

インシリコ創薬による  
ドラッグデザイン パッケージ ソフトウェア

# ***MolDesk Screening***

## ***Ver.1.1.89***

(モルデスク スクリーニング)

マニュアル



株式会社 情報数理バイオ

## 目次

1. スクリーニング計算.....	5
1.1. インストール・ライセンス認証.....	5
1.2. LigandBox の準備.....	6
1.3. ChEMBL sdfs の準備.....	8
1.4. ユーザ指定のスクリーニング用化合物 DB の準備.....	9
1.4.1. プロジェクトの保存.....	9
1.4.2. データベース作成条件の設定.....	10
1.4.3. データベース作成計算.....	18
1.4.4. データベース作成場所.....	19
1.5. スクリーニング用化合物 DB の再分割.....	22
1.5.1. プロジェクトの保存.....	22
1.5.2. データベース再分割条件の設定.....	23
1.5.3. データベース作成場所.....	25
1.6. スクリーニング計算の概要.....	27
1.7. スクリーニング計算の並列数・メモリ量・時間.....	28
1.8. スクリーニング計算の手順.....	29
1.9. mol2 ファイルの準備.....	31
1.10. MTS 法またはドッキングスコア順によるスクリーニング.....	32
1.10.1. プロジェクトの作成.....	32
1.10.2. スクリーニング計算のデータ入力.....	36
1.10.3. mol2 ファイルによる化合物の追加.....	39
1.10.4. LigandBox の検索対象化合物の入力.....	41
1.10.5. スクリーニング結果リストのサイズの入力.....	42
1.10.6. ドッキング計算方法の入力.....	43
1.10.7. スクリーニング計算の開始.....	44
1.10.8. スクリーニング計算結果の確認.....	45
1.10.9. ドッキングポーズの確認.....	49
1.10.10. スクリーニング計算結果のファイル出力.....	49
1.11. ML-MTS 法の計算手順.....	50
1.11.1. プロジェクトの作成.....	50
1.11.2. スクリーニング計算のデータ入力.....	50
1.11.3. mol2 ファイルによる化合物の追加.....	51
1.11.4. LigandBox の検索対象化合物の入力.....	56
1.11.5. スクリーニング計算結果の確認他.....	56

1.12.	ML-DSI 法の計算手順.....	57
1.12.1.	プロジェクトの作成.....	57
1.12.2.	スクリーニング計算のデータ入力.....	57
1.12.3.	mol2 ファイルによる化合物の追加.....	58
1.12.4.	LigandBox の検索対象化合物の入力.....	64
1.12.5.	スクリーニング計算結果の確認.....	64
1.13.	繰り返しスクリーニング計算 1.....	65
1.13.1.	受容体分子の選択.....	66
1.13.1.	スクリーニング計算のデータ入力.....	67
1.13.1.	mol2 ファイルによる化合物の追加.....	69
1.13.2.	スクリーニング計算結果の確認.....	71
1.14.	繰り返しスクリーニング計算 2.....	72
1.14.1.	受容体分子の選択.....	72
1.14.2.	スクリーニング計算のデータ入力.....	73
1.14.3.	スクリーニング計算結果の確認.....	75
2.	Molsite によるポケット探索・ドッキング計算.....	76
2.1.	PDB ファイルを読み水素原子・電荷を付加.....	76
2.2.	Molsite によるポケット探索.....	77
2.3.	Molsite によるポケット探索の並列数・時間.....	79
2.4.	ドッキング計算.....	79
3.	Docking Score QSAR (Predict Activity).....	83
3.1.	CHEMBL 実験値データの取得.....	84
3.2.	回帰パラメーターの計算.....	88
3.2.1.	概要説明.....	88
3.2.2.	プロジェクトの作成.....	88
3.2.3.	回帰パラメーターの計算実行.....	89
3.2.4.	回帰パラメーターの計算結果のグラフによる確認.....	92
3.2.5.	回帰パラメーターファイルの確認.....	93
3.3.	活性値の予測計算.....	96
3.3.1.	活性値予測計算の実行.....	96
3.3.2.	活性値予測計算の結果確認.....	99
4.	回帰分析による化合物の特性値予測 (Predict with Regression model).....	101
4.1.	CHEMBL 実験値データの取得.....	102
4.2.	入力実験データファイルの作成.....	106
4.2.1.	プロジェクトの作成.....	106
4.2.2.	ChEMBL データファイルからの作成法.....	107

4.2.3.	実験データファイルのデータ種を編集する必要がある例.....	111
4.2.4.	実験データファイルのデータ種を編集する必要がない例.....	112
4.2.5.	実験データファイルをユーザがすべて編集する作成法.....	113
4.3.	回帰パラメーターの計算.....	114
4.3.1.	回帰パラメーター計算のための入力項目.....	114
4.3.2.	回帰パラメーターの計算実行.....	118
4.3.3.	回帰パラメーターの計算結果のグラフによる確認.....	121
4.3.4.	回帰パラメーターファイルの確認.....	122
4.4.	特性値の予測計算.....	124
4.4.1.	特性値予測計算の実行.....	124
4.4.2.	特性値予測計算の結果確認.....	127
5.	MVO Screening.....	129
5.1.	クエリ分子の選択.....	129
5.2.	検索対象分子の選択.....	130
5.3.	結果表示.....	131
6.	類似構造検索.....	133
6.1.	クエリ分子の選択.....	133
6.2.	検索対象分子の選択・計算・結果表示.....	134
7.	部分構造検索.....	136
7.1.	クエリ分子の選択.....	136
7.2.	部分構造検索対象分子の選択・計算・結果表示.....	137
8.	MPI / GPU による MD 計算の高速並列計算.....	138
8.1.	MPI 動作環境の設定方法.....	140
8.1.1.	Windows 64bit.....	140
8.1.2.	Linux 64bit.....	140
8.2.	CUDA 動作環境の設定方法.....	141
8.2.1.	Windows 64bit.....	141
8.2.2.	Linux 64bit.....	141
8.3.	Preference の設定.....	142
8.3.1.	Molecular Dynamics.....	142
8.3.2.	Screening.....	146
8.3.3.	Molecule.....	151
8.4.	psygene / psygene-G による MD 計算.....	152

## 1. スクリーニング計算

MolDesk Basic で実行できる操作はすべて MolDesk Screening でも実行できます。

本マニュアルでは MolDesk Screening でのみ実行できる操作について説明します。

MolDesk Basic と共通の操作については「MolDesk Basic マニュアル」を参照してください。

Ligand Box は 200 万個の低分子化合物を含むデータベースです。MolDesk Screening のスクリーニング計算では、LigandBox およびユーザが指定した化合物から、数百～数千化合物の薬剤候補化合物を絞り込むことができます。

また、ユーザが指定した低分子化合物に対して、スクリーニング計算が可能になる様に加工して、スクリーニング計算できます。

### 1.1. インストール・ライセンス認証

MolDesk Screening のインストール方法、ライセンス認証の方法は、MolDesk Basic と同じです。MolDesk Basic のマニュアルを参照してください。

ただし、分子動力学計算を、MPI や、nVIDIA のグラフィックボードで高速に並列計算したい場合は、MPI のインストールや CUDA 動作環境などが別途必要になります。具体的な方法は、「MPI/GPU による MD 計算の高速並列計算」の章をご参照ください。

※ Mac は、分子動力学計算の MPI や CUDA による並列計算をサポートしません。

## 1.2. LigandBox の準備

LigandBox の準備は、LigandBox の化合物からスクリーニングする場合に必要です。

※ LigandBox はダウンロードサイトで配布しています。ダウンロードサイトに関する情報をご存知ない方は、(株) 情報数理バイオ [moldesk@imsbio.co.jp](mailto:moldesk@imsbio.co.jp) までメールでご連絡ください。ダウンロードサイトの URL と、アカウント、パスワードをご連絡いたします。

現時点で配布している LigandBox は、LigandBox ver.2210、LigandBox ver.2104、LigandBox ver.2004 と LigandBox ver.1906 です。

LigandBox ver.2210 は、8 個のデータ圧縮ファイルからなります。内容は以下の通りです。

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

それぞれ解凍展開すると、以下の 4 つのデータになります。

LB\_drug\_Namiki2204 : ナミキ商事・医薬向け 300 万化合物  
LB\_agri\_Namiki2204 : ナミキ商事・農薬向け 300 万化合物  
LB\_drug\_Kishida2210 : キシダ化学・医薬向け 100 万化合物  
LB\_agri\_Kishida2210 : キシダ化学・農薬向け 100 万化合物

※ LB\_drug\_Namiki2204.gz と LB\_drug\_Namiki2204.ligandImage.zip は 同じフォルダ（ディレクトリ）に解凍展開してください。  
その他、3 データも同様です。

フォルダ（ディレクトリ）に解凍展開すると、以下の構成になります。

LB_drug_Namiki2204	- ligand	: 化合物の mol2 ファイル
	- mts_data	: 蛋白質-化合物相互作用行列
	- protein	: 181 蛋白質
	- pro_list (ファイル)	: 181 蛋白質リスト
	- pro.list (ファイル)	: 181 蛋白質リスト
	- version (ファイル)	: DB 作成で使った sievgene のバージョン
	- ligandImage	: MolDesk Screening で使う画像ファイル

(他の 3 つも同様)

※ Linux マシンの場合、解凍コマンドは、例えば、

```
unzip LB_drug_Namiki2204.gz
```

です。

※ 解凍先のパスには、スペースを含まないようにしてください。スペースを含んだパスに解凍すると（例）C:\Program Files）、MolDesk Screening が正常に動作しません。

※ 解凍展開すると、それぞれ最大約 63G バイトになりますので、インストール先の記憶媒体の容量にご注意ください。

※ Windows で解凍ソフトを使う場合は、ファイルサイズが大きいため注意が必要です。解凍ソフトによっては、サイズが大きすぎて解凍できない場合があります。

（例えば、Explzh (x64) というフリーの解凍ソフトでは解凍できました。動作確認が取れているのは「Explzh」「7-Zip」です。エラーが出た場合はこれらの解凍ソフトを使ってください。）

Linux マシンをお持ちの場合は、Linux マシンで解凍した方が、解凍しやすいです。

最後に、「8.3.2 Screening」を参照し、LigandBox を MolDesk Screening に設定して、スクリーニング計算対象として使用できるようにします。

### 1.3. ChEMBL sdfs の準備

ChEMBL sdfs の準備は、化合物の各種特性の回帰分析による予測をする場合に必要です。  
1 個のデータ圧縮ファイルからなります。

内容は以下の通りです。

chembl_24_sdfs_moldesk.zip
----------------------------

フォルダ（ディレクトリ）に解凍展開すると、以下の構成になります。

chembl\_24\_sdfs\_moldesk - c000    : 化合物の sdf ファイル  
                                  ChEMBL ID がファイル名になっている  
- chembl\_id.lst (ファイル)    : 化合物リスト

最後に、「8.3.2 Screening」を参照し、ChEMBL sdfs を MolDesk Screening に設定して、回帰分析 ([Make Regression model] と [Predict with Regression model]) で参照する化合物として使用できるようにします。



## 1.4. ユーザ指定のスクリーニング用化合物 DB の準備

LigandBox を使用してスクリーニング計算を行う場合は、この節は読み飛ばしてください。



[Preparation] -



[Make DB for Screening] で、配布される LigandBox 以外に、

ユーザ指定の化合物ファイルを入力にして、スクリーニング用のデータベースを作成することができます。それにより、LigandBox 以外の、ユーザが指定した化合物をスクリーニング対象にできます。

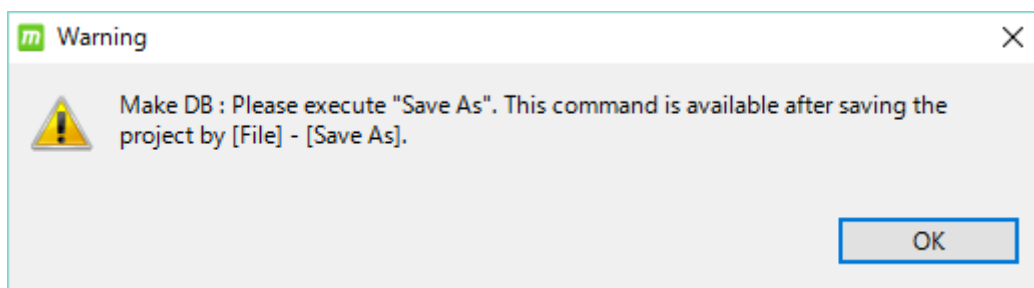
入力する化合物ファイルは、複数の sdf ファイルが可能です。


※ スクリーニングで使用するデータベースに含まれる化合物数は、最低でも数 100～数 1000 個以上にしてください。化合物数が少ないとスクリーニング結果テーブルにヒット化合物が出力されません。

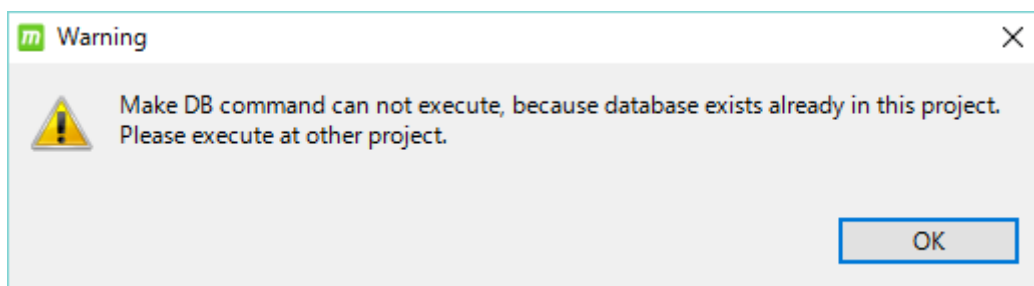
化合物スクリーニングの予測モデルで使用している重線形回帰式のパラメタを決定するためには、分子数が最低数 100～数 1000 程度以上必要なためです。

### 1.4.1. プロジェクトの保存

プロジェクトを保存していない場合は、保存を促すワーニング画面がでますので、プロジェクトを [File] - [Save as] メニューで保存してください。スクリーニング用の化合物データベースは、保存したフォルダ以下に作成しますので、容量に余裕のある場所に保存してください。10 万化合物あたり、約 6GB の容量が必要です。



また、すでに作成済のプロジェクトを、[File] - [Open Project] で開いた場合で、 [Make DB for Screening] をクリックしたときに、以下のワーニング画面が出る場合があります。



すでに、出力先になる **work ¥ database** フォルダが作成されているプロジェクトでデータベースを作成しようとしたため実行不可能となっておりますので、**work ¥ database** の存在しない別のプロジェクトで実行してください。

#### 1.4.2. データベース作成条件の設定

プロジェクトが保存されていて、**work ¥ database** フォルダが存在しない場合は、以下のデータベース作成条件の設定画面が出ます。

**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 files:

set supplier name if "<SUPPLIERNAME\_\*>" does not exists in sdf files (option):

Tag of molecular names (option):

SOURCE of input files (option):

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.)
```

**Filtering 1**

☒ by partial structure ☒ General ☐ Agricultural

**Filtering 2**

☒ by molecular weight Min  Max

2 行目の、  
 [set substitute property name to identify molecules, if "<NScode>" or  
 "<SUPPLIERID\_\*>" or "<IDNUMBER>" or "<idnumber>" does tag does not exists in  
 sdf file :]

の文字列の入力が重要なので以下で詳しく説明します。

初めに、入力 sdf ファイルの中身をテキストエディターなどで確認してください。

※ sdf ファイルを Windows で開いて中身を確認する場合は、フリーソフトの TeraPad が便利です。

sdf ファイルの付加情報として、property 名として、NScode または SUPPLIERID\_\* または IDNUMBER または, idnumber の記述がある場合、すなわち、

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

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

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

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

(ただし、\*は任意文字列)

が記述されている場合、MolDesk では、自動的に生成する mol2 ファイルのコメント行の IDNUMBER=に、sdf ファイルの上記 property 値の文字列を以下のように記述します。  
これにより出力分子の同定が可能になります。

(mol2 ファイル記述例)

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

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

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

. . .
```

生成する mol2 ファイルは、スクリーニング対象となりますので、スクリーニング計算結果リストにも、IDNUMBER 項としてこの値が記述され、入力 sdf 分子との紐付けが可能になります。

ここで、仮に、入力 sdf ファイルの分子記述が以下の通り、付加情報として、property 名として、NScode または SUPPLIERID\_\* または IDNUMBER または, idnumber の記述がなかったとします。

```
Mrv1622910011607582D

14 13 0 0 0 0          999 v2000
  0.2198    0.0635    0.0000 C    0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0.9343   -1.1740    0.0000 C    0 0 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 0
  0.2198    0.8885    0.0000 C    0 0 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 0 0
 -1.2091    0.0635    0.0000 C    0 0 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 0 0
 -1.2091    0.8885    0.0000 C    0 0 0 0 0 0 0 0 0 0 0 0 0 0
  1.6488   -1.5865    0.0000 O    0 0 0 0 0 0 0 0 0 0 0 0 0 0
  1.6488    0.0635    0.0000 N    0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0.2198   -1.5865    0.0000 O    0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -1.9236   -0.3490    0.0000 O    0 0 0 0 0 0 0 0 0 0 0 0 0 0
 -0.4946    2.1260    0.0000 O    0 0 0 0 0 0 0 0 0 0 0 0 0 0
  3.6524    0.0000    0.0000 Cl   0 0 0 0 0 0 0 0 0 0 0 0 0 0

2 3 1 0 0 0 0
3 1 1 0 0 0 0
4 1 2 0 0 0 0
5 1 1 0 0 0 0
6 5 2 0 0 0 0
7 4 1 0 0 0 0
8 6 1 0 0 0 0
9 2 2 0 0 0 0
```

```

10 3 1 0 0 0 0
11 2 1 0 0 0 0
12 6 1 0 0 0 0
13 7 1 0 0 0 0
 8 7 2 0 0 0 0
M END
> <SID>
NS-000000001-0001

$$$$
. . .

```

この sdf ファイルでは、このままでは、出力 mol2 ファイルに IDNUMBER= の記述ができないので、入力 sdf ファイルの分子と出力 mol2 ファイルの紐付けがなくなります。そこで、代わりに property 名 SID の記述を mol2 ファイルのコメント行の IDNUMBER= として記述することになります。

その場合に以下の通り、2 行目の、

[set substitute property name to identify molecules, if "<NScode>" or "<SUPPLIERID\_\*>" or "<IDNUMBER>" or "<idnumber>" does tag does not exists in sdf file :]

に、SID と上記 property 名を記述します。

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 tag does not exists in sdf file :

☒ make conformer

set supplier name if "<SUPPLIERNAME\_\*>" does not exists in sdf files (option) :

Tag of molecular names (option) :

SOURCE of input files (option) :

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.)

Filtering 1

☒ by partial structure

☒ General ☐ Agricultural

Filtering 2

☒ by molecular weight

Min 200 Max 400

OK Cancel

もし、入力 sdf ファイルの **property** 名が以下のように、空白が入っている場合は、**property** 名として認識できません。

```
> <entry name>
molecule.001
```

この場合は、例えば、以下のように sdf ファイルのタグ名をエディターなどで一括置換して、空白をなくしてからご使用ください。

```
> <entry_name>
molecule.001
```

各項目の内容は、以下の通りです。

[Convert to 3D]

項目	内容
2D → 3D	<p>チェックすると 3 次元化を行う。</p> <p>3 次元化は以下の手順で行う。AMBER GAFF2 力場によりエネルギー極小化計算による 3 次元化計算を行う。その際に、H 原子の付加と電荷の生成を行う。H 原子の付加は、酸性/塩基性官能基が水中で解離状態になるよう付加し、電荷は MOPAC7 AM1 で生成する。</p> <p>チェックを外すと 3 次元化を行わない。この際は、H 原子の付加、電荷の生成は行わないでオリジナルの構造をそのまま反映して、Mol2 ファイルを出力する。分子がすでに 3 次元化されているなどの理由で 3 次元化が不要の場合に、チェックを外す。</p>
make conformer	<p>3 次元化する際に、分子の <b>Conformer</b> を生成する場合はチェックする。4 員環以上の環構造の部分について生成。分子内にキラル中心が存在する場合には、光学異性体も同時に生成。</p>

<p>set substitute property name to identify molecules, if "&lt;NScode&gt;" or "&lt;SUPPLIERID_*&gt;" or "&lt;IDNUMBER&gt;" or "&lt;idnumber&gt;" does tag does not exists in sdf file :</p>	<p>入力 sdf ファイルの中に、&lt;NScode&gt; または &lt;SUPPLIERID_*&gt; または &lt;IDNUMBER&gt; または &lt;idnumber&gt; の property 名が存在しない時に、別の property 名の property 値を、IDNUMBER=として出力 mol2 ファイルに記述する。</p> <p>IDNUMBER として認識させたい別の property 名を記入する。記入がない場合は、上記 3 つの prpperty 名を自動判定して、IDNUMBER とする。</p> <p>上記 3 つの property 名が存在せず、ユーザ入力の別名 property 名も存在しない場合は、IDNUMBER= は出力 mol2 ファイルに付与されない（この場合も、3 次元化はできるが、入力 sdf ファイルの分子と出力 mol2 ファイルの分子の紐付けができなくなる）。</p>
---	---

[Convert to 3D] 以下はオプションです。指定は必須ではありません。

項目	内容
Set supplier name if “<SUPPLIERNAME_*>” does not exists in sdf files	<p>入力 sdf ファイルの中に、&lt;SUPPLIERNAME_*&gt; が存在しない時に、ここで入力した文字列を、SUPPLIER = として出力 Mol2 ファイルに記録できる（分子 1 個ごとの指定はできない）。</p> <p>入力 sdf ファイルの中に、&lt;SUPPLIERNAME_*&gt; が存在する場合は、そちらの記述が優先されて出力 Mol2 ファイルに、SUPPLIER=として記述される。</p> <p>（下記の <b>SUPPLIER</b> の部分）。</p>
Tag of molecular name	<p>分子名の先頭につけるタグを指定する。分子名とは、出力 Mol2 ファイルの @&lt;TRIPOS&gt;MOLECULE の次の行に記載される文字列。</p> <p>（下記の <b>MOLECULE</b> の部分）。</p> <p>これは、プログラムが独自に生成する分子識別 ID 番号です。</p>
SOURCE of input files	<p>入力ファイルの供給源を指定する。</p> <p>出力 Mol2 ファイルの COMMENT 行に、SOURCE=として記載される。</p> <p>（下記の <b>SOURCE</b> の部分）。</p>

(mol2 ファイル記述例)

```
@<TRIPOS>COMMENT
LIGANDBOX_ID = MOLECULE-000000001-01
SUPPLIER = SUPPLIER
SOURCE = SOURCE
IDNUMBER = NS-0000000001-0001
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

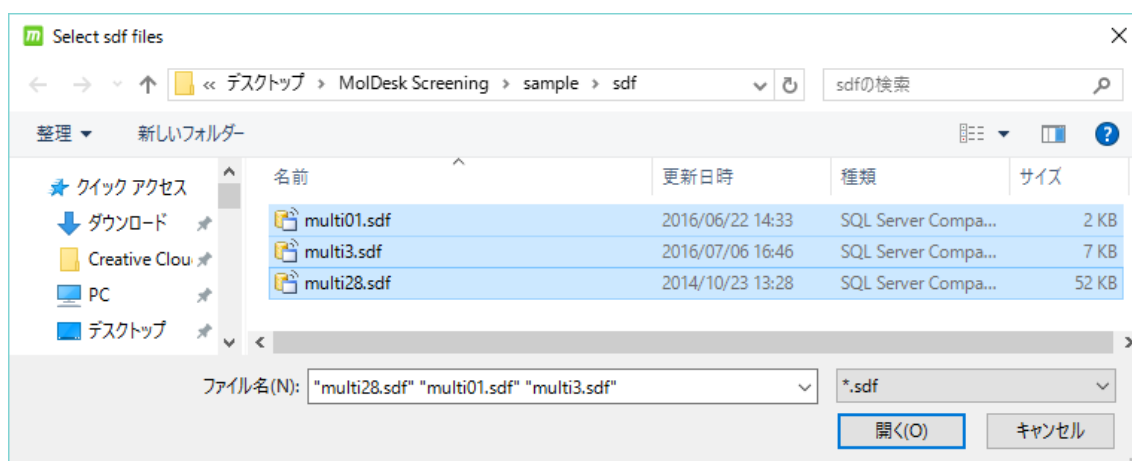
@<TRIPOS>MOLECULE
MOLECULE-000000001-01
```



[Filtering]

項目	内容
[by partial structure] 部分構造によるフィルタリングを行う場合はチェックする	一般的な薬剤にふさわしくない構造を除外するか、農薬にふさわしくない構造を除外するか選択する。
[by molecular weight] 分子量によるフィルタリングを行う場合はチェックする	最小分子量と、最大分子量を指定する。

[OK] をクリックすると、ファイルセクターが出るので、入力ファイルを選択して [開く] をクリックします。



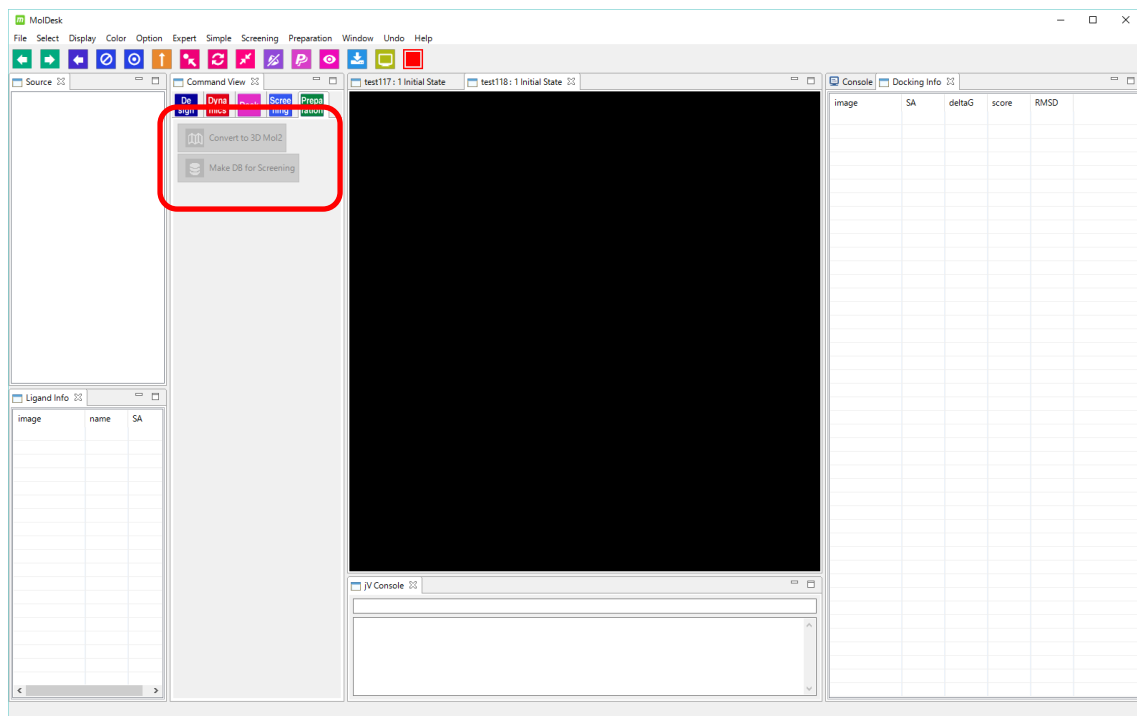
この例では MolDesk Screening フォルダに含まれる以下の 3 つの sdf ファイルを選択しました。

- MolDesk Screenng -> sample -> sdf -> multi01.sdf (化合物が 1 個含まれる)
- MolDesk Screenng -> sample -> sdf -> multi3.sdf (化合物が 3 個含まれる)
- MolDesk Screenng -> sample -> sdf -> multi28.sdf (化合物が 28 個含まれる)

[開く] をクリックすると、計算が始まります。

### 1.4.3. データベース作成計算

スクリーニング計算を開始するとコマンドボタンがグレーになります。コマンドボタンがグレーになっている間は計算中です。



スクリーニング用の化合物データベース作成の計算時間の目安は、以下の表の通りです。

計算法	Intel Core i7-4790K 4.0GHz / 16GB メモリ / windows8.1 <b>8 並列で実行</b>	Xeon E5-2697 v2 @ 2.70GHz x 2 (24 コア 48 プロセッサ) / 64GB メモリ / Linux CentOS6 <b>48 並列で実行</b>
計算時間	641 時間 (26 日 7 時間)	191 時間 (7 日 23 時間)

化合物数が 259,868 分子の計算例

実際の計算時間は、データベースを作成したい化合物数で比例倍してください。

並列計算で高速に計算しますが、化合物データベース作成時に必要なメモリ量は並列数が大きくなるほど増大します。

例えば、8 並列の場合は 16GB、48 並列の場合は 32GB が必要です。

並列数は [Help] - [Preference] - [Screening] の Thread number の設定で指定できます。

デフォルトではマシンの最大プロセッサ数が設定されています。

並列数が大きいほどメモリを消費しますので、メモリ量が少ないマシンでは **Thread number** の値を小さくしてください。

Window 32bit では計算できません (メモリ不足のため)。Windows 64bit または Linux 64bit のなるべくスペックの良いマシンをご用意ください。

30 万個の分子のスクリーニング計算用の化合物データベース作成で、約 4.5GB 記憶媒体を消費します。

途中で停止した場合は、初めから再計算が必要です。

#### 1.4.4. データベース作成場所

データベース作成の計算が終了するとコマンドボタンがグレーから使用できるように変わります。データベースの作成される場所は、保存したプロジェクトのフォルダを [PROJECT] とすると、

