥namiki_medi170313					
Previous input				-New input	
Proteins Added to 181	Target-protein Delete selected Reset Method of docking Flexible	Compounds - Added (test AUC) -	Active ligands	Target-protein Target-protein name for output (in ASCI without blank) cox2 Method of docking Image: Plexible Rigid remake grid files of protein Size of output list max No. of lines: 10500 Test by AUC	Add to D8 for screening (or for test AUC) Select mol2 files Compounds Compounds
Reset		Reset	Reset	screening with small DB	Delete selected
					OK Cancel

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 -> 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 -> cox2L31.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L34.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L34.mol2

Select mol2 files	to add to	DB for screening			:
÷ → • ↑ [« san	nple > screening > cox2_li	gs1 √ Ĉ) cox2_ligs1の検索	م
整理 ▼ 新しい	フォルダー				- 🔳 🔞
Creative Clo	u 🖈 🔨	名前 ^	更新日時	種類	サイズ
PC	*	1 cox2L1.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
📃 デスクトップ	*	1 cox2L2.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
USB×モリ	*	上 cox2L5.mol2	2015/05/11 0:24	ArgusLab Docume	3 KB
🔮 ドキュメント	*	🛓 cox2L7.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
■ ピクチャ	*	🛓 cox2L12.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
cox2		上 cox2L14.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
doc		cox2L18.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
		cox2L19.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
img2		cox2L23.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
mol2		cox2L26.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
🔜 デスクトップ		cox2L31.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
ConeDrive		1 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
画像		cox2L41.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
公開		<u> cox2L51.mol2</u>	2015/05/11 0:24	ArgusLab Docume	4 KB

🔟 Screening : ML-MTS (Mach	nine Learning - Multiple Tar	get Screening method))			×
Database for Screening : [H	Help] - [Preference] - [Scre	ening] - [LigandBox] :				
Previous input				New input		
Proteins Added to 181 Tarr	get-protein elete selected eset fethod of docking lexible	Compounds Added (test AUC)	Active ligands	Target-protein Target-protein name for output (in ASCI without blank) cox2 Method of docking Flexible Rigid remake grid files of protein Size of output list max. No. of lines : 10500	Compounds Add to DB for screening (or for test AUC) Select mol2 files Compounds Cox21.mol2 cox21.mol2 cox21.mol2 cox21.mol2 cox21.mol2 cox21.mol2 cox21.18.mol2 cox21.18.mol2 cox21.18.mol2 cox21.18.mol2 cox21.38.mol2 cox21.31.mol2 cox21.31.mol2 cox21.31.mol2 cox21.34.mol2 cox21.34.mol2	Active ligands as teacher Select mol2 files - Active ligands
Reset		Reset	Reset	Test by AUC screening with small DB	Delete selected	Delete selected
						OK Cancel

m Screening : N	/L-MTS (Machine Learning - Multiple	Target Screening method)		×
Database for Sc F:¥namiki_medi	creening : [Help] - [Preference] - [S i170313	icreening] - [LigandBox] :		
- Previous inpu	ıt		-New input	
Proteins		Compounds		Compounds
Added to	181 Target-protein	Added (test AUC) Active ligands	Target-protein	Add to DB for screening Active ligands as teacher
Delete sel	Delete selected Reset Method of docking Flexible	Delete selected Delete selected	Method of docking Method of docking Flexible Rigid Size of output list max No. of lines : 10500	Cord or test AUC) Select mol2 files Compounds Active ligands Cox2L1.mol2 cox2L3mol2 cox2L1smol2 cox2L12mol2 cox2L1smol2 cox2L13mol2 cox2L13mol2 cox2L13mol2 cox2L13mol2 cox2L13mol2 cox2L13mol2 cox2L13mol2 cox2L13mol2 cox2L3mol2 cox2L3mol2 cox2L3mol2 cox2L3mol2 cox2L3mol2 cox2L3mol2 cox2L3mol2 cox2L3mol2 cox2L3mol2
Reset		Reset	screening with small DB	
				OK Cancel

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 -> 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 -> cox2L17.mol2

Select mol2 files	Select mol2 files of active ligand as teacher								
← → • ↑	< sam	ple > screening > cox2_ligs2	~ Ō	cox2_ligs2の核	東タ				
整理 ▼ 新しい	フォルダー				::: •				
Creative Clo	ust ^	名前 ^	更新	日時	種類 ^				
PC	*	1 cox2L0.mol2	2015	5/05/11 0:24	ArgusLab Docume				
📃 デスクトップ	*	上 cox2L3.mol2	2015	5/05/11 0:24	ArgusLab Docume				
USBXEU	*	上 cox2L4.mol2	2015	6/05/11 0:24	ArgusLab Docume				
🚆 Γ΄キュメント	*	1 cox2L6.mol2	2015	6/05/11 0:24	ArgusLab Docume				
■ ピクチャ	*	上 cox2L8.mol2	2015	6/05/11 0:24	ArgusLab Docume				
cox2		cox2L9.mol2	2015	5/05/11 0:24	ArgusLab Docume				
doc		cox2L10.mol2	2015	5/05/11 0:24	ArgusLab Docume				
		cox2L11.mol2	2015	6/05/11 0:24	ArgusLab Docume				
img2		cox2L13.mol2	2015	5/05/11 0:24	ArgusLab Docume				
mol2		cox2L15.mol2		5/05/11 0:24	ArgusLab Docume				
デスクトップ		cox2L16.mol2		5/05/11 0:24	ArgusLab Docume				
C OneDrive		cox2L17.mol2		5/05/11 0:24	ArgusLab Docume				
		cox2L22.mol2		5/05/11 0:24	ArgusLab Docume				
		cox2L24.mol2		5/05/11 0:24	ArgusLab Docume				
画像		cox2L25.mol2		5/05/11 0:24	ArgusLab Docume				
公開		cox2L29.mol2	2015	5/05/11 0:24	ArgusLab Docume				

Screening : ML-MTS	(Machine Learning - Multiple	Target Screening methoc)			×
Database for Screening F:¥namiki_medi170313	g : [Help] - [Preference] - [S	icreening] - [LigandBox] :				
Previous input				-New input		
Proteins Added to 181 Delete selected Reset	Target-protein Delete selected Reset Method of docking Flexible	Compounds - Added (test AUC) -	Active ligands	Target-protein Target-protein name for output (in ASCII without blank) cox2 Method of docking Image: second	Compounds Add to DB for screening (or for test AUC) Select mol2 files Compounds Cox21.1.mol2 cox21.8.mol2 cox21.8.mol2 cox21.1.mol2 cox21.14.mol2 cox21.14.mol2 cox21.14.mol2 cox21.19.mol2 cox21.19.mol2 cox21.31.mol2 cox21.31.mol2 cox21.31.mol2 cox21.33.mol2 cox21.33.mol2 cox21.34.mol2 cox21.34.m	Active ligands as teacher Select mol2 files Active ligands cox2L0.mol2 cox2L3.mol2 cox2L6.mol2 cox2L6.mol2 cox2L6.mol2 cox2L10.mol2 cox2L1.mol2 cox2L11.mol2 cox2L15.mol2 cox2L15.mol2 cox2L15.mol2 cox2L15.mol2 cox2L15.mol2 cox2L15.mol2 cox2L15.mol2 cox2L15.mol2
					[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.

base for Screenin		Screening] - [LigandBox]	:			
amiki_medi170313	3					
evious input				-New input		
roteins		Compounds			Compounds	
Added to 181	Target-protein	Added (test AUC)	Active ligands	- Target-protein	Add to DB for screening	- Active ligands as teacher
				Target-protein name for output	(or for test AUC)	Select mol2 files
				(in ASCII without blank)	Select mol2 files	Active ligands
				cox2	Compounds	cox2L0.mol2
					cox2L1.mol2	cox2L3.mol2
	Delete selected				cox2L2.mol2	cox2L4.mol2 cox2L6.mol2
					cox2L5.mol2 cox2L7.mol2	cox2L8.mol2
	Reset				cox2L12.mol2	cox2L9.mol2
	Method of docking				cox2L14.mol2	cox2L10.mol2 cox2L11.mol2
	Flexible				cox2L18.mol2 cox2L19.mol2	cox2L13.mol2
				Method of docking	cox2L23.mol2	cox2L15.mol2 cox2L16.mol2
				Flexible Rigid	cox2L26.mol2 cox2L31.mol2	cox2L17.mol2
					cox2L34.mol2	cox2L22.mol2
					cox2L36.mol2	
				remake grid files of protein		
				Size of output list		
				max No. of lines : 10500 🌩		
Delete selected		Delete selected	Delete selected			
				Test by AUC	Delete selected	Delete selected
Reset		Reset	Reset	✓ screening with small DB	Derete Selected	Sciette Scietted

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



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.

m Screening : ML-DSI (Machine Learning - Docking S	Score Index method)		×
Database for Screening F:¥namiki_medi170313	g : [Help] - [Preference] - [S	icreening] - [LigandBox] :		
Previous input			- New input	
Proteins		Compounds		Compounds
Added to 181	Target-protein	Added (test AUC) Active	igands Target-protein name for outy (in ASCII without blank)	Add to DB for screening Active ligands as teacher
	Delete selected Reset Method of docking Flexible		Method of docking	
			remake grid files of prot Size of output list max No. of lines : 10500	
Delete selected Reset		Delete selected Delete Reset Reset	Test by AUC	Delete selected Delete selected OK Cancel

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.