[PROJECT] -> work -> database

です。database は、以下のフォルダ構成になります。

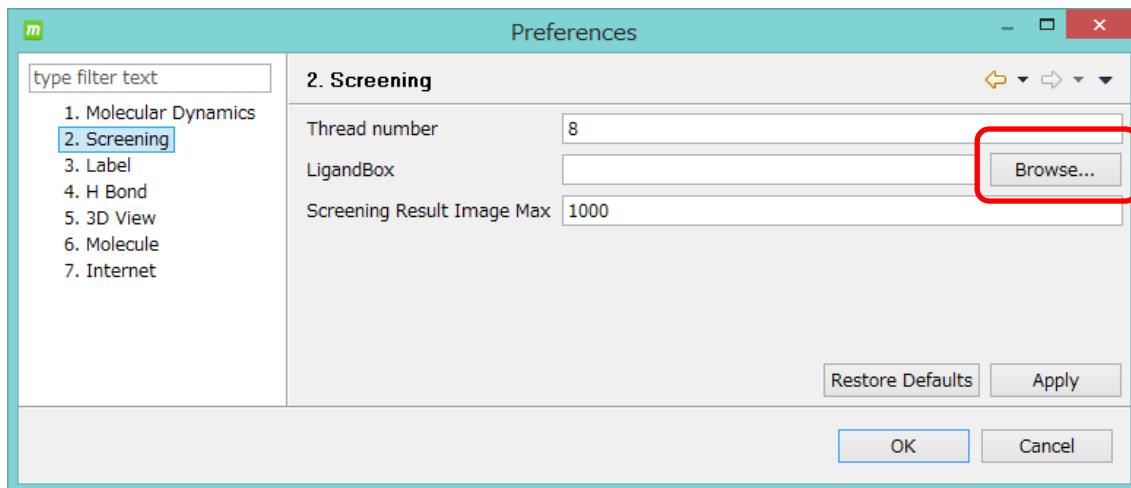
```
[PROJECT] - work - database - ligand
                             - ligandImage
                             - mol2_files
                             - mts_data
                             - protein
                             - all.mol2 (ファイル)
                             - all_exclude.mol2 (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

それぞれの内容は以下の通りです。

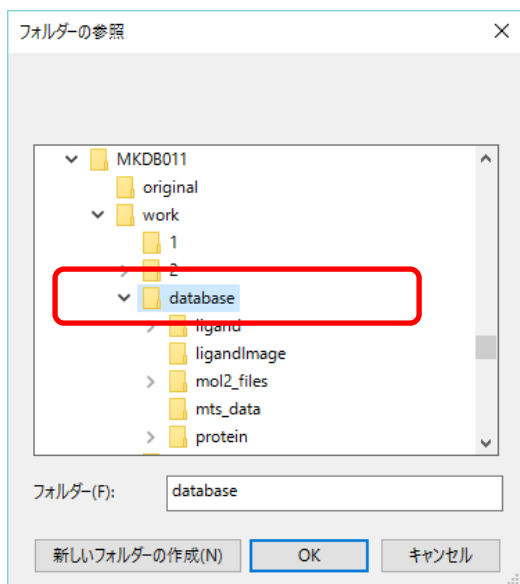
項目	内容
ligand	分子数 10 万個ごとに、c*** というフォルダに 3 次元化した mol2 ファイルを作成。フィルタリングした後。
ligandimage	2 次元図の画像ファイル類
mol2_files	入力ファイル単位で 3d*** というフォルダに 3 次元化した mol2 ファイルを作成。フィルタリングする前。
mts_data	化合物と 181 タンパク質の相互作用行列ファイル
protein	181 タンパク質のドッキング計算用の入力ファイル類を各タンパク質毎のフォルダに保存
all.mol2	mol2_files 以下のすべての mol2 ファイルをマージした multi mol2 ファイル。フィルタリングする前。
all_exclude.mol2	all.mol2 ファイルを部分構造によるフィルタリングした後の mol2 ファイル。
error.log	3 次元化計算時のエラーログ。エラーにより 3 次元構造が生成できなかった分子を確認できる。
exclude.info	部分構造によるフィルタリングを実行するときに使用する、各分子の部分構造存在の情報ファイル
pro.list	181 タンパク質のリスト
version	データベースのバージョンファイル。

「8.3.2 Screening」を参照し、ここで作成した database を設定すると、database を対象にしてスクリーニング計算ができるようになります。具体的には以下の通りです。

[Help] - [Preference] 画面を開いて、「2. Screening」を選択し、[Browse] をクリックします。





保存したプロジェクト（例では、MKDB011）の database フォルダを下図のように選択し、[OK]をクリックします。




database フォルダが指定されたのを確認し、[OK]をクリックします。


## 1.5. スクリーニング用化合物 DB の再分割

LigandBox を使用してスクリーニング計算を行う場合は、この節は読み飛ばしてください。

 [Preparation] -  [Remake DB for Screening] で、スクリーニング用のデータベースを再分割することができます。

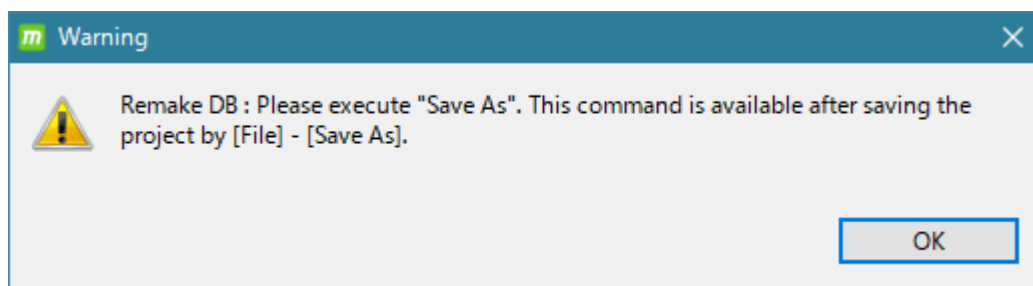
再分割する目的は、以下で説明する通り、スクリーニング計算の高速化です。


前節の、 [Make DB for Screening] で作成したデータベースや、LigandBox は、1 万化合物ごとに内部的に分割されています。LigandBox は 200 万化合物なので 200 分割、仮

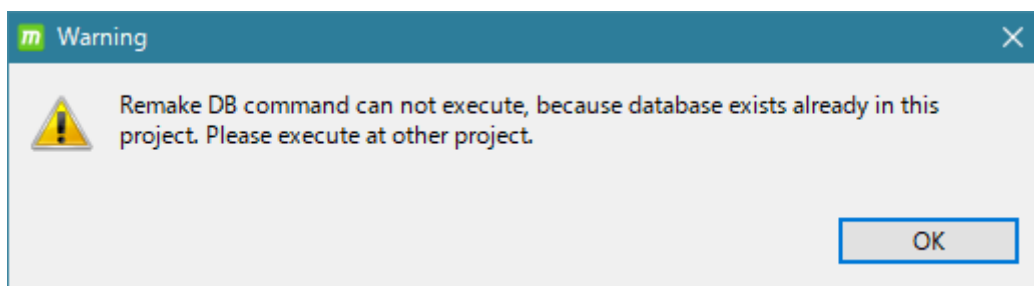
に、ユーザが  [Make DB for Screening] で作成したデータベースの化合物数が 30 万化合物だった場合 30 分割になります。並列計算によっては、これより大きなスレッド並列数で並列計算が可能な計算機がありえます。この場合に、より大きな分割数で再分割してデータベースを作成し直すと、スクリーニング計算が早く完了します。

### 1.5.1. プロジェクトの保存

プロジェクトを保存していない場合は、保存を促すワーニング画面がでますので、プロジェクトを [File] – [Save as] メニューで保存してください。再分割したスクリーニング用の化合物データベースは、保存したフォルダ以下に作成しますので、容量に余裕のある場所に保存してください。10 万化合物あたり、約 6GB の容量が必要です。



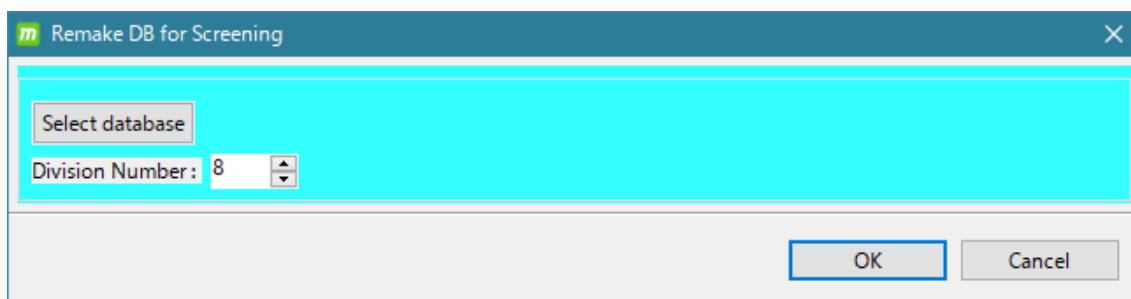
また、すでに作成済のプロジェクトを、[File] – [Open Project] で開いた場合で、 [Remake DB for Screening] をクリックしたときに、以下のワーニング画面が出ることがあります。



すでに、出力先になる **work ¥ database** フォルダが作成されているプロジェクトでデータベースを作成しようとしたため実行不可能となっておりますので、**work ¥ database** の存在しない別のプロジェクトで実行してください。

### 1.5.2. データベース再分割条件の設定

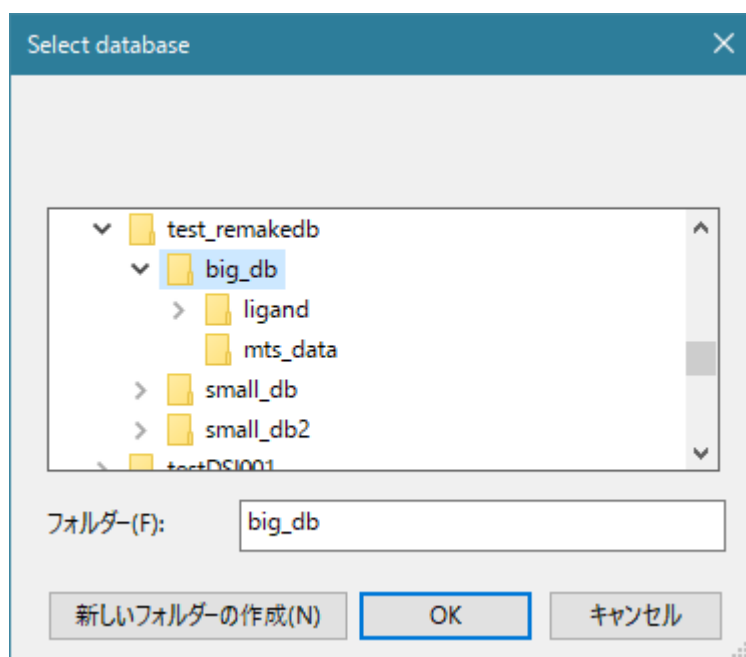
プロジェクトが保存されていて、**work ¥ database** フォルダが存在しない場合は、以下のデータベース作成条件の設定画面が出ます。



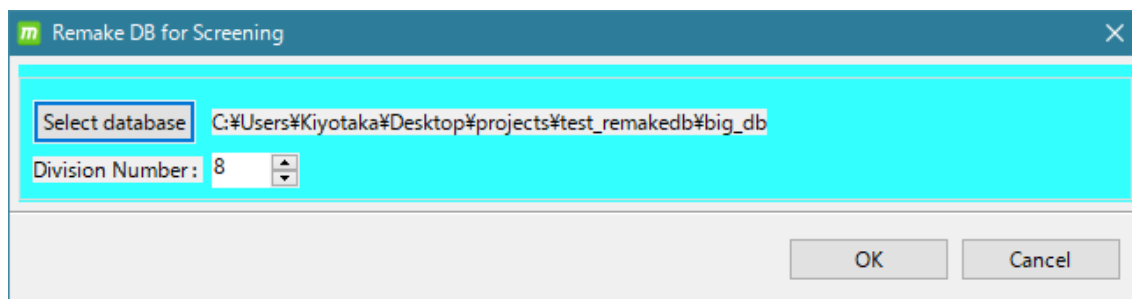
各項目の内容は、以下の通りです。

項目	内容
Select database	再分割するデータベースのフォルダ（ディレクトリ）を選択する。
Division Number	再分割するときの分割数

[Select database] をクリックすると、以下のフォルダ選択画面が表示するので、**ligand** と **mts\_data** フォルダが直下にあるフォルダを選択します。再分割計算に必須なフォルダは、**ligand** と **mts\_data** です。それ以外は、再分割で使用しませんので不要です。



[OK] をクリックすると、選択したフォルダが下図のように入力されます。



次に、[Division Number] を選択しますが、デフォルトで、計算機で可能な最大スレッド数が表示されています（上図の計算機では 8）。

[OK] をクリックすると、再分割が始まります。



### 1.5.3. データベース作成場所

データベース作成の計算が終了するとコマンドボタンがグレーから使用できるように変わります。データベースの作成される場所は、保存したプロジェクトのフォルダを [PROJECT] とすると、

[PROJECT] -> work -> database

です。database は、以下のフォルダ構成になります。

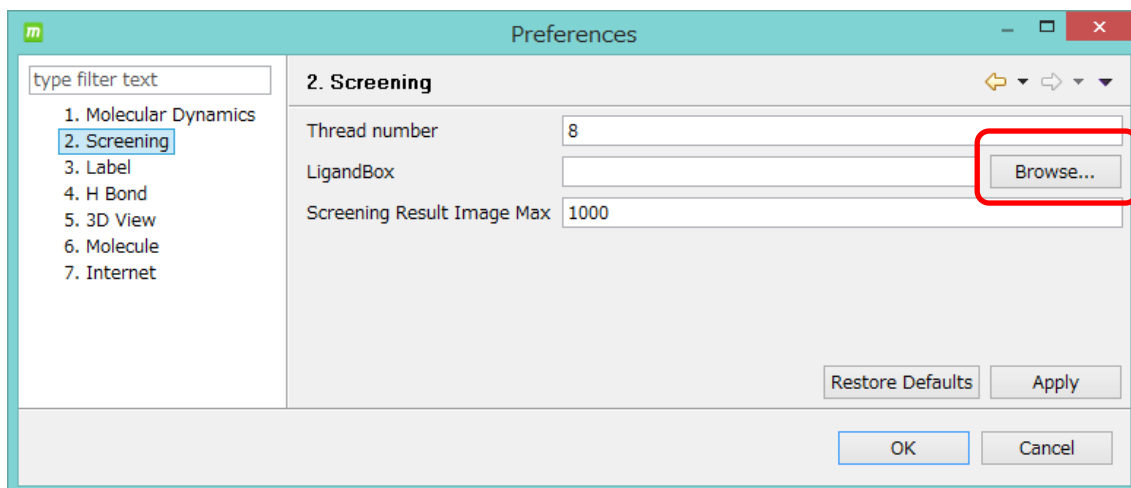
```
[PROJECT] - work - database - ligand
                                - ligandImage
                                - mts_data
                                - protein
                                - pro_list (ファイル)
                                - version (ファイル)
```

それぞれの内容は以下の通りです。

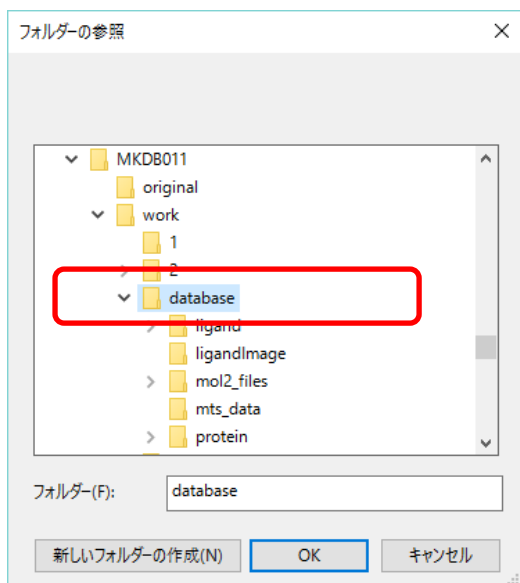
項目	内容
ligand	分子数 10 万個ごとに、c***というフォルダに 3 次元化した mol2 ファイルを作成。フィルタリングした後。
ligandimage	2 次元図の画像ファイル類
mts_data	化合物と 181 タンパク質の相互作用行列ファイル
protein	181 タンパク質のドッキング計算用の入力ファイル類を各タンパク質毎のフォルダに保存
pro.list	181 タンパク質のリスト
version	データベースのバージョンファイル。

「8.3.2 Screening」を参照し、ここで作成した database を設定すると、database を対象にしてスクリーニング計算ができるようになります。具体的には以下の通りです。

[Help] - [Preference] 画面を開いて、「2. Screening」を選択し、[Browse] をクリックします。



保存したプロジェクト（例では、MKDB011）の database フォルダを下図のように選択し、[OK]をクリックします。



database フォルダが指定されたのを確認し、[OK]をクリックします。

## 1.6. スクリーニング計算の概要

### MolDesk Screening データフロー図



※ 化合物データベースは、ユーザの分子、または、ご提供する LigandBox が可能

- myPresto の 以下の手法を用いたインシリコスクリーニングが実行できます。
  - ① ドッキングスコア順
  - ② MTS 法 (MTS)
  - ③ 機械学習 MTS 法 (ML-MTS)
  - ④ 機械学習 DSI 法 (ML-DSI)
- いずれの手法もドッキング計算をベースにしているため、最も精度が高くなる活性化化合物の分子量はおよそ 200~400 Da の範囲です。
- 機械学習に用いる活性化化合物の数は 5 個以上を推奨します。
- 検索対象は、LigandBox の 200 万化合物およびユーザが追加した化合物です。
- 各スクリーニング計算法の入力は以下の通りです。

計算法	ターゲット タンパク質 (PDB)	既知活性 リガンド (mol2)	ユーザ追加による 検索対象化合物 (mol2)
ドッキングスコア順	◎		○
MTS	◎		○
ML-MTS	◎	◎	○
ML-DSI		◎	○

◎必須 ○任意

※ ここでは、LigandBox からスクリーニングすることを前提で説明していますが、ユーザの化合物からスクリーニングすることも現行バージョンでは可能です。

## 1.7. スクリーニング計算の並列数・メモリ量・時間

スクリーニングの計算時間の目安を以下に示します。

計算法	Intel Core i7-4790K 4.0GHz / 16GB メモリ / windows8.1 <b>8 並列で実行</b>	Xeon E5-2697 v2 @ 2.70GHz x 2 (24 コア 48 プロセッサ) / 64GB メモリ / Linux CentOS6 <b>48 並列で実行</b>
ドッキング スコア順 または MTS	35 時間 31 分	10 時間 3 分
ML-MTS	45 時間 7 分	13 時間 12 分
ML-DSI	8 時間 26 分	2 時間 49 分

タンパク質を含む受容体側が 8,928 原子、LigandBox + 174 化合物の計算例

並列計算に関して特に設定を行う必要はありません（スレッド並列計算で行います）。

MolDesk Screening をインストール・アクティベーションした直後に、すぐ並列計算が行えます。

スクリーニング計算時に必要なメモリ量は並列数が大きくなるほど増大します。

例えば、8 並列の場合は 16GB、48 並列の場合は 32GB が必要です。

並列数は [Help] - [Preference] - [Screening] の Thread number の設定で指定できます。

デフォルトではマシンの最大プロセッサ数が設定されています。

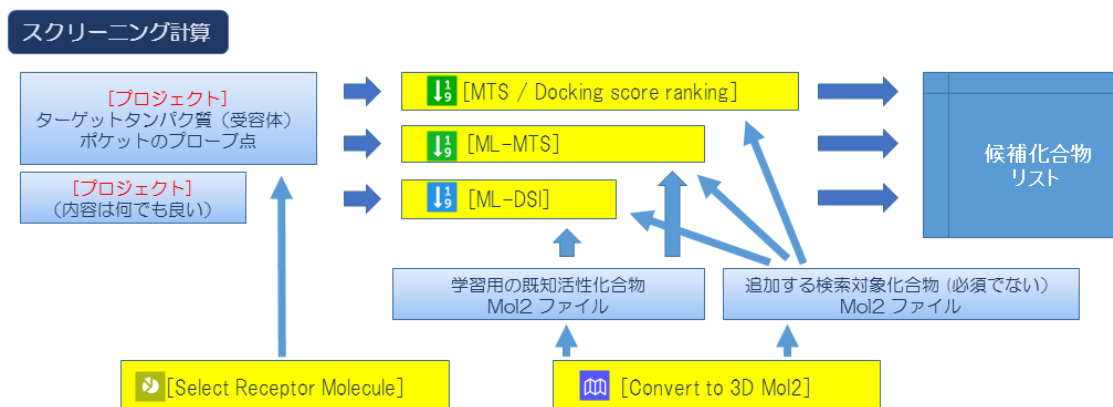
並列数が多いほどメモリを消費しますので、メモリ量が少ないマシンでは Thread number の値を小さくしてください。

Window 32bit の場合は、メモリ不足のためにスクリーニング計算が実行できません。

スクリーニング計算を実行する場合は、Windows 64bit または Linux 64bit のなるべくスペックの良いマシンをご用意ください。

1 回あたりのスクリーニング計算で、約 5GB 記憶媒体を消費します。

## 1.8. スクリーニング計算の手順



### 1. mol2 ファイルの準備

ML-MTS 法と ML-DSI 法では、学習用の既知活性リガンドの mol2 ファイルが必要です (MTS / Docking score ranking 法では不要です)。

また、追加したい検索対象化合物がある場合はそれらの mol2 ファイルも必要です。


mol2 ファイルの準備手順については「1.9 mol2 ファイルの準備」を参照してください。




### 2. ターゲットタンパク質とポケットの準備

MTS / Docking score ranking 法と ML-MTS 法では、ターゲットタンパク質とポケットをモデリングしておく必要があります (ML-DSI 法はリガンドベースの計算法のため、ターゲットタンパク質とポケットのモデリングは不要です)。

ターゲットタンパク質とポケットがモデリングされたプロジェクトを作成する方法は 2 通りあります。

- A) ターゲットタンパク質とポケットのプロープ点をモデリングしたプロジェクトを作成します。ポケットのプロープ点の作成方法は MolDesk Basic のマニュアルを参照してください。
- B) モデリングがすでに終了しているターゲットタンパク質の PDB ファイルと、ポケットのプロープ点の PDB ファイルをプロジェクトに読み込みます。PDB ファイルの読み込み方法は MolDesk Basic のマニュアルを参照してください。

ターゲットタンパク質とポケットをモデリングした後、受容体分子を  [Select Receptor Molecule] で選択します。このとき、受容体のポケットの空間は空けるように選択してください。

3.  [MTS / Docking score ranking]、 [ML-MTS]、 [ML-DSI] のいずれかをクリックすると、スクリーニング計算の入力ダイアログが表示されます。  
上記で作成した mol2 ファイルや、ターゲットタンパク質の名称を入力します。
4. [OK]をクリックするとスクリーニング計算が開始します。
5. ターゲットタンパク質、既知活性リガンド、検索対象化合物の入力方法は以下の通りです。

入力するもの	入力方法
ターゲットタンパク質	プロジェクトを作成し、ターゲットタンパク質とポケットを作成するか、既に作成したターゲットタンパク質の PDB ファイルとポケットの PDB ファイルを入力します。
既知活性リガンド および 検索対象化合物	1 化合物あたり 1 個の mol2 ファイルをファイル選択ダイアログから入力します。


本ドキュメントでは、MolDesk Screening のインストール時にデスクトップに作成された MolDesk Screening フォルダに含まれるサンプルデータを用いて、実際にスクリーニング計算を行いながら操作手順を説明します。

## 1.9. mol2 ファイルの準備

学習用の既知活性化合物がある場合や、ユーザが追加したい検索対象化合物がある場合は、それらの mol2 ファイル（3 次元構造化したもの）をあらかじめ用意する必要があります。

mol2 ファイルは、1 つのファイルに複数の分子が記述されているマルチ形式と、1 つのファイルに 1 つの分子が記述されたシングル形式の両方を入力できます。

ただし、化合物分子の 3 次元化と電荷の付加はスクリーニング計算では行わないので、3 次元化と電荷の付加がない化合物分子に対しては、3 次元化と電荷の付加を予め実行しておく必要があります。

化合物分子の 3 次元化と電荷の付加を実行した mol2 ファイルを生成するためには、 [Convert to 3D Mol2] を実行します。



[Convert to 3D Mol2] の詳細は MolDesk Basic のマニュアルを参照してください。

## 1.10. MTS 法またはドッキングスコア順によるスクリーニング

MTS 法またはドッキングスコア順によるスクリーニング計算では、1 回の計算で MTS 法とドッキングスコア順を同時に計算します。

### 1.10.1. プロジェクトの作成

プロジェクトを作成し、ターゲットタンパク質をモデリングし、ポケットを作成します。

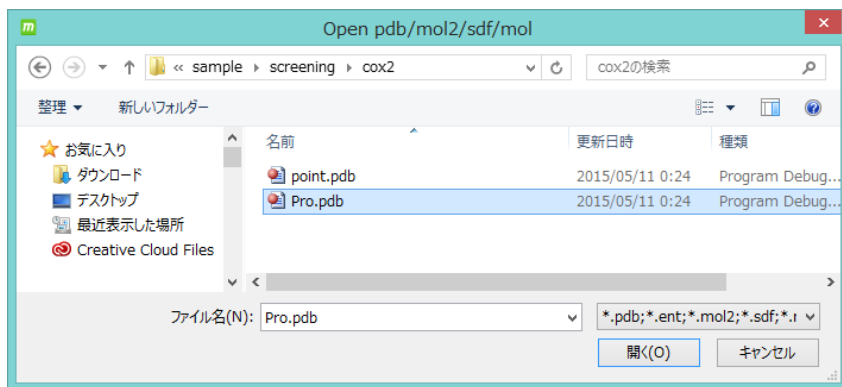
この例では、タンパク質とポケット情報をファイルから読み込みます。

ポケットの作成は  [Make Pocket] や  [Find Pocket] で行うことも可能です。  
ポケットを作成する方法については MolDesk Basic のマニュアルを参照してください。

[File] - [Open Molecular File] メニューでプロジェクトを作成します。

この例では MolDesk Screening フォルダに含まれる以下の PDB ファイルを選択して新規プロジェクトを作成します。

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

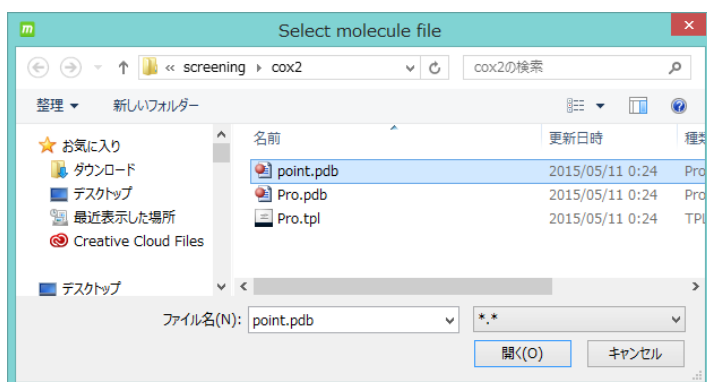
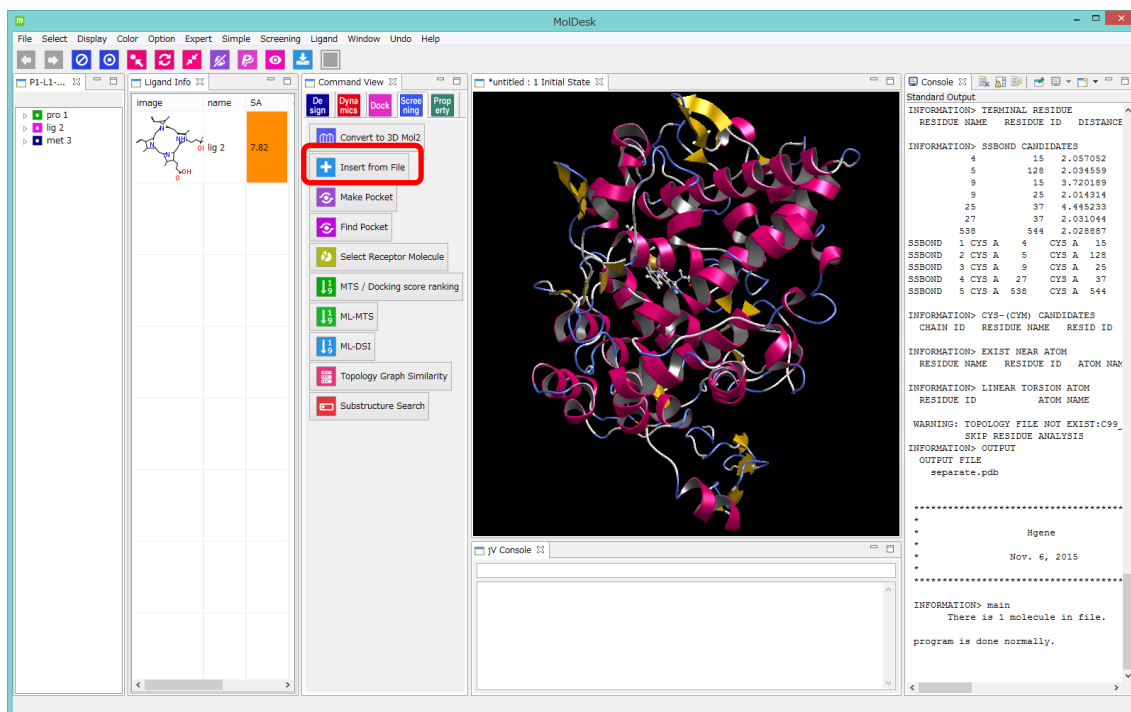


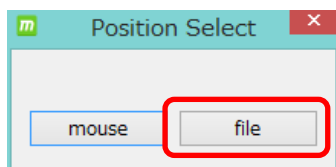




[Insert from File] をクリックし、以下のポケットファイルを選択します。

MolDock Screening -> sample -> screening -> cox2 -> point.pdb

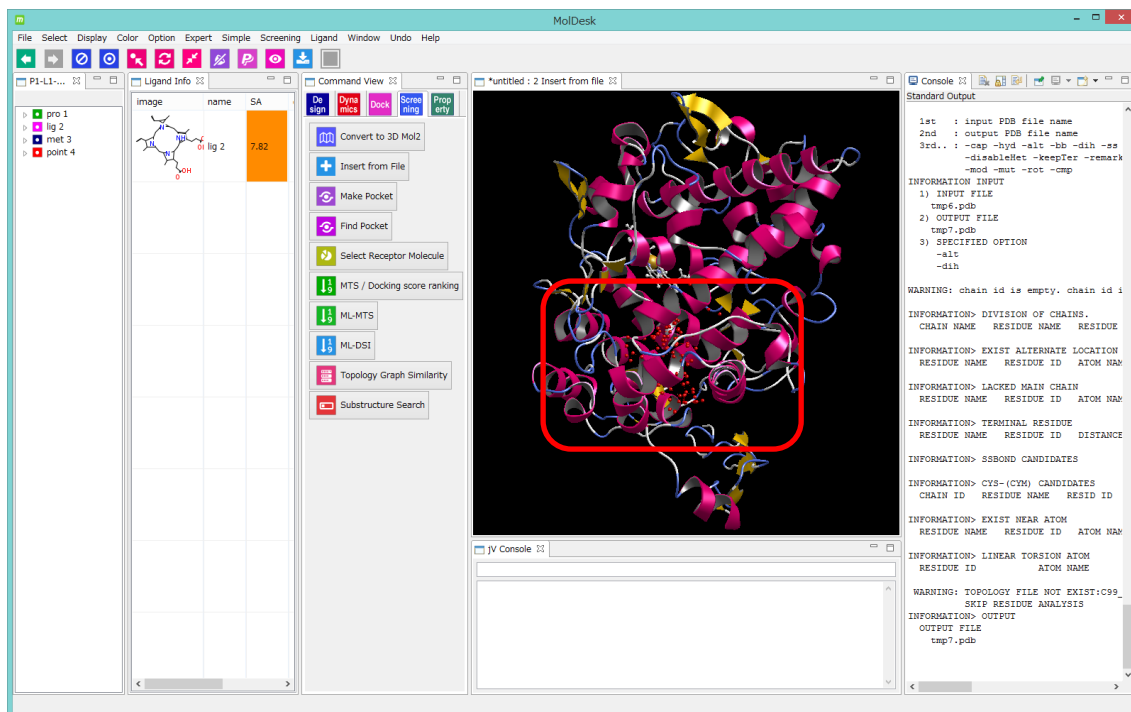




[file] を選択します。

[mouse] を選択するとユーザがマウスクリックした座標にポケットファイルが入力され、正確な計算できません。

ファイルで指定した位置にポケットが入力されます。

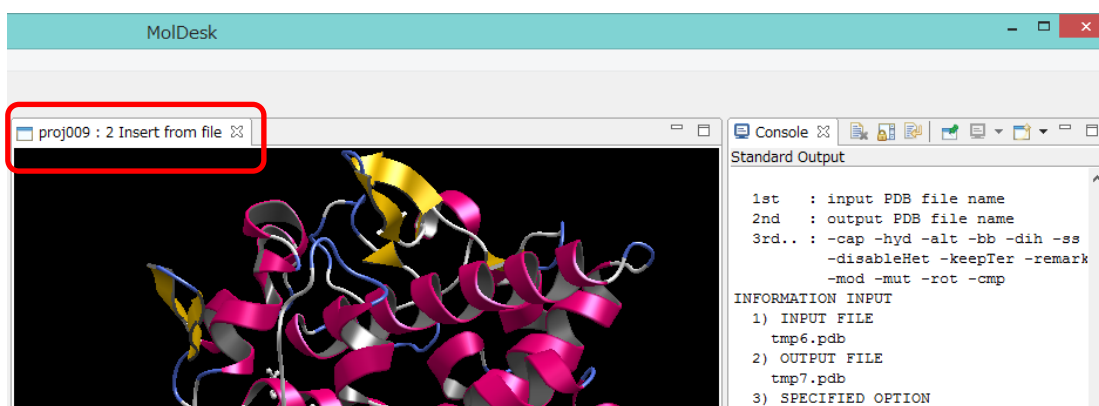


このプロジェクトを「proj009」という名前で保存します。

プロジェクトの保存方法は MolDesk Basic のマニュアルを参照してください。

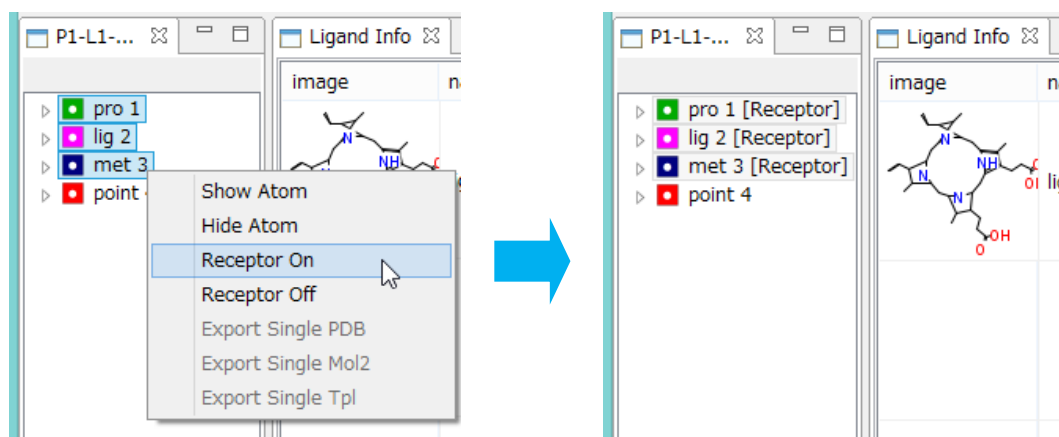
スクリーニング計算は 1 回の計算で生成するデータ量が多い（～数 GB）ため、計算前にプロジェクトの保存を行い、データを保存するフォルダを確定しておく必要があります。

プロジェクトを保存すると 3D 画面のタブ名がプロジェクト名に変わります。



ドッキング計算の受容体を指定します。

ここでは、ツリー表示画面で ■ pro 1、■ lig 2、■ met 3 を Ctrl + クリックで選択し、右クリックから[Receptor On] を選択しています。（lig2、met3 はポケットと関係ない場所にあるので、■ pro 1 だけを選択しても OK です）。受容体は、ポケットの空間を開けるように選択します。




受容体選択方法の詳細は MolDesk Basic のマニュアルを参照してください。

### 1.10.2. スクリーニング計算のデータ入力

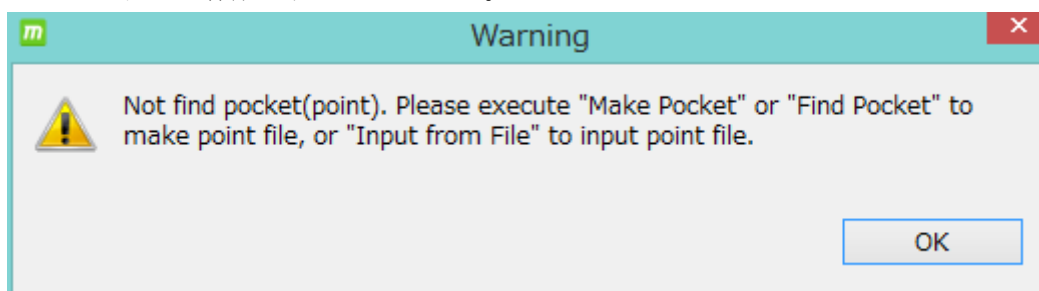


[MTS / Docking score ranking] をクリックします。

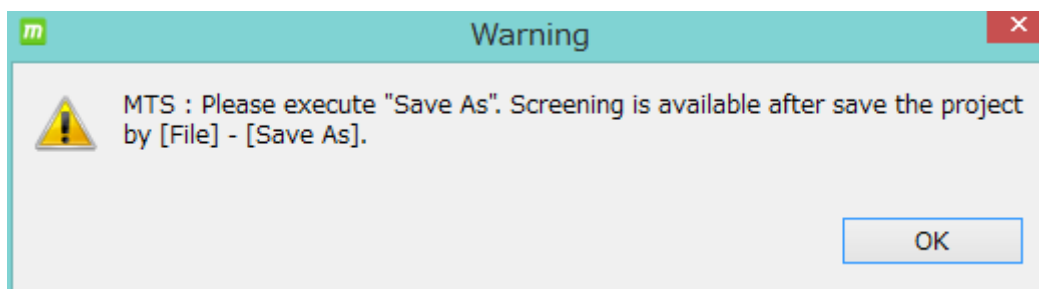
必要な作業が行われていないと以下のワーニングダイアログが表示されますので、必要な

作業を行った後、 [MTS / Docking score ranking] を再度クリックしてください。

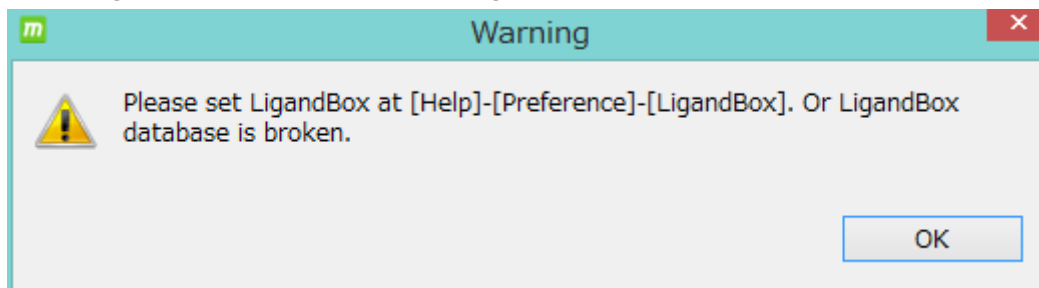
- ポケットを作成していない場合は以下のワーニングダイアログが表示されますので、ポケットの作成を行ってください。



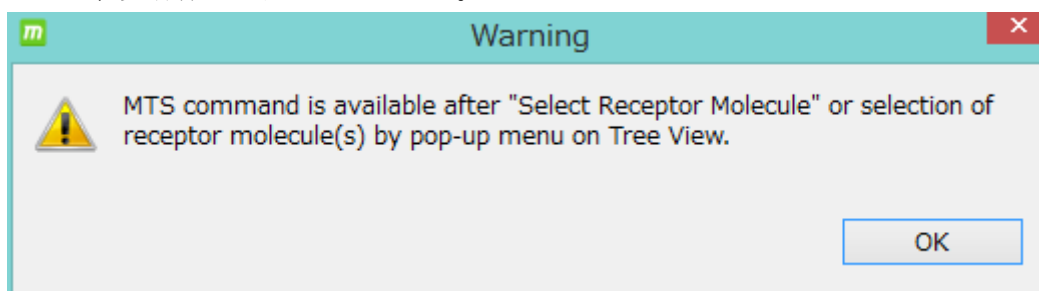
- プロジェクトが保存されていない場合は以下のワーニングダイアログが表示されますので、プロジェクトの保存を行ってください。



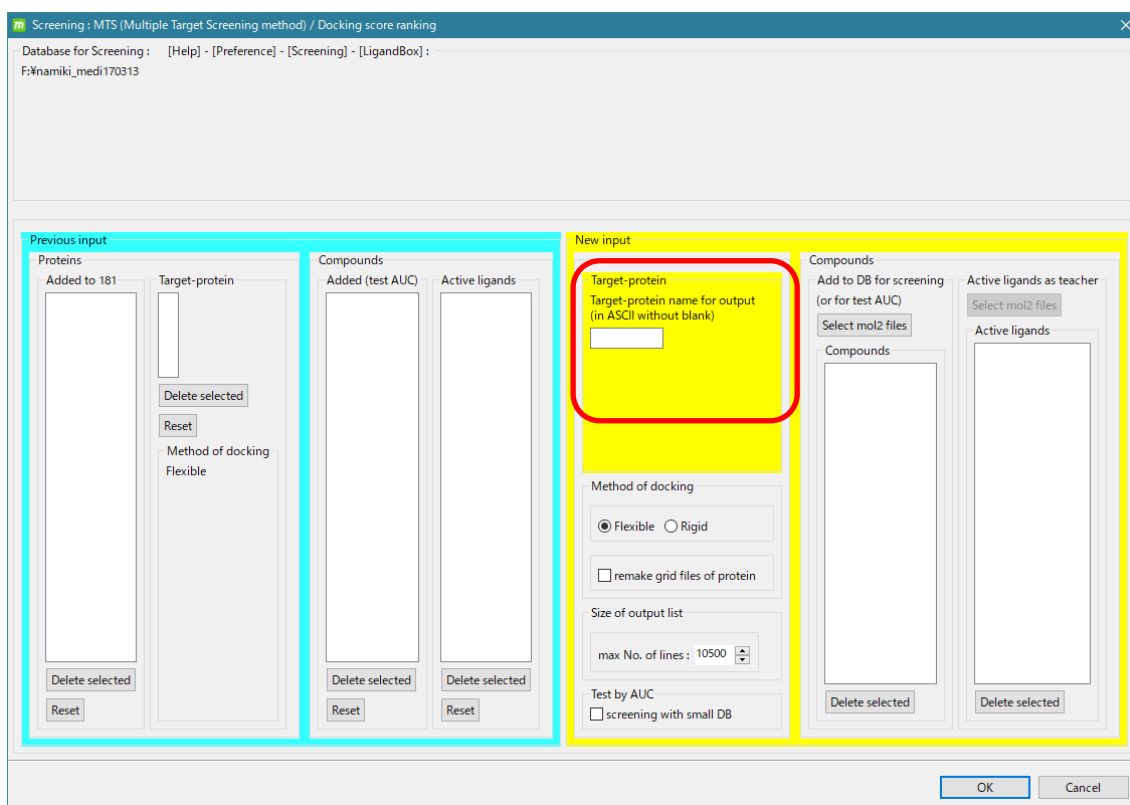
LigandBox が設定されていない場合は以下のワーニングダイアログが表示されますので、「1.2 LigandBox の準備」を参照し LigandBox を設定してください。



- 受容体分子が選択されていない場合は以下のワーニングダイアログが表示されますので、受容体を選択してください。



必要な作業がすべて行われている場合は、スクリーニング計算のデータ入力ダイアログが表示されます。



MTS 法 / ドッキングスコア順によるスクリーニング計算では、ターゲットタンパク質 ([Target protein]) の名前を入力する必要があります。

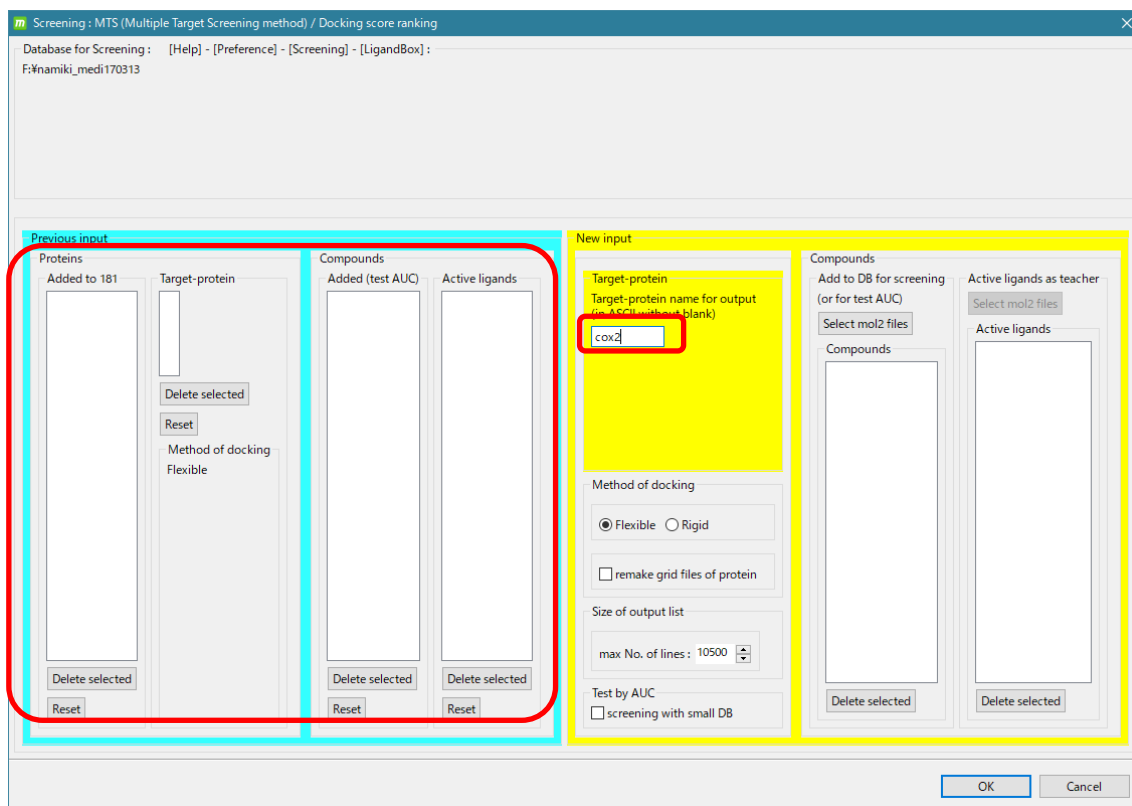
ターゲットタンパク質の名前は任意の名称を **空白を含まない英数字** で入力します。

他は入力必須ではありません。

既知活性リガンド ([Active ligands as teacher]) は MTS 法 / ドッキングスコア順によるスクリーニング計算では使用しないため、入力できないようになっています。