m Screening : ML-DSI (Machine Learning - Docking Score Index method)	×
- Database for Screening : [Help] - [Preference] - [Screening] - [LigandBox] : F:¥namiki_medi170313	
Previous input New input	
Proteins Compounds Com	mpounds
Target-protein name for output (in ASCII without blank)	dd to DB for screening of for test ALIC Select mol2 files Compounds
Delete selected Delete selected Test by AUC Reset Reset Screening with small D8	Delete selected OK Cancel

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 -> cox2L12.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L14.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 -> cox2L34.mol2

Select mol2 files	to add to	DB for screening			
• → • ↑	< san	nple > screening > cox2_l	igs1 🗸 (り cox2_ligs1の検索	م
整理 ▼ 新しい	フォルダー			=	- 🔳 🕜
Creative Clo	u * ^	へ 名前	更新日時	種類	サイズ
PC	*	cox2L1.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
📃 デスクトップ	*	1 cox2L2.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
USBXモリ	*	1 cox2L5.mol2	2015/05/11 0:24	ArgusLab Docume	3 KB
🔮 ドキュメント	*	上 cox2L7.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
ー ■ ピクチャ	*	🛓 cox2L12.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
cox2		<u> cox2L14.mol2</u>	2015/05/11 0:24	ArgusLab Docume	5 KB
doc		<u> cox2L18.mol2</u>	2015/05/11 0:24	ArgusLab Docume	4 KB
		上 cox2L19.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
img2		<u> cox2L23.mol2</u>	2015/05/11 0:24	ArgusLab Docume	5 KB
mol2		上 cox2L26.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
🗖 デスクトップ		上 cox2L31.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
ConeDrive		<u> cox2L34.mol2</u>	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
画像		cox2L41.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
公開		cox2L51.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB

m Screening : ML-DSI (Machine	e Learning - Docking Score Index m	thod)			×
– Database for Screening : [He F:¥hamiki_medi170313	elp) - [Preference] - [Screening] - [L	gandBox] :			
Previous input			New input		
Dela	ete selected	est AUQ Active ligands	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	(or for test AUC)	Active ligands as teacher Select mol2 files Active ligands
					OK Cancel

	Machine Learning - Docking 1: [Help] - [Preference] - [:			
Previous input			New input	
Proteins		Compounds		Compounds
Added to 181	Target-protein Delete selected Reset Method of docking Flexible	Added (test AUC) Active ligands	Target-protein Target-protein name for output (in ASCI without blank)	Add to DB for screening (or for test AUC) Select mol2 files Compounds cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2 cox21.tmol2
			Flexible O Rigid remake grid files of protein Size of output list max No. of lines: 10500	cox2123.mol2 cox2125.mol2 cox2131.mol2 cox2134.mol2 cox2136.mol2
Delete selected Reset		Delete selected Delete selected Reset Reset	Test by AUC	Delete selected

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 -> 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 -> cox2L17.mol2

Select mol2 files	Select mol2 files of active ligand as teacher								
← → ~ ↑	« san	nple > screening > cox2_ligs2	ٽ ×	cox2_ligs2ග	検索	م			
整理 ▼ 新しい	マォルダー								
Creative Clo	ni k n	名前 ^	更新	日時	種類	^			
PC	*	1 cox2L0.mol2	2015	/05/11 0:24	ArgusLab	Docume			
📃 デスクトップ	*	🚹 cox2L3.mol2	2015	/05/11 0:24	ArgusLab	Docume			
USBXモリ	*	🛓 cox2L4.mol2	2015	/05/11 0:24	ArgusLab	Docume			
🔮 ドキュメント	*	1 cox2L6.mol2	2015	/05/11 0:24	ArgusLab	Docume			
📰 ピクチャ	*	cox2L8.mol2	2015	/05/11 0:24	ArgusLab	Docume			
cox2		cox2L9.mol2	2015	/05/11 0:24	ArgusLab	Docume			
doc		上 cox2L10.mol2	2015	/05/11 0:24	ArgusLab	Docume			
		cox2L11.mol2		/05/11 0:24		Docume			
img2		cox2L13.mol2		/05/11 0:24	~	Docume			
mol2		cox2L15.mol2		/05/11 0:24	-	Docume			
📃 デスクトップ		cox2L16.mol2		/05/11 0:24	~	Docume			
a OneDrive		cox2L17.mol2		/05/11 0:24	-	Docume			
		cox2L22.mol2		/05/11 0:24	~	Docume			
画像		cox2L24.mol2		/05/11 0:24	-	Docume			
		cox2L25.mol2		/05/11 0:24	-	Docume			
公開		上 cox2L29.mol2	2015	/05/11 0:24	ArgusLab	Docume			

Screening : ML-DSI (Machine Learning - Docking	Score Index method)		×
Database for Screening F:¥namiki_medi170313	g: [Help] - [Preference] - [Screening] - [LigandBox] :		
Previous input			- New input	
Proteins Added to 181 Delete selected Reset	Target-protein Delete selected Reset Method of docking Flexible	Compounds Added (test AUC) Active ligands	Target-protein Target-protein name for output (in ASCI without blank) Method of docking Image: the state of the state o	Add to DB for screening (or for test AUC) Active ligands as teacher Select mol2 files Select mol2 files Compounds cox214 mol2 cox214 mol2 cox215 mol2 cox215 mol2 cox215 mol2 cox214 mol2 cox214 mol2 cox214 mol2 cox214 mol2 cox215 mol2 cox215 mol2 cox213 mol2 cox215 mol2
			n	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.

base for Screenin	g : [Help] - [Preference] - [Screening] - [LigandBox]	:			
amiki_medi 170313	1					
evious input roteins		Compounds		New input	Compounds	
Added to 181	Target-protein	Added (test AUC)	Active ligands	Target-protein	Add to DB for screening	- Active ligands as teacher
				Target-protein name for output (in ASCII without blank)	(or for test AUC)	Select mol2 files
					Select mol2 files	Active ligands
					Compounds	cox2L0.mol2 cox2L3.mol2
	Delete selected				cox2L1.mol2 cox2L2.mol2	cox2L4.mol2
					cox2L5.mol2 cox2L7.mol2	cox2L6.mol2 cox2L8.mol2
	Reset				cox2L12.mol2	cox2L9.mol2 cox2L10.mol2
	Method of docking Flexible				cox2L14.mol2 cox2L18.mol2	cox2L11.mol2
	Tiexible			Method of docking	cox2L19.mol2 cox2L23.mol2	cox2L13.mol2 cox2L15.mol2
				Flexible Rigid	cox2L26.mol2 cox2L31.mol2	cox2L16.mol2 cox2L17.mol2
					cox2L34.mol2 cox2L34.mol2	cox2L22.mol2
				remake grid files of protein	COX2L36.mol2	
				Size of output list		
				max No. of lines : 10500 🚔		
Delete selected		Delete selected	Delete selected			
Reset		Reset	Reset	Test by AUC creening with small DB	Delete selected	Delete selected

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

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

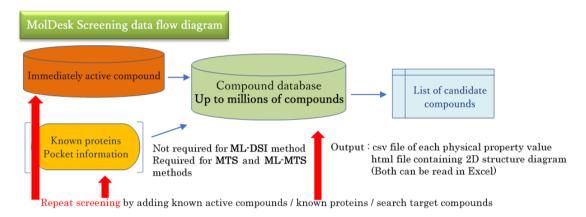
Click OK to start the screening calculation.

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.

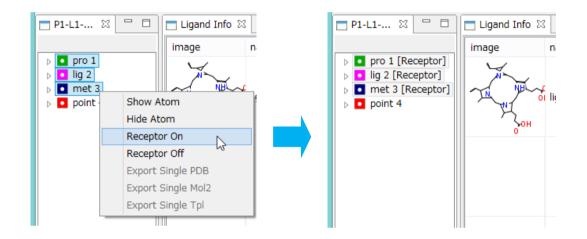
In addition, since the target protein was found later, it can be assumed that the screening calculation by the ML-DSI method is executed and then the screening calculation by the ML-MTS method is executed.

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, pig 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



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.

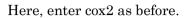
				New input		
teins		Compounds			Compounds	
dded to 181	Target-protein cox2 Delete selected Reset Method of docking Flexible	Added (test AUC) cox21.mol2 cox21.mol2 cox21.mol2 cox21.mol2 cox21.12.mol2 cox21.12.mol2 cox21.18.mol2 cox21.9.mol2 cox21.9.mol2 cox21.9.mol2 cox21.9.mol2	- Active ligands	Target-protein 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	Add to DB for screening- (or for test AUC) Select mol2 files Compounds	Active ligands as teach
Delete selected		Delete selected	Delete selected	Test by AUC	Delete selected	Delete selected
Reset		Reset	Reset	screening with small DB		

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.

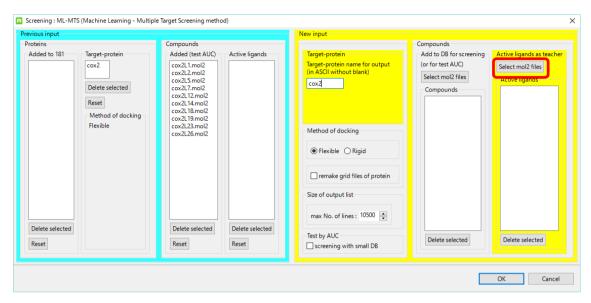
In Target-protein, enter the name of the target protein.



evious input				New input		
Inoteins Added to 181	Target-protein cox2 Delete selected Reset Method of docking Flexible	Compounds Added (test AUC) cox2L1.mol2 cox2L2.mol2 cox2L5.mol2 cox2L1.mol2 cox2L14.mol2 cox2L14.mol2 cox2L18.mol2 cox2L18.mol2 cox2L3.mol2 cox2L9.mol2	Active ligands	Target-protein Target-protein name for output (in ASCI without blank) cox2 Method of docking Flexible ○ Rigid remake grid files of protein Size of output list 	Compounds Add to DB for screening (or for test AUC) Select mol2 files Compounds	Active ligands as teacher Select mol2 files Active ligands
Delete selected Reset		Delete selected Reset	Delete selected	Test by AUC	Delete selected	Delete selected

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 -> cox2L54.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L55.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L62.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 -> cox2L67.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L67.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L71.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L71.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

	ample > screening > cox2_li	gs1 v (Cox2_ligs1の検索	م ر
整理 ▼ 新しいフォルダ-	-			- 🔳 🕻
Creative Clou 🖈 ^	名前 <u>1</u> cox2L34.mol2	更新日時 2015/05/11 0:24	種類 ArgusLab Docume	サイズ 4 KB
ー ニ デスクトップ 🖈	cox2L36.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
USBXEU 🖈	cox2L38.mol2	2015/05/11 0:24 2015/05/11 0:24	ArgusLab Docume ArgusLab Docume	4 KB 4 KB
🔮 ドキュメント 🛛 🖈	cox2L51.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
N 1057 オングン オンジョン マングン オンジョン マングン オンジョン マンジョン マンション マンション マンション マンション マンション マンション マンション マンシー マンシー マンシー マンシー マンシー マンシー マンシー マンシ	cox2L52.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
cox2	cox2L53.mol2	2015/05/11 0:24 2015/05/11 0:24	ArgusLab Docume ArgusLab Docume	5 KB 4 KB
img2	cox2L55.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
mol2	cox2L62.mol2	2015/05/11 0:24 2015/05/11 0:24	ArgusLab Docume ArgusLab Docume	4 KB 5 KB
デスクトップ	cox2L67.mol2	2015/05/11 0:24	ArgusLab Docume	6 KB
ConeDrive	cox2L68.mol2	2015/05/11 0:24 2015/05/11 0:24	ArgusLab Docume ArgusLab Docume	4 KB 5 KB
ドキュメント 画像	cox2L72.mol2	2015/05/11 0:24	ArgusLab Docume	5 KB
公開	cox2L73.mol2 cox2L74.mol2	2015/05/11 0:24 2015/05/11 0:24	ArgusLab Docume ArgusLab Docume	5 KB 5 KB
& Kiyotaka Misoo PC	cox2L79.mol2	2015/05/11 0:24	ArgusLab Docume	4 KB
- PC - V	<u> cox2L80.mol2</u>	2015/05/11 0:24	ArgusLab Docume	5 KB

evious input				- New input		
oteins		Compounds			Compounds	
Added to 181	Target-protein cox2 Delete selected Reset Method of docking Flexible	Added (test AUC) cxx21.mol2 cxx21.mol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2	Active ligands	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 10500 	Add to DB for screening (or for test AUC) Select mol2 files Compounds	Active ligands as teach Select mol2 files Active ligands cox2138.mol2 cox2141.mol2 cox2151.mol2 cox2153.mol2 cox2153.mol2 cox2153.mol2 cox2153.mol2 cox2153.mol2 cox2162.mol2 cox2166.mol2 cox2167.mol2 cox2167.mol2 cox2173.mol2 cox2173.mol2 cox2173.mol2 cox2173.mol2 cox2173.mol2
Delete selected Reset		Delete selected Reset	Delete selected	Test by AUC	Delete selected	Delete selected

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".

vious input			- New input		
oteins		Compounds		Compounds	
Added to 181	Target-protein cox2	cox2L1.mol2 cox2L2.mol2 cox2L5.mol2	re ligands Target-protein Target-protein name for output (in ASCII without blank) rox2	Add to DB for screening (or for test AUC) Select mol2 files	Active ligands as teache Select mol2 files Active ligands
	Delete selected Reset Method of docking Flexible	cox217.mol2 cox2112.mol2 cox2114.mol2 cox2118.mol2 cox2118.mol2 cox2128.mol2 cox2123.mol2 cox2123.mol2	Method of docking	Compounds	cox2L38.mol2 cox2L41.mol2 cox2L51.mol2 cox2L52.mol2 cox2L53.mol2 cox2L54.mol2 cox2L54.mol2 cox2L55.mol2 cox2L62.mol2 cox2L62.mol2
			□ remake grid files of protein Size of output list max No. of lines : 10500		cox2L67.mol2 cox2L68.mol2 cox2L71.mol2 cox2L72.mol2 cox2L72.mol2 cox2L73.mol2 cox2L74.mol2
Delete selected Reset		Delete selected Dele Reset Reset	ete selected Test by AUC	Delete selected	Delete selected

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.

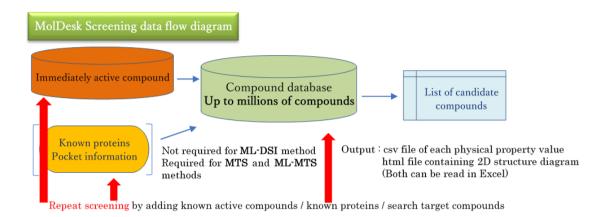
1.13.2. Review screening calculation results

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 proteins 1.13 Repeat screening calculation 1 toperform 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".

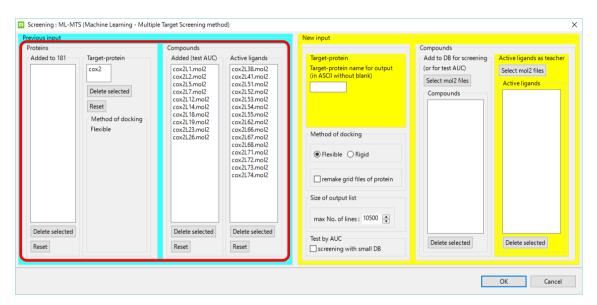
• The receptor setting is canceled every time you perform a screening calculation, so please set it every time.

1.14.2. Enter data for screening calculations



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.

vious input				New input		
oteins Added to 181	Target-protein cox2 Delete selected Reset Method of docking Flexible	Compounds Added (test AUC) cox21.mol2 cox212.mol2 cox21.7.mol2 cox21.12.mol2 cox21.14.mol2 cox21.14.mol2 cox21.18.mol2 cox21.19.mol2 cox21.19.mol2 cox21.29.mol2 cox21.29.mol2	Active ligands cox2138.mol2 cox2141.mol2 cox2151.mol2 cox2155.mol2 cox2155.mol2 cox2155.mol2 cox2155.mol2 cox2165.mol2 cox2165.mol2 cox2167.mol2 cox2171.mol2 cox2173.mol2 cox2174.mol2	Target-protein Target-protein name for output (in ASCII without blank) cox2a Method of docking	Compounds Add to DB for screening (or for test AUC) Select mol2 files Compounds	Active ligands as teach Select mol2 files Active ligands
Delete selected		Delete selected	Delete selected	Test by AUC	Delete selected	Delete selected

- 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].

Screening : ML-MT	S (Machine Learning - Multipl	e Target Screening metho	od)		×
Previous input				New input	
Proteins		Compounds			Compounds
Added to 181	Target-protein	Added (test AUC)	Active ligands	Target-protein	Add to DB for screening - Active ligands as teacher
cox2	cox2a Delete selected Reset Method of docking Flexible	cox211.mol2 cox212.mol2 cox215.mol2 cox217.mol2 cox2117.mol2 cox2114.mol2 cox2114.mol2 cox2118.mol2 cox2119.mol2 cox219.mol2 cox2126.mol2	Cox2138.mol2 Cox2138.mol2 Cox2151.mol2 Cox2151.mol2 Cox2153.mol2 Cox2153.mol2 Cox2154.mol2 Cox2165.mol2 Cox2165.mol2 Cox2166.mol2 Cox2166.mol2 Cox2172.mol2 Cox2173.mol2 Cox2173.mol2 Cox2173.mol2 Cox2174.mol2	Target-protein name for output (in ASCII without blank) Method of docking Image: state of the state o	(or for test AUC) Select mol2 files Compounds
Reset			Reset	Test by AUC	Delete selected Delete selected
reser		Reset	Neset	screening with small DB	
					OK Cancel

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

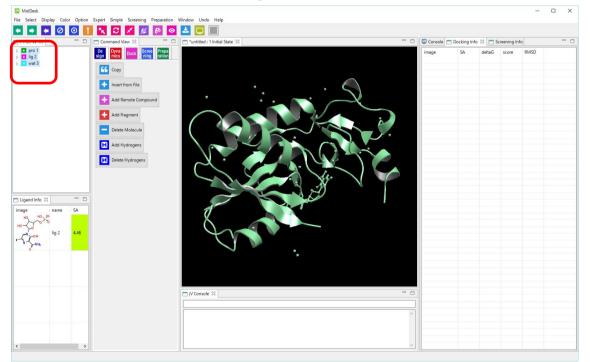
[File] – [Open Molecular File] loads the following files in the MolDesk Screening folder created on the desktop when MolDesk was installed.

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 purinemetabolizing 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.

🔟 Wa	rning >	
Â	Find Pocket : Please select molecule(s) including at least one protein/nucleotide before this command.	
	ОК	

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.