この例では、ターゲットタンパク質の名前に「cox2」と入力します。

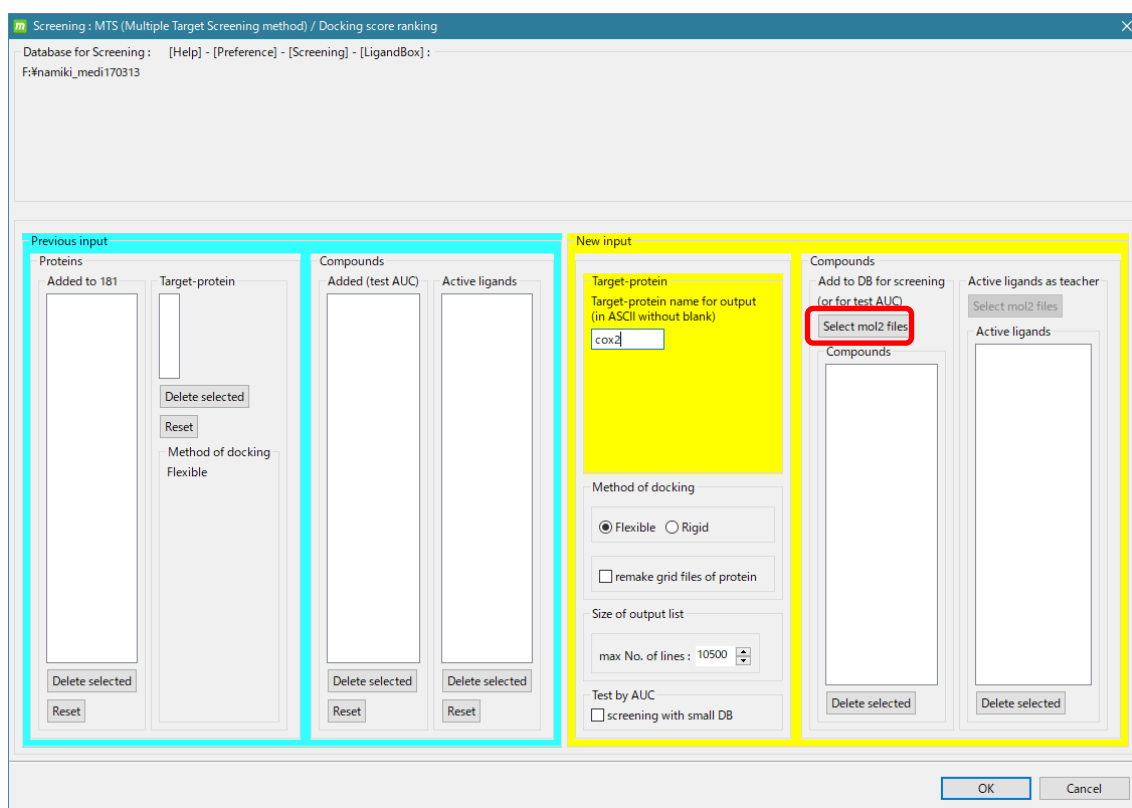
ダイアログ左の青背景の領域には、前回の計算の入力内容が表示されます。この例では初回のため空欄になっています。



### 1.10.3. mol2 ファイルによる化合物の追加

LigandBox に加えてスクリーニング対象とする化合物を mol2 ファイルで入力します。

この例では恣意的に既知活性リガンドを入力し、既知活性化合物がどれだけ上位にリスト化されるかを確認します。スクリーニング計算終了後にデータベースエンリッチメント曲線を表示し、AUC (Area under the curve) による計算精度を検証します。



[Add to DB for screening] の [Select mol2 files] を選択して以下の 13 ファイルを選択し「開く」をクリックします。

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

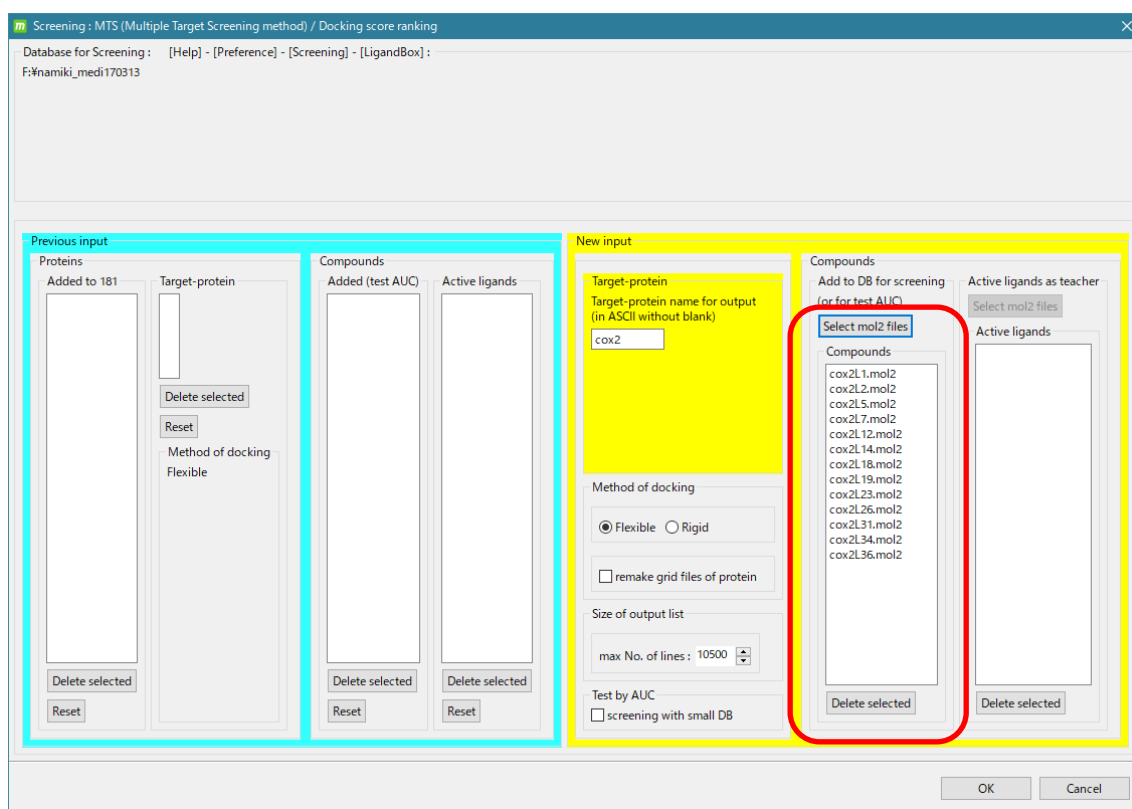
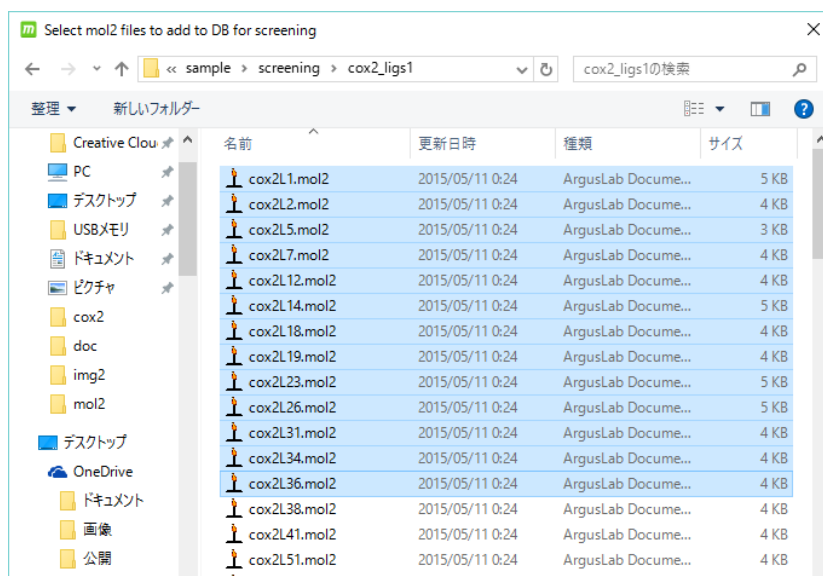
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L23.mol2

MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L26.mol2

MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L31.mol2

MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L34.mol2

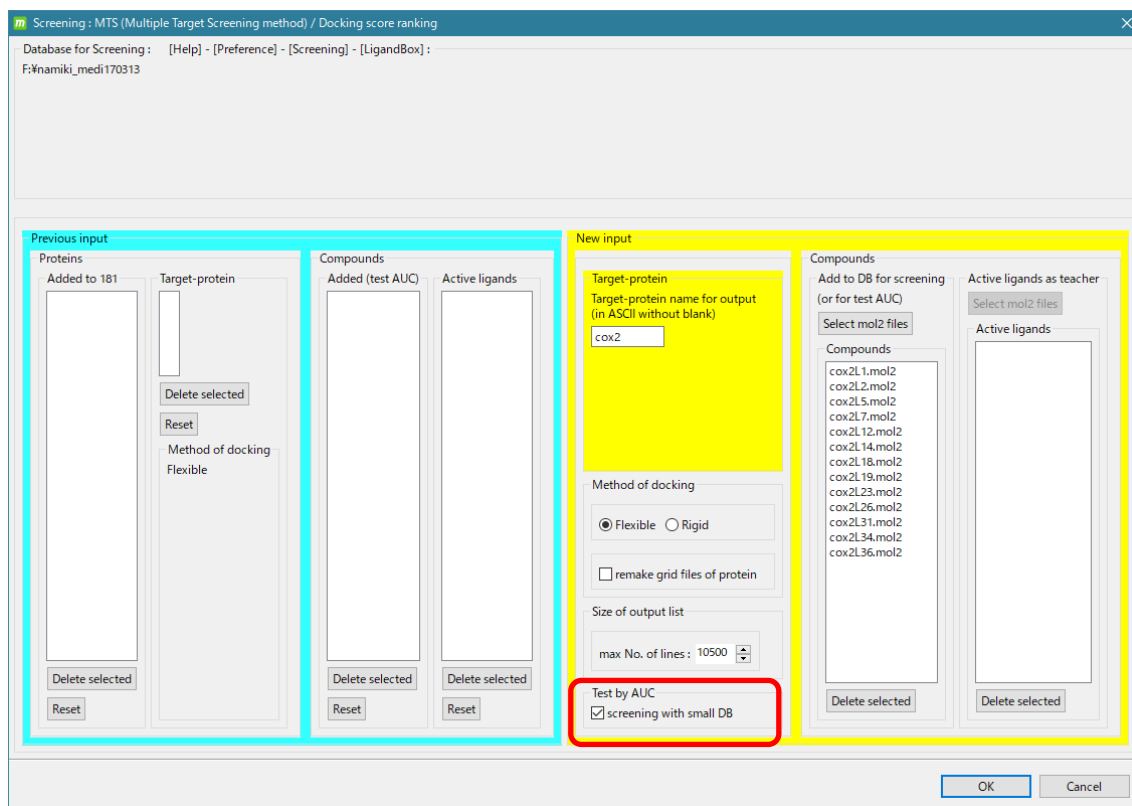
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L36.mol2





#### 1.10.4. LigandBox の検索対象化合物の入力

[screening with small DB] にチェックを入れると、LigandBox の検索対象化合物をテスト計算用の小さなグループ（2 万化合物以下）に限定することができます。



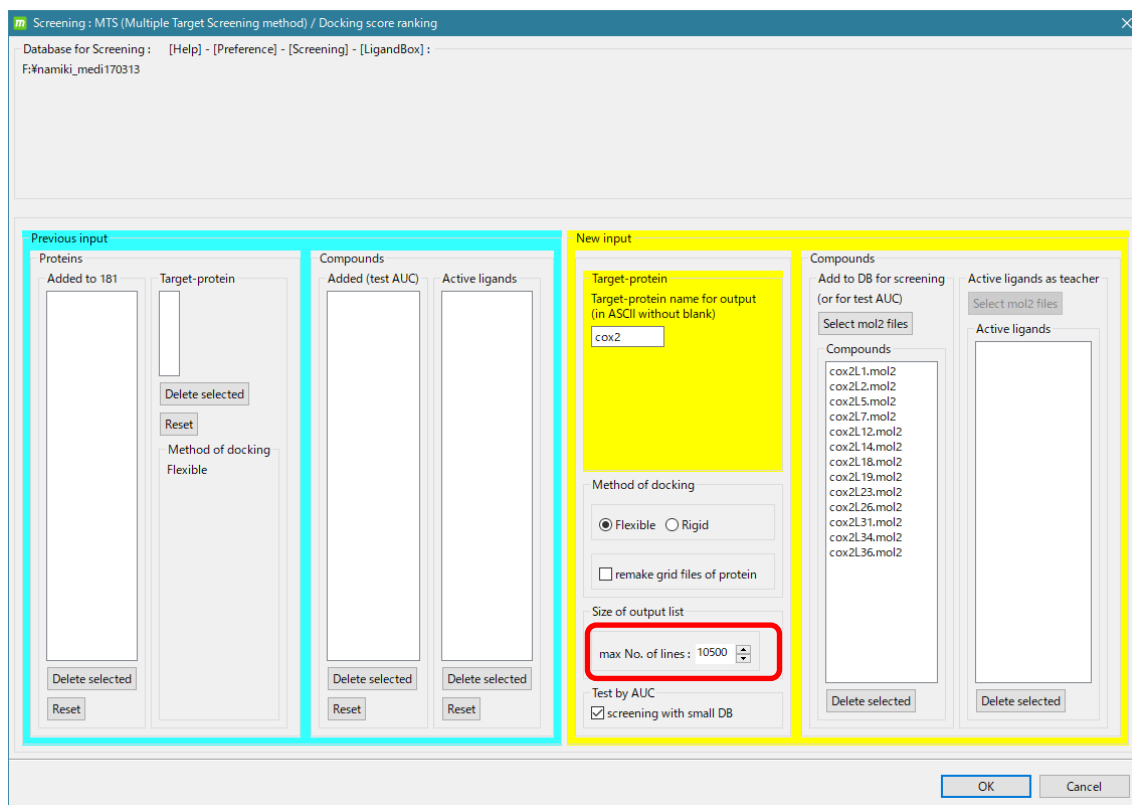
チェックしない場合はすべての化合物に対しスクリーニング計算を行います。

この例ではテスト時間短縮のため、[screening with small DB] をチェックします。

### 1.10.5. スクリーニング結果リストのサイズの入力

[max No. of lines] はスクリーニング計算後に表示される結果の数です。ヒット化合物をスコアが上位のものからこの数だけ表示します。

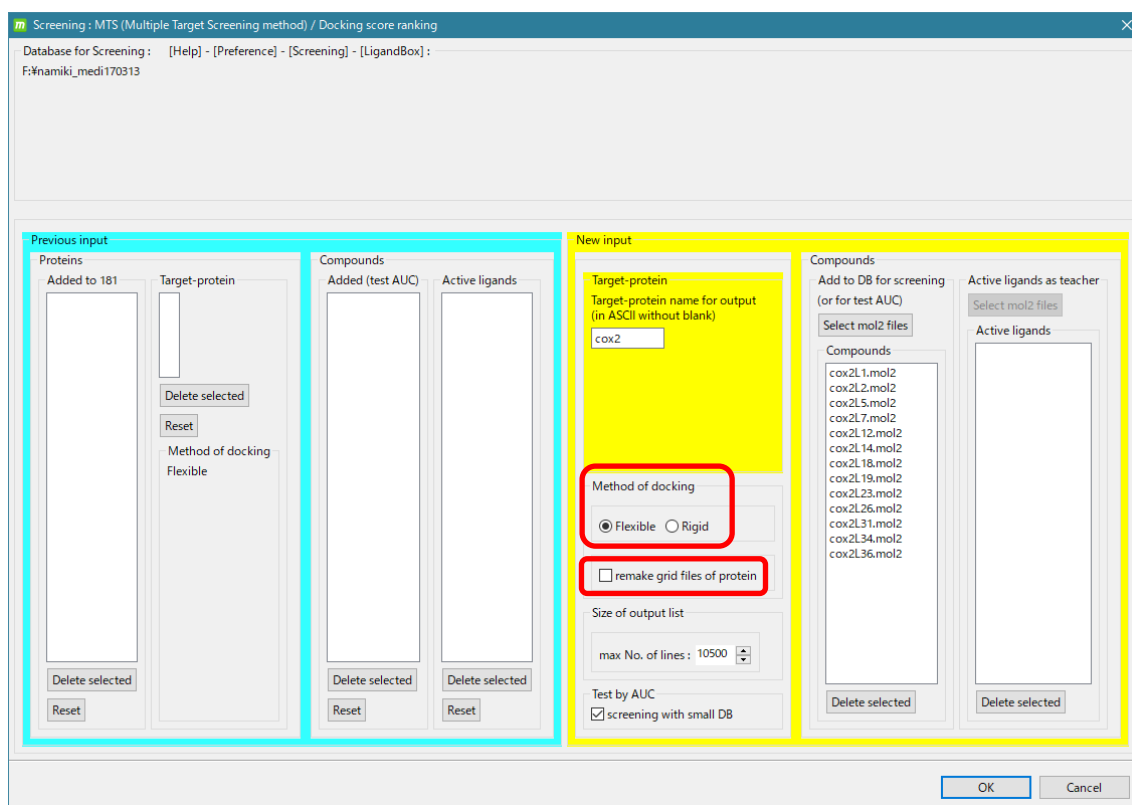
今回はデフォルト値 10,500 のまま実行します。



#### 1.10.6. ドッキング計算方法の入力

[Method of docking] では、追加で入力した化合物のドッキング計算を行う際、化合物の構造を **flexible**（多数の候補構造を生成する）で計算するか、**rigid**（入力した構造のまま剛体）で計算するかを選択します。通常は **flexible** を選択します。化合物の代表構造を入力し、リガンドを剛体としてドッキング計算する場合は **rigid** を選択してください。

[remake grid files of protein] では、MTS 法のスクリーニング計算で使用するリファレンスタンパク質のグリッドファイルについて、前回の計算で作成したグリッドファイルを使用するか、再計算してグリッドファイルを作成しなおすかを選択します。通常はデフォルトのままチェックせずに進めます。



リファレンスタンパク質：

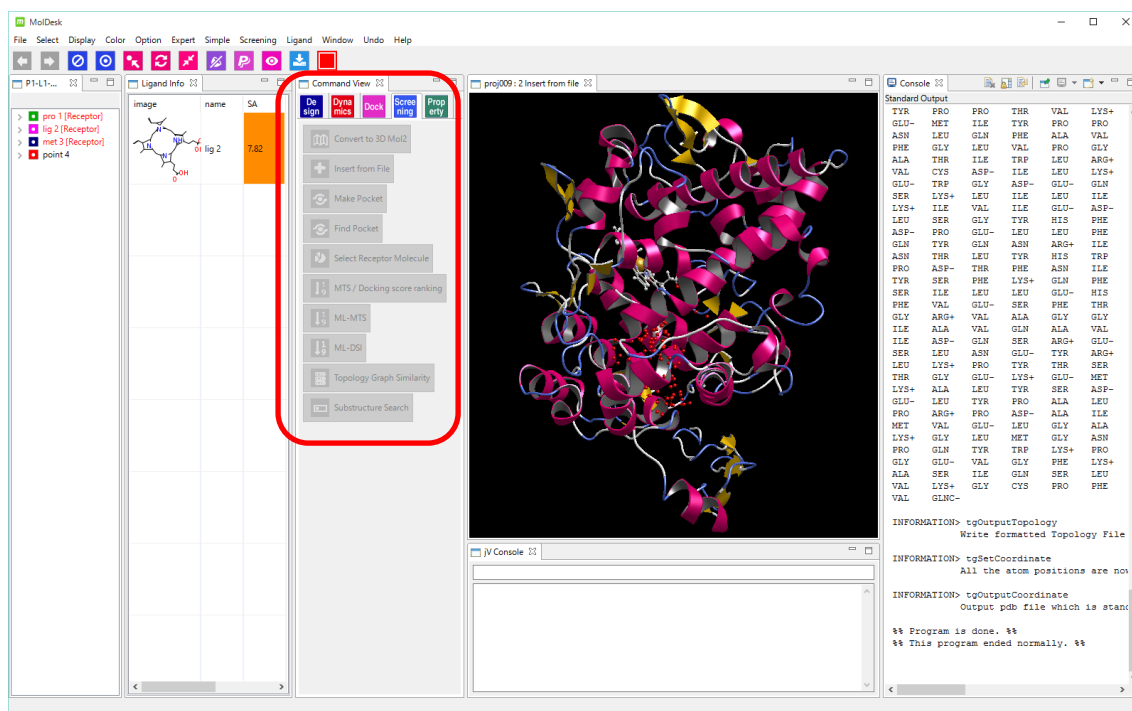
すべてのスクリーニング計算のドッキング計算で使用する 181 のリファレンスタンパク質を指します。詳細は myPresto のマニュアルを参照してください。

### 1.10.7. スクリーニング計算の開始

[OK]をクリックすると、スクリーニング計算が開始します。

この例では検索対象を約 2 万分子に限定しているため、通常の PC でも数時間で計算が終わります。

スクリーニング計算を開始するとコマンドボタンがグレーになります。コマンドボタンがグレーになっている間は計算中です。



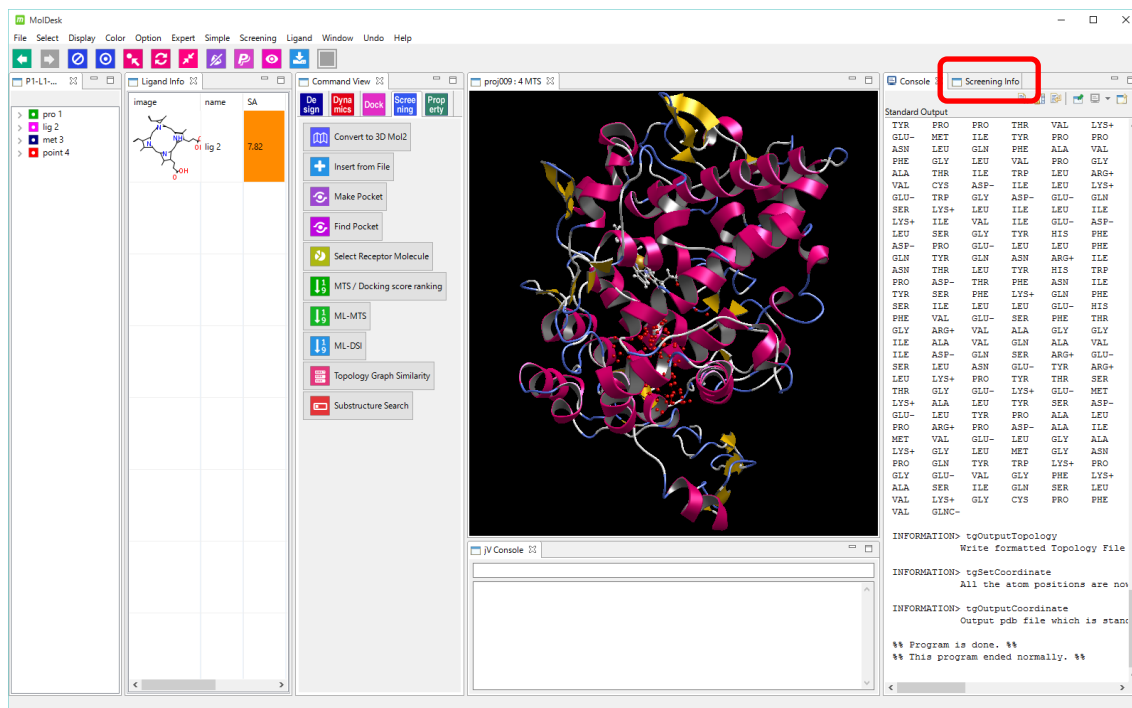
あらかじめ他のプロジェクトを開いていた場合、スクリーニング計算中でもそのプロジェクトを操作できます。ただしプロセッサの占有率によっては動作が極端に遅くなる場合があります。

スクリーニング計算で並列計算する際の並列数は、[Help] - [Preference] - [Screening] の **Thread number** で設定できます。詳細は「1.7 スクリーニング計算の並列数・メモリ量・時間」を参照してください。

### 1.10.8. スクリーニング計算結果の確認

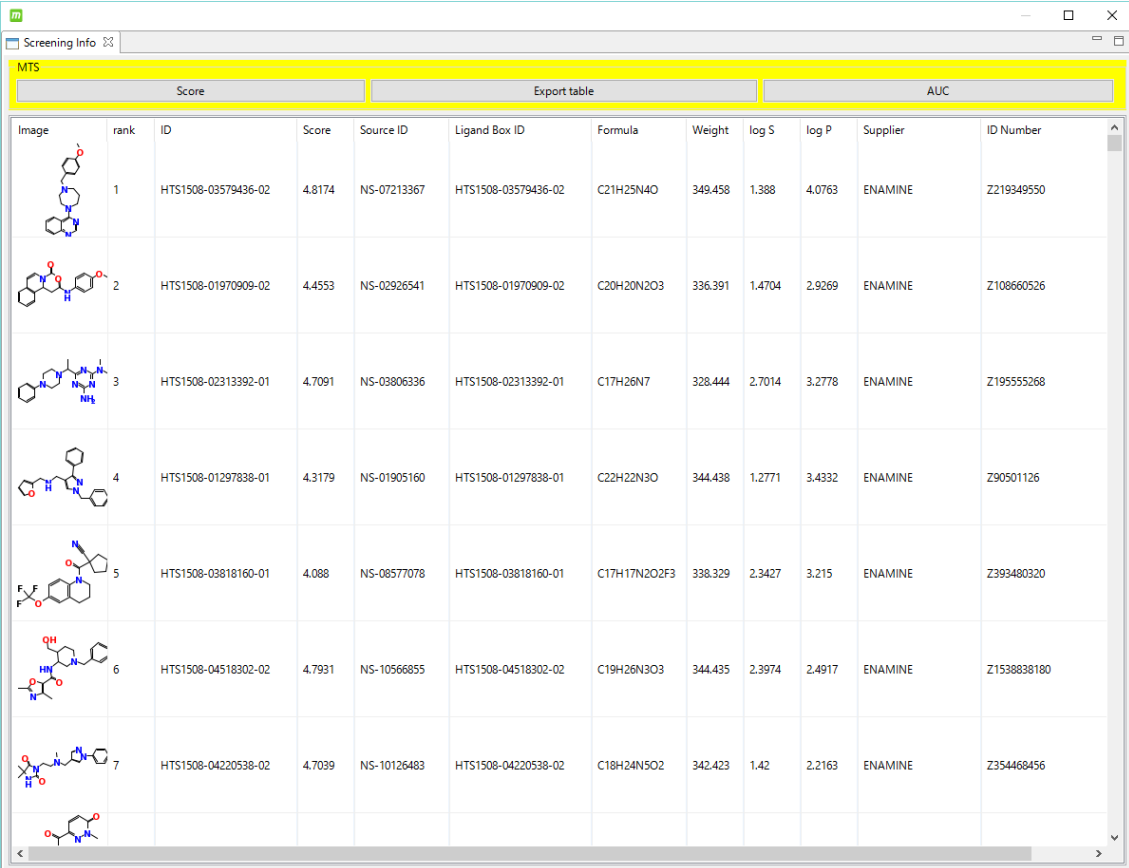
スクリーニング計算が終了するとコマンドボタンが元の表示に戻ります。

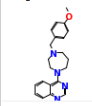






また、[Screening Info] のタブにスクリーニング結果がリスト表示されます。



- [Screening Info] タブが表示されていない場合は、[Windows] - [Screening Info] メニューをクリックして表示してください。

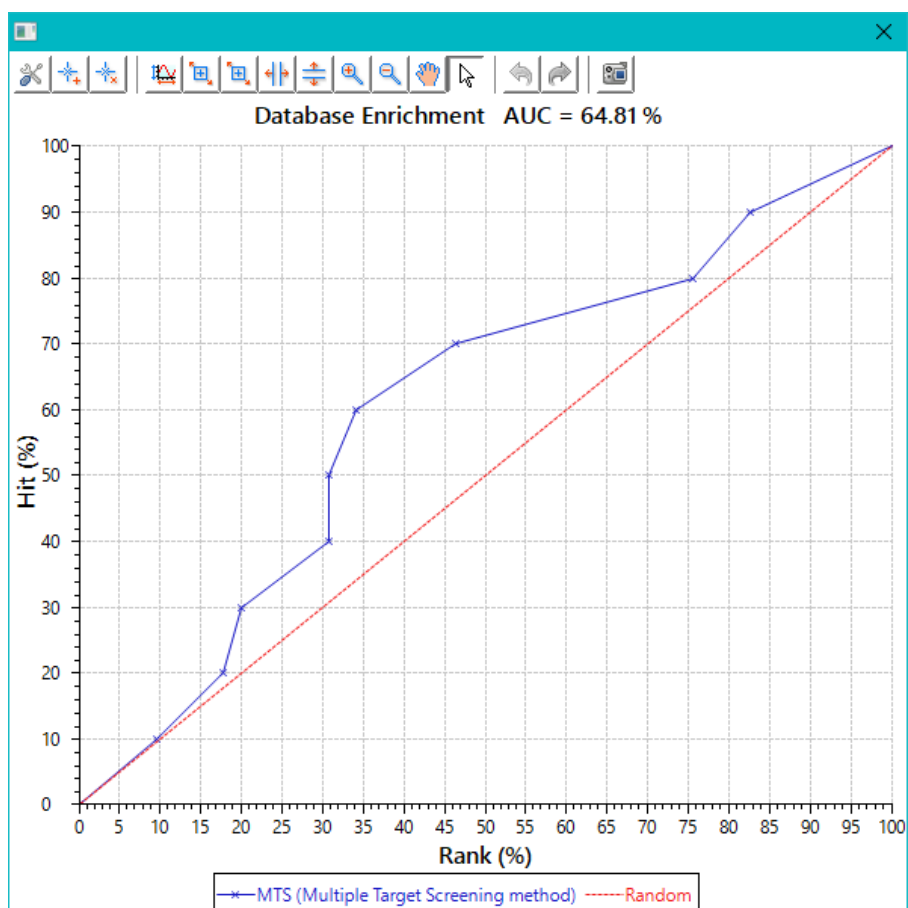
リストが見にくいときは、MolDesk のウィンドウを広げたり、[Screening Info] タブをマウスでドラッグしてウィンドウの外部に単独で表示させたりしてください。



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

デフォルトではランキング順に候補化合物が表示されています。リスト上部の各項目をクリックすることで、項目によるソートが可能です。

リスト上部にある [AUC] ボタンをクリックすると、データベースエンリッチメント曲線のグラフが表示されます。



このグラフは既知の活性リガンドを入力した場合に表示可能で、手法の精度を確認するために用います。

この例では AUC (Area under the curve) が 64.81 %であることが確認できました。  
※数値は実行ごとに少し変動します。

[Screening Info] リストの上部にある、[Score] ボタンをクリックすると、ドッキングスコア順のスクリーニング計算結果が表示されます。

リストのタイトルが、「MTS」から「Docking score ranking」に変わります。

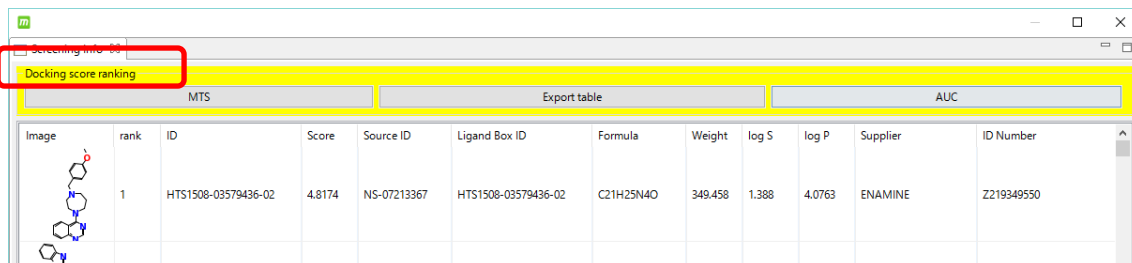
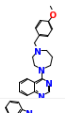
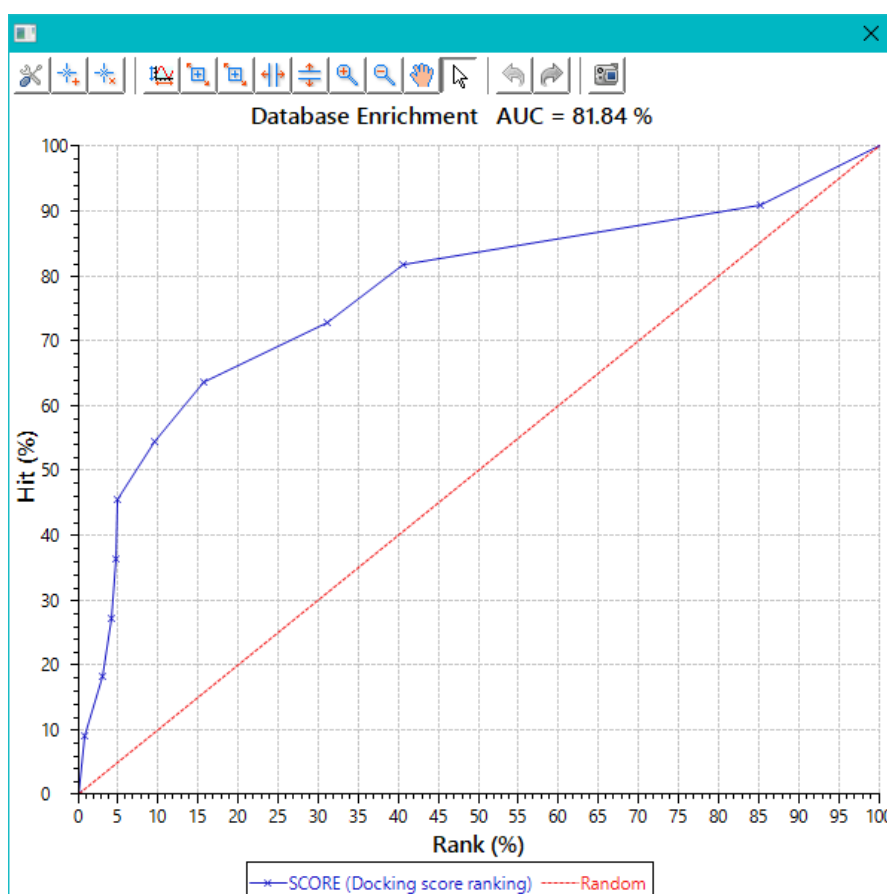


Image	rank	ID	Score	Source ID	Ligand Box ID	Formula	Weight	log S	log P	Supplier	ID Number
	1	HTS1508-03579436-02	4.8174	NS-07213367	HTS1508-03579436-02	C21H25N4O	349.458	1.388	4.0763	ENAMINE	Z219349550



[AUC] ボタンをクリックすると、ドッキングスコア順のスクリーニング計算の AUC を確認できます。この例では 81.84 %でした。

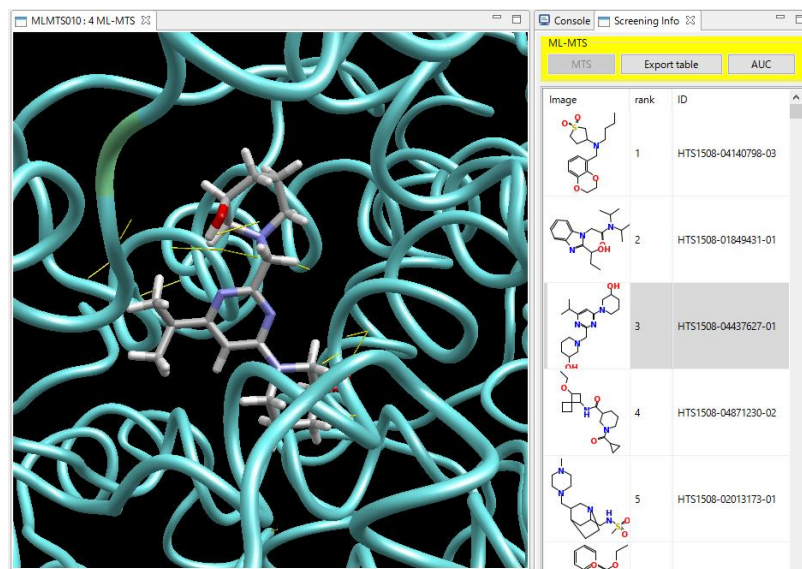
※数値は実行ごとに少し変動します。

一般的に、MTS 計算よりドッキングスコア順の計算の方が精度が良い傾向があります。



### 1.10.9. ドッキングポーズの確認

リストの化合物を選択すると、3D 画面でそのドッキングポーズを確認できます。



化合物の選択は ↑ ↓ キーで切り替えられ、連動して 3D 画面の表示も切り替わります。

### 1.10.10. スクリーニング計算結果のファイル出力

スクリーニング結果をファイル出力することができます。

リスト上部にある [Export table] ボタン をクリックすると、表示しているリストのデータを csv ファイル（カンマ区切り）または HTML ファイルで出力できます。

HTML ファイルで出力した場合は、ユーザ指定文字列.html\_image というフォルダが生成され、その中にすべての画像ファイルが id.png というファイル名で出力されます。出力順序はデフォルトのランキング順となります。この画像データ付き HTML ファイルは Excel に読み込むことが可能です。

## 1.11. ML-MTS 法の計算手順

ML-MTS 法のスクリーニング計算は、ターゲットタンパク質とポケットを準備するところまでは MTS 法 / ドッキングスコア順によるスクリーニング計算と手順は全く同じです。

### 1.11.1. プロジェクトの作成

MTS 法 / ドッキングスコア順と同様に、「1.10.1 プロジェクトの作成」を行います。

### 1.11.2. スクリーニング計算のデータ入力



[ML-MTS] をクリックします。

ワーニングダイアログが表示された場合は「1.10.2 スクリーニング計算のデータ入力」を参照して必要な作業を行ってください。

必要な作業がすべて行われている場合は、スクリーニング計算のデータ入力ダイアログが表示されます。

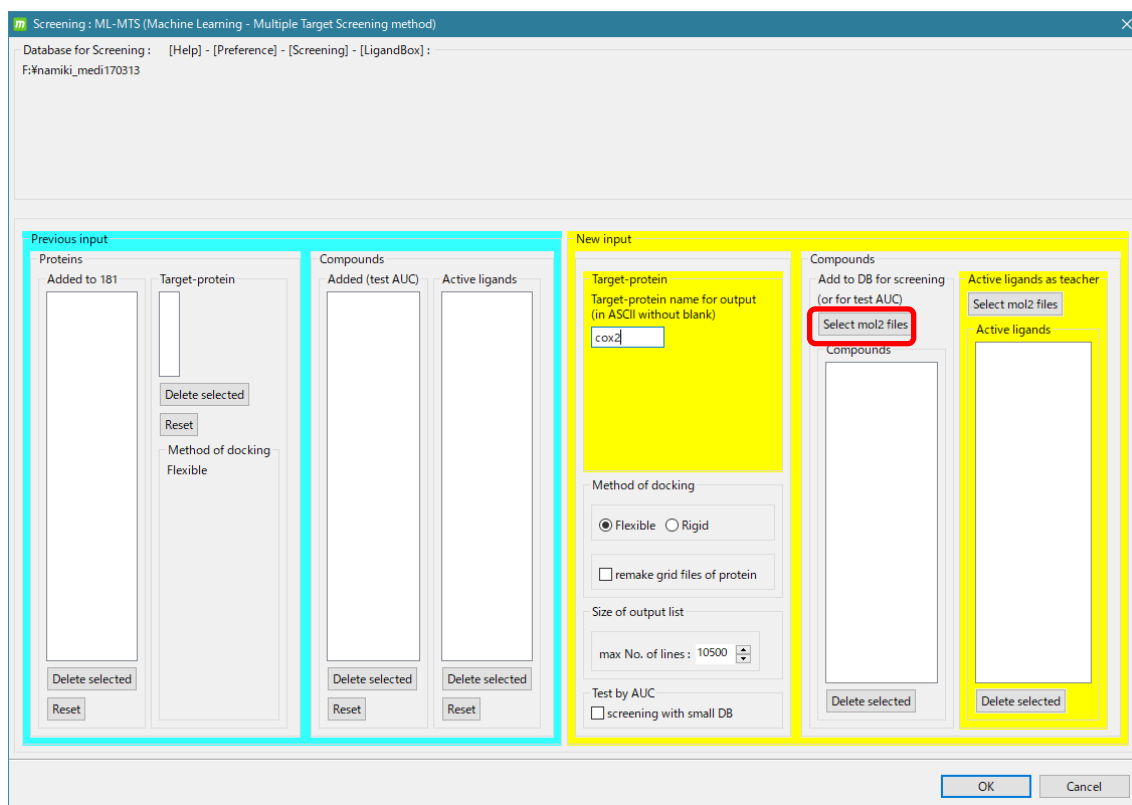
ML-MTS 法によるスクリーニング計算では、ターゲットタンパク質 ([Target protein]) の名前と既知活性リガンド ([Active ligands as teacher]) を入力する必要があります。ターゲットタンパク質の名前は任意の名称を空白を含まない英数字で入力します。

他は入力必須ではありません。

この例では、ターゲットタンパク質の名前に「cox2」と入力します。

### 1.11.3. mol2 ファイルによる化合物の追加

LigandBox に加えてスクリーニング対象とする化合物を mol2 ファイルで入力します。



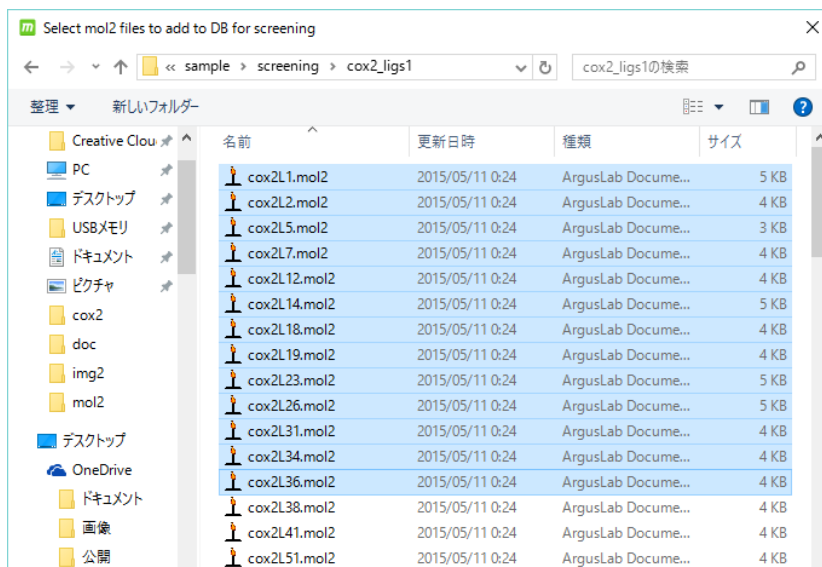
[Add to DB for screening] の [Select mol2 files] を選択して以下の 13 ファイルを選択し「開く」をクリックします。

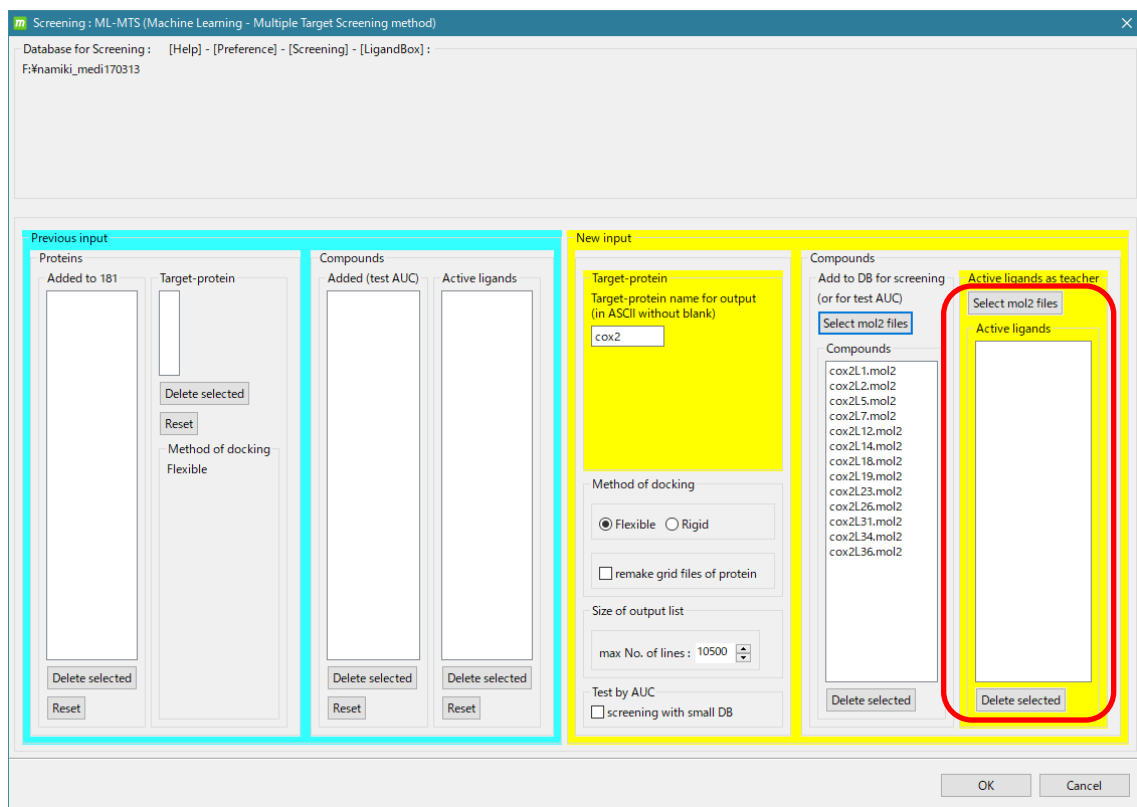
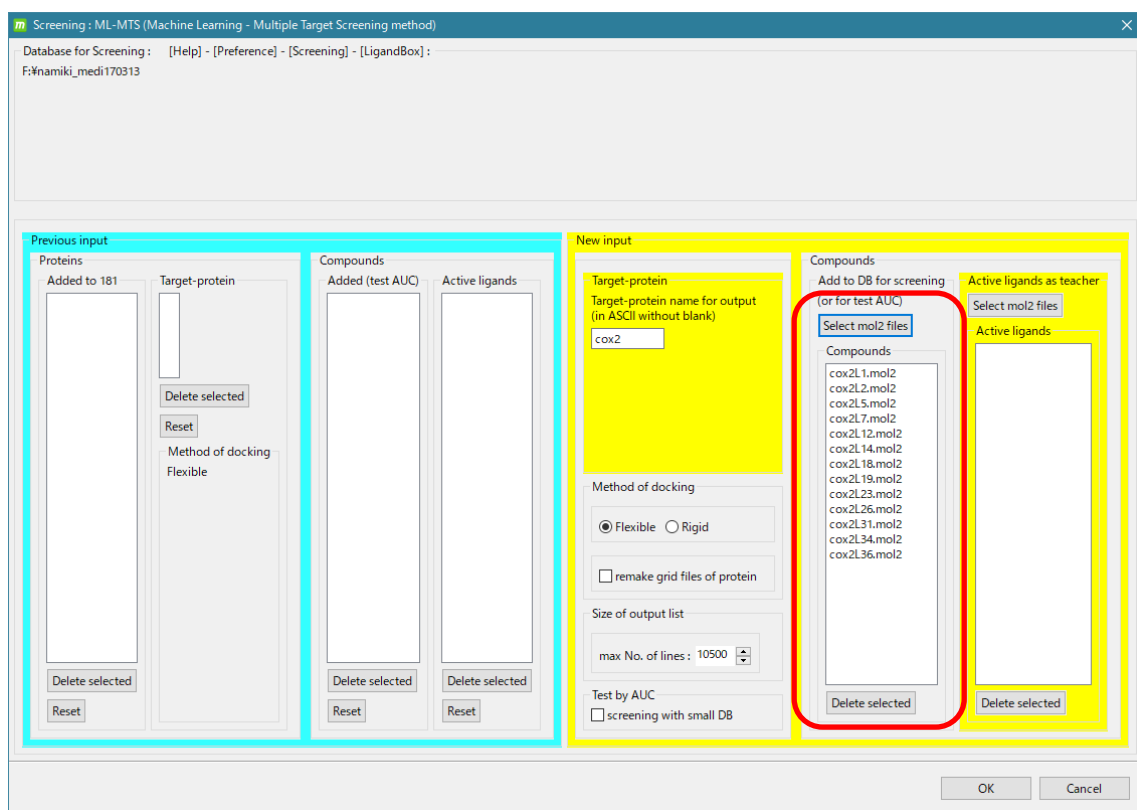
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L1.mol2

MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L2.mol2

MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L5.mol2

MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L7.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L12.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L14.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L18.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L19.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L23.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L26.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L31.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L34.mol2  
 MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L36.mol2

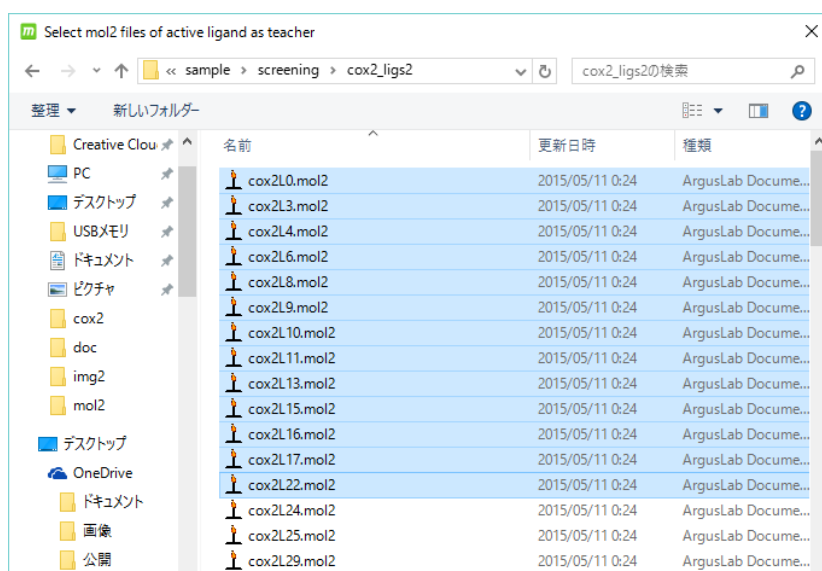




[Active ligands as teacher] の [Select mol2 files] を選択して以下の 13 ファイルを選択し

「開く」をクリックします。

MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L0.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L3.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L4.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L6.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L8.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L9.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L10.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L11.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L13.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L15.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L16.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L17.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L22.mol2



Screening : ML-MTS (Machine Learning - Multiple Target Screening method)

Database for Screening : [Help] - [Preference] - [Screening] - [LigandBox] :  
F:\namiki\_medi170313

---

**Previous input**

Proteins

Added to 181

Target-protein

Delete selected

Reset

Method of docking

Flexible

Delete selected

Reset

Compounds

Added (test AUC)

Active ligands

Delete selected

Reset

Delete selected

Reset

**New input**

Target-protein

Target-protein name for output (in ASCII without blank)

cox2

Method of docking

☒ Flexible ☐ Rigid

☐ remake grid files of protein

Size of output list

max No. of lines : 10500

Test by AUC

☐ screening with small DB

Compounds

Add to DB for screening (or for test AUC)

Select mol2 files

Compounds

cox2L1.mol2  
cox2L1.mol2  
cox2L2.mol2  
cox2L2.mol2  
cox2L5.mol2  
cox2L7.mol2  
cox2L12.mol2  
cox2L14.mol2  
cox2L18.mol2  
cox2L19.mol2  
cox2L23.mol2  
cox2L26.mol2  
cox2L31.mol2  
cox2L34.mol2  
cox2L36.mol2

Delete selected

Active ligands as teacher

Select mol2 files

Active ligands

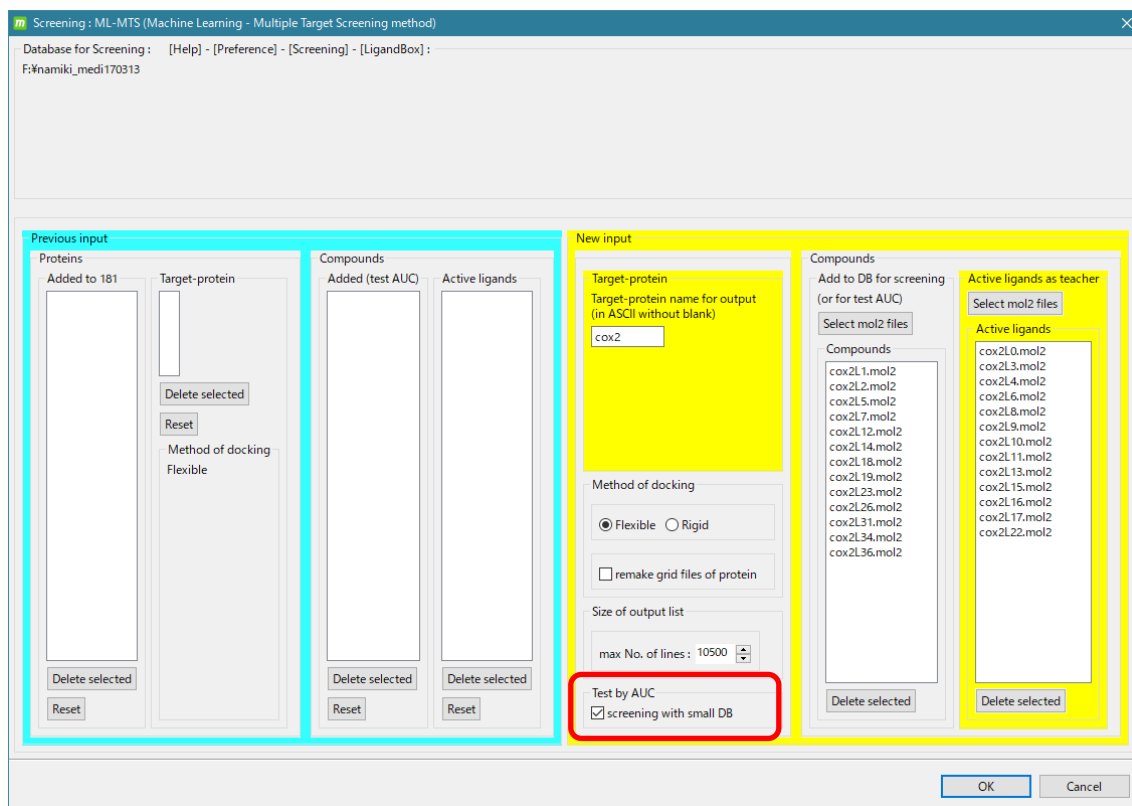
cox2L0.mol2  
cox2L3.mol2  
cox2L4.mol2  
cox2L6.mol2  
cox2L8.mol2  
cox2L9.mol2  
cox2L10.mol2  
cox2L11.mol2  
cox2L13.mol2  
cox2L15.mol2  
cox2L16.mol2  
cox2L17.mol2  
cox2L22.mol2

Delete selected

OK Cancel

#### 1.11.4. LigandBox の検索対象化合物の入力

[screening with small DB] にチェックを入れると、LigandBox の検索対象化合物をテスト計算用の小さなグループ（2 万化合物以下）に限定することができます。



チェックしない場合はすべての化合物に対しスクリーニング計算を行います。

この例ではテスト時間短縮のため、[screening with small DB] をチェックします。

[OK]をクリックすると、スクリーニング計算が開始します。

#### 1.11.5. スクリーニング計算結果の確認他

その他の設定項目、スクリーニング計算結果の確認方法、ファイル出力方法等は、MTS 法 / ドッキングスコア順によるスクリーニング計算の場合とほぼ同じです。

「1.10.5 スクリーニング結果リストのサイズの入力」～「1.10.10 スクリーニング計算結果のファイル出力」を参照してください。



## 1.12. ML-DSI 法の計算手順

ML-DSI 法は、リガンドベースのスクリーニング手法のため、ターゲットタンパク質は必要ありません。

### 1.12.1. プロジェクトの作成

[File] - [New Project] メニューで、空のプロジェクトを作成し、保存します。  
プロジェクトの保存方法は MolDesk Basic のマニュアルを参照してください。

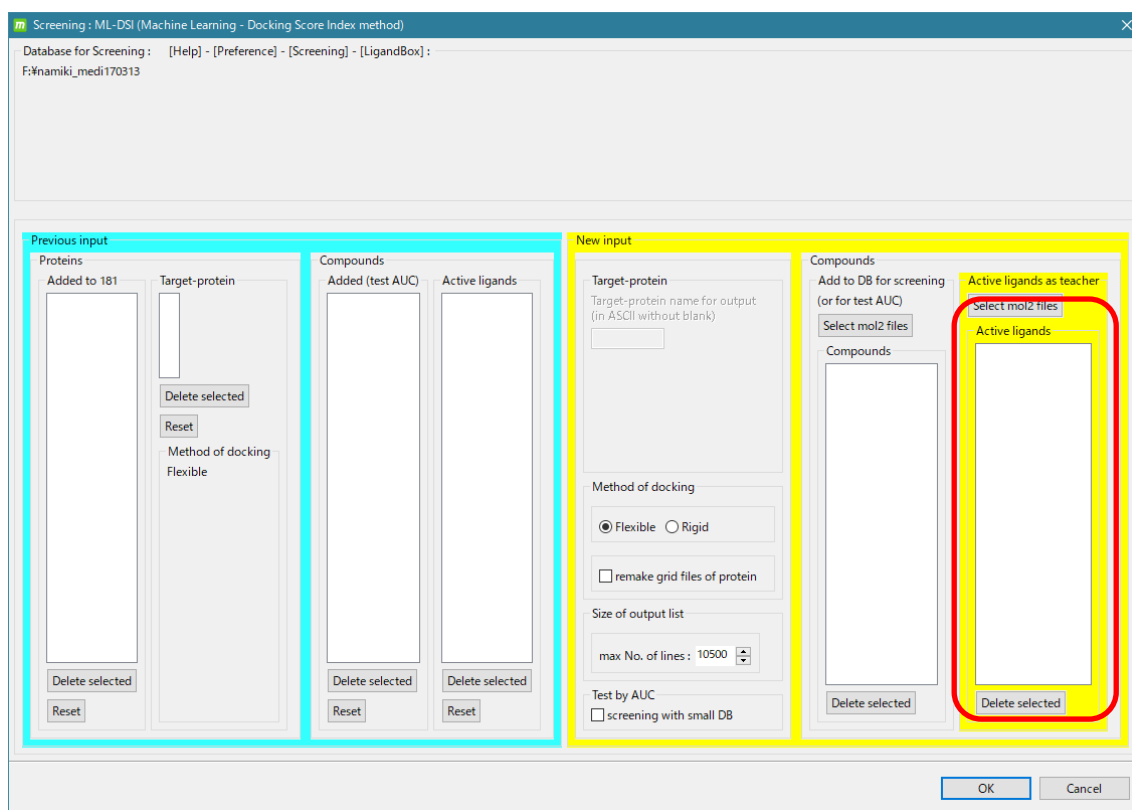
### 1.12.2. スクリーニング計算のデータ入力



[ML-DSI] をクリックします。

ワーニングダイアログが表示された場合は「1.10.2 スクリーニング計算のデータ入力」を参照して必要な作業を行ってください。

必要な作業がすべて行われている場合は、スクリーニング計算のデータ入力ダイアログが表示されます。



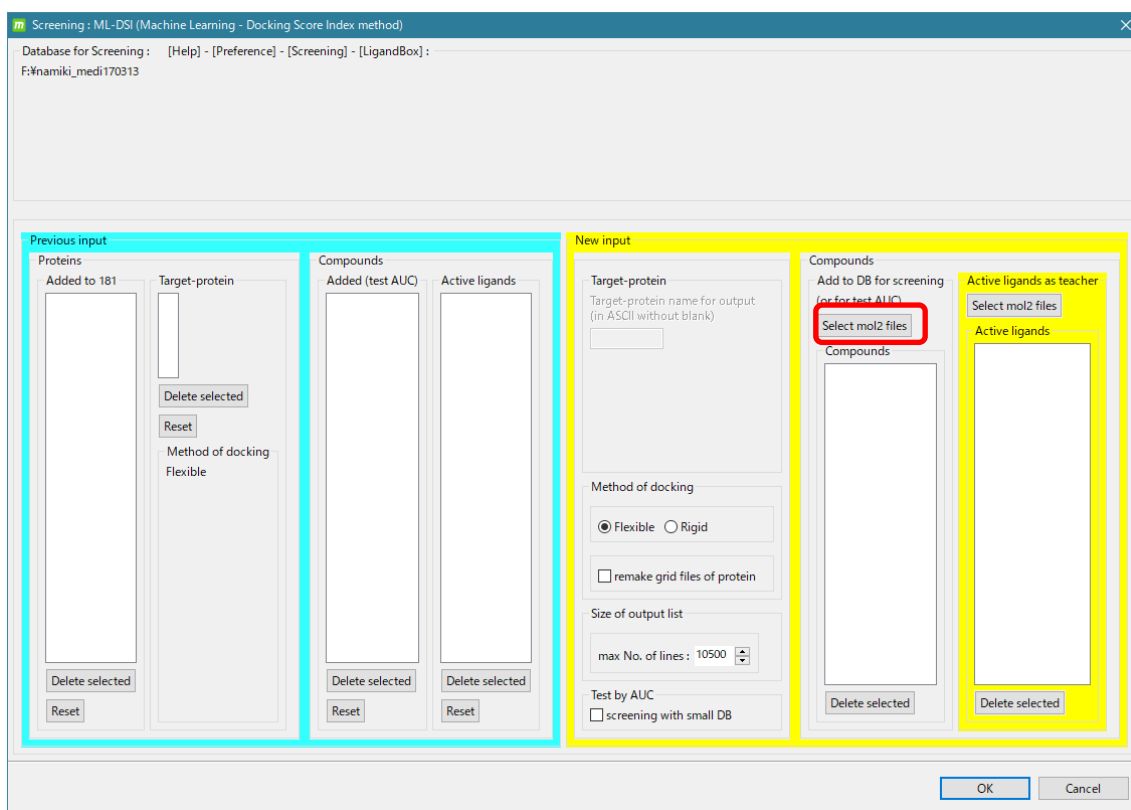
ML-DSI 法によるスクリーニング計算では、既知活性リガンド ([Active ligands as teacher]) を入力する必要があります。

他は入力必須ではありません。

ターゲットタンパク質 ([Target protein]) は ML-DSI 法によるスクリーニング計算では使用しないため、入力できないようになっています。

### 1.12.3. mol2 ファイルによる化合物の追加

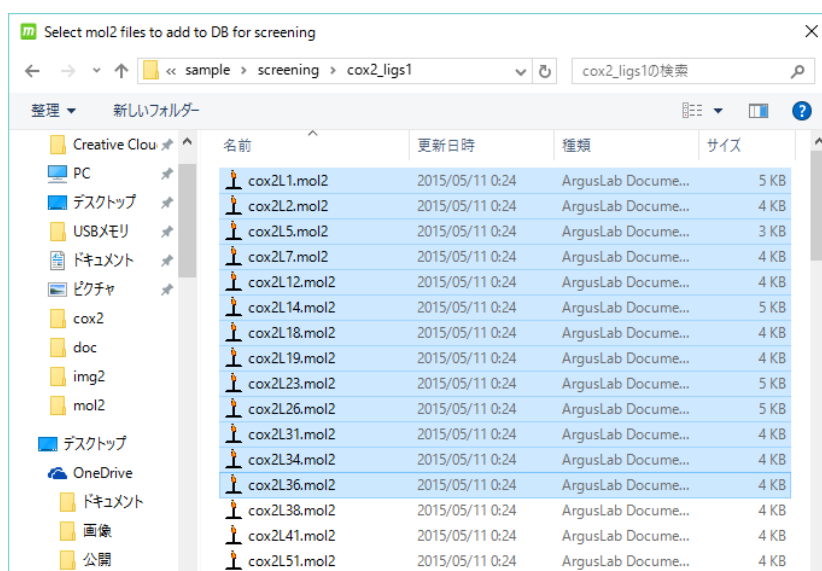
LigandBox に加えてスクリーニング対象とする化合物を mol2 ファイルで入力します。

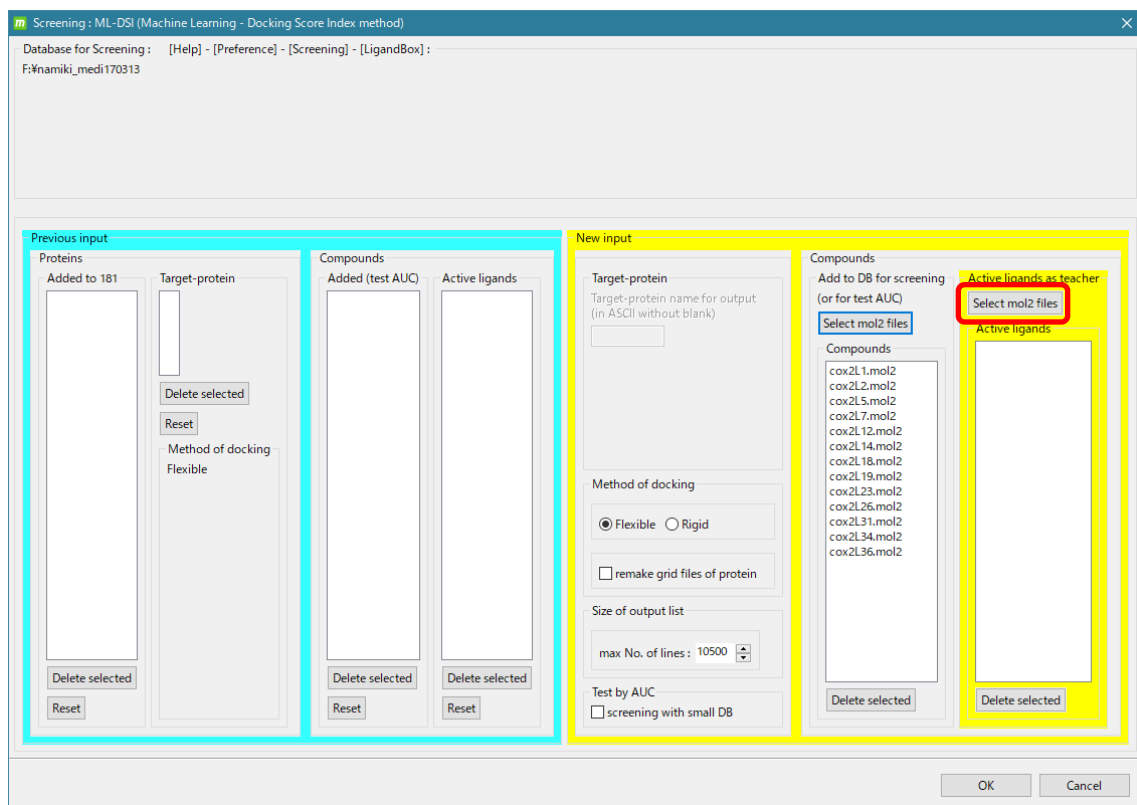
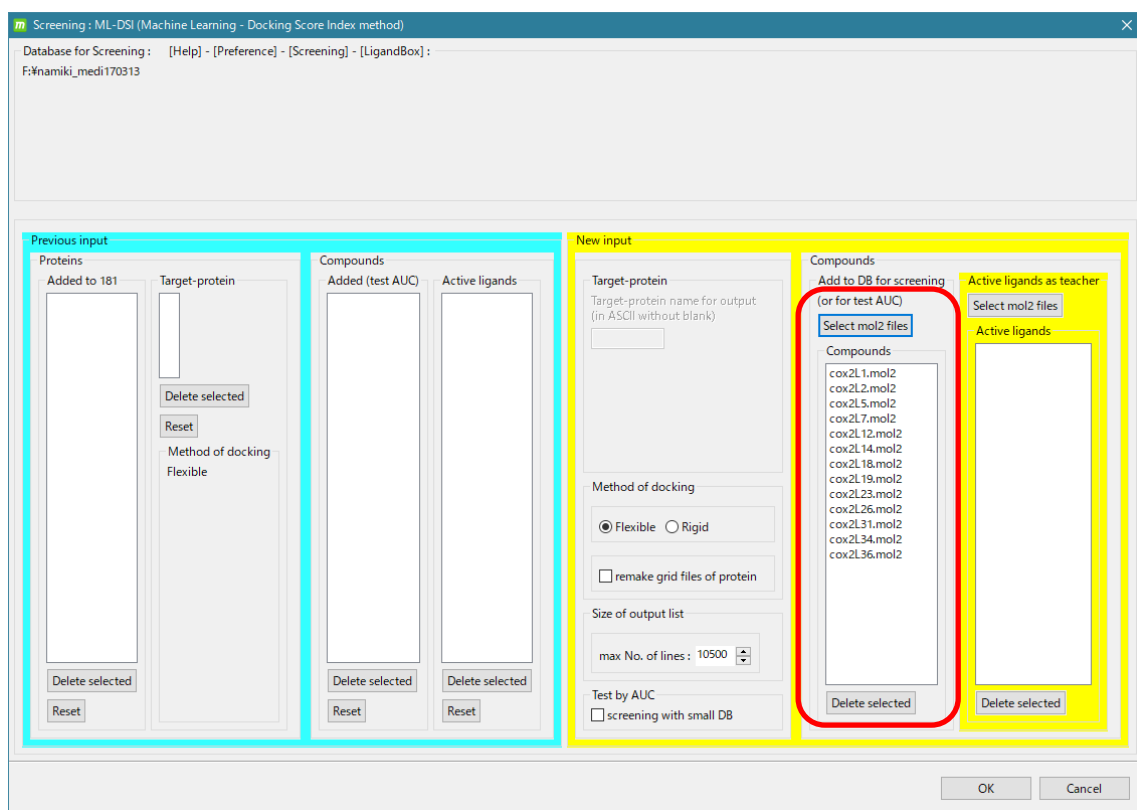


[Add to DB for screening] の [Select mol2 files] を選択して以下の 13 ファイルを選択し「開く」をクリックします。

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  
 (次のページにつづく)

MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L19.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L23.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L26.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L31.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L34.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L36.mol2