No. of Candidate:	10	*
Radius of Pocket(A):	4.00	×
Minimize no. of Points:	5	•
○ fast ○ fast (by sitegene)		Desk Screening)

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.

m Preferences		- 🗆	×	
type filter text	6. Molecule		•	
1. Molecular dynamics 2. Screening 4. H bond	Protein > Content of standard amino acid (%)			
4. H bond 5. 3D view 6. Molecule 7. Internet	MolSite UAP file path Force Field [TpIgeneX]:	C:¥workspace7¥moldesk¥MolDesk¥mypresto¥template¥uap.mol2 CHARMm22 (for Amino Acid Only) O CHARMm19 (for Amino Acid On ver.1.7 O AMBER parm99		
		Restore Defaults Apply		
		OK Cancel		

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.

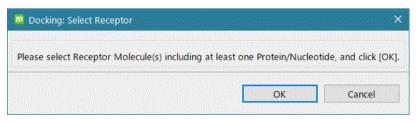
m I	MolDesk		
File	Select	Display	Color
-	•	0	0
Pi	2-L1-p	8 -	
2	pro 1	_	
\rightarrow	lig 2		
>	🛛 wat 3		
>	point	4	
>	point	5	
	point		
>	point	12	
-	noint	12	

2.4. Docking calculations

Next, perform the docking calculation.

Click [Docking].

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



	P2-L1-p	X	
>	💶 pro 1		
>	pro 1	6	
>	💽 wat 3		
>	point	4	
5	point	5	

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

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

<u> </u>	ocking: Choose method to select Ligands
One lig	o select Ligand(s). jand in Tree/3D view île(s) (sdf/mol/mol2)
	OK Cancel
m	Docking: Select Ligand
Please s	elect one Ligand molecule for docking, and click [OK].
	OK Cancel

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 **2** lig 2, and then click OK.

A dialog appears to enter docking calculation conditions.

Docking X							
Basic							
Docking Method moderate Make grid data							
Clustering Number Number of cluster: 10							
Show Customize Tabs							
OK Cancel							

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).

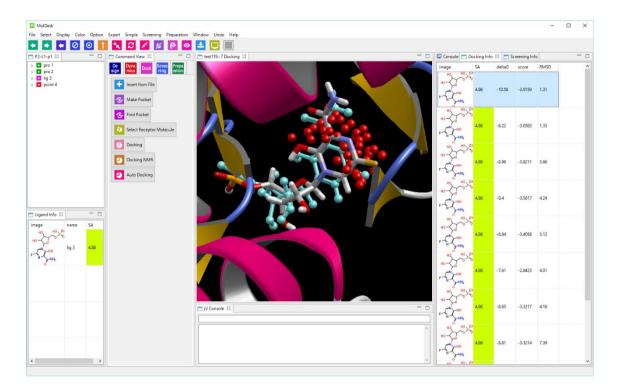
All other pockets will be deleted automatically.

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 :

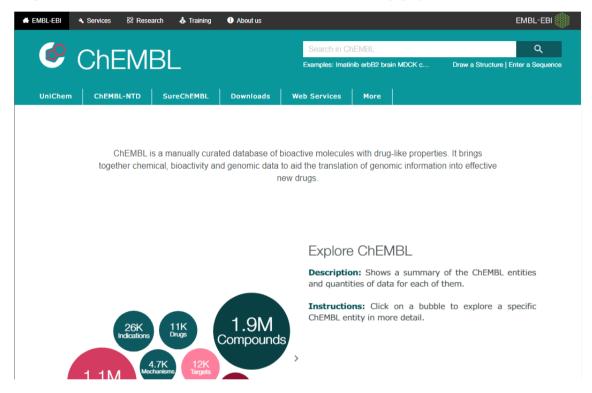
A method of estimating the binding free energy with a weighted average of docking scores for a large number of proteins. The pharmacophore is represented by 600 kinds of proteins, the 600 docking score of a certain compound is analyzed as a principal component, and the experimental data ΔG is subjected 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. Acquisition of 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.

<u>https://www.ebi.ac.uk/chembl/</u> (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.

				EMBL-EBI	
	Tyrosine-protein kinase ABL			٩	
	Examples: Imatinib erbB2 brain MDCK c1	cccc1N	Draw a Structure Er	nter a Sequence	
s	Old Interface More				

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

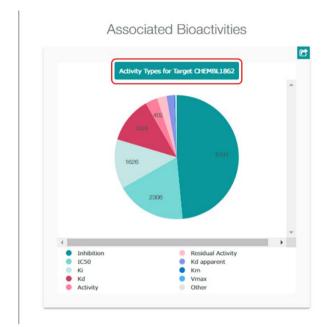
In this example, select CHEMBL1862.

EBI > D	EBI > Databases > Chemical Biology > ChEMBL Database > Taroets Search Results > Twrosine-protein kinase ABL							
Sea	rch Results							
All R	esults 299 Compounds 245	Fargets 7 Assays 13	Documents 23	Cells 11 Tissues	0			
Ta	rgets							
Show	Full Query 0							
Table H	### leatmap		0 Selected	rgets - Select All ctivities 😧		▲ c	SV 🛓 TSV	
₽	Filters 😋	Records per page: 20 •	Sho	ow/Hide Columns			٩	
*	 Organism Taxonomy L1 	Showing 1-7 out of 7 rec	cords			<	1 >	
	Eukaryotes 6	< C					•	
	Viruses 1 Organism Taxonomy L2	ChEMBL 🔶 Sear ID Hit	rch Name 🗘	UniProt Type Accessions		Compounds 🗘	Activities	
	Mammalia 6 retro-transcribing 1	CHEMBL5166	Tyrosine- protein kinase V-ABL	P00521 SINGLE PROTEIN		52 By Mol. Wt.:	68 By Std. Type	
	Primates 5 - N/A - 1 Rođentia 1	CHEMBL1862	Tyrosine- protein kinase ABL	P00519 SINGLE PROTEIN	Homo sapiens	4669 By Mol. Wt.:	12638 By Std. Type	

The following pages appear.

۲
C

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



The following pages appear:

EBI > Databases > Chemical E			vities > Querv								
Edit Querystring 😧 Show Full Query 😢											
III Table				E	12,638 Activ 0 Selected - Se Browse Compo	lect All				د csv د tsv	
₽ Filte	ers 🧠	Record 20	ds per page: ▼		Show/Hi	de Columns		*		٩)
▲ Standard	Туре	Sho	wing 1-20 out	of 12,638 reco	rds			< 1	2 3 4	5 ··· >	
Inhibition IC50 Ki Kd Activity	6131 2306 1626 1524 402	•	Molecule ChEMBL \$ ID	Compound Key	Standard Type	Standard Relation	Standard Value	Standard Units	pChEMBL 🔶 Value	Comment ≑	
Residual Activity Kd apparent Km Vmax Vmax(app) Other Categories	316 243 29 17 12 32		ංර [ං] දා ⁰¹ CHEMBL538507	18	IC50	>	100000	nM	No Data	No Data	C
Target Type SINGLE PROTEIN Organism	12638 Taxonomy L1		Wa.	Genistein	IC50	=	10000	nM	5	No Data	c
			CHEMBL44								

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

III Table	12,638 Activities 0 Selected - Select All Browse Compounds @	▲ CSV ▲ TSV
Clic	there to download your file. The download will expire on 2020-01-01T00:56:13.965907+00:00. Learn More	

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.

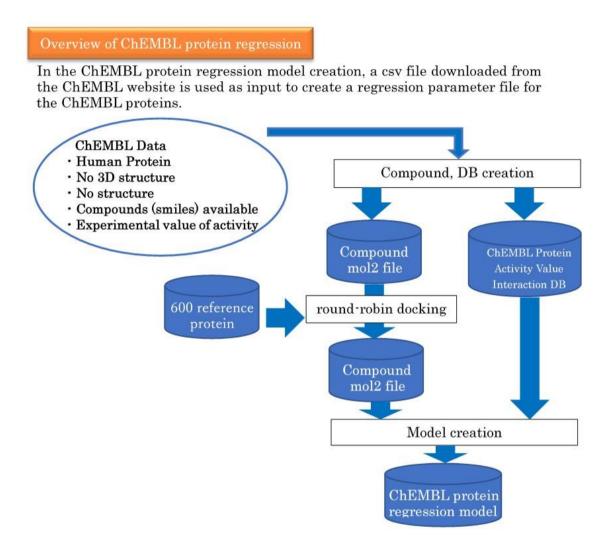
 $CHEMBL 25\-chembl_activity\-XXXX.tsv.gz$

 \downarrow

 $CHEMBL 25\-chembl_activity\-XXXX.tsv$

3.2. Calculating regression parameters

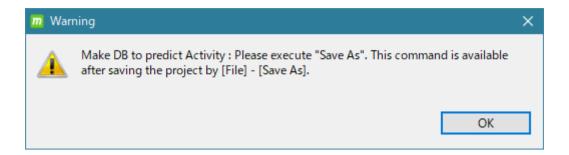
3.2.1. Overview



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.

🔟 Make DB to predict Activity		×
ChEMBL download data file (*.tsv)	_	\sim
1.	Browse	Name of protein:
2	Browse	Name of protein:
3.	Browse	Name of protein:
4.	Browse	Name of protein:
5.	Browse	Name of protein:
6.	Browse	Name of protein:
7.	Browse	Name of protein:
8.	Browse	Name of protein:
9.	Browse	Name of protein:
10.	Browse	Name of protein:
11.	Browse	Name of protein:
12	Browse	Name of protein:
13.	Browse	Name of protein:
14.	Browse	Name of protein:
15.	Browse	Name of protein:
16	Browse	Name of protein:
>>> Name of protein: in ASCII without blank		
Experiments used in calculation Image: With the state of the		
Filtering		
Ø by moleculer weight Min 100 € Max 600 €		
	[OK Cancel

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.