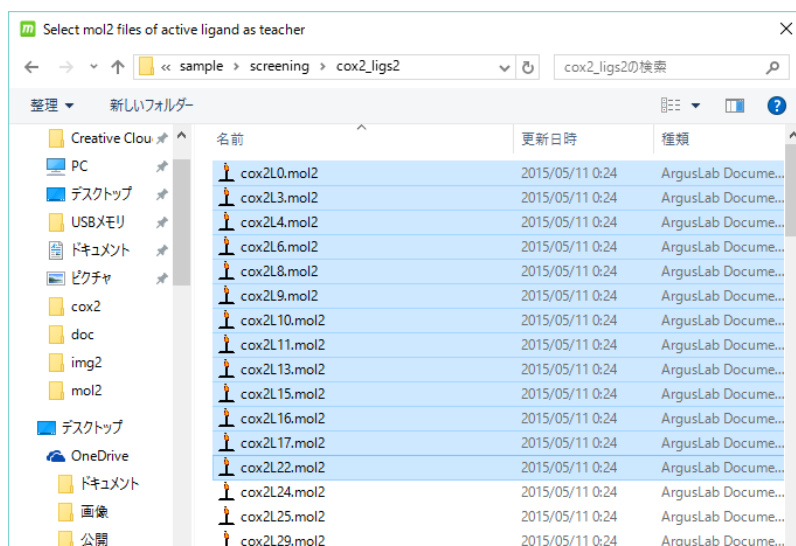




[Active ligands as teacher] の [Select mol2 files] を選択して以下の 13 ファイルを選択し

「開く」をクリックします。

MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L0.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L3.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L4.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L6.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L8.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L9.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L10.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L11.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L13.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L15.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L16.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L17.mol2  
MolDesk Screening -> sample -> screening -> cox2\_lig2 -> cox2L22.mol2



Screening : ML-DSI (Machine Learning - Docking Score Index method)

Database for Screening : [Help] - [Preference] - [Screening] - [LigandBox] :  
F:\namiki\_medi170313

---

**Previous input**

Proteins

Added to 181

Target-protein

Delete selected

Reset

Method of docking

Flexible

Delete selected

Reset

Compounds

Added (test AUC)

Active ligands

Delete selected

Reset

**New input**

Target-protein

Target-protein name for output  
(in ASCII without blank)

Method of docking

☒ Flexible ☐ Rigid

☐ remake grid files of protein

Size of output list

max No. of lines : 10500

Test by AUC

☐ screening with small DB

Compounds

Add to DB for screening  
(or for test AUC)

Select mol2 files

Compounds

cox2L1.mol2  
cox2L2.mol2  
cox2L3.mol2  
cox2L4.mol2  
cox2L5.mol2  
cox2L6.mol2  
cox2L7.mol2  
cox2L8.mol2  
cox2L9.mol2  
cox2L10.mol2  
cox2L11.mol2  
cox2L12.mol2  
cox2L13.mol2  
cox2L14.mol2  
cox2L15.mol2  
cox2L16.mol2  
cox2L17.mol2  
cox2L18.mol2  
cox2L19.mol2  
cox2L20.mol2  
cox2L21.mol2  
cox2L22.mol2  
cox2L23.mol2  
cox2L24.mol2  
cox2L25.mol2  
cox2L26.mol2  
cox2L27.mol2  
cox2L28.mol2  
cox2L29.mol2  
cox2L30.mol2  
cox2L31.mol2  
cox2L32.mol2  
cox2L33.mol2  
cox2L34.mol2  
cox2L35.mol2  
cox2L36.mol2

Delete selected

Active ligands as teacher

Select mol2 files

Active ligands

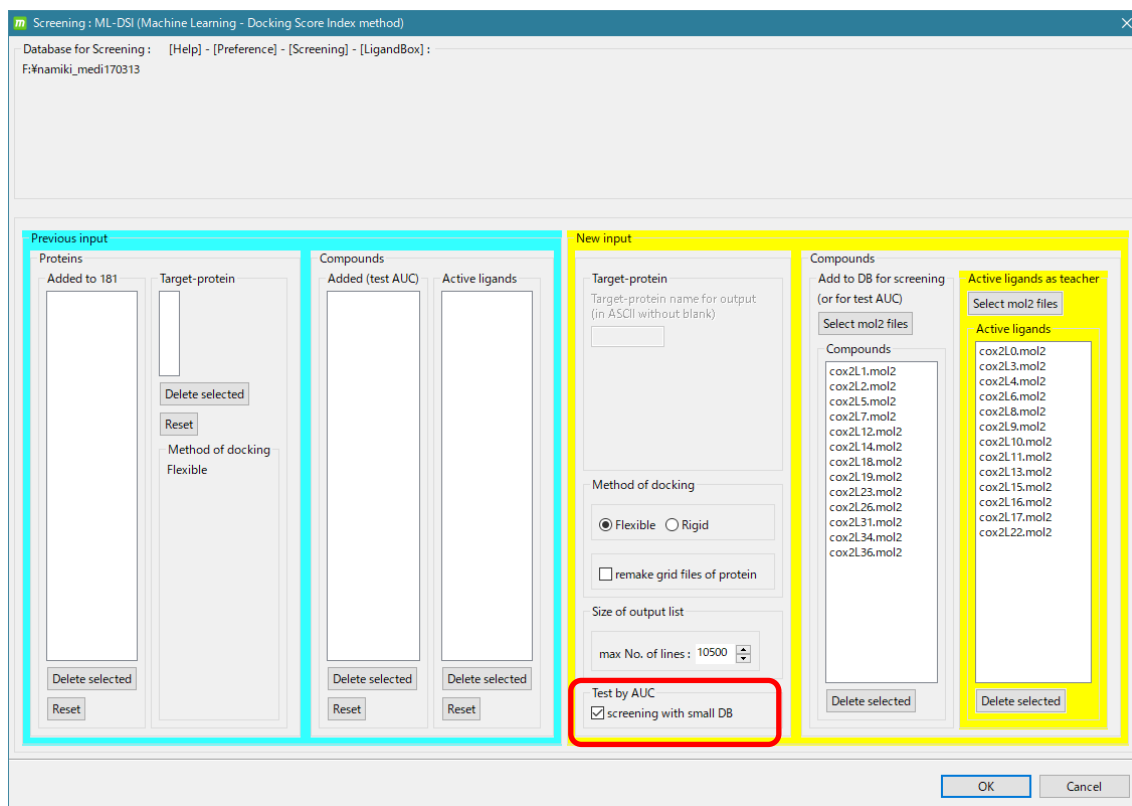
cox2L0.mol2  
cox2L1.mol2  
cox2L2.mol2  
cox2L3.mol2  
cox2L4.mol2  
cox2L5.mol2  
cox2L6.mol2  
cox2L7.mol2  
cox2L8.mol2  
cox2L9.mol2  
cox2L10.mol2  
cox2L11.mol2  
cox2L12.mol2  
cox2L13.mol2  
cox2L14.mol2  
cox2L15.mol2  
cox2L16.mol2  
cox2L17.mol2  
cox2L18.mol2  
cox2L19.mol2  
cox2L20.mol2  
cox2L21.mol2  
cox2L22.mol2  
cox2L23.mol2  
cox2L24.mol2  
cox2L25.mol2  
cox2L26.mol2  
cox2L27.mol2  
cox2L28.mol2  
cox2L29.mol2  
cox2L30.mol2  
cox2L31.mol2  
cox2L32.mol2  
cox2L33.mol2  
cox2L34.mol2  
cox2L35.mol2  
cox2L36.mol2

Delete selected

OK Cancel

#### 1.12.4. LigandBox の検索対象化合物の入力

[screening with small DB] にチェックを入れると、LigandBox の検索対象化合物をテスト計算用の小さなグループ（2 万化合物以下）に限定することができます。



チェックしない場合はすべての化合物に対しスクリーニング計算を行います。

この例ではテスト時間短縮のため、[screening with small DB] をチェックします。

[OK]をクリックすると、スクリーニング計算が開始します。

#### 1.12.5. スクリーニング計算結果の確認

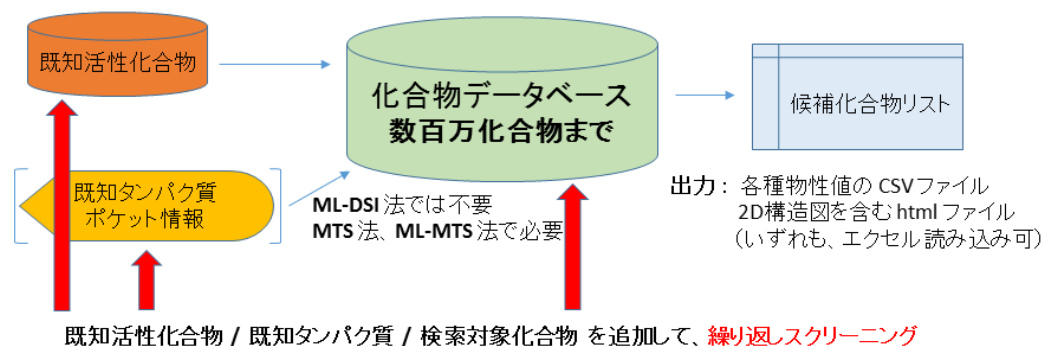
その他の設定項目、スクリーニング計算結果の確認方法、ファイル出力方法等は、MTS 法 / ドッキングスコア順によるスクリーニング計算の場合とほぼ同じです。

「1.10.5 スクリーニング結果リストのサイズの入力」～「1.10.10 スクリーニング計算結果のファイル出力」を参照してください。



### 1.13. 繰り返しスクリーニング計算 1

#### MolDesk Screening データフロー図



スクリーニング計算を繰り返して実行することができます。計算手法は異なってもかまいません。





この例では、受容体の活性リガンドが当初は不明で後に判明したケースを想定し、MTS 法 / ドッキングスコア順によるスクリーニング計算を実行した後に、ML-MTS 法によるスクリーニング計算を実行する手順を説明します。

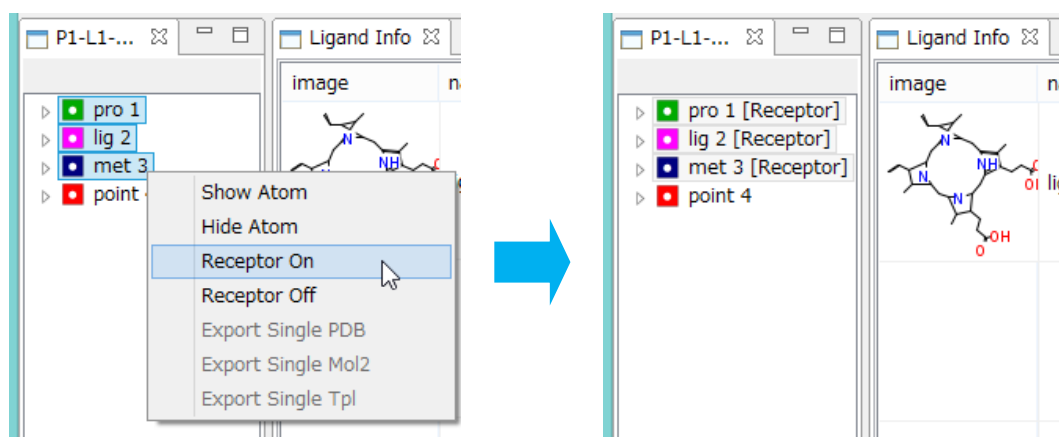
この他、ターゲットタンパク質が後に判明したために、ML-DSI 法によるスクリーニング計算を実行した後に、ML-MTS 法によるスクリーニング計算を実行する場合等も想定できます。

### 1.13.1. 受容体分子の選択

「1.10.1 プロジェクトの作成」で作成したプロジェクト（proj009）を使用します。

ドッキング計算の受容体を指定します。

ツリー表示画面で  pro 1、 lig 2、 met 3 を Ctrl + クリックで選択し、右クリックから[Receptor On] を選択します。（lig2、met3 はポケットと関係ない場所にあるので、 pro 1 だけを選択しても OK です）。受容体は、ポケットの空間を開けるように選択します。



受容体選択方法の詳細は MolDesk Basic のマニュアルを参照してください。

- 受容体の設定はスクリーニング計算を実行するごとに解除されますので、毎回設定してください。

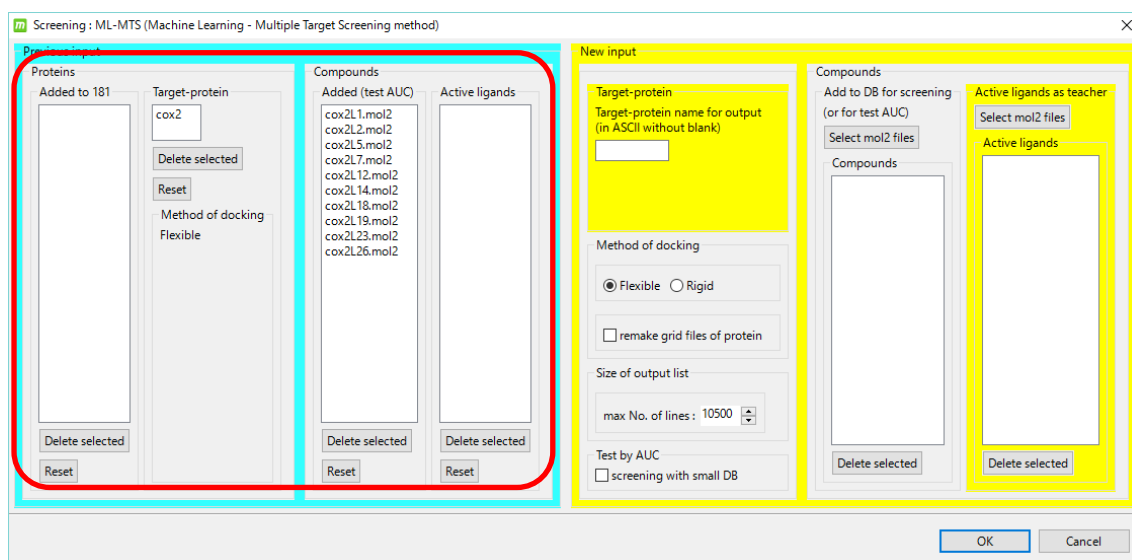
### 1.13.1. スクリーニング計算のデータ入力



[ML-MTS] をクリックします。

ワーニングダイアログが表示された場合は「1.10.2 スクリーニング計算のデータ入力」を参照して必要な作業を行ってください。

必要な作業がすべて行われている場合は、スクリーニング計算のデータ入力ダイアログが表示されます。



ダイアログ左の青背景の領域に、前回の計算の入力内容が表示されます。

前回入力した [Target-protein] や、前回入力した mol2 ファイルや、リガンド構造を Flexible で計算したことが確認できます

ここに mol2 ファイルを追加し、前回のスクリーニング計算に化合物を加えて、繰り返しスクリーニング計算を行います。

[Target-protein] にターゲットタンパク質の名前を入力します。  
ここでは前回と同じく **cox2** と入力します。

Screening: ML-MTS (Machine Learning - Multiple Target Screening method)

**Previous input**

Proteins

Added to 181

Target-protein

cox2

Delete selected

Reset

Method of docking

Flexible

Compounds

Added (test AUC)

cox2L1.mol2  
cox2L2.mol2  
cox2L5.mol2  
cox2L7.mol2  
cox2L12.mol2  
cox2L14.mol2  
cox2L18.mol2  
cox2L19.mol2  
cox2L23.mol2  
cox2L26.mol2

Delete selected

Reset

Active ligands

Delete selected

Reset

**New input**

Target-protein

Target-protein name for output  
(in ASCII without blank)

cox2

Method of docking

☒ Flexible ☐ Rigid

☐ remake grid files of protein

Size of output list

max No. of lines: 10500

Test by AUC

☒ screening with small DB

**Compounds**

Add to DB for screening  
(or for test AUC)

Select mol2 files

Compounds

Delete selected

**Active ligands as teacher**

Select mol2 files

Active ligands

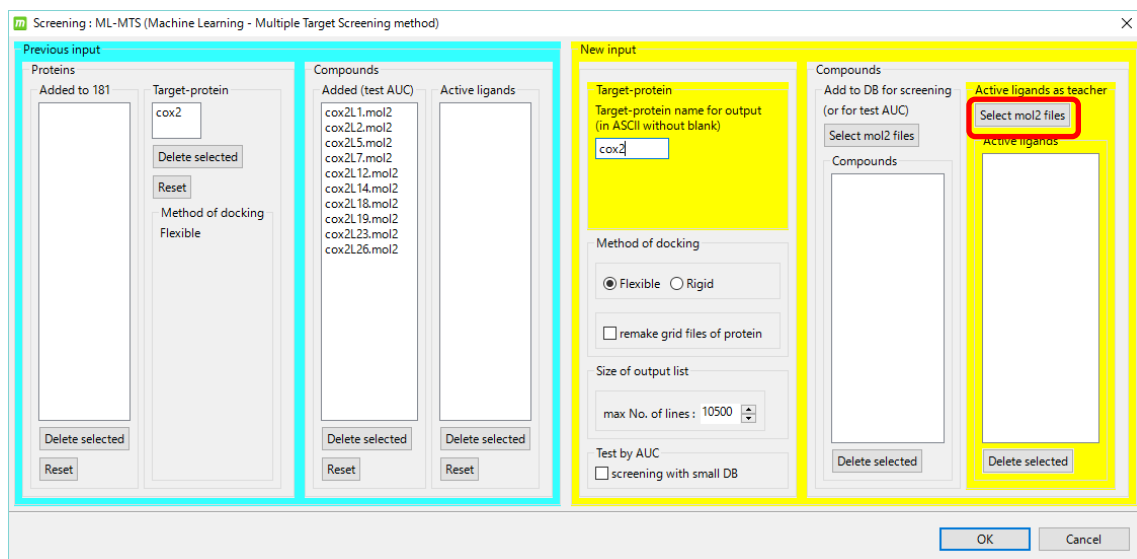
Delete selected

Reset

OK Cancel

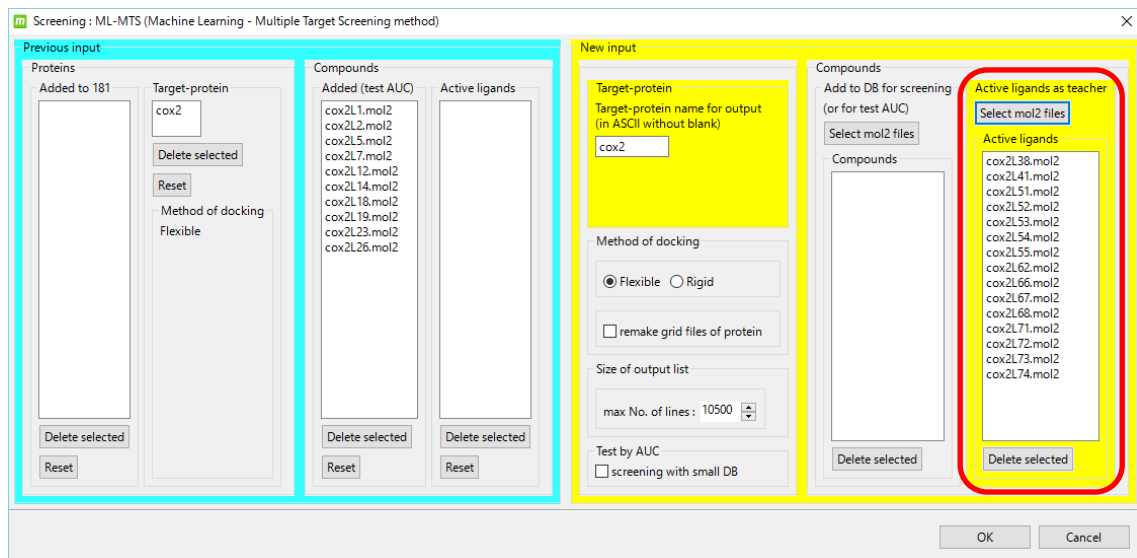
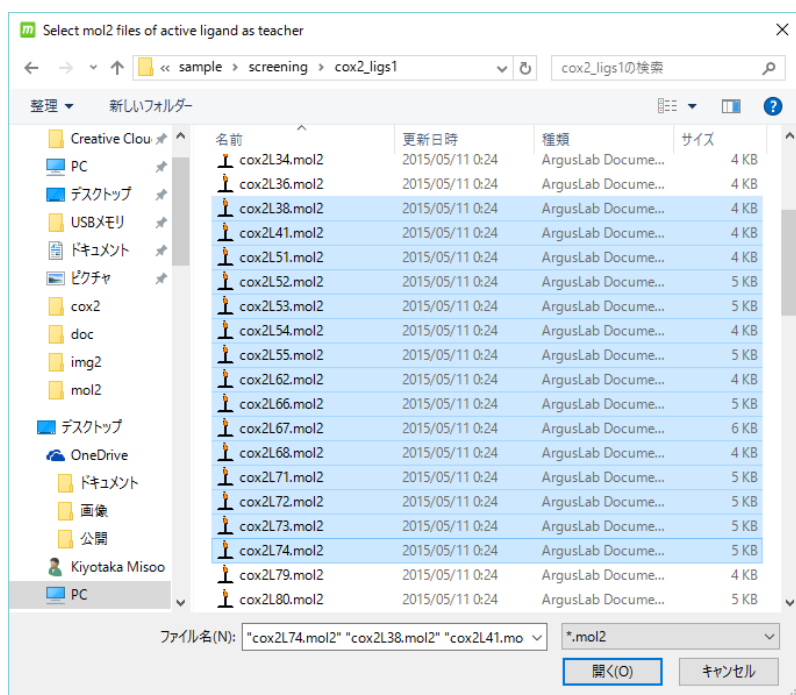
別のターゲットタンパク質を追加することも可能です。別のターゲットタンパク質を追加する場合の実行例は「1.14 繰り返しスクリーニング計算 2」を参照してください。

### 1.13.1. mol2 ファイルによる化合物の追加



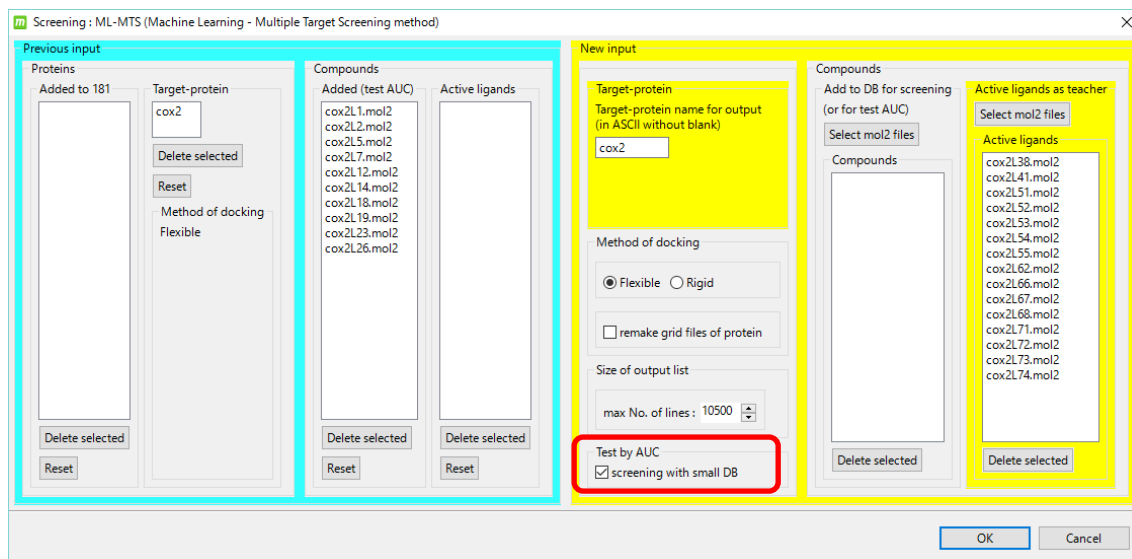
[Active ligands as teacher] の [Select mol2 files] を選択して以下の 15 ファイルを選択し「開く」をクリックします。

MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L38.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L41.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L51.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L52.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L53.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L54.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L55.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L62.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L66.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L67.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L68.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L71.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L72.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L73.mol2  
MolDesk Screening -> sample -> screening -> cox2\_ligs1 -> cox2L74.mol2



この例ではテスト時間短縮のため、[screening with small DB] をチェックします。  
詳細は「1.12.4 LigandBox の検索対象化合物の入力」を参照してください。

[OK]をクリックすると、ML-MTS 法によるスクリーニング計算が開始します。



1 回目の計算結果は再計算せずに再利用できるようになっているため、前回よりは少ない時間で計算が完了します。

### 1.13.2. スクリーニング計算結果の確認

その他の設定項目、スクリーニング計算結果の確認方法、ファイル出力方法等は、MTS 法 / ドッキングスコア順によるスクリーニング計算の場合とほぼ同じです。

「1.10.5 スクリーニング結果リストのサイズの入力」～「1.10.10 スクリーニング計算結果のファイル出力」を参照してください。

今回の ML-MTS 法によるスクリーニング計算では既知活性化合物を入力して機械学習したため、前回の MTS 法 (AUC=64.81%) およびドッキングスコア順 (AUC=81.84%) によるスクリーニング計算よりも格段に AUC が良くなります。

## 1.14. 繰り返しスクリーニング計算 2

### MolDesk Screening データフロー図



「1.13 繰り返しスクリーニング計算 1」にさらにターゲットタンパク質を追加し、ML-MTS 法によるスクリーニング計算を行います。

- 過去に計算したターゲットタンパク質についてより精緻な構造が判明し、再計算するケースを想定しています。

### 1.14.1. 受容体分子の選択

「1.13 繰り返しスクリーニング計算 1」のプロジェクトを使用し、「1.13.1 受容体分子の選択」の手順で受容体分子の選択を行います。

- 受容体の設定はスクリーニング計算を実行するごとに解除されますので、毎回設定してください。



### 1.14.2. スクリーニング計算のデータ入力



[ML-MTS] をクリックします。

ワーニングダイアログが表示された場合は「1.10.2 スクリーニング計算のデータ入力」を参照して必要な作業を行ってください。

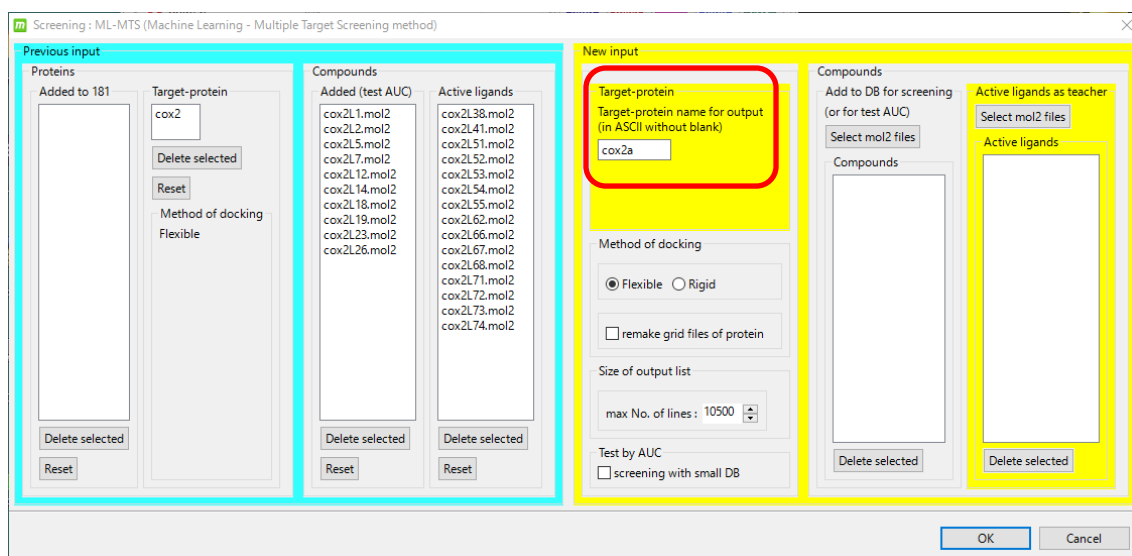
必要な作業がすべて行われている場合は、スクリーニング計算のデータ入力ダイアログが表示されます。

ダイアログ左の青背景の領域に、前回の計算の入力内容が表示されます。

「1.13 繰り返しスクリーニング計算 1」で入力した mol2 ファイルが[Active ligands] で確認できます

[Target-protein] にターゲットタンパク質の名前を入力します。

ここでは前回・前々回と異なる名前を指定することにし **cox2a** と入力します。




- ここでは説明を省略しましたが、別のターゲットタンパク質を追加する場合は、そのプロジェクトのモデリングが必要です。具体的には、そのターゲットタンパク質を入力し、新しいポケット生成を行い、受容体の選択を行います。
- 繰り返しスクリーニングで追加できるターゲットタンパク質の個数は繰り返し 1 回につき 1 つです。複数のターゲットタンパク質を追加したい場合は 1 つずつ追加を行ってください。
- 繰り返しスクリーニング計算は何回でも実行可能です。化合物やタンパク質のデータを蓄積して計算することができます。

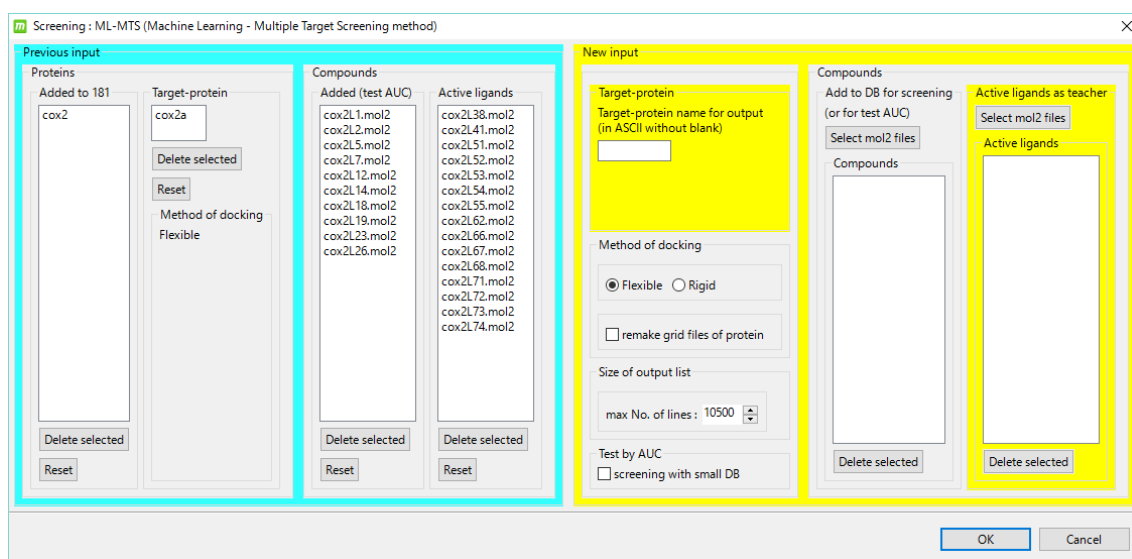
[OK] をクリックすると、ML-MTS 法による繰り返しスクリーニング計算が開始します。

### 1.14.3. スクリーニング計算結果の確認

その他の設定項目、スクリーニング計算結果の確認方法、ファイル出力方法等は、MTS 法 / ドッキングスコア順によるスクリーニング計算の場合とほぼ同じです。

「1.10.5 スクリーニング結果リストのサイズの入力」～「1.10.10 スクリーニング計算結果のファイル出力」を参照してください。

本操作後に再度  [ML-MTS] をクリックすると、前々回のターゲットタンパク質 **cox2** が [Added to 181] に、前回のターゲットタンパク質 **cox2a** が [Target-protein] に表示されています。



このように、ターゲットタンパク質を追加してスクリーニング計算を繰り返す毎に [Added to 181] にもタンパク質が追加されていきます。  
[Added to 181] に追加されたタンパク質はリファレンスタンパク質と同様に扱われます。

#### リファレンスタンパク質：

すべてのスクリーニング計算のドッキング計算で使用する 181 のリファレンスタンパク質を指します。詳細は myPresto のマニュアルを参照してください。

## 2. Molsite によるポケット探索・ドッキング計算

MolDesk Screening では、Molsite による高精度なポケット探索を実行できます。

### 2.1. PDB ファイルを読み水素原子・電荷を付加

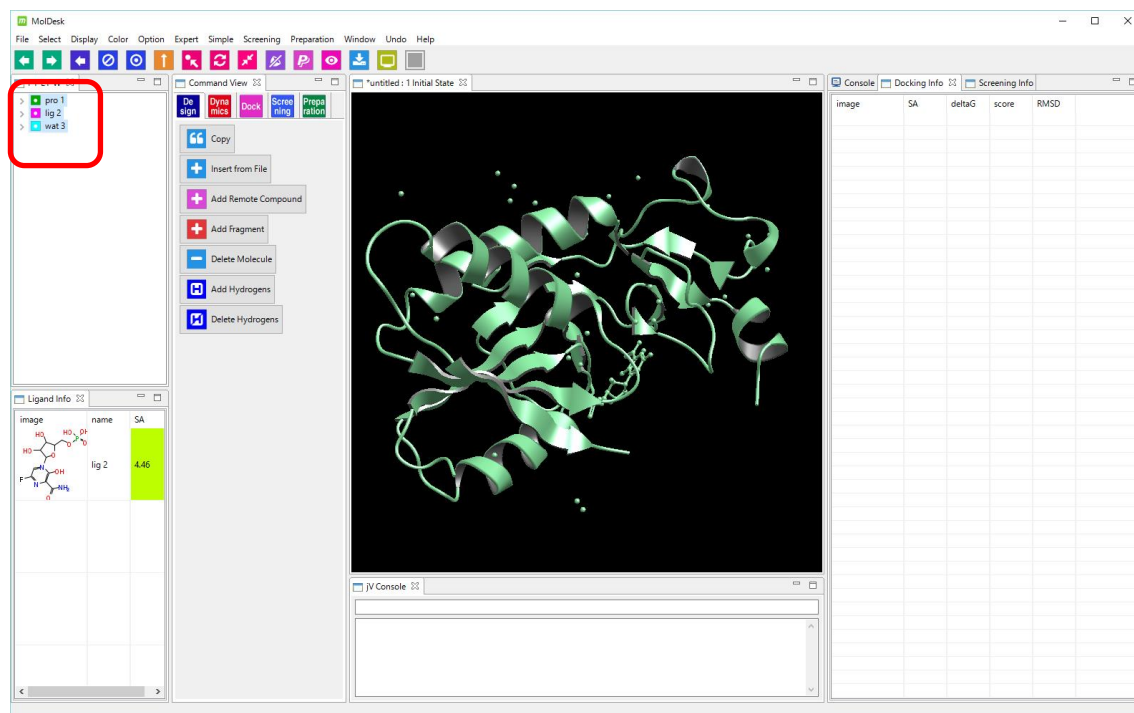
[File] - [Open Molecular File] で、MolDesk のインストール時にデスクトップに作成された MolDesk Screening フォルダの中にある以下のファイルを読み込みます。

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

[Open Molecular File] については MolDesk Basic のマニュアルを参照してください。

例に用いた「4kn6」は、プリン代謝酵素の一つである HGPRT（ヒポキサンチン・グアニンホスホリボシルトランスフェラーゼ）と、エボラ熱治療薬候補ファビピラビルに、リボース-5'-1 リン酸が結合した化合物を含んだものです。

ツリー表示画面で、タンパク質、化合物、結晶水の 3 つをマウスで選択（Shift+クリックで複数選択）します。





[Add Hydrogens] で、欠落しているすべての水素原子を付加します。

この際、タンパク質には電荷も付加されます。

※ 化合物の水素原子付加は、-p / -h / -m の3通りのオプションを選択できますが、内容の詳細は「MolDesk Basic マニュアル」をご参照ください。ここでは、デフォルトの -p オプションを選択します。

次に、 lig2 を選択した状態で  [Partial Charge] を実行し、MOPAC7 AM1 で電荷計算をして化合物に電荷を付与します。


## 2.2. Molsite によるポケット探索

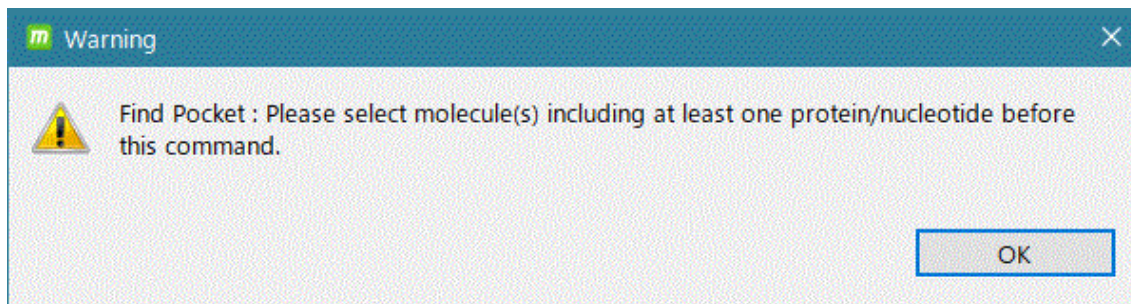