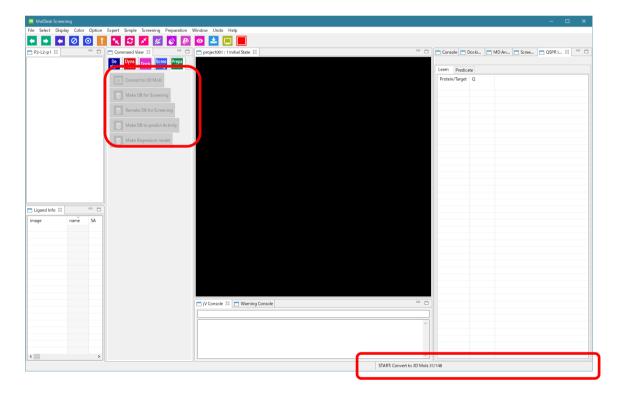
[Filtering]

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.

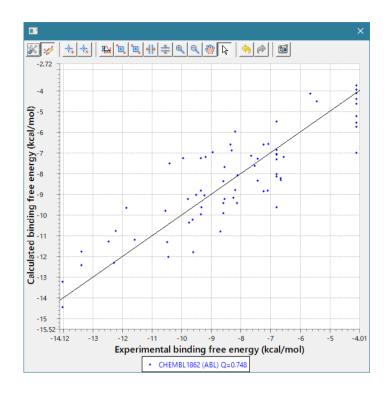
In addition, on the right side of the screen, display the Learn tab of the QSPR Info screen as shown below.

Lists the protein name entered when the regression model was created and the correlation coefficient (Q value) of the regression model created.

Consc	ole 🛅 D	ocki	<u> </u>	MD An	Scree	QSPR I.	23	
Learn	Predict							
Protein	/Target	Q						
ABL		0.7478						

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.

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 converted to 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 \times 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.

m Predict Activity Input	×
PCA parameter	select
Input Mol2 Files	
select	
	OK Cancel

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

m Select PCA parameter file.		×
\leftarrow \rightarrow \checkmark \uparrow \checkmark work \Rightarrow database_qsar \Rightarrow 09.param \checkmark \circlearrowright	09.paramの検索	م
整理 ▼ 新しいフォルダー		•
omp ^ 名前	更新日時 崔類	
hread ABL.param	2018/09/03 15:33 PARAM	M ファイル
pleiades		
project-jsch-test		
projects		
protein 🗸 <		>
ファイル名(N): ABL.param ~	*.param	\sim
	開く(0) 🛨 キャン	セル :

m Predict Activ	vity Input X
PCA parameter	C:¥data¥QSAR003¥work¥database_qsar¥09.param¥ABL.param select
Input Mol2 Files	;
select	
select	
	OK Cancel

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.

Select Input mol2 files.				×
\leftarrow \rightarrow \checkmark \uparrow \square \ll sample \Rightarrow TGS \Rightarrow ligand	ڻ ~	ligandの検索		Q
整理 ▼ 新しいフォルダー			•	?
Ligand 个 名前		更新日時	種類	^
MolDesk Basic <u>0c_3.mol2</u>		2009/07/01 17:12	ArgusLab D)oc
MolDesk Screeni 0c_4.mol2		2009/07/01 17:12	ArgusLab D)oc
molmil-gh-page <u>1</u> 0c_8.mol2		2009/07/01 17:12	ArgusLab D)oc
mypresto <u>1</u> 0c_13.mol2		2009/07/01 17:12	ArgusLab D)oc
myPresto Portal		2009/07/01 17:12	ArousLab D)oc ¥
ファイル名(N): "0c_4.mol2" "0c_3.mol2"	~	*.mol2		~
		開<(O) ▼	キャンセル	

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.

m Predict Activity Input	×
PCA parameter C:¥data¥QSAR003¥work¥database_qsar¥09.param¥ABL.param	select
Input Mol2 Files	
C:¥data¥0c_3.mol2 C:¥data¥0c_4.mol2	
select	
ОК	Cancel

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.

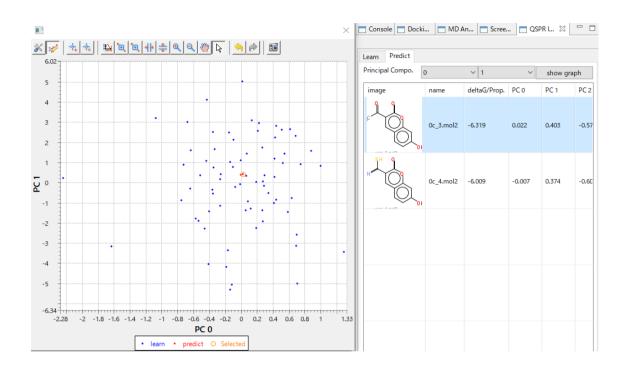
3.3.2. Confirmation of the results of the active value prediction calculation

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.

Console Docki	🗖 MD A	An 📑 Scree.	🔄 🗖 QSI	PR I 🛛	
	0	× 1	~	show gra	ph
image	name	deltaG/Prop.	PC 0	PC 1	PC 2
	0c_3.mol2	-6.319	0.022	0.403	-0.57
	0c_4.mol2	-6.009	-0.007	0.374	-0.60

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.

, Д [Screening]-[Predict with Regression m

lode	elJ	

There are two buttons.

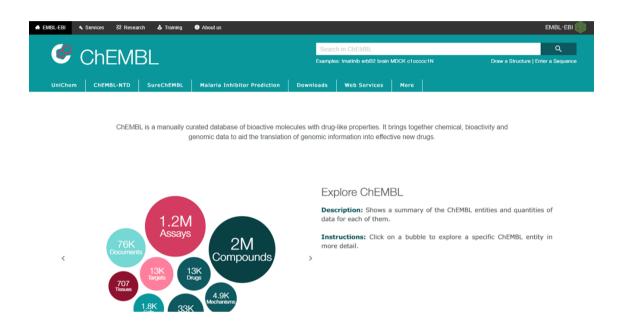
In order to predict various characteristic values with [Predict with Regression model], it is necessary to create a data file created by learning regression parameters with [Make Regression model] in advance.

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. Acquisition of CHEMBL experimental value data

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

<u>https://www.ebi.ac.uk/chembl/</u> (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.

			EMBL-EBI 🧶
aque	ous solubility		٩
Example	es: Imatinib erbB2	brain MDCK c1ccccc1N	Draw a Structure Enter a Sequence
ownloads	Web Servic	es More	

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".

EBI > Databases > Chemical Biology > ChEMBL Database > Assays Search Results > aqueous solubility							
Search Results							
All Results 2656 Compounds 19 Targets 67 Assays 2054 Documents 515 Cells 0 Tissues 1							
Assays							
Show Full Query 🛛							
III Table		2,054 Assays 0 Selected - Select All Browse Activities @			± CSV ± TSV		
Filters of	Records per page:	Show/Hide Columns		•	٩		
Type Label	Showing 1-20 out of 2,054 records			(1 2 3	4 5 ··· >		
P - Physicochemical 2031 A - ADME 19 B - Binding 4	□ ChEMBL	Description	Compounds 👻	Document ChEMBL	0 ≑ Source ≑ mat		
 Classifications L1 		L	148				
- N/A - 2054	CHEMBL631962 P	Aqueous solubility No Data	By Mol. Wt.:	CHEMBL1132890 phys	II-molecule sicochemical Literature		
 Classifications L2 				forn	hat		
- N/A - 2054	CHEMBL1034535 P	Aqueous solubility in phosphate	147 By Mol. Wt.:		II-molecule Scientific		
 Classifications L3 	CHEMBL1034535 P	buffered saline by No Data multi-screen solubility assay		CHEMBL1151935 phys form			
2054			110				

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

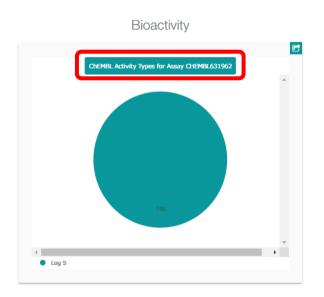
EBI > Databases > Chemical Biology > ChEMBL Database > CHEMBL631962

Assay Report Card

-**Basic Information** C Assay ID: CHEMBL631962 Type: Physicochemical Description: Aqueous solubility Format: BAO_0000100 Journal: Bioorg. Med. Chem. Lett. (2000) 10:1155-1158 Organism: Strain: Tissue: Cell Type: Subcellular Fraction: Target: CHEMBL2362975 CHEMBL1132890 Document: Cell: Tissue:

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

Activity Charts



The following pages appear:

EBI > D	EBI > Databases > Chemical Biology > ChEMBL Database > Activities > Query											
Br	owse Activitie	S										
	Querystring 😧 Full Query 😧											
⊞ Table						148 Activ 0 Selected - S Browse Comp	elect All				▲ CSV	± TSV
₽	Filters	¢	Record 20	ls per page:	•	Show	/Hide Columns					٩
*	 Standard Type 		Sho	wing 1-20 out	– t of 148 records				<	< 1 2 3		·· >
	Log S	148	4									•
	 Target Type 			Molecule ChEMBL \$ ID	Compound Key 👙	Standard 💂	Standard Relation	Standard Value	Standard 💂	pChEMBL 🔶 Value	Comment ≑	Assay ChEMBL
	NO TARGET	148										ID
	 Organism Taxonomy L1 											
	- N/A -	148		ota	DDT	Log S	-	-7.15	No Data	No Data	No Data	CHEMBL6319
	 Organism Taxonomy L2 			CHEMBL416898								
	- N/A -	148			Hexamethylbenzene	Log S	=	-5.23	No Data	No Data	No Data	CHEMBL6319
	 Organism Taxonomy L3 			CHEMBL16347								

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

III Table	148 Activities 0 Selected - Select All Browse Compounds ₽	± CSV ± TSV
Click	here to download your file. Learn More	

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.

m Warning					
<u>^</u>	Make DB to predict Activity : Please execute "Save As". This command is available after saving the project by [File] - [Save As].				
	ОК				

4.2.2. How to create from ChEMBL data file

[Preparation] - [Make Regression model]

Click. Then the following screen will be displayed.

m Make regression model Input	×
Make Experiment Data File < ChEMBL Download File	
Input Experiment Data Files	
select	
Input Mol2 Files (Optional)	
select	
Input Regression Parameter Files (Already learned) (Optional)	
select	
Regression method : O SIMP RBST O WEIT	
Descriptor : I Fluctuation Physical ASA Charge polarization MACCSKey Radius	
Output Regression Parameter File F:#project001#work#regression#regression.param	select
	OK Cancel
	OK Cancel

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.