pro 1 をツリー表示画面で選択し



[Find Pocket] をクリックします。

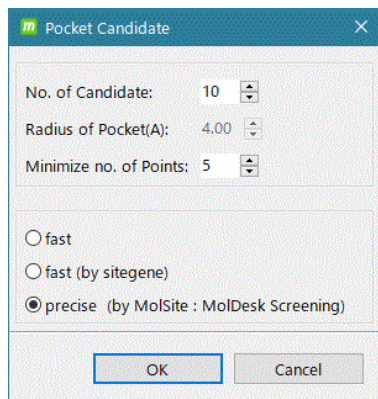
分子の選択を行わずに  [Find Pocket] を実行すると以下のワーニングダイアログが表示されます。その場合は分子を選択してから再実行してください。



ここでは、少なくとも1つのタンパク質分子を選択します。タンパク質分子を少なくとも1つ選択していれば、タンパク質以外の複数の分子も含めることが可能です。選択した分子の表面上でポケット探索を行います。このため、ポケットの空間は空けるように分子を選択してください。

※ Molsite はタンパク質分子のポケット探索に対応していて、核酸分子は探索対象としていません。

ポケット探索条件を設定するダイアログが表示されます。



この例では

[precise (by MolSite : MolDesk Screening)]

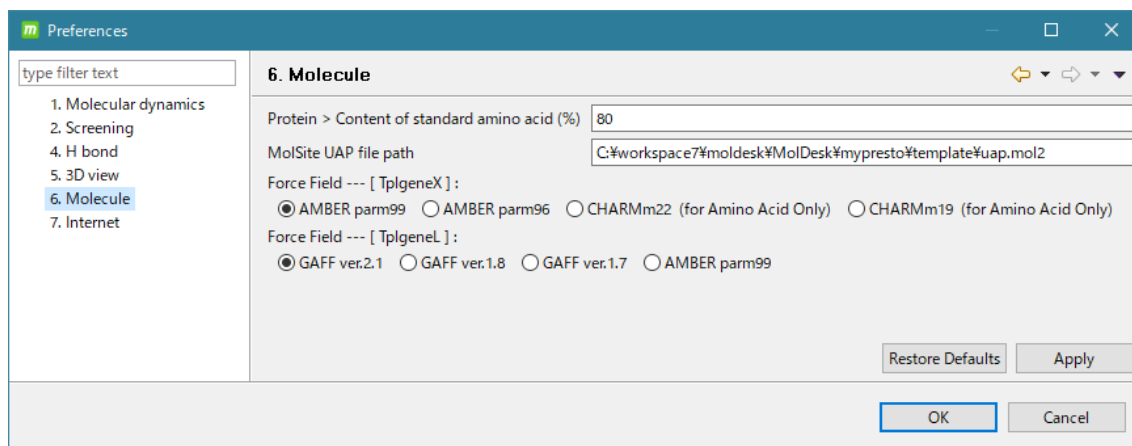
を選択します。

各設定項目の説明は

MolDesk Basic のマニュアルを参照してください。

※ MolSite では、UAP 分子という多数の ligand の候補分子を次々ポケット候補にドッキング計算をしますが、UAP 分子は、[Help]-[Preference]-[Molecule] で表示する以下の画面で、デフォルトで設定されている分子以外の任意の分子を、ユーザが設定することができます。

UAP 分子を変更したい場合は、以下の[MolSite UAP file path] に変更したい分子の mol2 ファイルパスを記入して、[Apply]または[OK] をクリックして確定してから計算してください。



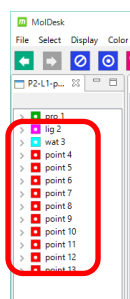
### 2.3. Molsite によるポケット探索の並列数・時間

Molsite によるポケット探索の計算時間の目安は以下の通りです。

並列計算に関して特に設定を行う必要はありません（スレッド並列計算で行います）。

	<b>Intel Core i7-4790K 4.0GHz / 16GB メモリ / windows8.1 8 並列で実行</b>
PDB 4kn6 (1555 原子)	11 分
PDB 1m17 (4744 原子)	15 分

この例では、計算終了後、10 のポケット候補がスコア順に表示されます。



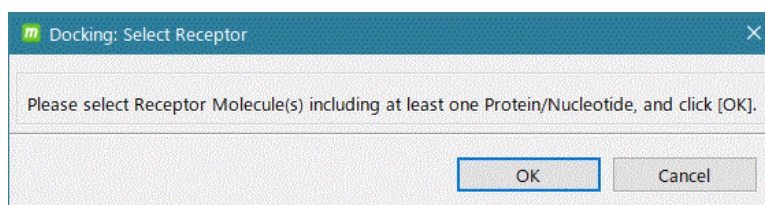
### 2.4. ドッキング計算

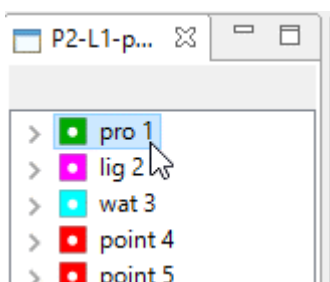
次にドッキング計算を行います。



[Docking] をクリックします。

受容体の選択を促すメッセージダイアログが表示されます。

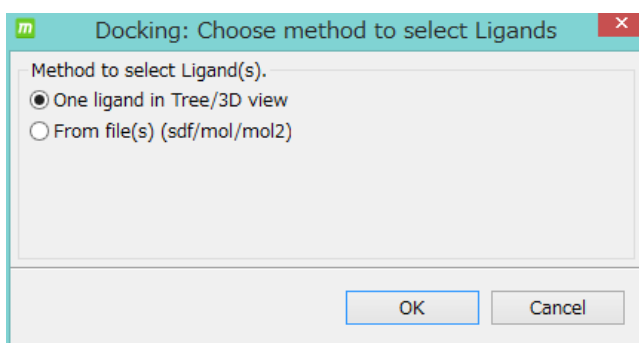




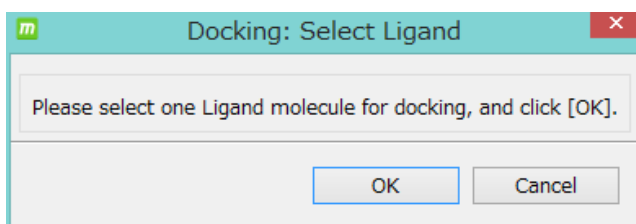
この例では  pro1 を選択して [OK] をクリックします。

受容体分子は、タンパク質分子を最低 1 つ含む必要があります。化合物や金属を含んでいてもかまいません。

リガンドの選択方法を指定するダイアログが表示されます。

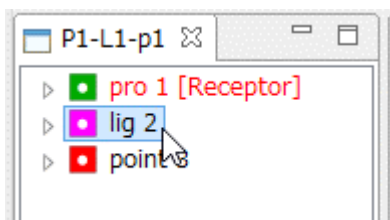



この例では、デフォルトの  
[One ligand in Tree/3D view]  
のまま [OK] をクリックします。



リガンドの選択を促すダイアログが  
表示されます。

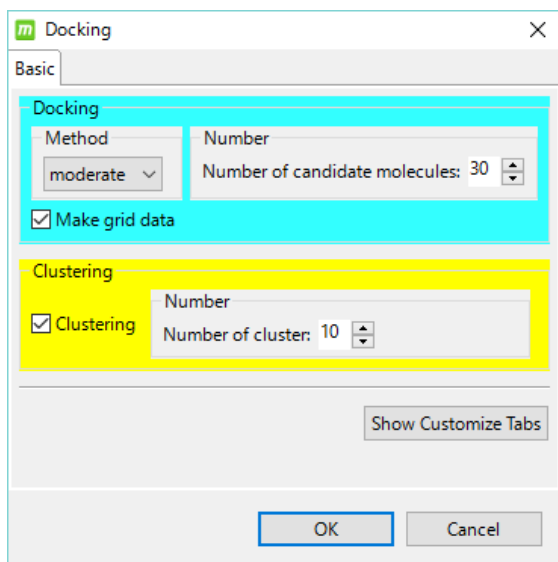
リガンドとして用いる化合物を  
ツリー表示画面または 3D 画面から  
1 つ選択します。



この例では  lig 2 を選択して[OK]をクリックしま  
す。

ドッキング計算条件を入力するダイアログが表示されます。





ドッキング計算の精度、構造クラスタリングの有無などを入力します。

この例ではデフォルトで計算します。  
このまま [OK] をクリックします。

設定項目の意味は MolDesk Basic のマニュアルを参照してください。

もっともスコアの良いポケット（ポケットの中でツリー表示画面で一番上にあるもの）を使用してドッキング計算が実行されます。

それ以外のポケットは自動的に削除されます。

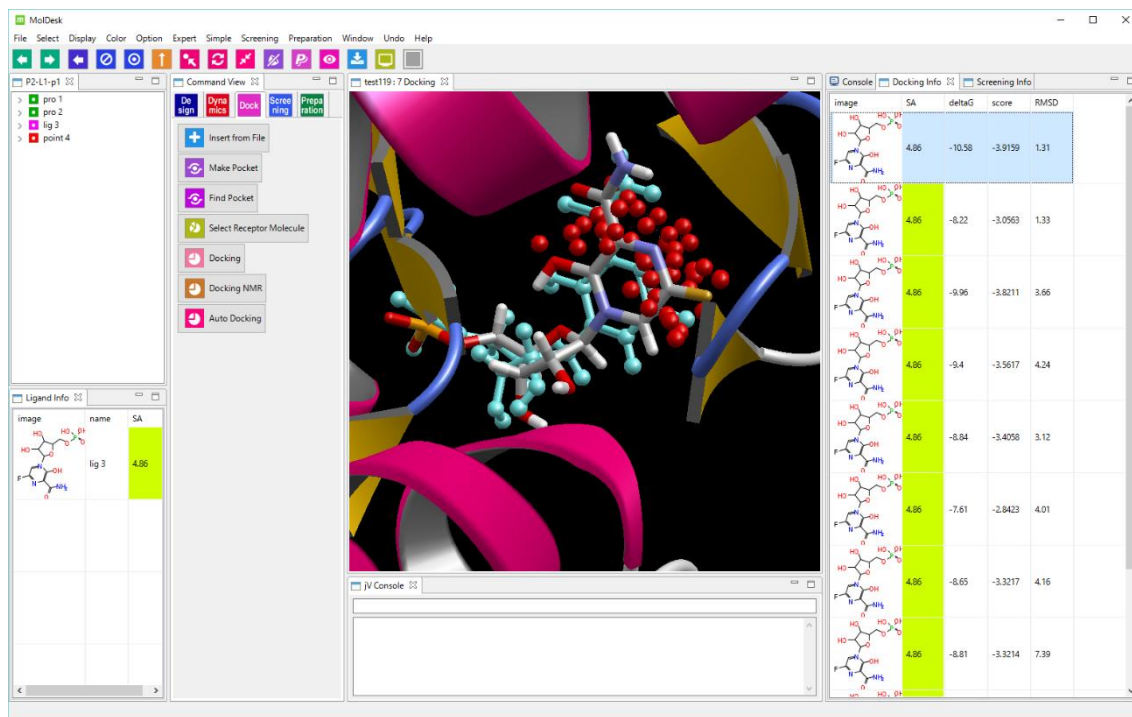
別のポケットを使用してドッキング計算を行いたい場合は



[Delete Molecule] でそれよりスコアの良いポケットをすべて削除してからドッキング計算を実行してください。

ドッキング結果の確認と保存は MolDesk Basic のマニュアルを参照してください。

この例では、最もスコアの良かった予測構造と正解構造 (lig3, [Color] –[Atom] –[Any]で水色で表示) を比較した結果、RMSD = 1.31 Å となりました。



### 3. Docking Score QSAR (Predict Activity)

Docking Score QSAR 法を用いた化合物分子の活性予測をします。指定した化合物の、特定のタンパク質に対する活性値を予測します。

Docking Score QSAR は、特定のタンパク質に対する回帰パラメーターを作成するための

[Preparation] - [Make DB to predict Activity]



と、上記で作成された回帰パラメーターを使って、特定のタンパク質に対する複数の化合物の活性値を一度に計算する

[Screening] - [Predict Activity]



の2つのボタンがあります。

[Predict Activity] で活性値の予測をするためには、[Make DB to predict Activity] で、回帰パラメタを学習して作成したデータファイルを予め作成する必要があります。

[Make DB to predict Activity] [Predict Activity] とともに、入力した化合物とプログラム内部の約 600 個のタンパク質とのドッキング計算を総当たりで実行するため、長時間の計算時間が必要です。(500 化合物で、通常の 8 スレッド並列の CPU マシンで約 3 時間)。

#### Docking Score QSAR :

多数のタンパク質に対するドッキングスコアの重み付き平均で結合自由エネルギーを推算する方法。ファーマコフォアを 600 種類のタンパク質で代表し、ある化合物の 600 ドッキングスコアを主成分解析し、実験データ  $\Delta G$  に、最小二乗法で回帰分析する。推算モデルは、リッジ回帰を用いた記述子ベースの重み付き PCR で計算しており、ロバスト推定 (M 推定) を利用して外れ値の除外を行っている。回帰に利用する親和性データおよび構造データは、ChEMBL および PDB (公的データベース) より取得している。ChEMBL より得られた親和性データ (IC<sub>50</sub> 値, %阻害値, 活性値など) は、全て結合自由エネルギー  $\Delta G$  に換算している。但し、ChEMBL には換算に必要な実験情報が不足していた為、幾つかの仮定を置いている (K<sub>d</sub>=K<sub>i</sub> 等)。

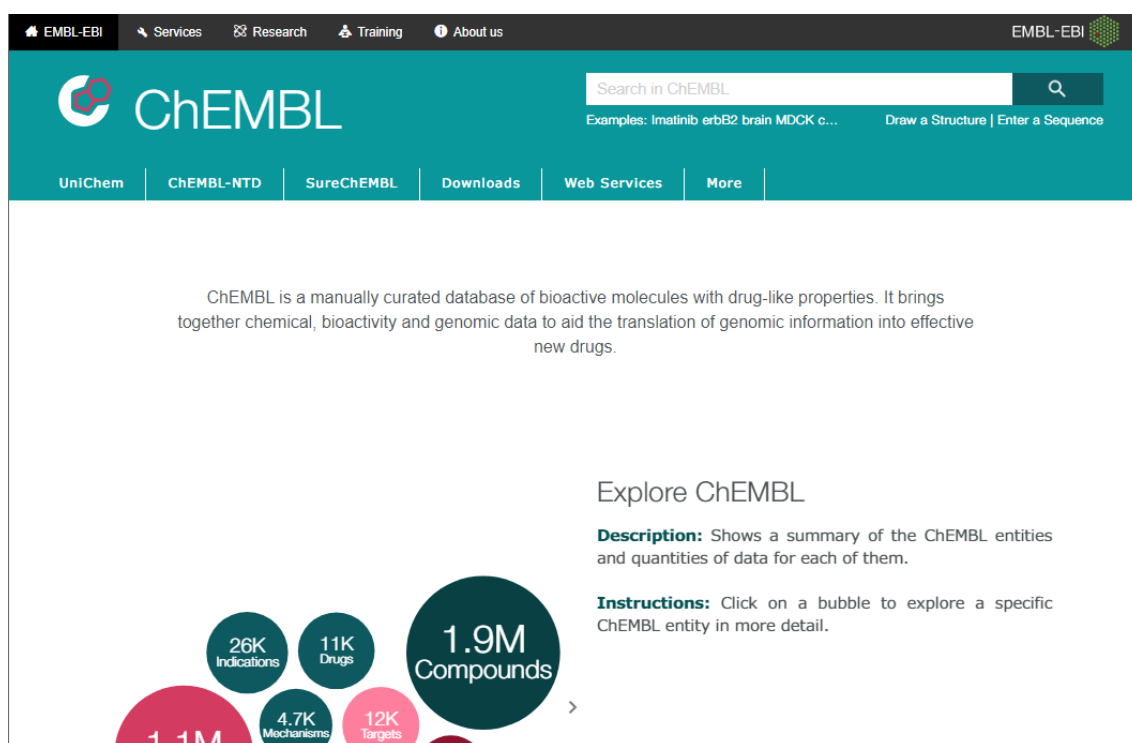
### 3.1. ChEMBL 実験値データの取得

対象となるタンパク質は、ユーザが自由に ChEMBL からダウンロードして選択できます。

初めに、ChEMBL から特定のタンパク質に対する様々な化合物の親和性データ (IC<sub>50</sub> 値, % 阻害値, 活性値など) を以下の手順でファイルとして入手します。

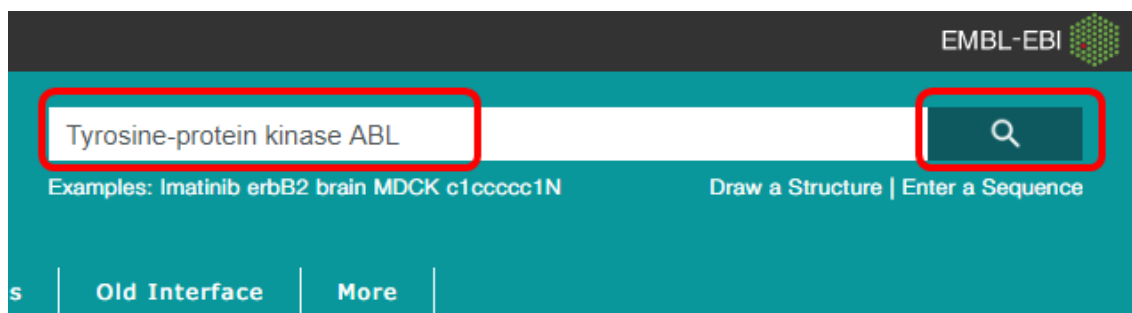
<https://www.ebi.ac.uk/chembl/> (ChEMBL トップページ)

にアクセスします。



予測対象のタンパク質について入力して、検索ボタンを選択 (クリック) します。

この例では、[Tyrosine-protein kinase ABL] と入力します。



表示された検索結果の中から「Targets」グループを選択（クリック）して表示し、適当なタンパク質を選びます。

この例では、CHEMBL1862 を選択します。

EBI > Databases > Chemical Biology > ChEMBL Database > Targets Search Results > Tyrosine-protein kinase ABL

### Search Results

All Results 299 Compounds 245 **Targets 7** Assays 13 Documents 23 Cells 11 Tissues 0

## Targets

Show Full Query ?

7 Targets  
0 Selected - Select All  
Browse Activities ?

Table Heatmap

Records per page: 20 Show/Hide Columns

Showing 1-7 out of 7 records

<input type="checkbox"/>	ChEMBL ID	Search Hit	Name	UniProt Accessions	Type	Organism	Compounds	Activities
<input type="checkbox"/>	CHEMBL5166		Tyrosine-protein kinase V-ABL	P00521	SINGLE PROTEIN	Abelson murine leukemia virus	By Mol. Wt.: 52	By Std. Type: 68
<input type="checkbox"/>	<b>CHEMBL1862</b>		Tyrosine-protein kinase ABL	P00519	SINGLE PROTEIN	Homo sapiens	By Mol. Wt.: 4669	By Std. Type: 12638

以下のページが表示されます。

EBI > Databases > Chemical Biology > ChEMBL Database > CHEMBL1862

## Target Report Card

### Name And Classification

ID: CHEMBL1862

Type: SINGLE PROTEIN

Preferred Name: Tyrosine-protein kinase ABL

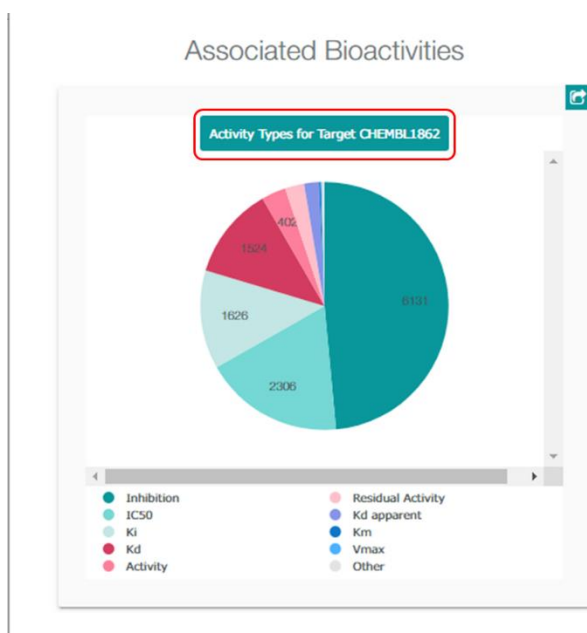
Synonyms: ABL ABL1 Abelson murine leukemia viral oncogene homolog 1 Abelson tyrosine-protein kinase 1 JTK7 Proto-oncogene c-Abl Tyrosine-protein kinase ABL1 p150

Organism: Homo sapiens

Species Group: No

Protein Target Classification: - Enzyme > Kinase > Protein Kinase > TK protein kinase group > Tyrosine protein kinase Abl family

ページ中の以下の円グラフで、[Activity Types for Target CHEMBL1862]を選択（クリック）します。



以下のページが表示されます。

EBI > Databases > Chemical Biology > ChEMBL Database > Activities > Query

## Browse Activities

[Edit Querystring](#) [Show Full Query](#)

12,638 Activities  
0 Selected - Select All  
[Browse Compounds](#) [Download CSV](#) [Download TSV](#)

Records per page: 20 [Show/Hide Columns](#)

Showing 1-20 out of 12,638 records

**Filters**



Standard Type

- Inhibition 6131
- IC50 2306
- Ki 1626
- Kd 1524
- Activity 402
- Residual Activity 316
- Kd apparent 243
- Km 29
- Vmax 17
- Vmax(app) 12
- Other Categories 32

Target Type

- SINGLE PROTEIN 12638

Organism Taxonomy L1

Molecule ChEMBL ID	Compound Key	Standard Type	Standard Relation	Standard Value	Standard Units	pChEMBL Value	Comment	As Ch ID
 CHEMBL538507	18	IC50	>	100000	nM	No Data	No Data	CHE
 CHEMBL44	Genistein	IC50	=	10000	nM	5	No Data	CHE

ページ上部の[TSV]を選択（クリック）するとタブ区切りテキストファイルのダウンロードリンクが生成されるので、生成された[here]を選択（クリック）します。



CHEMBL25-chembl\_activity-XXXX.tsv.gz というファイルがダウンロードされます。（※XXXX は長いランダム英数記号）

ファイルは gz 形式で圧縮されているため適当な解凍ソフトで tsv 形式のタブ区切りテキストファイルに解凍展開します。

CHEMBL25-chembl\_activity-XXXX.tsv.gz

↓

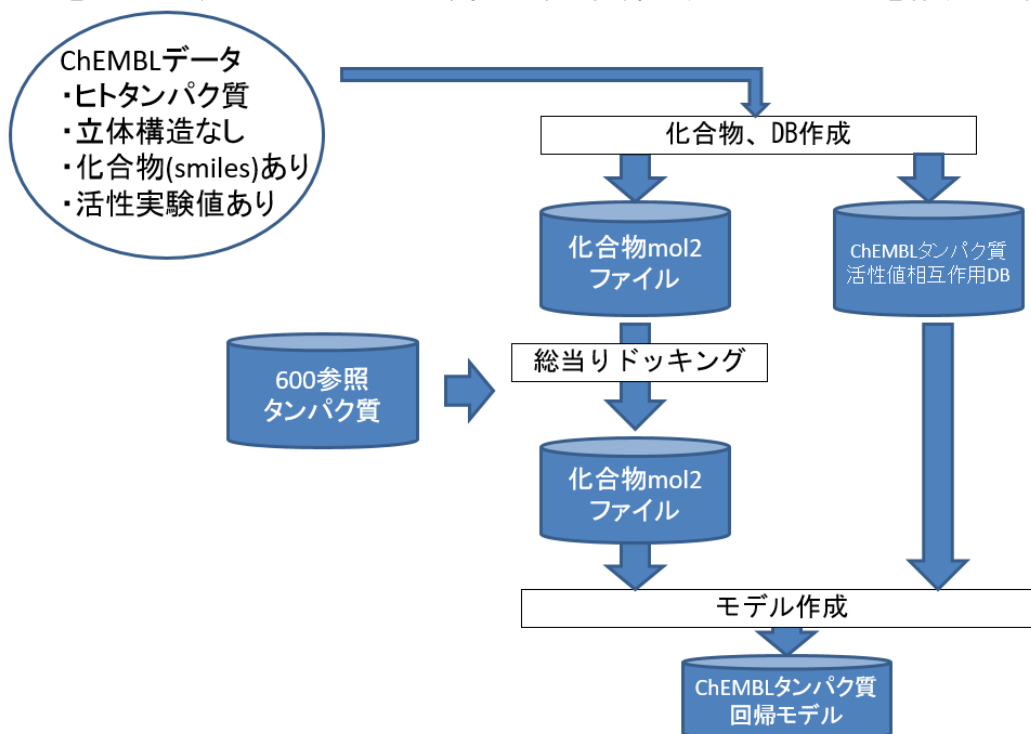
CHEMBL25-chembl\_activity-XXXX.tsv

## 3.2. 回帰パラメータの計算

### 3.2.1. 概要説明

#### ChEMBLタンパク質回帰モデル作成の概要

ChEMBLタンパク質回帰モデル作成では、ChEMBLのサイトからダウンロードしたcsvファイルを入力とし、ChEMBLのタンパク質に対する回帰パラメータファイルを作成します。



### 3.2.2. プロジェクトの作成

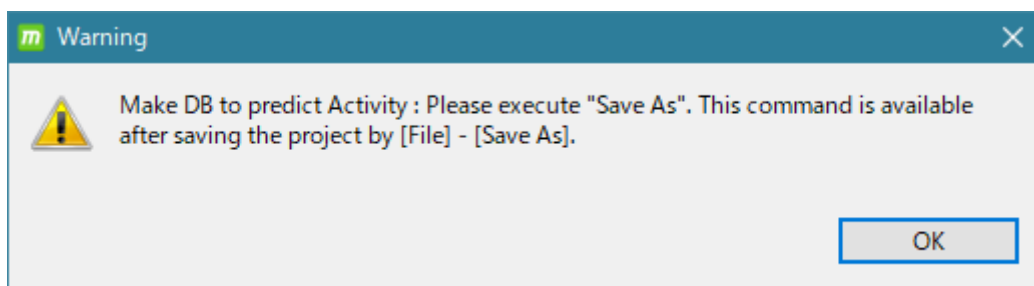
ここで、MolDesk Screening での操作に戻ります。

[File] - [New Project] メニューで、空のプロジェクトを作成し、保存します。

プロジェクトの保存方法は MolDesk Basic のマニュアルを参照してください。

- プロジェクトが保存されていない場合は以下のワーニングダイアログが表示されますので、プロジェクトの保存を行ってください。





### 3.2.3. 回帰パラメターの計算実行

特定のタンパク質に対する回帰パラメターを作成するための

[Preparation] - [Make DB to predict Activity]



をクリックします。すると以下の画面が表示されます。

[Browse] ボタンをクリックして、ファイルセレクトで前節で ChEMBL からダウンロードした、特定のタンパク質の活性値の実験データファイルを選択します。選択が終了すると、下図のようにファイルのパスが表示されます。

さらに、[Name of protein:] の欄に、特定のタンパク質の名称を入力します。この入力名が出力ファイルの名称など、計算に使われますので、**スペースを含まない英数字**で入力します。この例では、**ABL** と入力します。

Make DB to predict Activity

ChEMBL download data file (\*.tsv)

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

☒ Ki ☒ Kd ☒ ICS0 ☒ Activity ☒ Residual Activity ☒ Inhibition ☒ Potency >>> Omit other experiments.

Filtering

☒ by molecular weight Min 100 Max 600

OK Cancel

このページでは、最大、16 個の ChEMBL からダウンロードした実験データファイルを入力できます。すなわち、特定のタンパク質 16 個分の回帰パラメーターの計算を一度に並列で実行できます。ただし、一般には、計算時間がかかります（48 並列のマシンで、1 個あたり数日が目安）。

#### [Experiments used in calculation]

ChEMBL からダウンロードしたデータファイルの中で、回帰モデル作成で使用する実験データを選択します。チェックした項目以外の実験データは無視します。

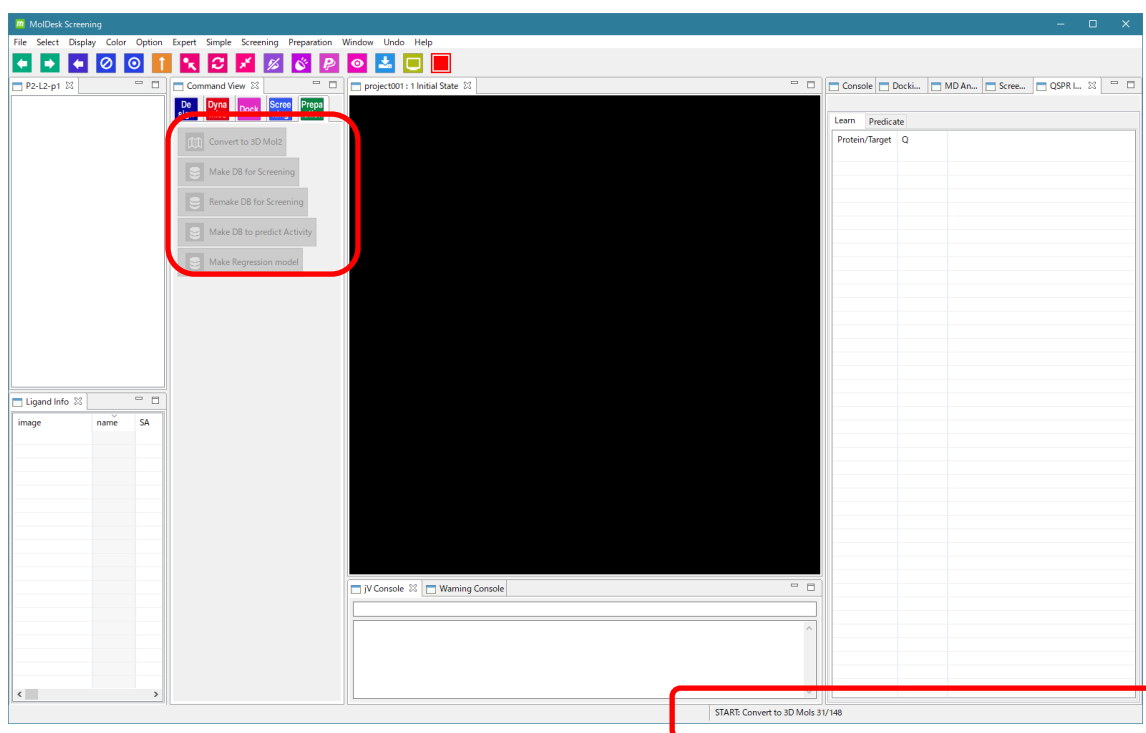
### [Filtering]

化合物の分子量によるフィルタリングを実行するか否かを、[Filtering] – [by molecular weight] で選択可能です。

MolDesk Screening では、ドッキング計算に myPresto の sievgene を使用している関係上、分子量 600 以上のドッキング計算の精度が落ちるため、デフォルトで 100 – 600 に設定されています。通常は、デフォルトのまま計算します。

[OK] をクリックすると（並列）計算が始まります。

計算を開始するとコマンドボタンがグレーになります。コマンドボタンがグレーになっている間は計算中です。また、右下の赤枠内に、計算中の簡単な計算状況を表示します。



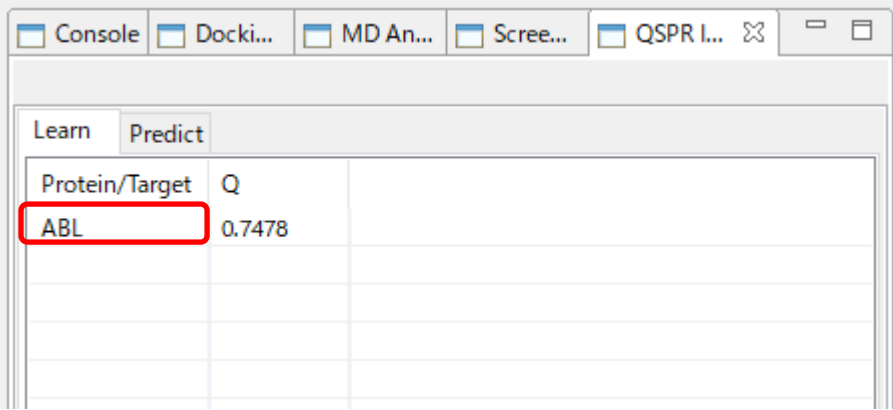
計算中でも他のプロジェクトを操作できますが、プロセッサの占有率によっては動作が極端に遅くなる場合がありますのでご注意ください。

並列計算する際の並列数は、[Help] - [Preference] - [Screening] の Thread number で設定できます。詳細は「1.7 スクリーニング計算の並列数・メモリ量・時間」を参照してください。

### 3.2.4. 回帰パラメターの計算結果のグラフによる確認

回帰パラメターの計算が終了するとコマンドボタンがグレーから使用できるようになります。また、画面の右下に [END: Make DB to predict Activity] と表示します。

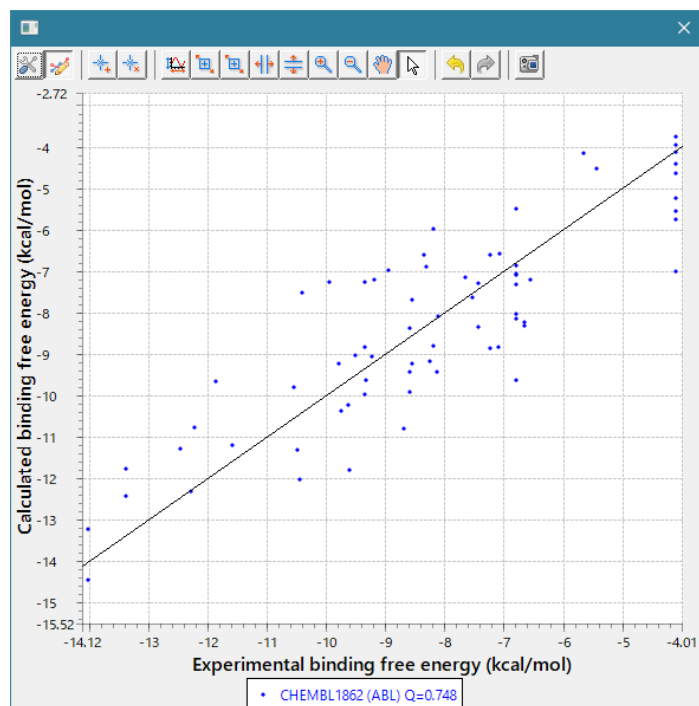
さらに、画面の右側に 下図のように QSPR Info 画面の [Learn] タブを表示します。回帰モデルを作成したときに入力したタンパク質名と、作成した回帰モデルの相関係数(Q値)をリスト表示します。



The screenshot shows a software window titled 'QSPR I...' with several tabs: 'Console', 'Docki...', 'MD An...', 'Scree...', and 'QSPR I...'. The 'QSPR I...' tab is active and contains two sub-tabs: 'Learn' and 'Predict'. The 'Learn' sub-tab is selected, displaying a table with two columns: 'Protein/Target' and 'Q'. The first row of the table has 'ABL' in the 'Protein/Target' column and '0.7478' in the 'Q' column. This row is highlighted with a red rectangular border.

Protein/Target	Q
ABL	0.7478

ここで、上の赤枠の [Protein/Target] カラムのタンパク質名をダブルクリックします。すると、回帰パラメタを作成したときに使った ChEMBL 由来の実験データ  $\Delta G$  と、回帰パラメタ計算時の  $\Delta G$  の計算値のグラフを次のように表示して、学習の信頼性を確認することができます。



### 3.2.5. 回帰パラメーターファイルの確認

計算結果の回帰パラメーターファイルが作成される場所は、保存したプロジェクトのフォルダを [PROJECT] とすると、

[PROJECT] -> work -> database\_qsar -> 09.param

です。ファイル名は、前節で、[Name of protein:] の欄に入力した、特定のタンパク質の名称を [PROTEIN] とすると、

**[PROTEIN].param**

となります。このファイルは、次節で説明する活性値予測の計算で使用するので重要です。

なお、database\_qsar には以下のフォルダとファイルが生成されます。  
 これらの内容については、ユーザは気にする必要はありませんが、内容を説明すると以下の通りです。

```

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

```

項目	内容
00.log	[PROTEIN] 毎に、[PROTEIN] フォルダ/error.log ファイルが出力される。各 [PROTEIN] の計算で、計算中で発生したエラーが出力される。
01.download	前節で入力した、ChEMBL からダウンロードした実験データファイルが保存される。
02.mol2	[PROTEIN] 毎に、[PROTEIN] フォルダ以下に、実験データに記載された化合物の mol2 ファイルを出力する。実験データファイルでは SMILES で記録されているが、mol2 に変換、3次元化と電荷生成も行う。
03.topology	[PROTEIN] 毎に、[PROTEIN] フォルダ以下に、各化合物フォルダを作成して、myPresto のトポロジーファイル生成。
04.optimize	上記各化合物の 3 次元構造のエネルギー極小化計算結果
05.db	ChEMBL 実験データファイルの加工データファイル
06.logand	ChEMBL 実験データファイルの化合物の mol2 ファイル
07.work	600 タンパク質×mol2 ファイルのドッキング計算をするフォルダ

08.score	600 タンパク質×mol2 ファイルのドッキング計算の全スコア 保存ファイル
09.param	回帰パラメーターのデータファイル
protein_qsar	600 タンパク質のドッキング計算用グリッドファイル作成フォルダ
ChEMBL.list	ChEMBL からダウンロードした実験データファイル名リスト
pro_list	600 タンパク質の PDB ID リスト

### 3.3. 活性値の予測計算

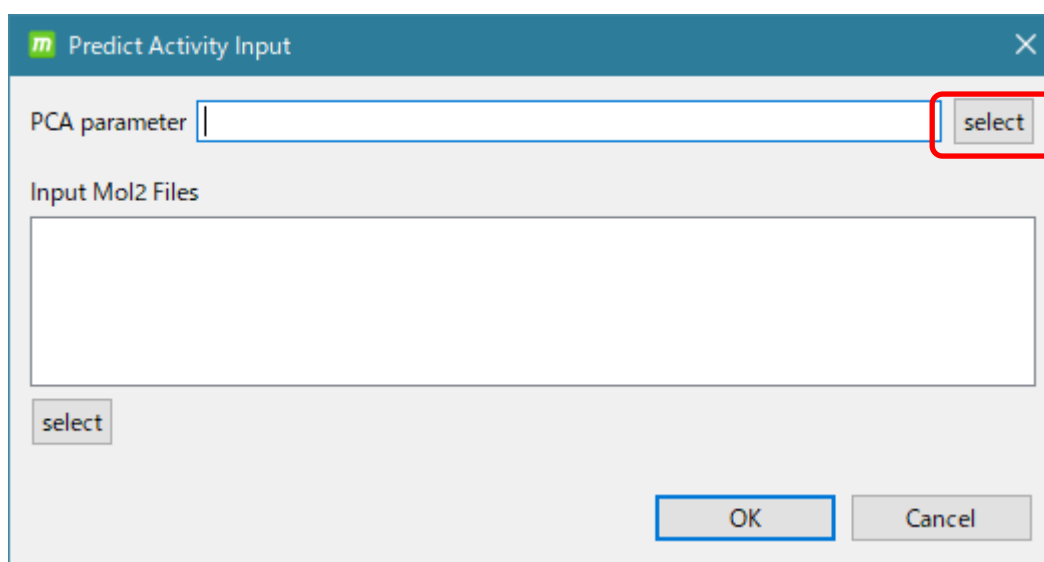
前節で作成した回帰パラメーターファイルを使って、特定のタンパク質に対する複数の化合物の活性値を一度に（並列）計算します。

#### 3.3.1. 活性値予測計算の実行

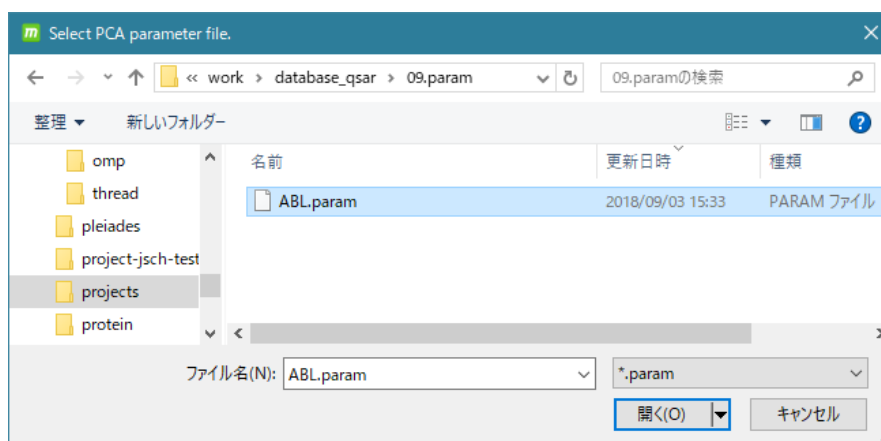
[Screening] - [Predict Activity]



をクリックします。以下の入力画面が出ます。

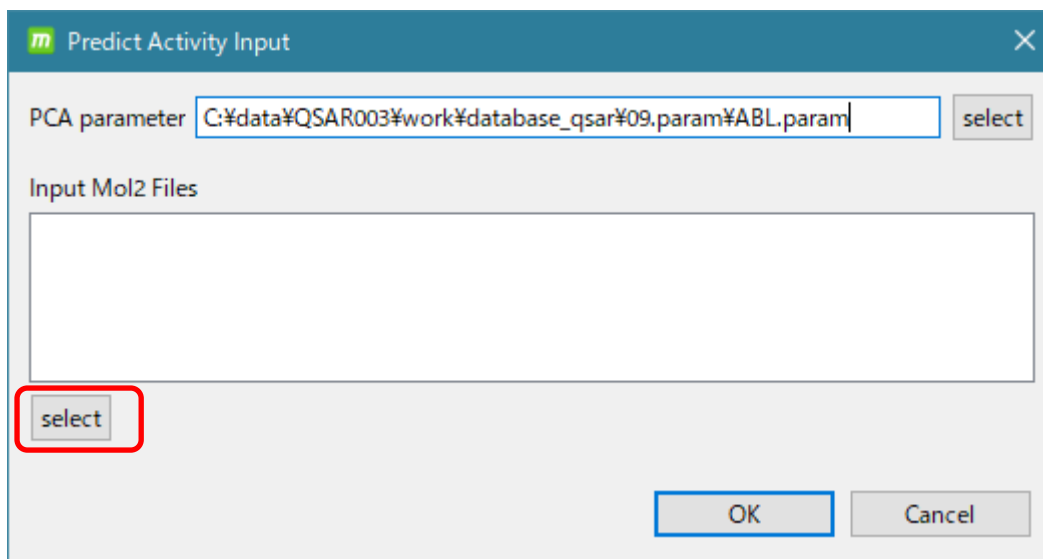


上図の、赤枠の [select] をクリックして、前節で計算した回帰パラメーターファイルを下図で選択します。

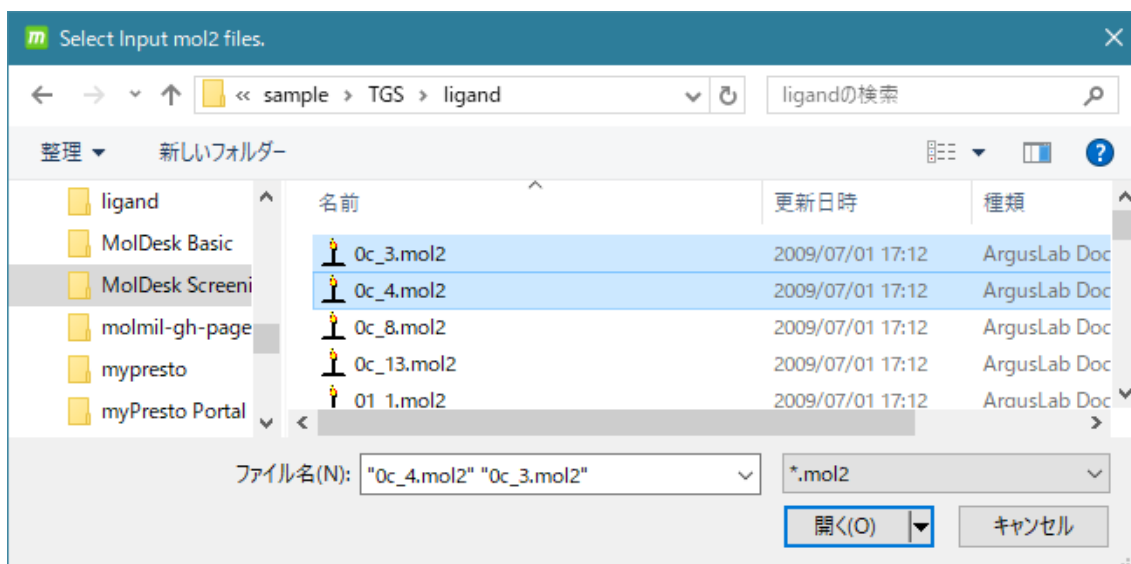




ここでは例として、[PROJECT] -> work -> database\_qsar -> 09.param -> ABL.param  
を選択します。

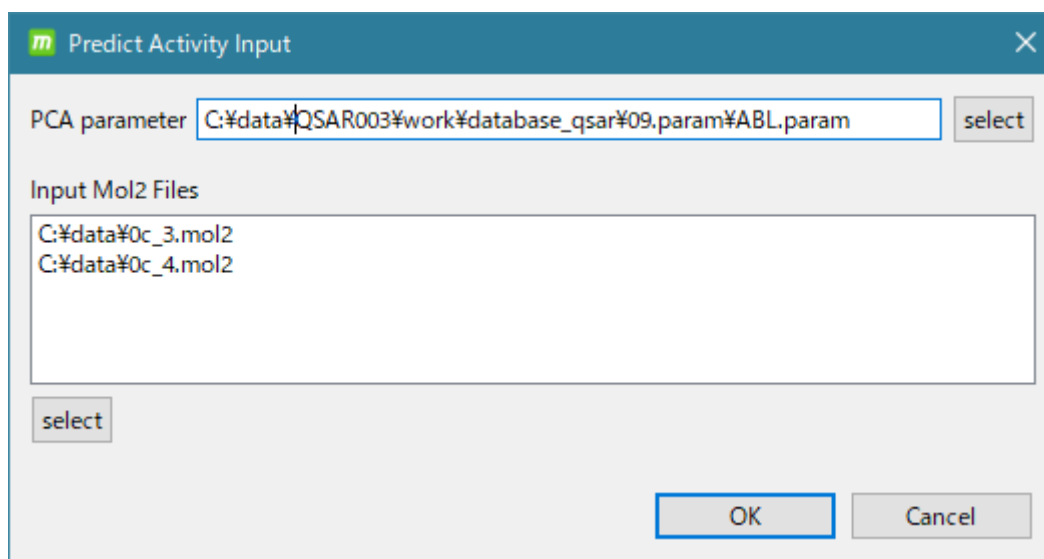


次に、上図の赤枠の [select] をクリックして、活性値を予測したい化合物の mol2 ファイルを下図で選択します。



ここでは例として、上の 2 個を選択しました。

- ※ 化合物の mol2 ファイルは、必ず 1 分子だけを含むようにしてください。MolDesk Screening の、[Preparation] – [Convert to 3D Mol2] で作成した Mol2 ファイルを入力として使用できます。
- ※ 近い将来に機能追加して、sdf ファイルで複数の分子を一括入力可能にする予定です。



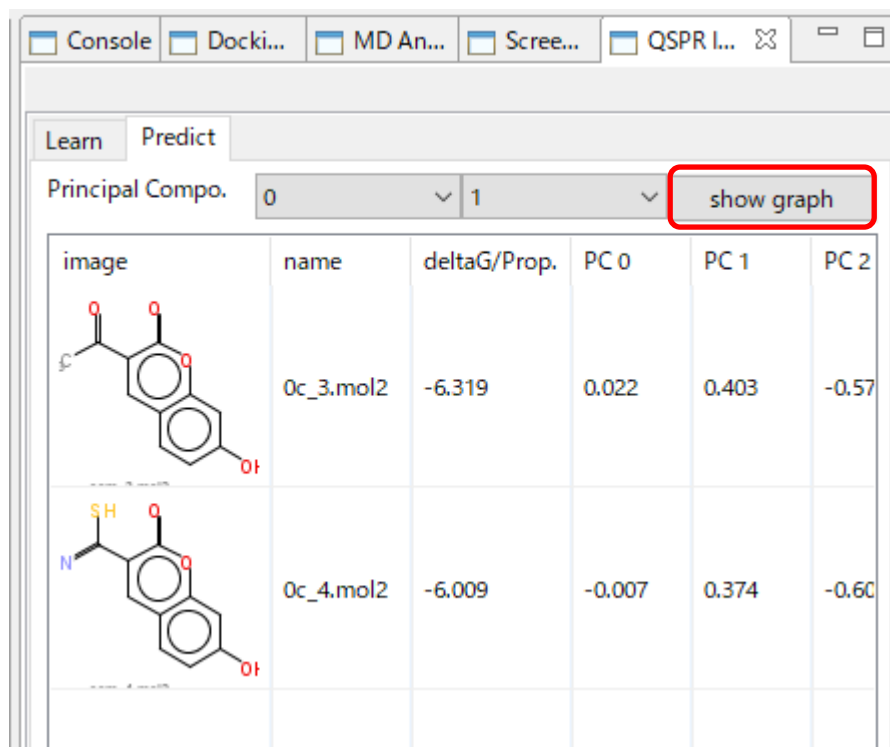
入力が終わると上図のようになりますので、[OK] をクリックすると、予測計算が始まります。

予測計算中は、画面の右下に計算の状況が表示されます。

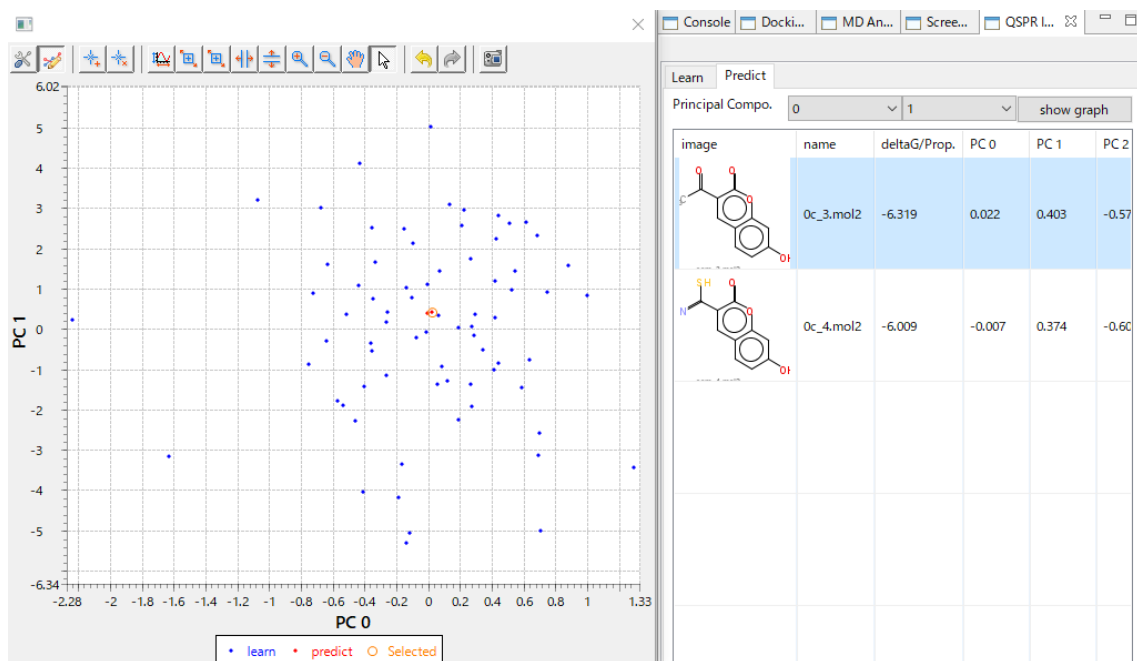
### 3.3.2. 活性値予測計算の結果確認

活性値予測の計算が終了するとコマンドボタンがグレーから使用できるようになります。また、画面の右下に [END: Predict Activity] と表示します。

さらに、画面の右側に 下図のように QSPR Info 画面の [Predict] タブを表示します。



ここで、上の赤枠の [show graph] ボタンをクリックすると、PCA グラフを表示します。図の例では、0 軸と 1 軸の PCA グラフを表示します。軸は、0～9 まで任意に選択できます。



リストの化合物をクリックすると、PCA グラフの中で、予測した化合物の位置を赤丸で確認できます。

逆に、グラフの赤丸をクリックすると、リストの化合物にフォーカスします。

青点は、学習に使用した ChEMBL 由来のデータです。予測した化合物が、ChEMBL 由来の化合物とかけ離れてないか確認できるので、計算の信頼性を評価できます。

予測した活性値(対数換算)は、テーブルの deltaG/Prop. (kcal/mol) のコラムに表示します。

#### 4. 回帰分析による化合物の特性値予測 (Predict with Regression model)

回帰分析による化合物の各種特性値を予測します。

化合物の各種特性値の実験データファイルから回帰パラメーターを作成するための

[Preparation] - [Make Regression model]



と、上記で作成された回帰パラメーターを使って、複数の化合物の各種特性値を一度に計算する

[Screening] - [Predict with Regression model]



の2つのボタンがあります。

[Predict with Regression model] で各種特性値の予測をするためには、[Make Regression model] で、回帰パラメタを学習して作成したデータファイルを予め作成する必要があります。

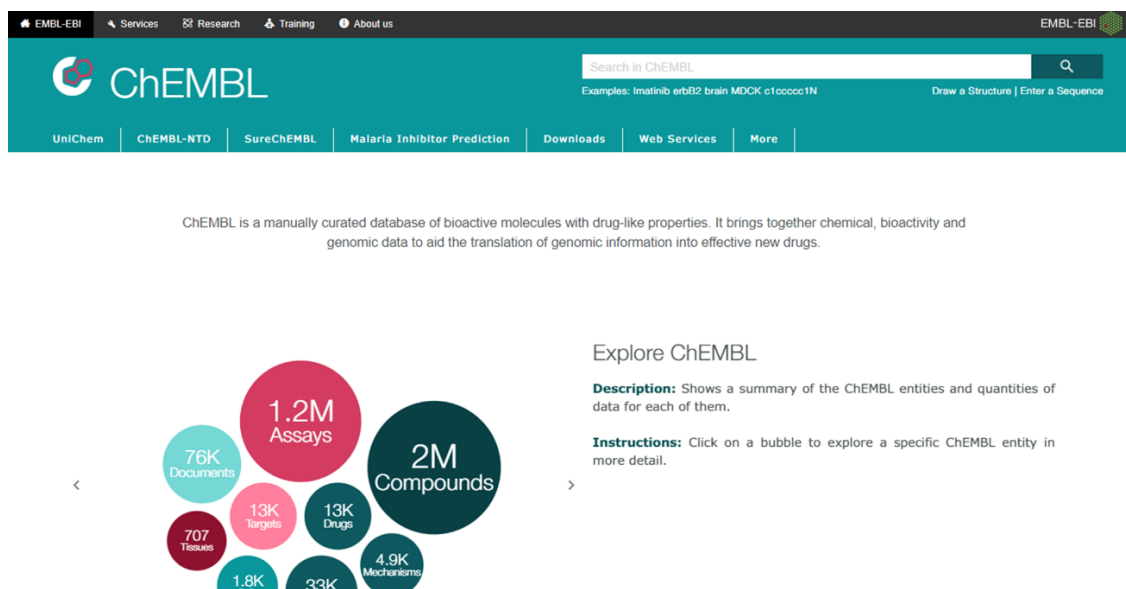
[Make Regression model] [Predict with Regression model] とともに、入力した化合物の descriptor を作成するため、比較的長時間の計算時間が必要です。

#### 4.1. ChEMBL 実験値データの取得

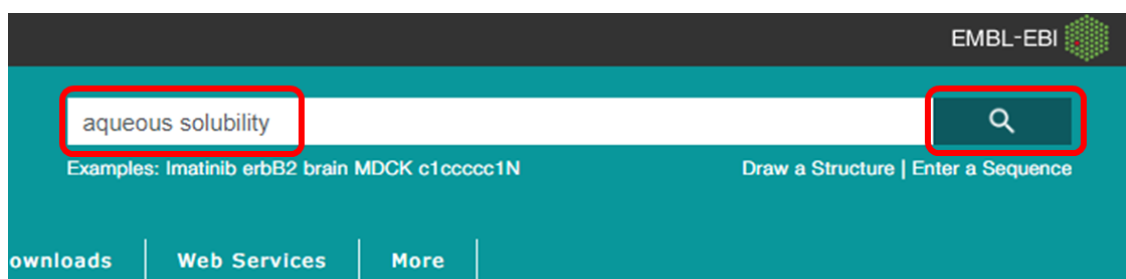
ChEMBL から各種特性値の実験値データを取得する手順を説明します。

<https://www.ebi.ac.uk/chembl/> (ChEMBL トップページ)

にアクセスします。



検索するときのキーワードとして "Search in ChEMBL" 欄に、「aqueous solubility」「permeability」など興味のある物性名を入力して、検索ボタンを選択 (クリック) します。この例では、[aqueous solubility] と入力します。



表示された検索結果の中から「Assays」グループを選択（クリック）して表示し、ターゲットリストを "Compounds" でデータ数の多い順に並べ替えます。

リストから適当なタンパク質を選びます。

この例では、"CHEMBL631962" を選択します。

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


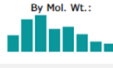
## Assays

Show Full Query

Table 2,054 Assays  
0 Selected - Select All  
Browse Activities

Records per page: 20 Show/Hide Columns

Showing 1-20 out of 2,054 records

ChEMBL ID	Search Hit	Assay Type	Description	Organism	Compounds	Document ChEMBL ID	BAO Format	Source
<input type="checkbox"/> CHEMBL631962		P	Aqueous solubility	No Data		CHEMBL1132890	small-molecule physicochemical format	Scientific Literature
<input type="checkbox"/> CHEMBL1034535		P	Aqueous solubility in phosphate buffered saline by multi-screen solubility assay	No Data		CHEMBL1151935	small-molecule physicochemical format	Scientific Literature

以下の画面が表示されます。("Assay Report Card"ページが表示される)

EBI > Databases > Chemical Biology > ChEMBL Database > CHEMBL631962

## Assay Report Card

### Basic Information

**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  
**Document:** CHEMBL1132890  
**Cell:**  
**Tissue:**

ページ中の以下の円グラフで、[Activity Types for Target CHEMBL631962]を選択（クリック）します。

## Activity Charts



以下のページが表示されます。

EBI > Databases > Chemical Biology > ChEMBL Database > Activities > Query

## Browse Activities

[Edit Querystring](#)

[Show Full Query](#)

148 Activities  
0 Selected - Select All  
[Browse Compounds](#)



Table

Records per page: 20 [Show/Hide Columns](#)

Showing 1-20 out of 148 records

Filters

- Standard Type
  - Log S 148
- Target Type
  - NO TARGET 148
- Organism Taxonomy L1
  - N/A - 148
- Organism Taxonomy L2
  - N/A - 148
- Organism Taxonomy L3

<input type="checkbox"/>	Molecule ChEMBL ID	Compound Key	Standard Type	Standard Relation	Standard Value	Standard Units	pChEMBL Value	Comment	Assay ChEMBL ID
<input type="checkbox"/>	 DDT		Log S	=	-7.15	No Data	No Data	No Data	<a href="#">CHEMBL631962</a>
<input type="checkbox"/>	 Hexamethylbenzene		Log S	=	-5.23	No Data	No Data	No Data	<a href="#">CHEMBL631962</a>



ページ上部の[TSV]を選択（クリック）するとタブ区切りテキストファイルのダウンロードリンクが生成されるので、生成された[here]を選択（クリック）します。



DOWNLOAD-XXXX.tsv.gz というファイルがダウンロードされます。（※XXXX は長いランダム英数記号）

ファイルは gz 形式で圧縮されているため適当な解凍ソフトで tsv 形式のタブ区切りテキストファイルに解凍展開します。

DOWNLOAD-XXXX.tsv.gz

↓

DOWNLOAD-XXXX.tsv

## 4.2. 入力実験データファイルの作成

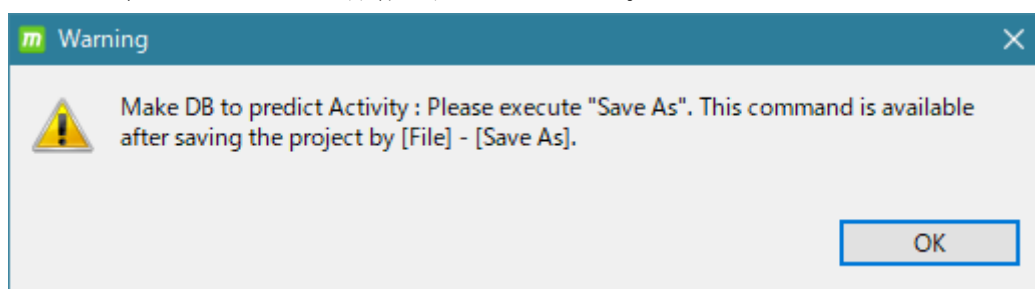
### 4.2.1. プロジェクトの作成

ここで、MolDesk Screening での操作に戻ります。

[File] - [New Project] メニューで、空のプロジェクトを作成し、保存します。

プロジェクトの保存方法は MolDesk Basic のマニュアルを参照してください。

- プロジェクトが保存されていない場合は以下のワーニングダイアログが表示されますので、プロジェクトの保存を行ってください。

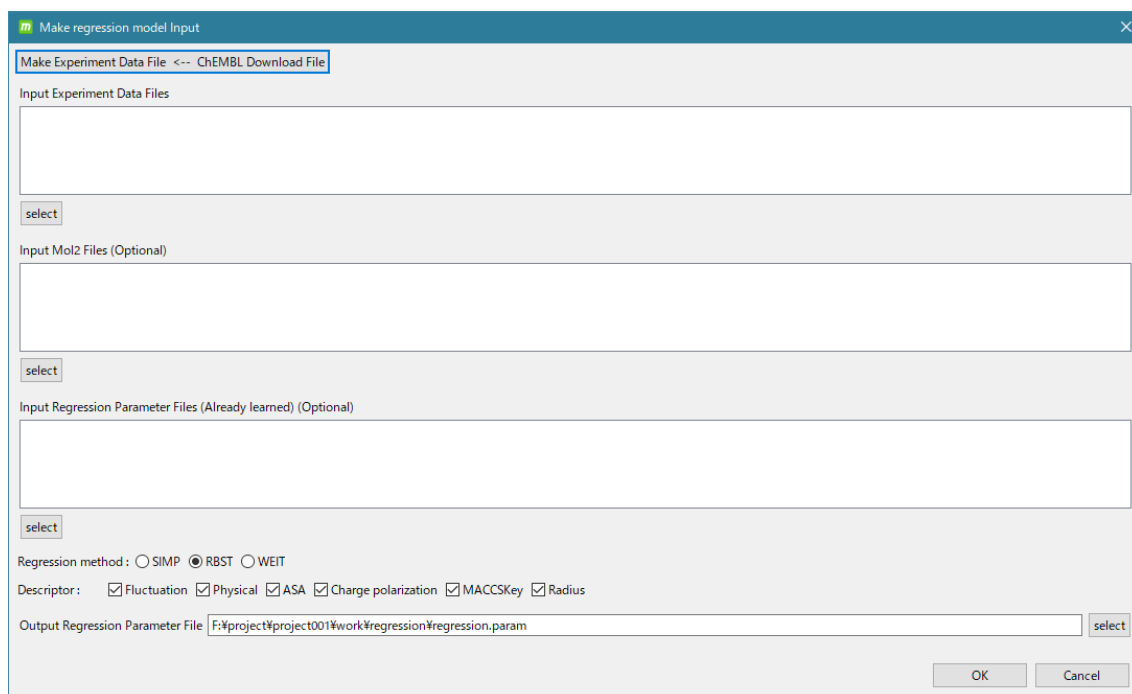


#### 4.2.2. ChEMBL データファイルからの作成法

[Preparation] - [Make Regression model]



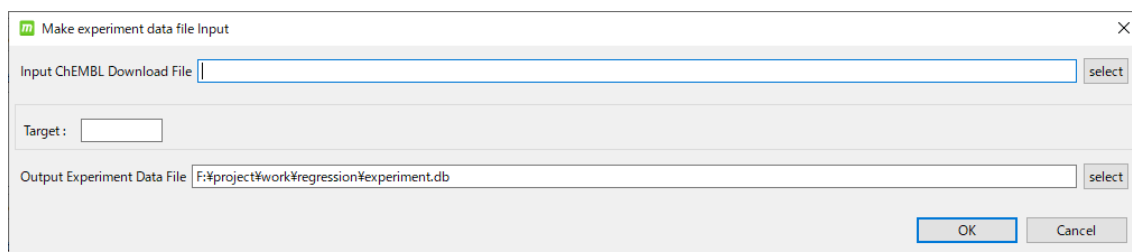
をクリックします。すると以下の画面が表示されます。

A screenshot of a software dialog box titled "Make regression model Input". At the top, there is a tab bar with two tabs: "Make Experiment Data File" (which is selected and highlighted in blue) and "ChEMBL Download File". Below the tabs, the dialog is divided into several sections. The first section, "Input Experiment Data Files", contains a large empty text box and a "select" button. The second section, "Input Mol2 Files (Optional)", also contains an empty text box and a "select" button. The third section, "Input Regression Parameter Files (Already learned) (Optional)", contains an empty text box and a "select" button. Below these sections, there are radio buttons for "Regression method": SIMP, RBST (which is selected), and WEIT. Underneath, there are checkboxes for "Descriptor": Fluctuation, Physical, ASA, Charge polarization, MACCSKey, and Radius, all of which are checked. At the bottom, there is a text field for "Output Regression Parameter File" with the path "F:\project\project001\work\regression\regression.param" and a "select" button. Finally, at the bottom right, there are "OK" and "Cancel" buttons.

ここで、実験データファイルを作成するために、一番上のボタン

[Make Experiment Data File ← ChEMBL Download File] をクリックします。

以下の画面が出ます。

A screenshot of a software dialog box titled "Make experiment data file Input". It features a single tab labeled "Make experiment data file Input". The dialog has three main input areas. The first is "Input ChEMBL Download File" with an empty text box and a "select" button. The second is "Target:" with an empty text box. The third is "Output Experiment Data File" with the path "F:\project\work\regression\experiment.db" and a "select" button. At the bottom right, there are "OK" and "Cancel" buttons.

入力項目は以下の通りです。

項目	内容
Input ChEMBL Download File	ChEMBL でダウンロードしたデータファイル (入力必須)
Target	実験の種類やターゲット蛋白質についての文字列 (入力必須) (例) LogS, LogP, ChEMBL1785 など。
Output Experiment Data File	出力する実験データファイルのパス (入力必須)

今回の例では、以下の通り入力して [OK] をクリックします。

([Input ChEMBL Download File] の入力は、[Select] をクリックしたときに出るファイル選択画面でダウンロードしたファイルを選択して入力します。)

このとき、experiment.db ファイルとして以下の内容のテキストファイルが出力されます。上記入力した “LogS” は以下の赤で出力されます。各カラムはスペース区切りです。

```

TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL15844  Log_S  -1.21  -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL15891  Log_S  -4.6   -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL14687  Log_S  0.62   -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL275626 Log_S  0.58   -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL278489 Log_S  -3.05  -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL279816 Log_S  -1.96  -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL103    Log_S  -4.42  -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL211456 Log_S  -0.47  -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL504760 Log_S  -1.96  -  1.0
TARGET  LogS  VER  ChEMBL631962  COMP  ChEMBL72     Log_S  -3.66  -  1.0
. . .

```

各カラム（スペース区切り）の意味は以下の通りです。

カラム	内容	計算で使用する
1	文字列 “TARGET” (固定)	×
2	入力画面の Target で入力した文字列	×
3	文字列 “VER” (固定)	×
4	ASSAY_ID	×
5	文字列 “COMP” (固定)	×
6	CMPD_CHEMBLID (化合物 ID)	○
7	STANDARD_TYPE (データ種別)	○
8	STANDARD_VALUE (データ値)	○
9	STANDARD_UNITS (単位)	×
10	重み (デフォルト 1.0)	○

※ ここで、データ種別に相当する第7カラムの文字列は、ユーザによる編集が必要になる場合があります。エクセルなどで編集してください。

計算プログラムでは、対数で回帰計算するために、対数になっていないデータ値に対して対数に変換する必要があるためです。

対数になっていないデータ値のデータ種別は、以下のテーブルのデータ種別に編集する必要があります。(上記の例 (Log\_S) のように、対数になっているデータについては編集しないで OK です。)

データ種別	変換式
S	sc = log(sc)/log(10.0)
P	
D	
Pa	if(sc .lt. 0.001) sc = 0.001
Papp	sc = log(sc)/log(10.0) -6.0
Pe	if (sc .le. -30.0 ) sc = -30.0
Peff	if (sc .ge. 30.0 ) sc = 30.0

対数になっていないデータ値の場合、  
溶解度・脂溶性などの物性値の場合  
**S, P, D** のいずれかのデータ種に変更してください。

膜透過に関するデータ値の場合、  
**Pa, Papp, Pe, Peff** のいずれかのデータ種に変更してください。

#### 4.2.3. 実験データファイルのデータ種を編集する必要がある例

前節の方法で作成した実験データファイルのデータ種を、ユーザが編集する必要がある例を以下で説明します。

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

この実験データファイルは、ChEMBL で検索するときのキーワードとして "Search ChEMBL" 欄に、「permeability」と入力してダウンロードした ChEMBL ダウンロードファイルから実験データファイルを作成した例ですが、データ種別が“permeability”になっていて、データ値は対数変換されていません。(Pa はユーザが Target で入力した文字列です。)

計算プログラムで精度良く計算するために、以下のようにエクセルなどの編集プログラムで、Pa, Papp, Pe, Peff のいずれかのデータ種に編集する必要があります。

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

なお、原理的には、溶解度・脂溶性、膜透過以外の特性値の回帰予測も計算できます。その場合は、もし対数変換されてない特性値の場合は、S, P, D または、Pa, Papp, Pe, Peff のいずれかのデータ種別に編集してください。対数変換されている特性値の場合は編集なしでそのまま計算できます。

#### 4.2.4. 実験データファイルのデータ種を編集する必要がない例

前節の方法で作成した実験データファイルのデータ種を、ユーザが編集する必要がない例を以下で説明します。

例 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
...
```

この実験データファイルの場合は、データ種別として“Papp”と“logPapp”が混在していますが、“logPapp”のデータ値はすでに対数化されていて、“Papp”のデータ値は、計算プログラム中で対数化されますので、編集する必要はありません。(Pa はユーザが Target で入力した文字列です。)

例 2)

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

この実験データファイルの場合は、データ種別が“LogD7.4”となっていますが、データ値はすでに対数化されているので、編集する必要はありません。(LogD はユーザが Target で入力した文字列です。)



#### 4.2.5. 実験データファイルをユーザがすべて編集する作成法

ChEMBL のダウンロードファイルを元にしないで、ユーザが自分で編集して、自分で実験したデータを使って、実験データファイルを作成することもできます。

その場合は、前節で説明した実験データファイルを一から編集してください。

各カラム（スペース区切り）の意味は以下の通りです。

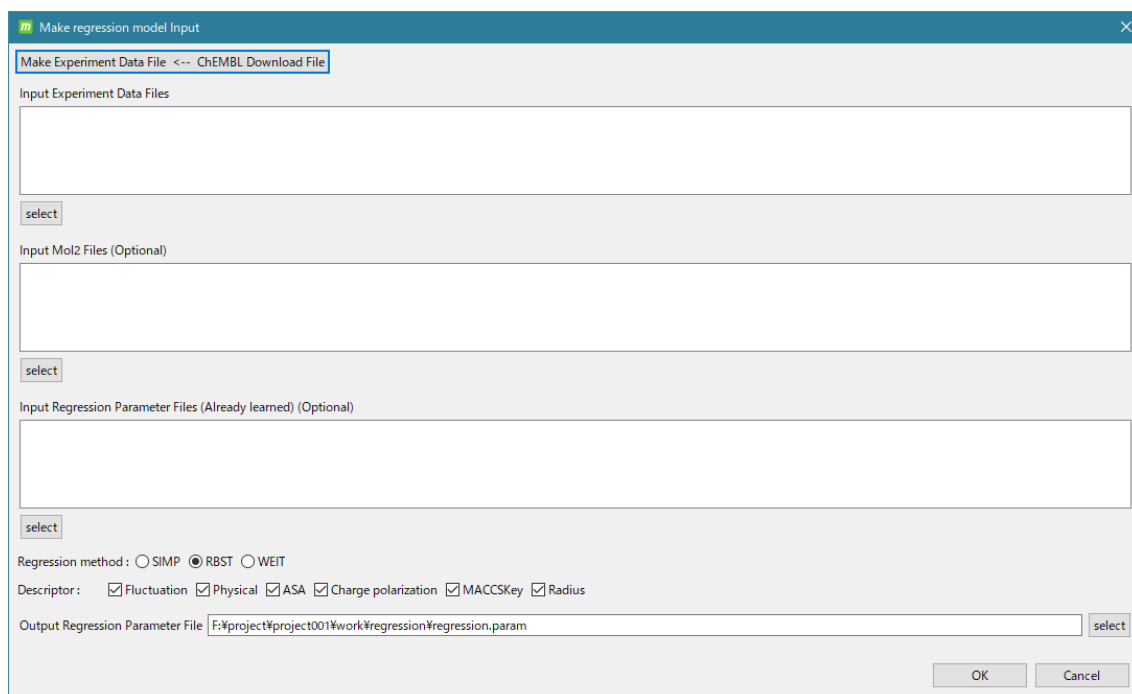
カラム	内容	計算で使用する
1	文字列 “TARGET” (固定)	×
2	入力画面の Target で入力した文字列	×
3	文字列 “VER” (固定)	×
4	ASSAY_ID	×
5	文字列 “COMP” (固定)	×
6	CMPD_CHEMBLID (化合物 ID)	○
7	STANDARD_TYPE (データ種別)	○
8	STANDARD_VALUE (データ値)	○
9	STANDARD_UNITS (単位)	×
10	重み (デフォルト 1.0)	○

この場合、第 6,7,8,10 カラムを重視して作成することになります。

## 4.3. 回帰パラメターの計算

### 4.3.1. 回帰パラメター計算のための入力項目

実験データファイルの作成が終了したら、[Preparation] - [Make Regression model] をクリックしたときの以下の画面に戻ります。

A screenshot of a software dialog box titled "Make regression model Input". The dialog has a blue header bar with a close button (X) in the top right. Below the header, there is a tabbed interface with two tabs: "Make Experiment Data File" (selected) and "ChEMBL Download File". The main area contains three input sections, each with a text box and a "select" button below it: "Input Experiment Data Files", "Input Mol2 Files (Optional)", and "Input Regression Parameter Files (Already learned) (Optional)". Below these sections, there are radio buttons for "Regression method": SIMP, RBST (selected), and WEIT. Underneath, there are checkboxes for "Descriptor": Fluctuation, Physical, ASA, Charge polarization, MACCSKey, and Radius, all of which are checked. At the bottom, there is a text field for "Output Regression Parameter File" containing the path "F:\project\project001\work\regression\regression.param" and a "select" button. Finally, there are "OK" and "Cancel" buttons at the bottom right.

各入力項目の内容は以下の通りです。

項目	内容
[Make Experiment Data File ← ChEMBL Download File] ボタン	ChEMBL でダウンロードしたデータファイルから、[Input Experiment Data Files] で入力する実験データファイルを作成するための画面を開く（前節で説明）。
Input Experiment Data Files	入力する実験データファイル（複数選択可）（入力必須）
Input Mol2 Files (Option)	<p>入力する mol2 ファイル（複数選択可）。</p> <p>入力がない場合は、ChEMBL sdfs を使用して計算。入力する場合は、実験データファイルに記載されている化合物の mol2 ファイルを入力する。</p> <p>（入力は必須でない）</p>
Input Regression Parameter Files (Already learned) (Option)	すでに過去に計算して得た回帰パラメーターファイル（複数選択可）（入力は必須でない）
Regression method	<p>回帰計算の方法</p> <ul style="list-style-type: none"> <li>・SIMP: 全データ同じ重み</li> <li>・RBST: ロバスト推定で重みを自動調整（デフォルト）</li> <li>・WEIT: 実験データファイルに記載した重みを使用</li> </ul>
Descriptor	<p>計算で使用する記述子の種類 （デフォルトはすべて ON）</p> <ul style="list-style-type: none"> <li>・Fluctuation: 物理量のゆらぎ・分散</li> <li>・Physical: 物理的なもの</li> <li>・ASA: ASA</li> <li>・Charge polarization: 水素結合数や水素結合しうる原子の電荷分極</li> <li>・MACCSKey: MACCSKey</li> <li>・Radius: 分子の半径（平均半径、3 極での Rgyr）</li> </ul>
Output Regression Parameter File	出力する回帰パラメーターファイル（入力必須）

この画面でファイル入力の項目は、

[Input Experiment Data Files]

[Input Mol2 Files (Option)]

[Input Regression Parameter Files (Already learned) (Option)]

[Output Regression Parameter File]

の4つですが、ファイル入力が必要の項目は、

[Input Experiment Data Files]

[Output Regression Parameter File]

の2つだけです。

[Input Experiment Data Files] では、前節の方法で作成した実験データファイルを入力します。複数入力可能です。

[Input Mol2 Files (Option)] では、ユーザが独自に作成した 化合物の mol2 ファイルを複数個入力して、化合物の descriptor 作成の計算に使用することができます。

この入力がない場合は、[Help] - [Preference] - [2.Screening] で設定した CheMBL sdfs を用いて化合物の descriptor を計算します。

ここで入力する mol2 ファイルには以下の制限があります。

- 1) ファイル名が、実験データファイルの第6カラムの “化合物 ID”.mol2 となっていること。
- 2) 1分子1ファイルであること。

[Input Regression Parameter Files (Already learned) (Option)] では、すでに本機能で過去に作成した回帰パラメーターファイルを入力できます。複数入力可能です。

[Output Regression Parameter File] では、出力する回帰パラメーターファイルのパスを設定します。デフォルトで

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

が設定されていますが、ファイル名を変えたい場合など、パスを変更したい場合は編集してください。

[Regression method] では、回帰計算の方法を選択します。WEIT (実験データファイルに記載した重みを使用) を選択した場合は、実験データファイルの第 10 カラムの重みを計算に使用しますので、ユーザが適宜、実験データファイルの重みを編集してください (以下の例の赤の部分)。

TARGET	Pa	VER	CHEMBL3430218	COMP	CHEMBL1294	Papp	28.65	10 <sup>-6</sup> _cm/s	1.0
TARGET	Pa	VER	CHEMBL3430218	COMP	CHEMBL421362	Papp	0.93	10 <sup>-6</sup> _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 <sup>-6</sup> _cm/s	1.0
TARGET	Pa	VER	CHEMBL3430218	COMP	CHEMBL3425629	logPapp	-1.05	-	2.0
. . .									