Make experiment data file Input	×
Input ChEMBL Download File	select
Target :	
Output Experiment Data File F:#project#work#regression#experiment.db	select
[OK Cancel

The input items are as follows.

item	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].)

D Make experiment data file Input	×
Input ChEMBL Download File F:¥project¥DOWNLOAD-G4FECiO7DAYOu8DwOkLcVC7OFsc85w4Dx1YPtuqqWxk=.tsv	select
Target : Log\$	
Output Experiment Data File F:¥project¥work¥regression¥experiment.db	select
	OK Cancel

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 LogSVER CHEMBL631962COMP CHEMBL14687Log_S0.62 - 1.0
TARGET LogS VER CHEMBL631962 COMP CHEMBL275626 Log_S 0.58 - 1.0
TARGET LogSVER CHEMBL631962COMP CHEMBL278489Log_S-3.05 - 1.0
TARGET LogSVER CHEMBL631962COMP CHEMBL279816Log_S-1.96 - 1.0
TARGET LogSVER CHEMBL631962COMP CHEMBL103Log_S-4.42 - 1.0
TARGET LogSVER CHEMBL631962COMP CHEMBL211456Log_S-0.47 - 1.0
TARGET LogSVER CHEMBL631962COMP CHEMBL504760Log_S-1.96 - 1.0
TARGET LogSVER CHEMBL631962COMP CHEMBL72Log_S-3.66 - 1.0
•••

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)	0
7	STANDARD_TYPE(data type)	0
8	STANDARD_VALUE(data value)	0
9	STANDARD_UNITS (units)	×
10	Weights (default 1.0)	0

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

* 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	$sc = \log(sc)/\log(10.0)$
Р	
D	
Pa	if(sc .lt. 0.001) sc = 0.001 sc = log(sc)/log(10.0) -6.0
Papp	if (sc .le30.0) sc =-30.0
Pe	if (sc. ge. 30.0) sc = 30.0
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 PaVER CHEMBL1034536 COMP CHEMBL572342permeability770.0 10'-6_cm/s 1.0TARGET PaVER CHEMBL1034536 COMP CHEMBL550752permeability1060.0 10'-6_cm/s 1.0TARGET PaVER CHEMBL1034536 COMP CHEMBL550761permeability380.0 10'-6_cm/s 1.0TARGET PaVER CHEMBL1034536 COMP CHEMBL550905permeability1.0 10'-6_cm/s 1.0TARGET PaVER CHEMBL1034536 COMP CHEMBL551385permeability30.0 10'-6_cm/s 1.0TARGET PaVER CHEMBL1034536 COMP CHEMBL551385permeability30.0 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 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 CH	HEMBL3301363 COMP	CHEMBL638	LogD7.4	1.7 - 1.0
TARGET LogD VER CH	HEMBL3301363 COMP	CHEMBL639	LogD7.4	2.46 - 1.0
TARGET LogD VER CH	HEMBL3301363 COMP	CHEMBL642	LogD7.4	-0.27 - 1.0
TARGET LogD VER CH	HEMBL3301363 COMP	CHEMBL645	LogD7.4	0.1 - 1.0
TARGET LogD VER CH	HEMBL3301363 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.

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)	0
7	STANDARD_TYPE(data type)	0
8	STANDARD_VALUE(data value)	0
9	STANDARD_UNITS (units)	×
10	Weights (default 1.0)	0

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

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]

m Make regression model Input	×
Make Experiment Data File < ChEMBL Download File	
Input Experiment Data Files	
select	
Input Mol2 Files (Optional)	
select	
Input Regression Parameter Files (Already learned) (Optional)	
select	
Regression method : O SIMP ®RBST O WEIT	
Descriptor: Pluctuation Physical ASA Charge polarization MACCSKey Radius	
Output Regression Parameter File Fi¥project¥project001¥work¥regression.param sele	ect
OK Cancel	

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 experimental data file to 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 is compound ID .mol2 in column 6of the experimental 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 CHEMBL3425624 Papp 5.95 10'-6_cm/s 1.0

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.)

 Delect Input experiment data ← → ∨ ↑	ject > project001 > work > regression	~ č	・ regressionの検	× م
整理 ▼ 新しいフォルダー				
MolDesk_sup ^	名前 ^	更新日時	種類	サイズ
MolDesk_wor	experiment.db	2018/12/06 18:42	Data Base File	10 KB
myPrestoPort	experiment2.db	2018/12/06 20:08	Data Base File	11 KB
namiki_medi1	experiment3.db	2018/12/06 20:09	Data Base File	11 KB
project	🗟 experiment4.db	2018/12/06 20:09	Data Base File	8 KB
project001	🚳 experiment5.db	2018/12/06 20:10	Data Base File	267 KB
System Volun 🗸				
ファイル	名(N): experiment.db		✓ *.db	~
			開く(O)	キャンセル

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

m Make regression model Input	×
Make Experiment Data File < ChEMBL Download File	
Input Experiment Data Files	
F:¥project¥project01¥work¥regression¥experiment.db	
select	
Input Mol2 Files (Optional)	
select	
Input Regression Parameter Files (Already learned) (Optional)	
select	
Regression method : O SIMP	
Descriptor : 🛛 Fluctuation 🖓 Physical 🖓 ASA 🖓 Charge polarization 🖓 MACCSKey 🖓 Radius	
Output Regression Parameter File F:¥project¥project001¥work¥regression¥regression.param	select
	OK Cancel

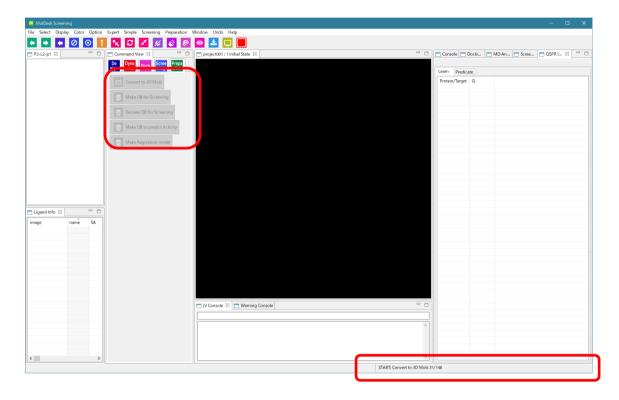
If you click [OK], the following warning screen will appear if ChEMBL sdfs is not set in Preference.

📶 Warr	ning	×
<u>^</u>	Please set correct ChEMBL sdfs database for regression at [Help]-[Preference]-[Screening]-[ChEMBL sdfs directory for regression].	
		ОК

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.

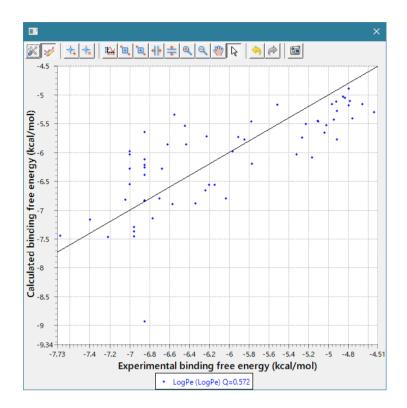
In addition, on the right side of the screen, display the Learn tab of the QSPR Info screen as shown below.

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.

Console Docki MD An Scree QSPR I	x	
Learn Predict		
Protein/Target Q		
LogPe 0.5721		

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.

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.

m Predict with Regression model	×
PCA parameter F:¥project¥logPe_001¥work¥regression.param	select
Learn stdout F:¥project¥logPe_001¥work¥regression¥learn.out	select
Input Mol2 Files	
select	
OK C	Cancel

	m Predict with Regression model	×
	PCA parameter [F:¥project¥logPe_001¥work¥regression.rparam	select
	Learn stdout F:¥project¥logPe_001¥work¥regression¥learn.out	select
	Input Mol2 Files	
	select	
-	ОК Са	incel

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.