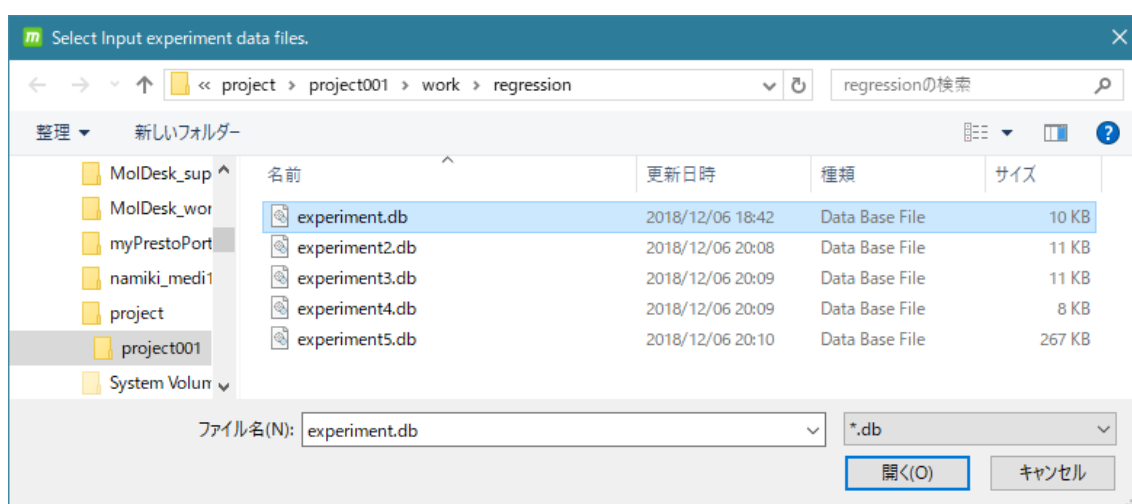
[Desriptor] では、化合物の記述子の計算で使用する種類を、ON/OFF で指定します。デフォルトでは、すべて使用 (ON) します。

### 4.3.2. 回帰パラメターの計算実行

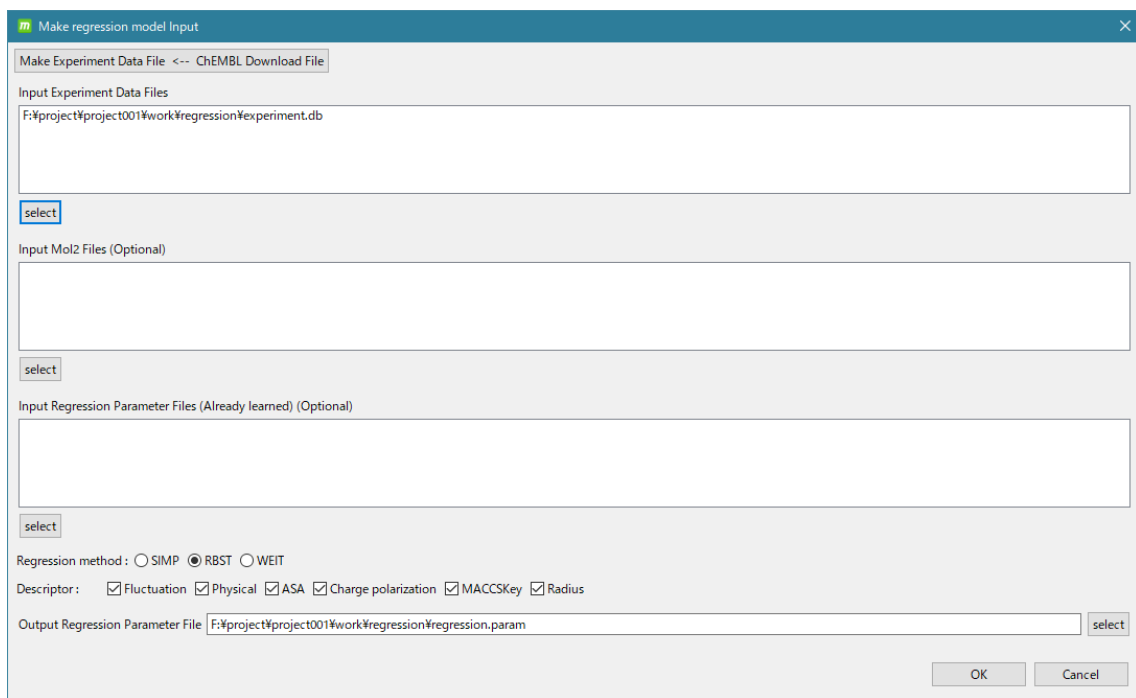
この例では、[Input Experiment Data Files] に、1 個の実験データファイルを入力して、それ以外はデフォルト値で計算を実行する場合を説明します。

[Input Experiment Data Files] の [Select] をクリックすると、以下のファイル選択画面が出るので、すでに前節の方法で作成した実験データファイルを選択します。

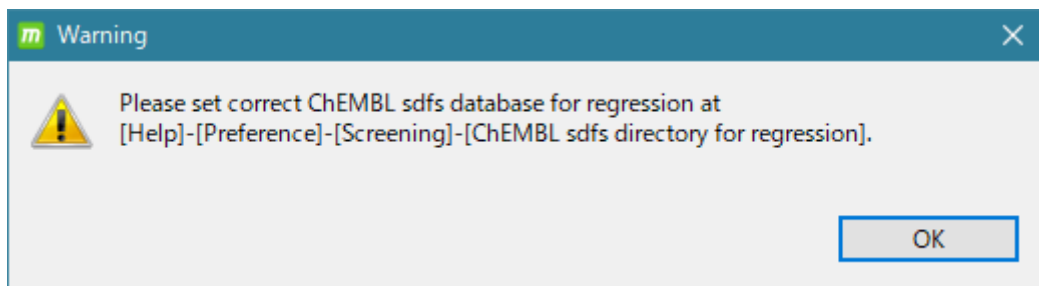
(複数選択できますが、ここでは 1 個だけ選択します。)



すると、以下の通り、選択した実験データファイルが取り込まれます。



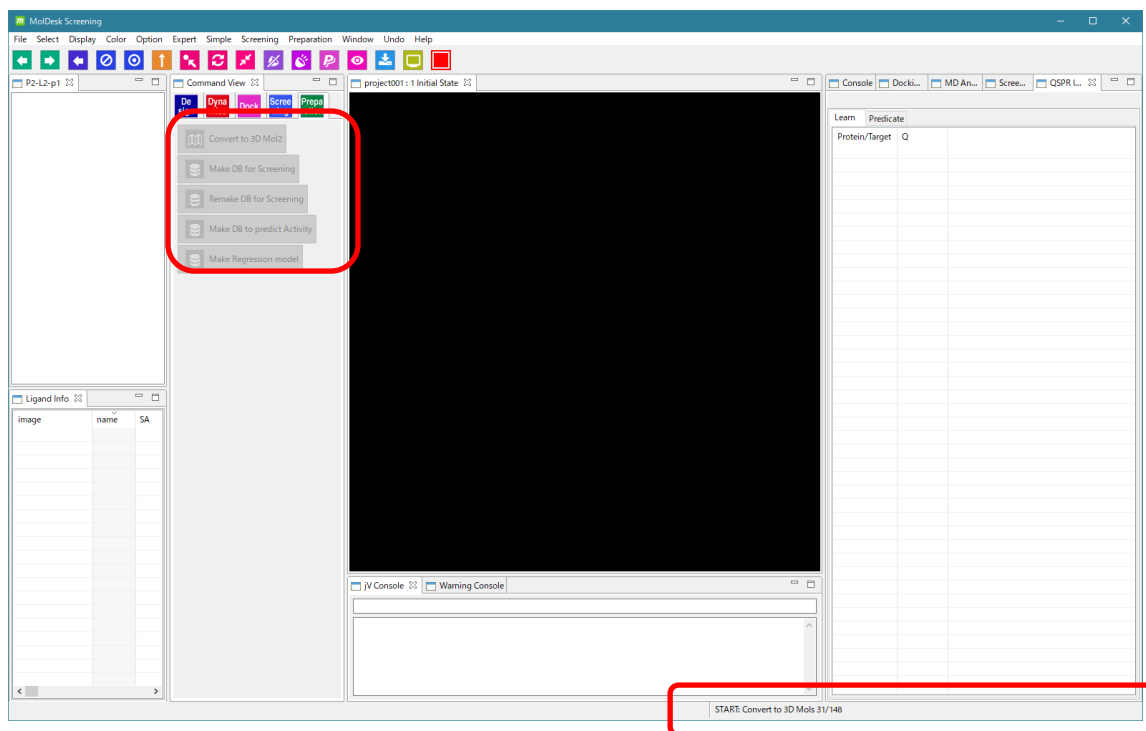
[OK] をクリックすると、ChEMBL sdfs を Preference で設定していない場合は、以下のワーニング画面が出ます。



[Input Mol2 Files (Option)] を入力していないので、化合物の descriptor を ChEMBL sdfs を使って計算しますので、ChEMBL sdfs の設定は必須になります。「[Screening](#)」での設定方法を参考にして、Preference で設定してください。

ChEMBL sdfs が Preference で設定されていると、[OK] をクリックした後に、(並列) 計算が始まります。

計算を開始するとコマンドボタンがグレーになります。コマンドボタンがグレーになっている間は計算中です。また、右下の赤枠内に、計算中の簡単な計算状況を表示します。



計算中でも他のプロジェクトを操作できますが、プロセッサの占有率によっては動作が極端に遅くなる場合がありますのでご注意ください。

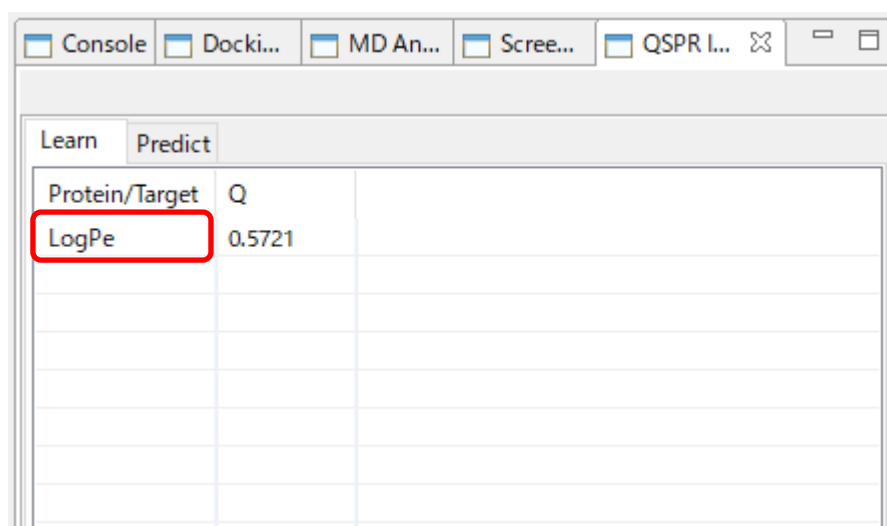
並列計算する際の並列数は、[Help] - [Preference] - [Screening] の Thread number で設定できます。詳細は「1.7 スクリーニング計算の並列数・メモリ量・時間」を参照してください。



### 4.3.3. 回帰パラメターの計算結果のグラフによる確認

回帰パラメターの計算が終了するとコマンドボタンがグレーから使用できるようになります。また、画面の右下に [END: Make Regression model] と表示します。

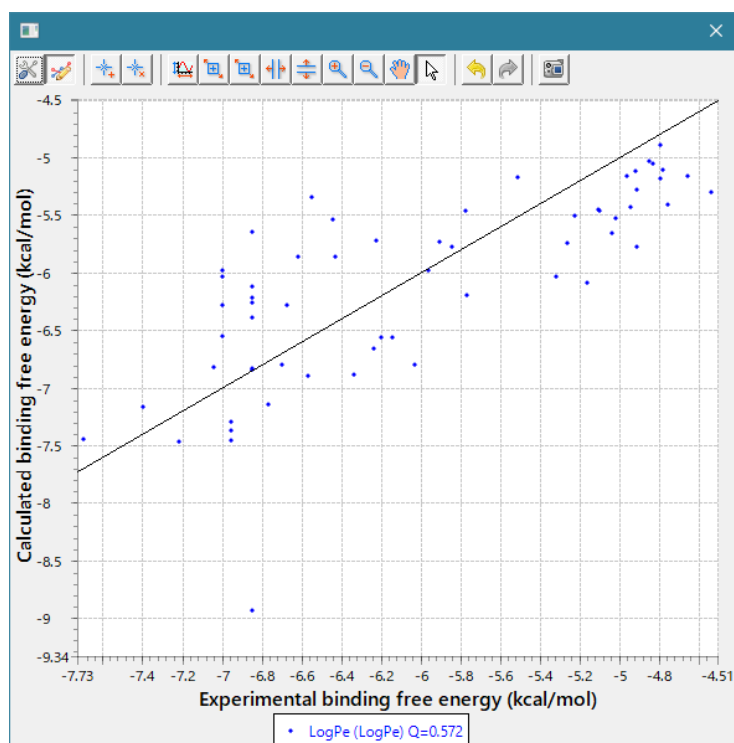
さらに、画面の右側に 下図のように QSPR Info 画面の [Learn] タブを表示します。回帰モデルを作成したときに入力した実験データファイルのデータ種別と、作成した回帰モデルの相関係数(Q 値)をリスト表示します。



The screenshot shows a software window titled 'QSPR Info' with several tabs: 'Console', 'Docki...', 'MD An...', 'Scree...', and 'QSPR I...'. The 'Learn' tab is active, displaying a table with two columns: 'Protein/Target' and 'Q'. The first row of the table has 'LogPe' in the 'Protein/Target' column and '0.5721' in the 'Q' column. The 'LogPe' cell is highlighted with a red rectangle.

Protein/Target	Q
LogPe	0.5721

ここで、上の赤枠の [Protein/Target] カラムのデータ種別名をダブルクリックします。すると、回帰パラメタを作成したときに使った実験データ値と、回帰パラメタ計算時の計算値のグラフを次のように表示して、学習の信頼性を確認することができます。



#### 4.3.4. 回帰パラメーターファイルの確認

計算結果の回帰パラメーターファイルの作成される場所は、[Output Regression Parameter File] で指定したパスになります。デフォルトでは、

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

です。このファイルは、次節で説明する特性値予測の計算で使用するので重要です。

なお、**regression** には以下のフォルダとファイルが生成されます。

これらの内容については、ユーザは気にする必要はありませんが、内容を説明すると以下の通りです。

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

項目	内容
input	入力した実験データファイル（必須）と回帰パラメーターファイル（入力があった場合だけ）が保存される。
mol2	化合物の mol2 ファイルと descriptor ファイルが保存される。descriptor 計算時の入力ファイル (*.inp) と、標準出力ファイル (*.stdout) など保存される。
mol2list_***	化合物ファイル名（拡張子なし）のリスト
error_MakeRegressionModel.log	回帰パラメーター計算時のエラー出力（エラーがあった場合）
learn.inp	学習計算時の入力ファイル
learn.out	学習計算時の標準出力ファイル
regression.param	学習計算で生成した回帰パラメーターファイル

## 4.4. 特性値の予測計算

前節で作成した回帰パラメーターのデータファイルを使って、複数の化合物の特性値を一度に（並列）計算します。

### 4.4.1. 特性値予測計算の実行

[Screening] - [Predict with Regression model]



をクリックします。以下の入力画面が出ます。

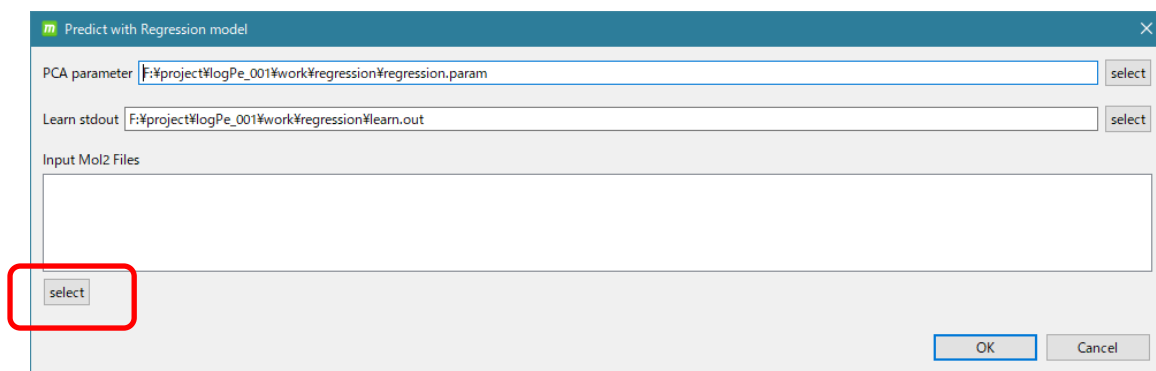
回帰パラメーターファイル（前節の方法で作成した）

学習計算時の標準出力ファイル（前節の方法で作成した）

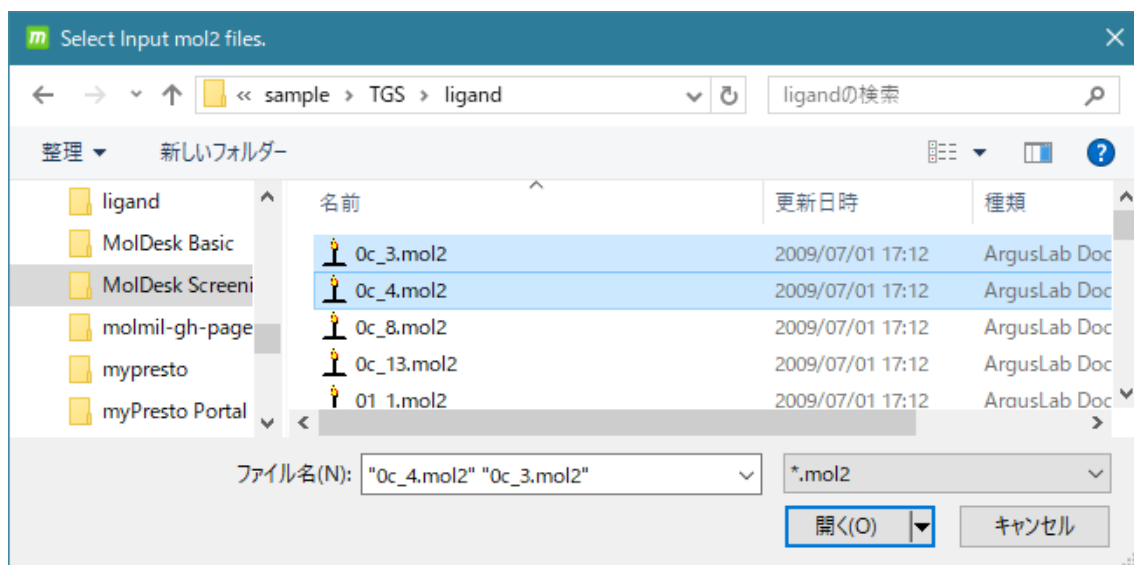
予測したい化合物の **mol2** ファイル

を入力して、[OK] をクリックします。

回帰パラメーターファイルと学習計算時の標準出力ファイルはデフォルトのパスがすでに記入されていますが、もし異なる場合は赤枠の [select] をクリックして表示されるファイル選択画面で、正しいファイルを選択して変更してください。

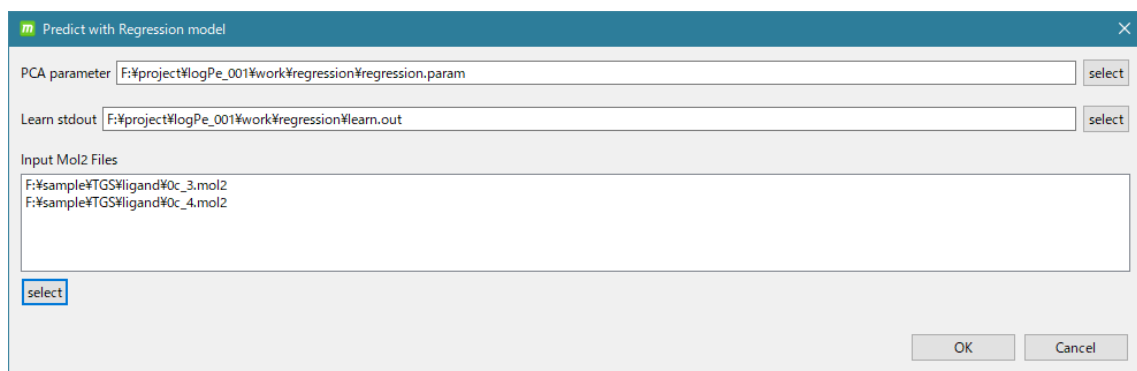


次に、上図の赤枠の [select] をクリックして、活性値を予測したい化合物の mol2 ファイルを下図で選択します。



ここでは例として、上の 2 個を選択しました。

- ※ 化合物の mol2 ファイルは、必ず 1 分子だけを含むようにしてください。MolDesk Screening の、[Preparation] – [Convert to 3D Mol2] で作成した Mol2 ファイルを入力として使用できます。
- ※ 近い将来に機能追加して、sdf ファイルで複数の分子を一括入力可能にする予定です。



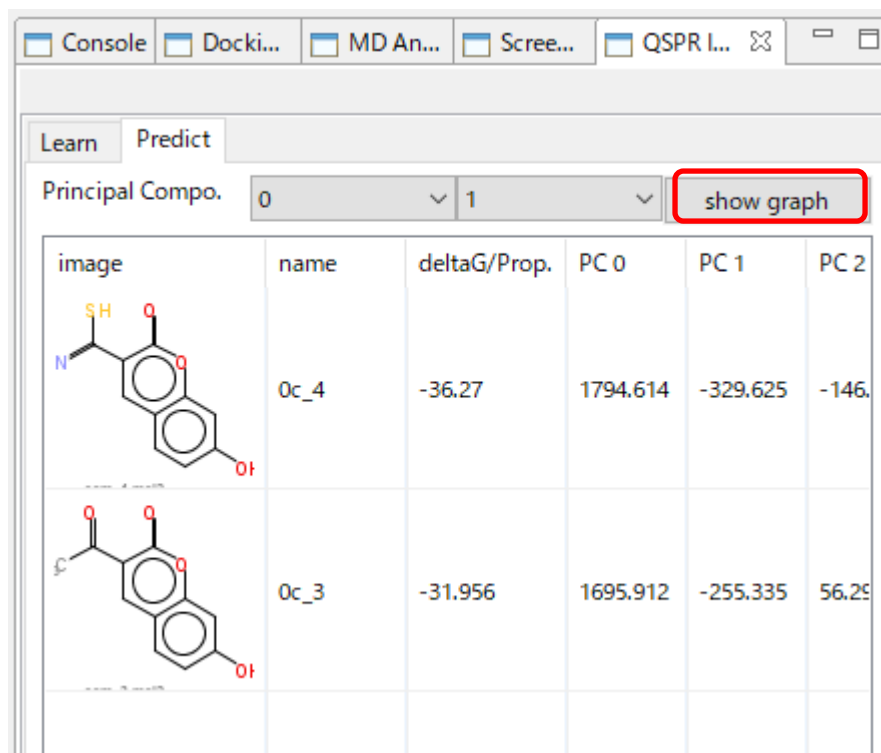
入力が終わると上図のようになりますので、[OK] をクリックすると、予測計算が始まります。

予測計算中は、画面の右下に計算の状況が表示されます。

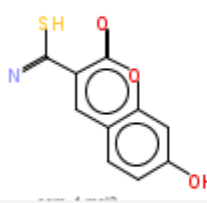
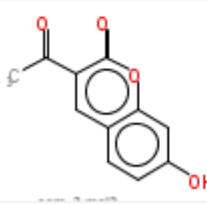
#### 4.4.2. 特性値予測計算の結果確認

特性値予測の計算が終了するとコマンドボタンがグレーから使用できるようになります。また、画面の右下に [END: Predict with Regression model] と表示します。

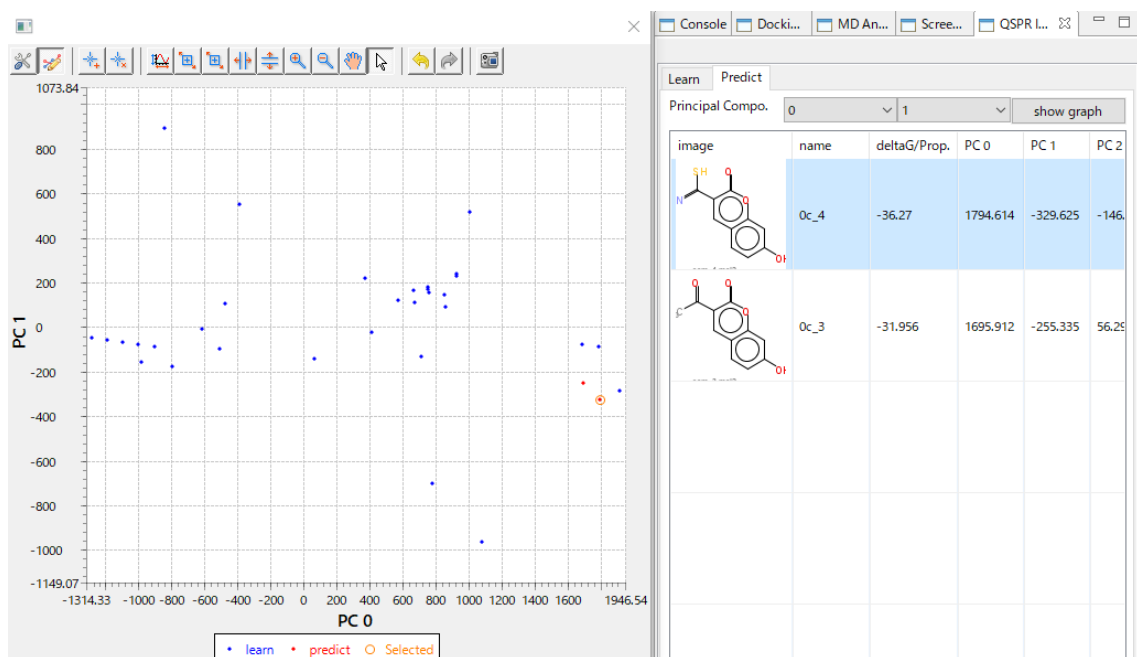
さらに、画面の右側に 下図のように QSPR Info 画面の [Predict] タブを表示します。



The screenshot shows the 'Predict' tab in the QSPR Info window. It features a table with columns for chemical structure (image), name, deltaG/Prop., and the first three principal components (PC 0, PC 1, PC 2). Two chemical structures are listed: 'Oc\_4' and 'Oc\_3'. A red box highlights the 'show graph' button in the top right corner of the table area.

image	name	deltaG/Prop.	PC 0	PC 1	PC 2
	Oc_4	-36.27	1794.614	-329.625	-146.
	Oc_3	-31.956	1695.912	-255.335	56.25

ここで、上の赤枠の [show graph] ボタンをクリックすると、PCA グラフを表示します。図の例では、0 軸と 1 軸の PCA グラフを表示します。軸は、0～9 まで任意に選択できます。



リストの化合物をクリックすると、PCA グラフの中で、予測した化合物の位置を赤丸で確認できます。

逆に、グラフの赤丸をクリックすると、リストの化合物にフォーカスします。

実験データファイル由来の化合物とかけ離れてないか確認できるので、計算の信頼性を評価できます。

予測した特性値（対数値換算）は、テーブルの deltaG/Prop. のカラムに表示します。



## 5. MVO Screening

MVO Screening 法を用いた化合物分子の類似化合物探索を実施します。  
指定した化合物に類似する化合物を、mol2 ファイルで入力した化合物の中から探索します。

### MVO Screening : (旧名 : MD MVO, 別名 : MIN-MVO)

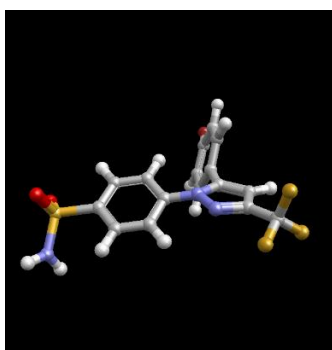
2つの分子の立体的な重ね合わせにより、重なり大きいものを類似性が高いとします。  
重ね合わせでは、分子の配座の発生と、原子電荷の類似も考慮して、エネルギー最小化を用いた重ね合わせを行い、スコアは体積重なり率の値です。

#### 5.1. クエリ分子の選択


クエリとして1つの化合物を選択します。この分子が検索のクエリになります。

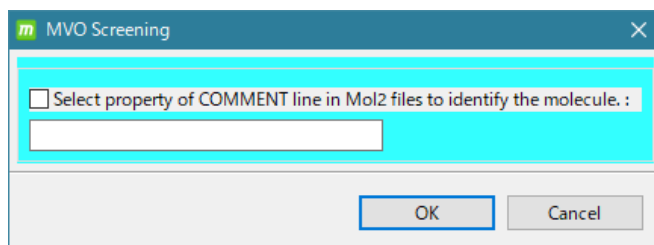
[File] - [Open Molecular File] で、MolDesk のインストール時にデスクトップに作成された MolDesk Screening フォルダの中にある以下のファイルを読み込みます。

MolDesk Screening -> sample -> MVO\_screening -> query -> 1cx2\_1.mol2



## 5.2. 検索対象分子の選択

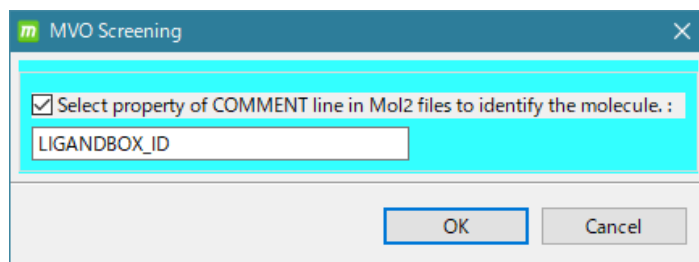
ここで、 [MVO Screening] をクリックすると、以下の画面を表示します。



ここでは、mol2 ファイルの COMMENT 行（もし存在すれば）の **property** を記入します。この **property** 値が、計算結果リストの ID となるので、分子を紐付けすることができるようになりますので、分子の区分に使いたい **property** 値がある場合はチェックして **property** 値を記入して [OK] をクリックしてください。

COMMENT 行が存在しない場合は、そのまま [OK] をクリックして次に進んでください。その場合はデフォルトで、@<TRIPOS>MOLECULE 行の次の行の分子名が ID となります。

※ mol2 ファイルは、sdf ファイルなどから、[Convert to 3D Mol2] コマンドで作成してください。

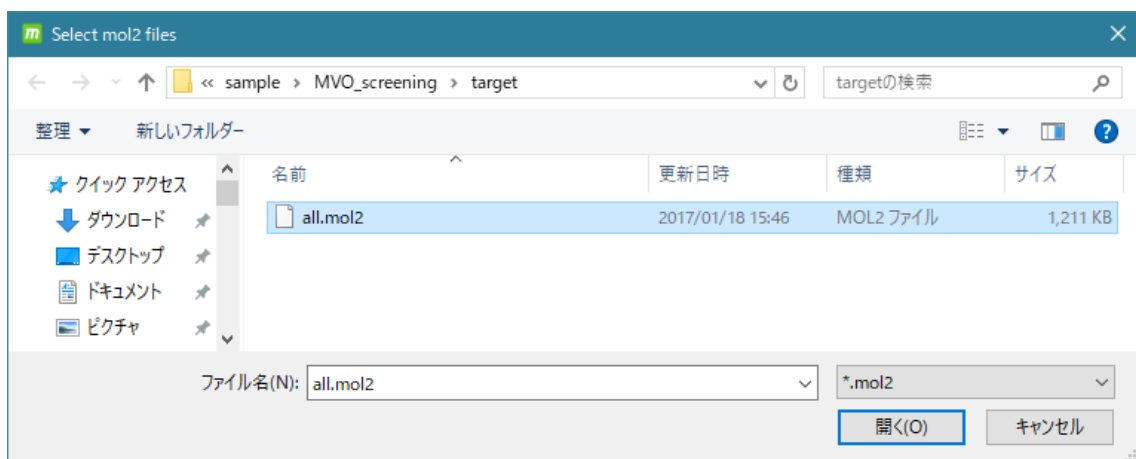


記入例は、上図のとおりです。（ID としたい **property** として、LIGANDBOX\_ID を記入し、チェックボックスにチェックを入れた。）

[OK] をクリックすると、次図のファイル選択画面を表示します。

ここでは、以下のフォルダ内の mol2 ファイルを選択します。「開く」をクリックすると MVO Screening が開始します。

MolDesk Screening -> sample -> MVO\_screening -> target



※ 例では、1 ファイルですが、複数ファイルも選択できます。

### 5.3. 結果表示

計算が終了すると [Screening Info] 画面に結果が表示されます。

スコアは、SMVO-Q、SMVO-D、SMVO-T の 3 通りで、それぞれソート可能です。

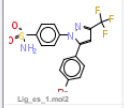
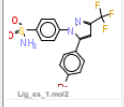
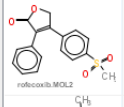
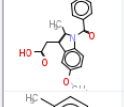
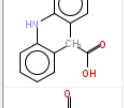
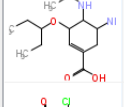
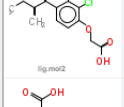
-1.0 が完全一致、値が大きい（絶対値が小さい）ほど類似性が高くなります。

それぞれのスコアの特徴は以下の通りです。

- |          |                             |
|----------|-----------------------------|
| SMVO-Q : | データベース中の小さい分子が選ばれる          |
| SMVO-D : | データベース中の大きな分子が選ばれる          |
| SMVO-T : | クエリー分子に大きさの近いデータベースの分子が選ばれる |

Console Docki... MD An... Scree... QSPR L...

Export table

Image	ID	S_MVO-D	S_MVO-Q	S_MVO-T
	Lig_es_1.mol2	-1.0	-1.0	-1.0
	Lig_es_1.mol2	-0.9972	-0.9917	-0.989
	rofecoxib.M	-0.7668	-0.8197	-0.6561
	Lig_es_1.mol2	-0.6311	-0.7549	-0.5238
	Lig_es_1.mol2	-0.7371	-0.6313	-0.5153
	2qwk.mol2	-0.5445	-0.7083	-0.4448
	lig.mol2	-0.6519	-0.5707	-0.4374

その他、以下の分子特性も表示します。

Formula, Weight, Charge, Donor, Acceptor, Chiral atoms

また、[Export table] ボタンをクリックすると、csv ファイル、html ファイルに結果を保存できます。

## 6. 類似構造検索

TGS (Topology Graph Similarity) 法を用いた化合物分子の類似構造検索を実施します。指定した化合物に構造が類似する化合物を、mol2 ファイルで入力した化合物の中から検索します。

### Topology Graph Similarity :

分子の共有結合をエッジとした分子グラフをエッジ行列表示とし、その行列固有値を指標として化合物の類似性を探索する手法です。

分子の構造情報を実数値のベクトルへ変換し、ベクトルの距離から類似性を計算します。非常に高速ですが、光学異性体、配座を区別することはできません。

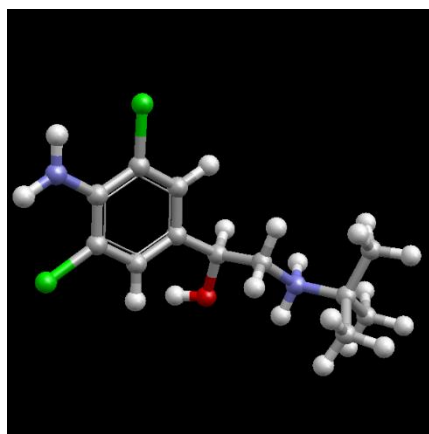
### 6.1. クエリ分子の選択

クエリとして1つの化合物を選択します。この分子が検索のクエリになります。

[File] - [New Project] で、MolDesk のインストール時にデスクトップに作成された MolDesk Screening フォルダの中にある以下のファイルを読み込みます。

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

[New Project] については MolDesk Basic のマニュアルを参照してください。

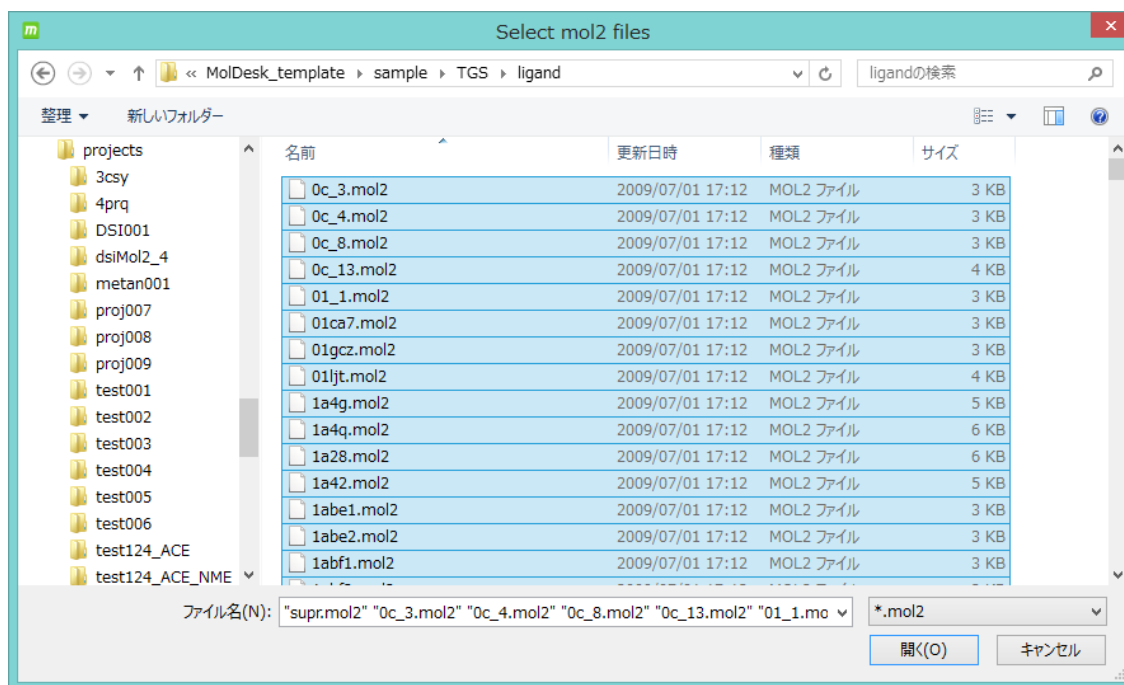


## 6.2. 検索対象分子の選択・計算・結果表示



[Topology Graph Similarity] をクリックし、以下のフォルダ内にあるすべての mol2 ファイルを選択します。「開く」をクリックすると類似構造検索が開始します。

MolDesk Screening -> sample -> TGS -> ligand



計算が終了すると [Console] 画面に結果が表示されます。

左から順位、化合物名、スコア (0.0 が完全一致、値が小さいほど類似性が高くなります) が表示されます。

The screenshot displays a docking software interface. On the left, a 3D molecular model shows a ligand (green and white) docked into a protein's binding pocket (grey and white). The protein structure includes a benzene ring and various side chains. On the right, a console window titled 'Standard Output' shows the following text:

```
calculate DB enrichment? ( y or n )
readmol3=
..\1\inp0.mol2

numatom          36          36
=== predicted compounds ===
1 .\ligand\abclen.mol2          0.0000000000
2 .\ligand\absalb.mol2          0.000026141
3 .\ligand\0303_1.mol2          0.0000161131
4 .\ligand\0303_2.mol2          0.0000161131
5 .\ligand\ickp.mol2            0.0000189301
6 .\ligand\01_1.mol2            0.0000233246
7 .\ligand\s_Sohdpa.mol2        0.0000235670
8 .\ligand\3cpa.mol2            0.0000236342
9 .\ligand\lphg.mol2            0.0000239249
10 .\ligand\0igcz.mol2          0.0000267501
11 .\ligand\lmif1.mol2          0.0000268643
12 .\ligand\ldrl.mol2           0.0000272029
13 .\ligand\icle.mol2           0.0000281120
14 .\ligand\abpropr.mol2        0.0000318040
15 .\ligand\0c_3.mol2           0.0000347880
16 .\ligand\2ada.mol2           0.0000356047
17 .\ligand\0c_4.mol2           0.0000362421
18 .\ligand\2ifb.mol2           0.0000368928
19 .\ligand\d_quinpi.mol2       0.0000370373
20 .\ligand\0c_8.mol2           0.0000374566
21 .\ligand\abmetop.mol2        0.0000387219
22 .\ligand\abateno.mol2        0.0000421454
23 .\ligand\keto.mol2           0.0000434973
24 .\ligand\livb.mol2           0.0000444297
25 .\ligand\abterb.mol2         0.0000450286
26 .\ligand\abalpre.mol2        0.0000450969
27 .\ligand\d_apomor.mol2       0.0000458711
28 .\ligand\abpindo.mol2        0.0000484559
29 .\ligand\ld3h.mol2           0.0000485247
30 .\ligand\lok1.mol2           0.0000486159
31 .\ligand\d_skf383.mol2        0.0000489636
32 .\ligand\Di1unisal.mol2       0.0000496398
33 .\ligand\d_dobuta.mol2        0.0000548639
34 .\ligand\ldala.mol2           0.0000555880
35 .\ligand\llala.mol2           0.0000555880
36 .\ligand\s_sumatr.mol2        0.0000567821
37 .\ligand\lf3d.mol2           0.0000573121
38 .\ligand\01ca7.mol2          0.0000580679
39 .\ligand\3cla.mol2           0.0000595256
40 .\ligand\lfen.mol2           0.0000595422
41 .\ligand\DHICA.mol2          0.0000627675
42 .\ligand\D_dopachrome.mol2    0.0000635609
```

Below the console window, a 'JV Console' window shows the following text:

```
Distance #1-#2: 0.0
Atom #1: CA 258 Residue: PHE 688 Chain(auth/label): A/? File: 1
Coordinates: ( 16.854, 15.448, 189.056 )
Atom #2: CA 258 Residue: PHE 688 Chain(auth/label): A/? File: 1
Coordinates: ( 16.854, 15.448, 189.056 )
Distance #1-#2: 0.0
Atom #1: CA 258 Residue: PHE 688 Chain(auth/label): A/? File: 1
Coordinates: ( 16.854, 15.448, 189.056 )
```

## 7. 部分構造検索

部分構造検索 (Substructure Search) では、指定した化合物に部分構造が類似する化合物を、mol2 ファイルで入力した化合物の中から検索します。

分子を、化学結合をエッジとするエッジ行列に変換し、ウルマンの定理により部分構造を比較します。分子の配座や光学異性体は考慮しません。

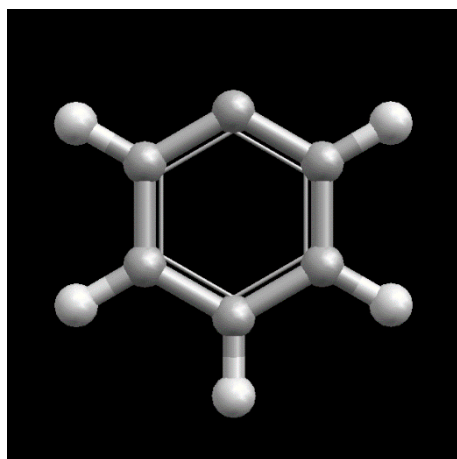
### 7.1. クエリ分子の選択

クエリとして 1 つの化合物を選択します。この分子が検索のクエリになります。

[File] - [New Project] で、MolDesk のインストール時にデスクトップに作成された MolDesk Screening フォルダの中にある以下のファイルを読み込みます。

MolDesk Screening -> sample -> substructure\_search -> query -> lig1.mol2

[New Project] については MolDesk Basic のマニュアルを参照してください。



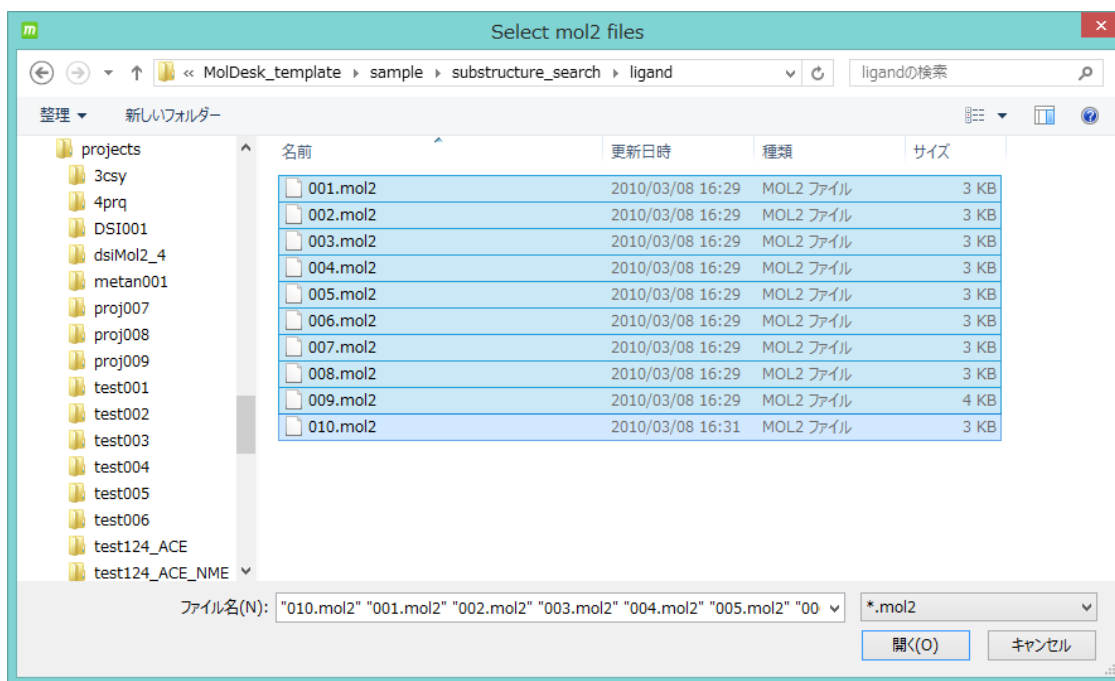


## 7.2. 部分構造検索対象分子の選択・計算・結果表示



[Substructure Search] をクリックし、以下のフォルダ内にあるすべての mol2 ファイルを選択します。「開く」をクリックすると類似構造検索が開始します。

MolDesk Screening -> sample -> substructure\_search -> ligand



計算が終了すると [Console] 画面に結果が表示されます。

左から内部番号、ファイル名、原子数、検索した部分構造の番号、見つかった部分構造の数が表示されます。

INFORMATION>				
WILDCARD ATOMS				
SUBSTRUCTURE		ATOMS		
1	..\inp0.mol2			11
NO	FILE_NAME	ATOM	SUB_NO	PATTERN
#	1 .\ligand_SSS\001.mol2	21	1	1
#	2 .\ligand_SSS\002.mol2	19	1	1
#	3 .\ligand_SSS\003.mol2	20	1	1
#	4 .\ligand_SSS\004.mol2	20	1	1
#	5 .\ligand_SSS\005.mol2	18	1	1
#	6 .\ligand_SSS\006.mol2	24	1	1
#	7 .\ligand_SSS\007.mol2	18	1	1
#	8 .\ligand_SSS\008.mol2	24	1	1
#	9 .\ligand_SSS\009.mol2	26	1	1
#	10 .\ligand_SSS\010.mol2	21	1	1

## 8. MPI / GPU による MD 計算の高速並列計算

myPresto の以下 4 つの分子動力学計算プログラムと GROMACS を実行することができます。

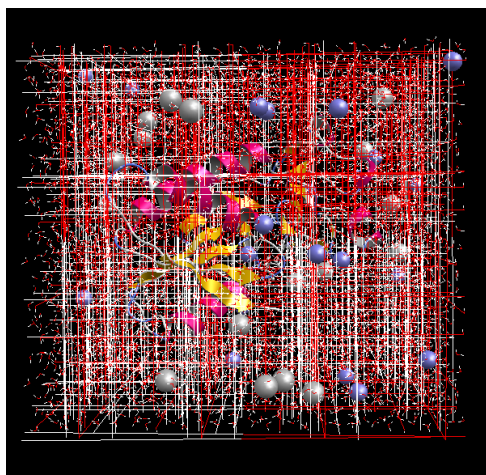
※ Mac は、myPresto による分子動力学計算の MPI や CUDA による並列計算をサポートしません。

MD プログラム	必要な動作環境	MD 計算の機能
cosgene	・ 特になし	・ すべての MD 計算が可能
cosgene_MPI	・ MPI ・ 64bit 限定	・ すべての MD 計算が可能
psygene	・ MPI ・ 64bit 限定	・ 周期的境界条件 (※) 以外の計算は不可 ・ 水溶媒の直方体の 1 辺のサイズが 54 Å 未満は不可
psygene-G	・ MPI ・ CUDA ・ 64bit 限定	・ 周期的境界条件 (※) 以外の計算は不可 ・ 水溶媒の直方体の 1 辺のサイズが 54 Å 未満は不可
GROMACS	Windows 64bit Linux / MAC はユーザのインストール・実行環境に依存	・ 周期的境界条件 (※) 以外の計算は不可

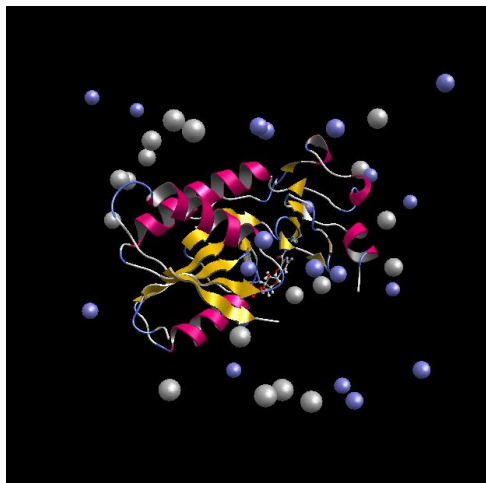
※Cube で水溶媒を生成

- cosgene\_MPI および psygene では、MPI を設定する必要があります。
- psygene-G では、MPI と CUDA を設定する必要があります。
- 現在 MPI は多ノードに非対応です。1 ノード上のマルチコアで動作します。
- psygene-G では、1 ノードあたり 4 つまでのマルチ GPU に対応します。
- psygene-G には NVIDIA のグラフィックスボードが必要です。
  - GF100 世代以降 (GTX460 以上、Compute Capability 2.0 以上) が必要です。

- ビデオメモリが多いほど大規模問題が計算できます。
- psygene-G では、SHAKE を使わない場合に、周期的境界にある水分子の表示が下記のように線状になりますが、計算の異常ではありません。



この場合は、水分子をツリー表示画面で右クリックで選択して、[Hide Atom] メニューで表示を隠してください。以下のように、その他の分子がきれいに表示します。



## 8.1. MPI 動作環境の設定方法

### 8.1.1. Windows 64bit

マイクロソフトの MS-MPI をインストールします。

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

から、[Download]をクリックして `msmpisetup.exe` をダウンロードします。ダウンロードした `msmpisetup.exe` をダブルクリックするとインストールが完了します。環境変数の設定も同時に行われます。

### 8.1.2. Linux 64bit

Open MPI (<https://www.open-mpi.org/>) または MPICH (<https://www.mpich.org/>) をインストールします。手順はそれぞれのマニュアルを参照してください。

Open MPI のインストールコマンドは以下の通りです。

Debian 系 64bit Linux の場合

```
$ sudo apt-get install openmpi-bin libopenmpi-dev
```

Redhat 系 64bit Linux の場合

```
$ yum install openmpi openmpi-devel
```

Ubuntu の場合は、環境設定も同時に完了しますが、CentOS の場合は、`~/.bash_profile` ファイルなどに、`export PATH=$PATH:/usr/lib64/openmpi/bin/` とパス設定が必要です。

## 8.2. CUDA 動作環境の設定方法

NVIDIA のグラフィックスボードが PC に装備されていることが必要です。

グラフィックスボードについては、GF100 世代以降（GTX460 以上、Compute Capability 2.0 以上）が必要です。ビデオメモリが多いほど大規模問題が計算できます。

### 8.2.1. Windows 64bit

NVIDIA のグラフィックドライバの最新版がインストールされていれば動作します。

グラフィックドライバは以下からダウンロードしてください。

<https://www.nvidia.co.jp/Download/index.aspx?lang=jp>

### 8.2.2. Linux 64bit

NVIDIA のグラフィックドライバの最新版がインストールされていれば動作します。

グラフィックドライバは以下からダウンロードしてください。

<https://www.nvidia.co.jp/Download/index.aspx?lang=jp>

### 8.3. Preference の設定

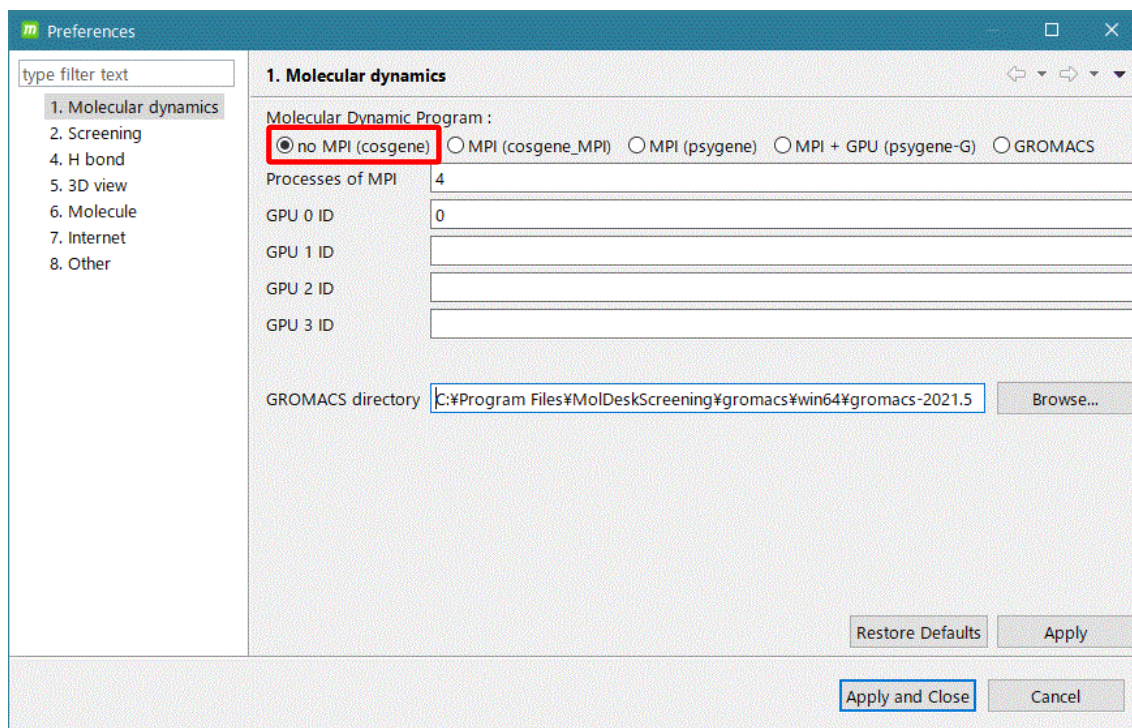
[Help] - [Preference] で各種 Preference 値を設定できます。

ここでは、MolDesk Screening でのみ設定が必要な「Molecular Dynamics」および「Screening」の項目についてのみ説明します。この他の項目については MolDesk Basic のマニュアルを参照してください。

#### 8.3.1. Molecular Dynamics

MD 計算の環境設定を行います。各選択肢の説明を以下に示します。

- not MPI (cosgene)

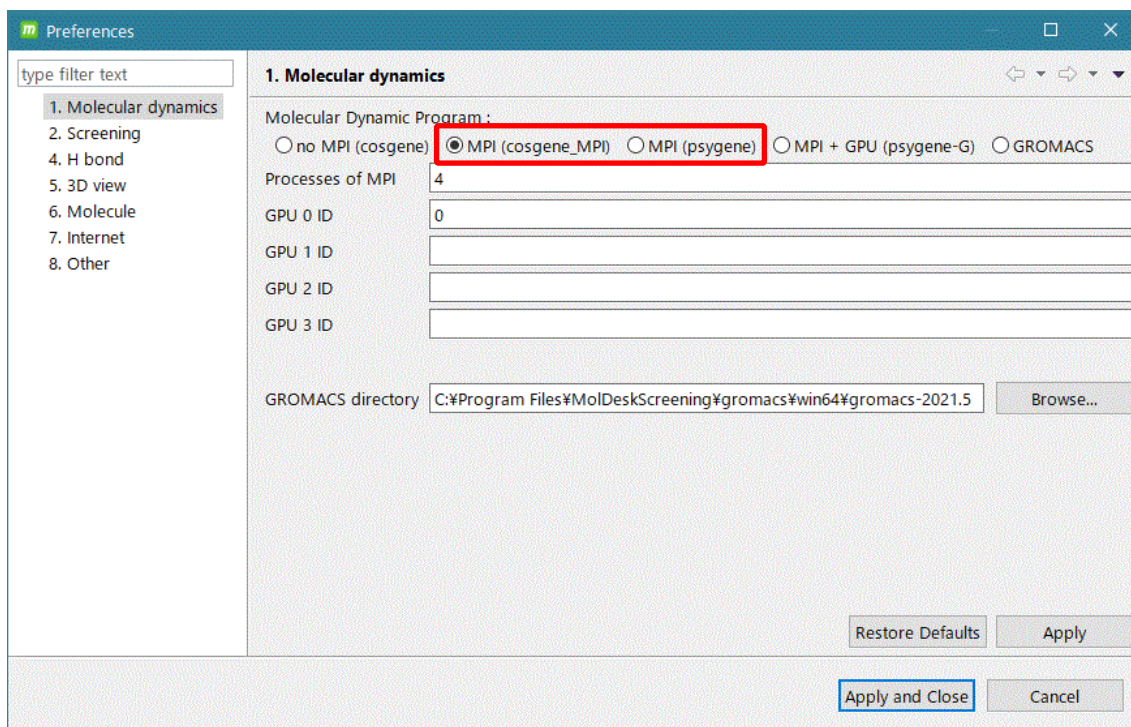


[not MPI (cosgene)] を選択したときは、並列計算は行いません。MPI などの環境設定も不要です。

[Processes of MPI] や  
[GPU 0 ID] [GPU 1 ID] [GPU 2 ID] [GPU 3 ID]  
の設定は無視されます。



- MPI (cosgene\_MPI) または MPI (psygene)



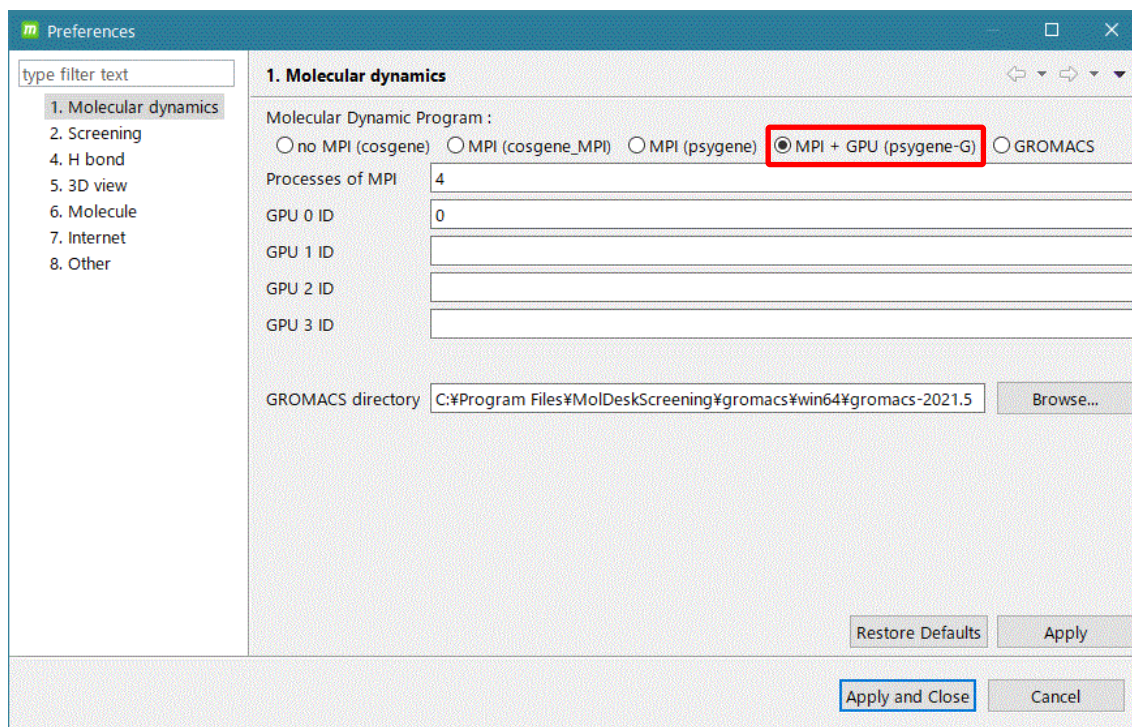
[MPI (cosgene\_MPI)] または [MPI (psygene)] を選択したときは、MPI による並列計算を実行します。

[Processes of MPI] で MPI 並列数を設定します。デフォルト値は、インストールしたシステムの最大物理プロセッサ数です。通常はこの値を変更する必要がありません。

[GPU 0 ID] [GPU 1 ID] [GPU 2 ID] [GPU 3 ID]  
の設定は無視されます。

※ Mac または Windows 32bit では、MPI, CUDA による MD 計算の並列計算をサポートしてないので、この設定画面は表示しません。

- MPI + GPU (sievgene-G)



[MPI + GPU (sievgene-G)] を選択したときは、MPI + GPU による並列計算を実行します。

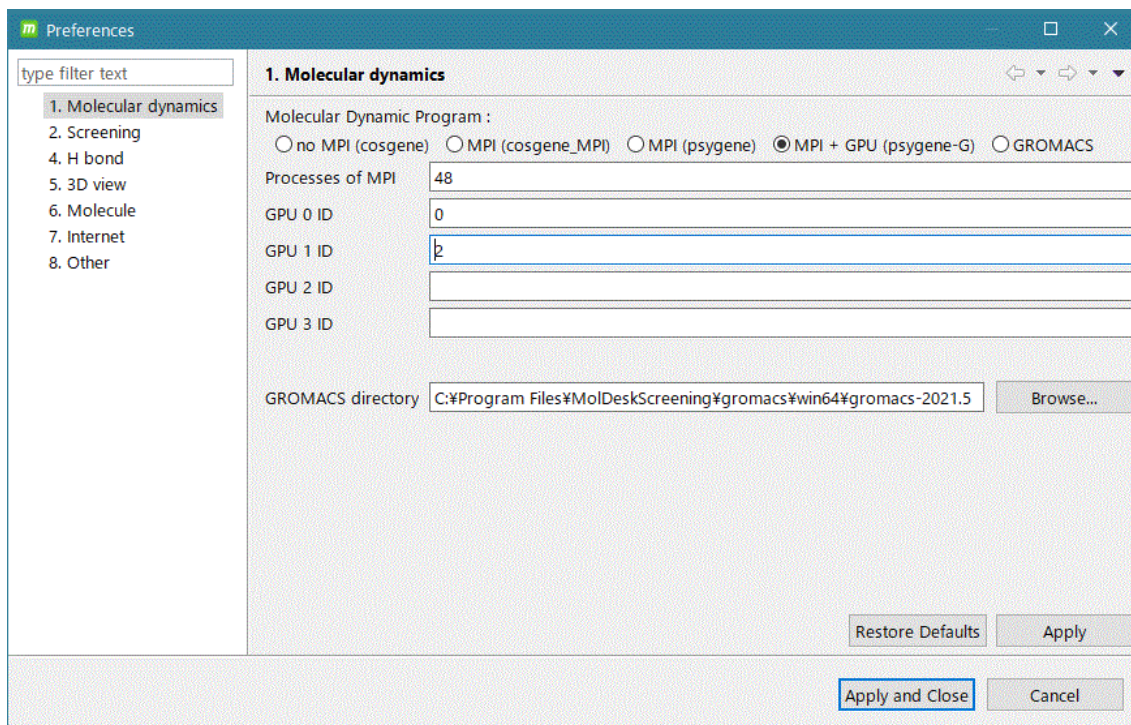
[Processes of MPI] で MPI 並列数を設定します。デフォルト値は、インストールしたシステムの最大物理プロセッサ数です。通常はこの値を変更する必要はありません。

[GPU 0 ID] [GPU 1 ID] [GPU 2 ID] [GPU 3 ID]

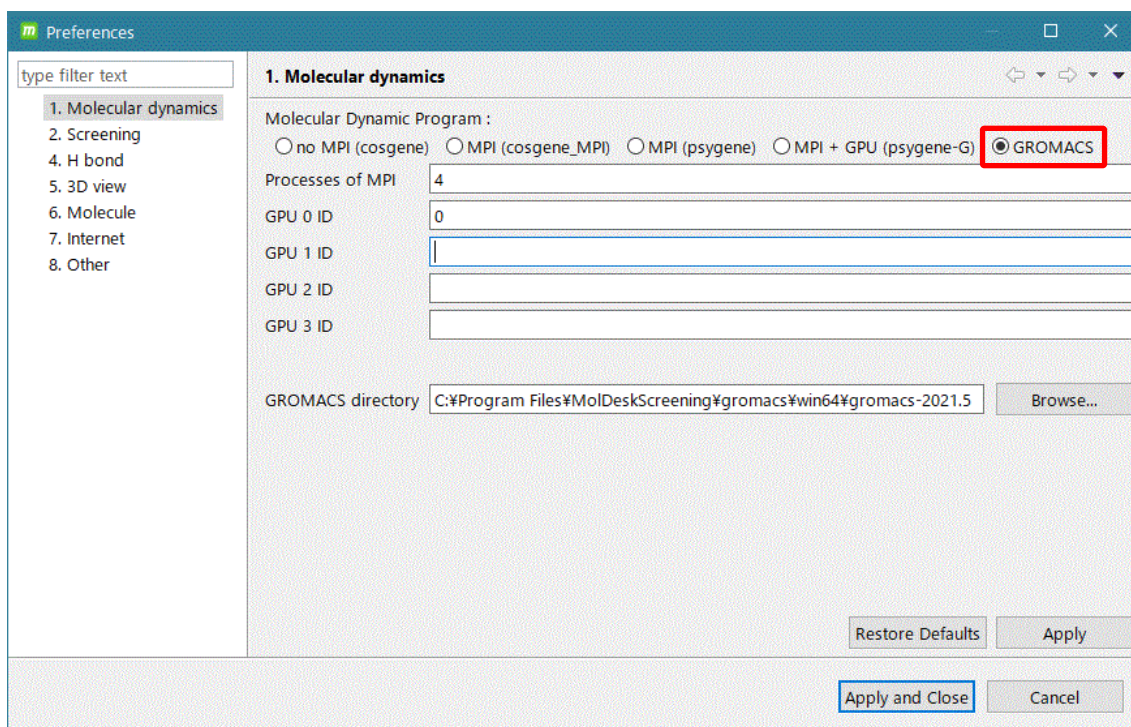
で使用する GPU の Device ID を設定します。使用する GPU の Device ID は 4 個まで設定可能で、最大 4 台のマルチ GPU 計算が可能です。

Device ID が 0 と Device ID が 2 の 2 個の GPU ボードを使用する例です (MPI 並列数は 48)。





## ● GROMACS

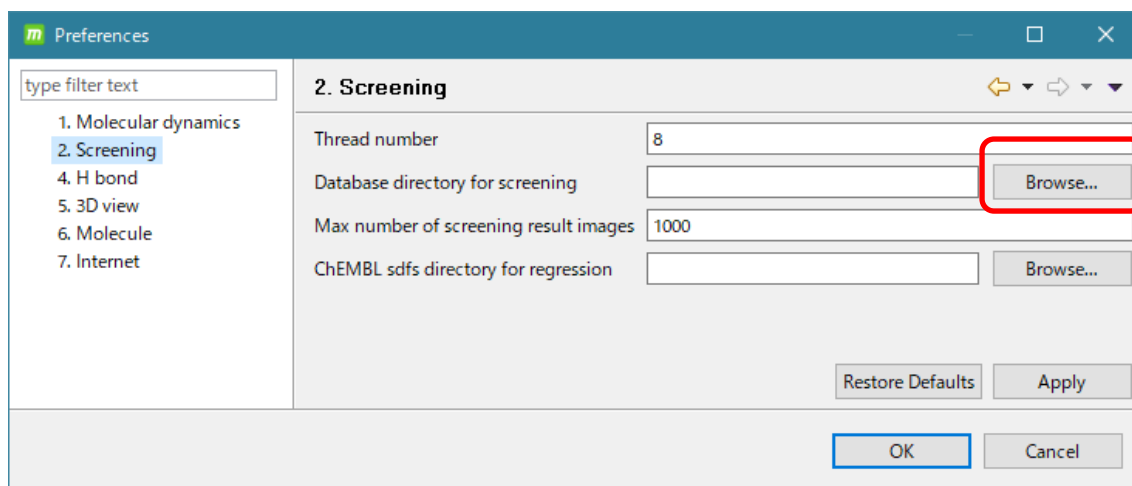


[GROMACS] を選択したときは、GROMACS による MD 計算を実行します。  
 使用法の詳細は、MolDesk Basic マニュアルを参照してください。

### 8.3.2. Screening

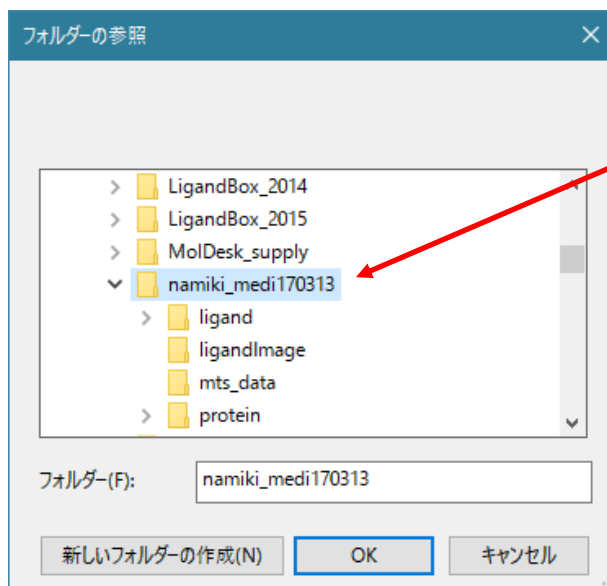
スクリーニング計算に用いる **LigandBox** またはユーザ作成のデータベース設定を行います。

[Help] - [Preference] 画面を開いて、「2. Screening」を選択し、[Browse] をクリックします。

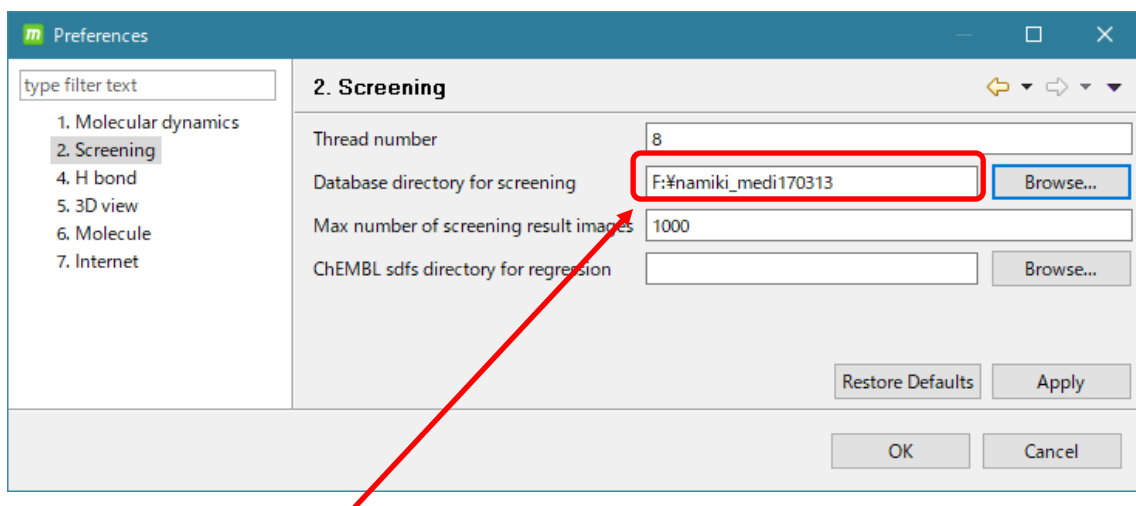


「1.2 LigandBox の準備」で解凍した **LigandBox**（下の例では、**namiki\_medi170313**）の

ligand  
ligandImage  
mts\_data  
protein

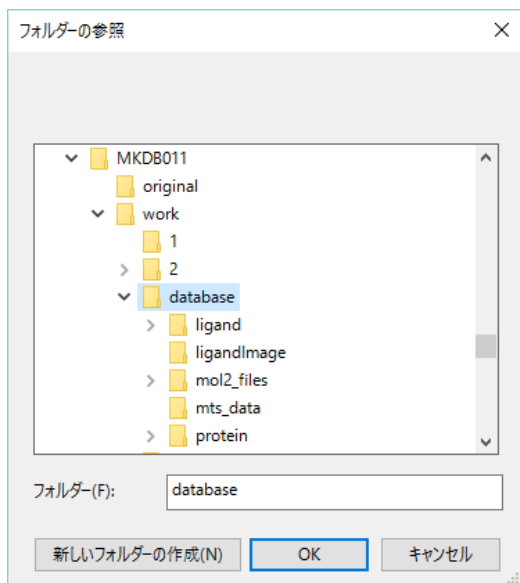


フォルダ（ディレクトリ）のある  
すぐ上のフォルダ（ディレクトリ）  
を選択して [OK] します。



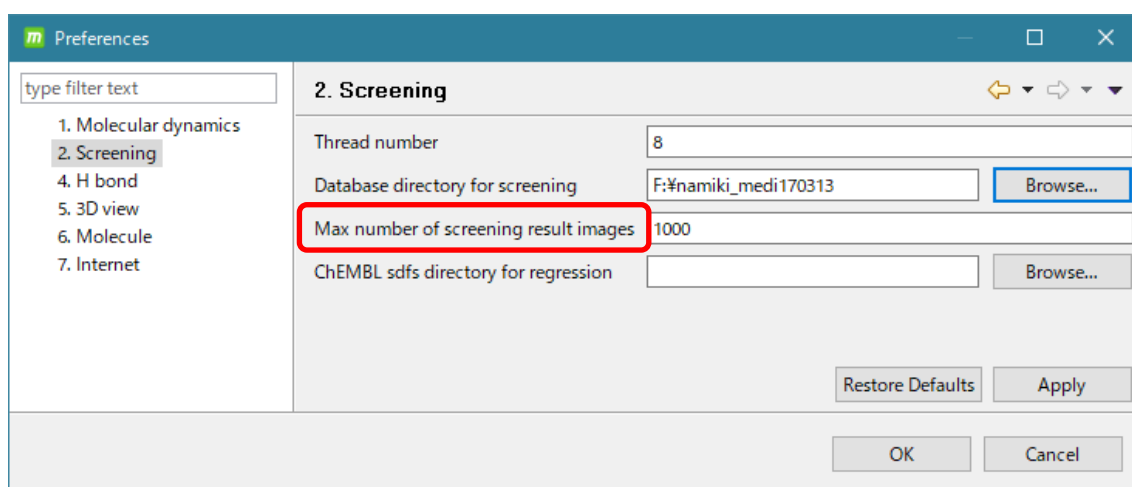
すると、**LigandBox** が上図のように設定されますので、[OK]をクリックします。  
 以上で、スクリーニング計算が実行できるようになります。

**LigandBox** でなく、「1.3 ユーザ指定のスクリーニング用化合物 DB の準備」で作成したデータベースを使用する場合は、保存したプロジェクトの **database** フォルダを下図のように選択し、[OK]をクリックします。



**database** フォルダが指定されたのを確認し、[OK]をクリックします。

2D 化学構造図の数は、[Help] - [Preference] 画面の[Max number of screening result images] で変更できます。



[Screening Info] 画面に表示するスクリーニング結果 2D 化学構造図の数を変更できます。

例えばこの値を 1000 にした場合、スクリーニング結果が 1000 以上あったとしても、1001 番目以降の結果の 2D 化学構造図は表示されません。

特に変える必要が無ければデフォルト（1000）のままで問題ありません。

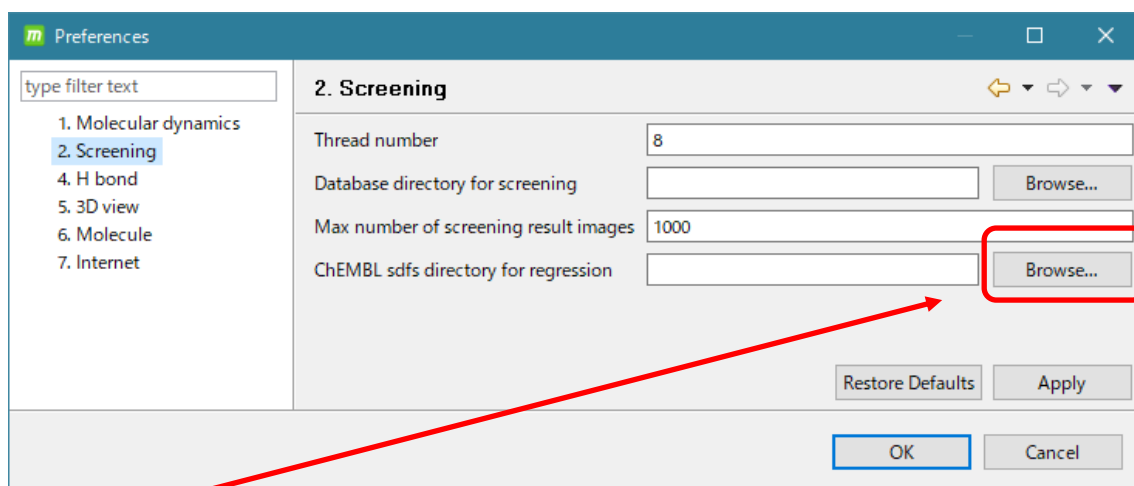
この値を 1000 より小さくすると計算機の使用リソースを削減できます。

大きくすると、プロジェクトを開いたときに 2D 化学構造図の生成に時間がかかるようになります。

値を変更したときは、プロジェクトを一回閉じてから開き直してください。スクリーニング結果リストの 2D 図の数が変更されて表示します。

[Help] - [Preference] - [2.Screening] 画面を選択して、ChEMBL sdfs の設定を行います。

化合物の各種特性の回帰分析による予測 ([Make Regression model] と [Predict with Regression model]) を実行するための設定です。



[Browse]をクリックします。

「1.3 ChEMBL sdfs の準備」で解凍した ChEMBL sdfs

(下の例では、chembl\_24\_sdfs\_moldesk) の

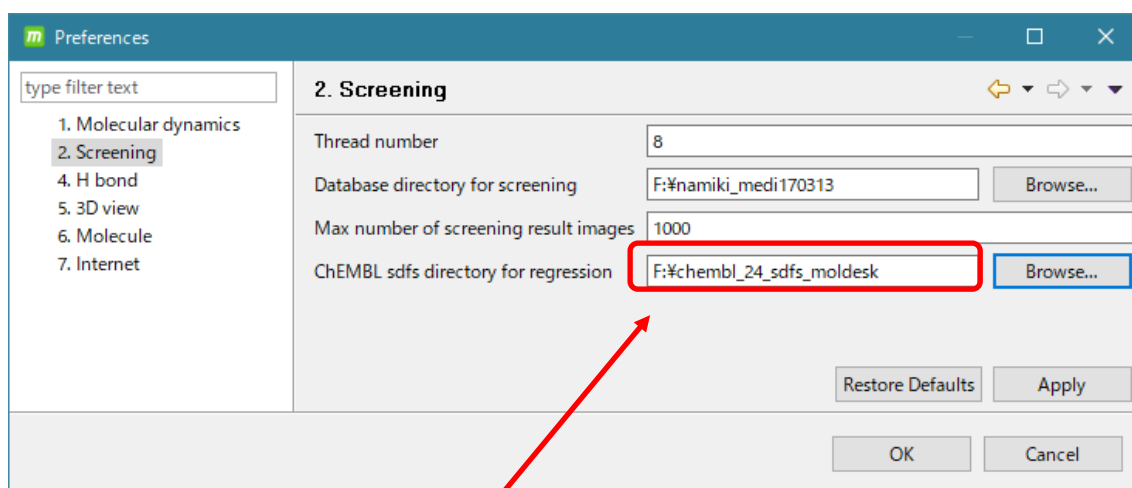
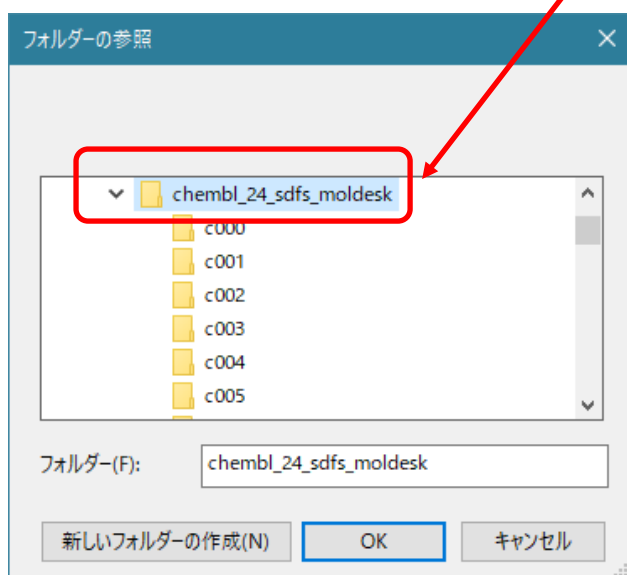
c000

c001

c002

...

フォルダ（ディレクトリ）のあるすぐ上のフォルダ（ディレクトリ）を選択して [OK] します。



すると、LigandBox が上図のように設定されますので、[OK]をクリックします。

以上で、化合物の各種特性の回帰分析による予測 ([Make Regression model] と [Predict with Regression model]) が実行できるようになります。



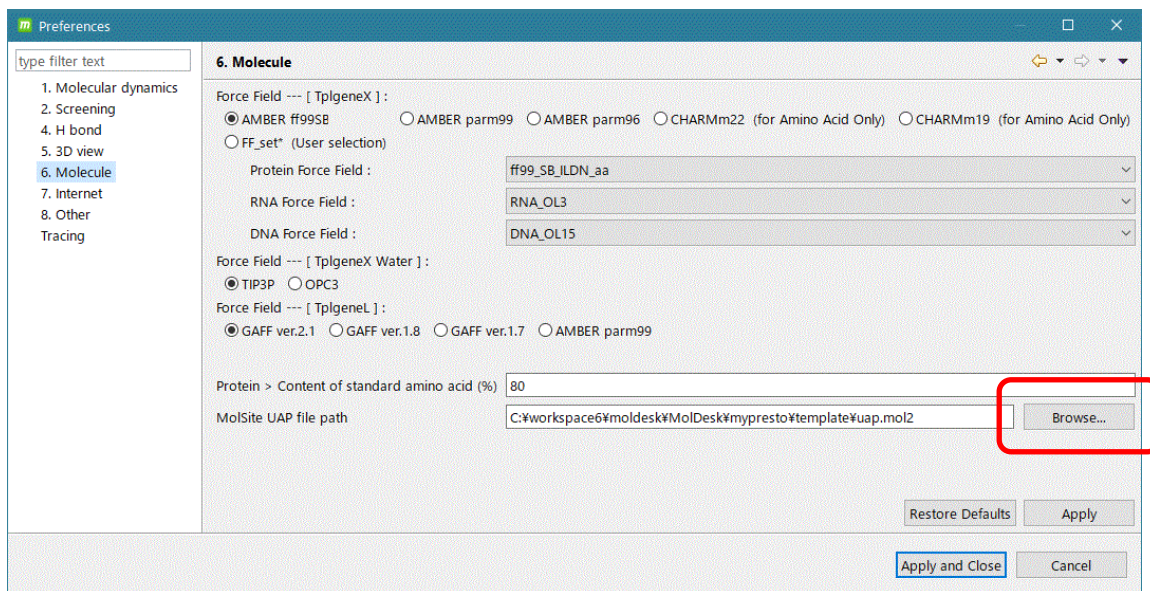
### 8.3.3. Molecule

MolSite によるポケット探索で使用する候補リガンドをデフォルトから変更したいときに設定します。

デフォルトで用意されている候補リガンドがすでに設定されていますが、候補リガンドを変更したいときにユーザが変更することができます。変更すると候補リガンドが結合しやすいポケットを探索するようになります。

ここでは、必ず **mol2** ファイルを設定してください。

[Help] - [Preference] 画面を開いて、「6.Molecule」を選択し、[MolSite UAP file path:] の [Browse] をクリックします。



#### 8.4. psygene / psygene-G による MD 計算

[Help] - [Preference] - 「Molecular Dynamics」

において、[MPI (sievgene)] または [MPI + GPU (sievgene-G)] を選択したときは、psygene または psygene-G による分子動力学の並列計算を実行します。

計算方法は、cosgene / cosgene\_MPI の場合と基本的は同じなので、計算方法の詳細は MolDesk Basic マニュアルをご参考ください。

ただし、psygene 系 と cosgene 系の MD 計算プログラムでは以下の機能の違いがあります。

- psygene / psygene-G は、水溶媒の形状が、Cap (球状) は計算できません。Cube (立方体) だけです。cosgene / cosgene\_MPI はどちらも計算できます。
- 真空中の溶質の計算は、どちらも可能です。  
(Cube water を形成しなくとも psygene 系で計算できるようになりました。)
- Generalized Born 法の計算が、cosgene / cosgene\_MPI ではできますが、psygene / psygene-G ではできません。