Select Input mol2 files.				×
\leftarrow \rightarrow \checkmark \uparrow \square \ll sam	ple > TGS > ligand	~ Ū	ligandの検索	Q
整理 ▼ 新しいフォルダー				• 🔳 🕐
ligand ^	名前 ^		更新日時	種類 ^
MolDesk Basic	1 0c_3.mol2		2009/07/01 17:12	ArgusLab Doc
MolDesk Screeni	1 0c_4.mol2		2009/07/01 17:12	ArgusLab Doc
🔒 molmil-gh-page	1 0c_8.mol2		2009/07/01 17:12	ArgusLab Doc
mypresto	1 0c_13.mol2		2009/07/01 17:12	ArgusLab Doc
myPresto Portal 🗸 🗸	01 1.mol2		2009/07/01 17:12	ArgusLab Doc ¥
ファイルキ	名(N): "0c_4.mol2" "0c_3.mol2"	~	*.mol2	~
			開<(O) ▼	キャンセル

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.

m Predict with Regression model	×
PCA parameter F:¥project¥logPe_001¥work¥regression.param	select
Learn stdout F:¥project¥logPe_001¥work¥regression¥learn.out	select
Input Mol2 Files	
F:¥sample¥TGS¥ligand¥0c_3.mol2 F:¥sample¥TGS¥ligand¥0c_4.mol2	
select	
ОК Са	incel

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.

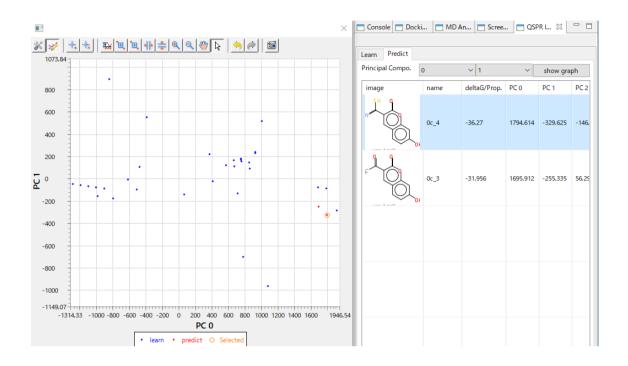
4.4.2. Characteristic value Check the result of the forecast calculation

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.

🗖 Console 🔲 Dock	i 🗖 MD A	n 📑 Scree.	🔄 QSP	'R I ⊠	
Learn Predict					
Principal Compo.	0	✓ 1	~	show gra	ph
image	name	deltaG/Prop.	PC 0	PC 1	PC 2
	0c_4	-36.27	1794.614	-329.625	- 146.
	0c_3	-31.956	1695.912	-255.335	56.29

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)

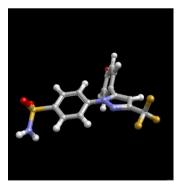
Due to the three-dimensional superposition of two molecules, the one with a large overlap is considered to have high similarity. In superposition, the generation of molecular conformation and the similarity of atomic charges are taken into consideration, and superposition is performed using energy minimization,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

m MVO Screening		×
Select property of COMMENT line in	n Mol2 files to identify	/ the molecule. :
	OK	Cancel

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.

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

m MVO Screening		×
Select property of COMMENT line in Mol2 fi LIGANDBOX_ID	iles to identif	y the molecule. :
	OK	Cancel

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

\leftarrow \rightarrow \checkmark \Uparrow (\checkmark sample \rightarrow MVO_screening \rightarrow target	5 V	targetの検索	م
整理 ▼ 新しいフォルダー		: = : =	- 🔳 💡
★ クイック アクセス ^ヘ 名前 ^ヘ	更新日時	種類	サイズ
↓ ダウンロード ★ all.mol2	2017/01/18 15:46	MOL2 ファイル	1,211 KB
🔜 デスクトップ 🖈			
الالتحالية المحالية ا محالية محالية المحالية محالية محالية محالية محالية المحالية المحالية محالية المحالية المحالية محالية محالية محالية محالية محالية محالية محالي محالية محالية محالية محالية محالية محالية محالية محالية محالية محالي محالية محالية محاليمحالية محاليم محالي محالي محالية م			
New York State St			
ファイル名(N): all.mol2	~	*.mol2	~
		開く(O)	キャンセル

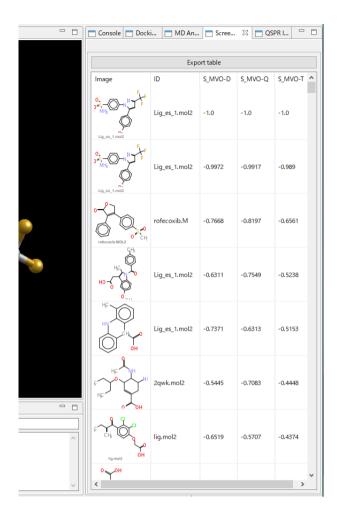
X In the example, it is one file, but you can also select multiple files.

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
- $S_{MVO\cdot T} \qquad \qquad Database \ molecules \ close \ to \ query \ molecules \ are \ selected$



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

You canalsoclick the [Export table] button to save the esults to 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.

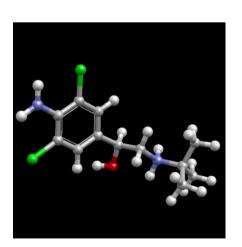
6.1. Select query molecules

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.



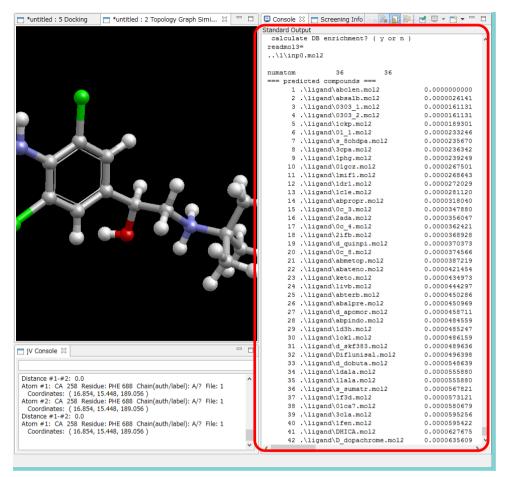
6.2. Select, calculate, and display results of search target molecules

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

m	Select mol	2 files			×
€ 🦻 ▾ ↑ 퉬 « MolDesk_	template → sample → TGS → ligand		v Ċ lig	andの検索	م
整理▼ 新しいフォルダー				•== •	
🍌 projects 🔨 🔨	名前	更新日時	種類	サイズ	^
Jacsy Anna	0c_3.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
4prq	0c_4.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
bSI001 dsiMol2 4	0c_8.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
metan001	0c_13.mol2	2009/07/01 17:12	MOL2 ファイル	4 KB	
proj007	01_1.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
	01ca7.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
🎍 proj008	01gcz.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
test001	🗋 01ljt.mol2	2009/07/01 17:12	MOL2 ファイル	4 KB	
test002	1a4g.mol2	2009/07/01 17:12	MOL2 ファイル	5 KB	
test002	1a4q.mol2	2009/07/01 17:12	MOL2 ファイル	6 KB	
test003	1a28.mol2	2009/07/01 17:12	MOL2 ファイル	6 KB	
test005	1a42.mol2	2009/07/01 17:12	MOL2 ファイル	5 KB	
test005	1abe1.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
test124 ACE	1abe2.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
test124_ACE_NME ¥	1abf1.mol2	2009/07/01 17:12	MOL2 ファイル	3 KB	
		/ /	····· · · ·		
ファイル名(N):	"supr.mol2" "0c_3.mol2" "0c_4.mol2" "0c_8	3.mol2" "0c_13.mol2"	"01_1.mo v *	.mol2	~
				開<(O) 4	ヤンセル

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.

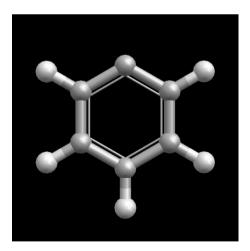
7.1. Query molecular selection

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.



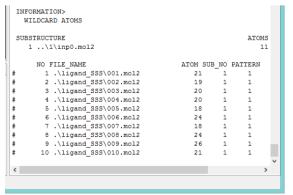
7.2. Selecting, calculating, and displaying results for substructure search target molecules Click Substructure Search and select all mol2 files in the following folders: Click "Open" to start the similar structure search.

MolDesk Screening-> sample-> substructure_search-> ligand

-) -> - 🕇 📗 « MolDesk	_template > sample > substructure	_search → ligand	v 🖒 liga	andの検索	م
	- 1 1 -	- 5			
整理 ▼ 新しいフォルダー				•== •	
projects	名前	更新日時	種類	サイズ	
JCsy	001.mol2	2010/03/08 16:29	MOL2 ファイル	3 KB	
🌡 4prq	002.mol2	2010/03/08 16:29	MOL2 ファイル	3 KB	
bSI001 dsiMol2 4	003.mol2	2010/03/08 16:29	MOL2 ファイル	3 KB	
metan001	004.mol2	2010/03/08 16:29	MOL2 ファイル	3 KB	
proj007	005.mol2	2010/03/08 16:29	MOL2 ファイル	3 KB	
proj008	006.mol2	2010/03/08 16:29) MOL2 ファイル	3 KB	
proj008	007.mol2	2010/03/08 16:29	MOL2 ファイル	3 KB	
test001	008.mol2	2010/03/08 16:29	MOL2 ファイル	3 KB	
test002	009.mol2	2010/03/08 16:29	MOL2 ファイル	4 KB	
test003	010.mol2	2010/03/08 16:31	L MOL2 ファイル	3 KB	
test004					
test005					
test006					
best124_ACE					
test124_ACE_NME ¥					
ファイルタ(N)・	"010.mol2" "001.mol2" "002.mol2" "(103 mol2" "004 mol2" "004	5 mol2" "00 v * r	mol2	
27 17041(14)	010.11012 001.11012 002.11012 0	003.111012 004.111012 00.		11012	

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.



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

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

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

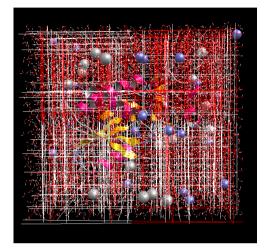
MD Program	Required operating environment	MD calculation capabilities
cosgene	- No particular	All MD calculations are possible
cosgene_MPI	MPI 64bit Limited	All MD calculations are possible
psygene	MPI 64bit Limited	 Calculations other than periodic boundary conditions (*) are not allowed. The size of 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

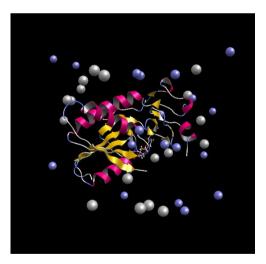
- cosgene_MPI and psygene, MPI must be set.
- psygene Grequires MPI and CUDA to be set .
- Currently, MPI is not multi-node compatible. Works multicore on one node.
- psygene -G supports up to four multi-GPUs per node.
- psygene -G requires an NVIDIA graphics board.
 - > GF100 generation or higher (GTX460 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. <u>https://www.nvidia.co.jp/Download/index.aspx?lang=jp</u>

8.2.2. Linux 64bit

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

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)

	– o x
1. Molecular dynamics	(⇒ = ⇒ + +
Molecular Dynamic Program : ono MPI (cosgene) OMPI (cosgene_MPI) OMPI (psygene) OMPI + GPU (psygene-G)	O gromacs
Processes of MPI 4 GPU 0 ID 0 GPU 1 ID 0 GPU 2 ID 0 GPU 3 ID 0 GROMACS directory C:¥Program Files¥MolDeskBasic¥gromacs¥win64¥gromacs-2021.5	Browse
	Molecular Dynamic Program : no MPI (cosgene) MPI (cosgene_MPI) MPI (psygene) MPI + GPU (psygene-G) Processes of MPI 4 GPU 0 ID 0 GPU 1 ID 0 GPU 2 ID 0 GPU 3 ID 0 GROMACS directory C:¥Program Files¥MolDeskBasic¥gromacs¥win64¥gromacs-2021.5

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.

n Preferences				- 🗆 X
type filter text	1. Molecular dynamic	5		<p th="" •="" •<="" ⇒=""></p>
1. Molecular dynamics 2. Screening 4. H bond 5. 3D view 6. Molecule 7. Internet 8. Other	Processes of MPI GPU 0 ID GPU 1 ID GPU 2 ID GPU 3 ID	Ogram : MPI (cosgene_MPI) O MPI (psygene) 4 0 C:¥Program Files¥MolDeskBasic¥gromacs¥	win64¥gromacs-2021.5	Browse
			Restore Defaults	Apply
			Apply and Close	Cancel

• 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)

type filter text 1. Molecular dynamics 1. Molecular dynamics Molecular Dynamic Program : 2. Screening Ono MPI (cosgene) OMPI (cosgene_MPI) OMPI (psygene) 4. H bond Processes of MPI 5. 3D view GPU 0 ID 0. Molecule 0	$(\neg \cdot \neg \cdot \bullet)$
2. Screening O no MPI (cosgene) O MPI (cosgene_MPI) O MPI (psygene) 4. H bond Processes of MPI 5. 3D view Processes of MPI 6. Molecule GPU 0 ID	
8. Other GPU 1 ID GPU 2 ID GPU 3 ID GROMACS directory C:#Program Files#MolDeskScreening#gromacs#win64#gromacs-2021.5 Restore Defaults	Browse

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).

m. Preferences		– 🗆 X
type filter text	1. Molecular dynamics	(; • ; • • •
1. Molecular dynamics 2. Screening 4. H bond 5. 3D view 6. Molecule 7. Internet 8. Other	Molecular Dynamic Program : O no MPI (cosgene) O MPI (cosgene_MPI) O MPI (psygene) O MPI + GPU (psygene-G) Processes of MPI 48 GPU 0 ID 0 GPU 1 ID 2 GPU 2 ID 0 GPU 3 ID 0 GROMACS directory C:¥Program Files¥MolDeskScreening¥gromacs¥win64¥gromacs-2021.5	Browse
	Apply and Close	Cancel

• GROMACS

Preferences		– 🗆 X
type filter text	1. Molecular dynamics	$\langle \neg \neg \neg \neg \bullet$
1. Molecular dynamics 2. Screening 4. H bond 5. 3D view 6. Molecule 7. Internet 8. Other	Molecular Dynamic Program : O no MPI (cosgene) O MPI (cosgene_MPI) O MPI (psygene) O MPI + GPU (psygene-G) Processes of MPI 4 GPU 0 ID 0 GPU 1 ID GPU 1 ID GPU 2 ID GPU 3 ID GPU 3 ID GROMACS directory C:¥Program Files¥MolDeskScreening¥gromacs¥win64¥gromacs-2021.5	Browse
	Apply and Close	Cancel

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].

Deferences			o x
type filter text	2. Screening		← ▼ ⇒ ▼
1. Molecular dynamics 2. Screening 4. H bond	Thread number Database directory for screening	8	Browse
5. 3D view 6. Molecule	Max number of screening result images	1000	
7. Internet	ChEMBL sdfs directory for regression		Browse
		Restore Defaults	Apply
		OK	Cancel

Ligand

ligandImage

mts_data

protein

of LigandBox (namiki_medi170313 in the example below) unzipped in "1.2 Preparing LigandBox"

フォルダーの参照	×	Select the folder (directory)
		immediately above the folder
		(directory) and click [OK]
> LigandBox_2014		
LigandBox_2015		
> MolDesk_supply		
✓ namiki_medi170313		
> 📙 ligand		
🔒 ligandImage		
mts_data		
> protein	~	
フォルダー(F): namiki_medi170313		
新しいフォルダーの作成(N) OK キャンセ	الا	

m Preferences		-	
type filter text 1. Molecular dynamics 2. Screening 4. H bond 5. 3D view 6. Molecule 7. Internet	2. Screening Thread number Database directory for screening	8 F:¥namiki_medi170313	o ▼ ⊂> ▼ ▼ Browse
	Max number of screening result images ChEMBL sdfs directory for regression		Browse
		Restore Defaults OK	Apply Cancel

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].

フォルダーの参照	\times
v original v work	^
> 2 database ligand ligandImage mol2_files mts_data protein	
フォルダー(F): database 新しいフォルダーの作成(N) OK キャンセル	•

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.

m Preferences		-	□ ×
type filter text 1. Molecular dynamics 2. Screening 4. H bond 5. 3D view 6. Molecule 7. Internet	2. Screening		← - ⇒
	Thread number	8	
	Database directory for screening	F:¥namiki_medi170313	Browse
	Max number of screening result images	1000	
	ChEMBL sdfs directory for regression		Browse
		Restore Defaults	Apply
		ОК	Cancel

[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.

m Preferences		-	
type filter text 1. Molecular dynamics 2. Screening 4. H bond 5. 3D view 6. Molecule 7. Internet	2. Screening		
	Thread number Database directory for screening	8	Browse
	Max number of screening result images ChEMBL sdfs directory for regression	1000	Browse
		Restore Defaults	Apply
		ОК	Cancel

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].

フォルダーの参照	/	×
v ch	embl_24_sdfs_moldesk	-
	2000	L
	c001	
	c002	
	c003	
	c004	
	c005	
フォルダー(F):	chembl_24_sdfs_moldesk	
新しいフォルダーの)作成(N) OK キャンセル	

Deferences			
type filter text 1. Molecular dynamics 2. Screening 4. H bond 5. 3D view 6. Molecule 7. Internet	2. Screening		⇔ - ⇔
	Thread number Database directory for screening	8 F:¥namiki_medi170313	Browse
	Max number of screening result images	1000	browse
	ChEMBL sdfs directory for regression	F:¥chembl_24_sdfs_moldesk	Browse
		Restore Defaults	Apply
		OK	Cancel
		<u> </u>	curren

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:].

		- 🗆 X
6. Molecule		< → → < <
Force Field [TplgeneX] :	99 OAMBER parm96 OCHARMm22 (for Amino Acid Only) OCHARM	Im19 (for Amino Acid Only)
Protein Force Field :	ff99_SB_ILDN_aa	~
RNA Force Field :	RNA_OL3	~
DNA Force Field :	DNA_OL15	~
Force Field [TplgeneL] :		
MolSite UAP file path	$\label{eq:c:workspace6} C: \verb"workspace6" moldesk" MolDesk" mypresto" template" ap.mol2 mol2 mol2 mol2 mol2 mol2 mol2 mol2 $	Browse
	Restore Apply and	
	Force Field [TplgeneX]: (Force Field [TplgeneX]: AMBER ff99SB AMBER parm99 AMBER parm96 CHARMm22 (for Amino Acid Only) CHARM FF_set* (User selection) Protein Force Field : [ff99_SB_ILDN_aa RNA_Force Field : [DNA_OL3 DNA Force Field [TplgeneX Water]: TIP3P OOPC3 Force Field [TplgeneL]: @ GAFF ver.2.1 OGAFF ver.1.8 OGAFF ver.1.7 AMBER parm99 Protein > Content of standard amino acid (%) B0 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.