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.1.1.105

(モルデスク スクリーニング)





目次

| 1. スクリーニ | ニング計算 | 5 |
|------------|----------------------------|----|
| 1.1. インス | ストール・ライセンス認証 | 5 |
| 1.2. Ligan | ndBox の準備 | 6 |
| 1.3. ChEN | ABL sdfs の準備 | 8 |
| 1.4. ユーナ | げ指定のスクリーニング用化合物 DB の準備 | 9 |
| 1.4.1. 7 | プロジェクトの保存 | 9 |
| 1.4.2. ラ | データベース作成条件の設定 | 10 |
| 1.4.3. ラ | データベース作成計算 | 18 |
| 1.4.4. ラ | データベース作成場所 | 19 |
| 1.5. スクリ | リーニング用化合物 DB の再分割 | 22 |
| 1.5.1. 7 | プロジェクトの保存 | 22 |
| 1.5.2. ラ | データベース再分割条件の設定 | 23 |
| 1.5.3. ラ | データベース作成場所 | 25 |
| 1.6. スクリ | リーニング計算の概要 | 27 |
| 1.7. スクリ | リーニング計算の並列数・メモリ量・時間 | 28 |
| 1.8. スクリ | リーニング計算の手順 | 29 |
| 1.9. mol2 | ファイルの準備 | 31 |
| 1.10. MT | 'S 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. M | L-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. Docking | Score QSAR (Predict Activity) | 76 |
| 2.1. CHI | EMBL 実験値データの取得 | 77 |
| 2.2. 回帰 | パラメターの計算 | 81 |
| 2.2.1. | 概要説明 | 81 |
| 2.2.2. | プロジェクトの作成 | 81 |
| 2.2.3. | 回帰パラメターの計算実行 | |
| 2.2.4. | 回帰パラメターの計算結果のグラフによる確認 | |
| 2.2.5. | 回帰パラメターファイルの確認 | |
| 2.3. 活性 | :値の予測計算 | |
| 2.3.1. | 活性値予測計算の実行 | |
| 2.3.2. | 活性値予測計算の結果確認 | |
| 3. 回帰分析 | による化合物の特性値予測(Predict with Regression model) | 94 |
| 3.1. CHI | EMBL 実験値データの取得 | 95 |
| 3.2. 入力 | 実験データファイルの作成 | |
| 3.2.1. | プロジェクトの作成 | |
| 3.2.2. | ChEMBL データファイルからの作成法 | 100 |
| 3.2.3. | 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 | |
| 3.2.4. | 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 | |
| 3.2.5. | 実験データファイルをユーザがすべて編集する作成法 | |
| 3.3. 回帰 | パラメターの計算 | |
| 3.3.1. | 回帰パラメター計算のための入力項目 | |

| 3.3. | .2. | 回帰パラメターの計算実行 | |
|-------|------|-----------------------------|-----|
| 3.3. | .3. | 回帰パラメターの計算結果のグラフによる確認 | |
| 3.3. | .4. | 回帰パラメターファイルの確認 | |
| 3.4. | 特性 | 生値の予測計算 | |
| 3.4. | .1. | 特性値予測計算の実行 | |
| 3.4. | .2. | 特性値予測計算の結果確認 | |
| 4. MV | O So | creening | 122 |
| 4.1. | クコ | ニリ分子の選択 | 122 |
| 4.2. | 検索 | 索対象分子の選択 | 123 |
| 4.3. | 結果 | 表示 | |
| 5. 類似 | 以構造 | き検索 | 126 |
| 5.1. | クコ | ニリ分子の選択 | 126 |
| 5.2. | 検索 | 友対象分子の選択・計算・結果表示 | 127 |
| 6. 部分 | う構 道 | き検索 | 129 |
| 6.1. | クコ | ニリ分子の選択 | |
| 6.2. | 部分 | ∂構造検索対象分子の選択・計算・結果表示 | 130 |
| 7. MP | PI/G | PU による MD 計算の高速並列計算 | 131 |
| 7.1. | MP | I 動作環境の設定方法 | 133 |
| 7.1. | .1. | Windows 64bit | 133 |
| 7.1. | .2. | Linux 64bit | 133 |
| 7.2. | CU | DA 動作環境の設定方法 | 134 |
| 7.2. | .1. | Windows 64bit | 134 |
| 7.2. | .2. | Linux 64bit | 134 |
| 7.3. | Pre | ference の設定 | 135 |
| 7.3. | .1. | Molecular Dynamics | 135 |
| 7.3. | .2. | Screening | |
| 7.3. | .3. | Molecule | 144 |
| 7.4. | psy | rgene / psygene-G による MD 計算 | |

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 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.2310、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 作成で使用した sievgene のバージョ |
| | | \succ |
| | - 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ので注意が必要です 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、「7.3.2 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、「7.3.2 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。

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

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> ***</nscode> | |
|---|--|
| | |
| > <supplierid_*> ***</supplierid_*> | |
| | |

| > | <idnumber></idnumber> | |
|----|-----------------------|--|
| ** | * | |

| * * * | > <idn< th=""></idn<> |
|-------|-----------------------|
| | * * * |

(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 propery 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</tripos> | |
|------------------------------|-------------|
| LIGANDBOX_ID = MOLECULE-0000 | 0001-01 |
| SUPPLIER = SUPPLIER | |
| SOURCE = SOURCE | |
| IDNUMBER = NS-000000001-0001 | |
| MOLECULAR_FORMULA = C8H9N0 |)4 |
| MOLECULAR_WEIGHT = 183.163 | |
| $MOLECULAR_CHARGE = 0$ | |
| SUM_OF_ATOMNUMBER = 96 | |
| SUM_OF_ATOMNUMBER_MINUS_0 | CHARGE = 96 |
| NUM_OF_DONOR = 5 | |
| NUM_OF_ACCEPTOR = 4 | |
| HOMO = -9.2167 | |
| LUMO = -0.5693 | |
| NUM_OF_CHIRAL_ATOMS = 1 | |

| @ <tripos>MOLE MOLECULE-00000 22 22 0 0 0 SMALL USER_CHARGES</tripos> | CULE 001-01 | | | | |
|---|----------------|---------|--------------|-------|---------|
| @ <tripos>ATOM</tripos> | | | | | |
| 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 |
| 0.9343 -1.1740 | 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 |
| 0.9343 -0.3490 | 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 |
| 0.2198 0.8885 | 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 |
| -0.4946 -0.3490 | 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 |
| -1.2091 0.0635 | 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 |
| -0.4946 1.3010 | 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 |
| -1.2091 0.8885 | 0.0000 C 0 0 0 0 0 0 0 0 0 0 0 0 |
| 1.6488 -1.5865 | 0.0000 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| 1.6488 0.0635 | 0.0000 N 0 0 0 0 0 0 0 0 0 0 0 0 |
| 0.2198 -1.5865 | 0.0000 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| -1.9236 -0.3490 | 0.0000 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| -0.4946 2.1260 | 0.0000 0 0 0 0 0 0 0 0 0 0 0 0 0 |
| 3.6524 0.0000 | 0.0000 Cl 0 0 0 0 0 0 0 0 0 0 0 0 |
| 2 3 1 0 0 0 0 | |
| 3 1 1 0 0 0 0 | |
| 4 1 2 0 0 0 0 | |
| 5 1 1 0 0 0 0 | |
| 6520000 | |
| 7 4 1 0 0 0 0 | |
| 8 6 1 0 0 0 0 | |
| 9 2 2 0 0 0 0 | |

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します。

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。

| m Make DB for Screening | × |
|--|----------------|
| - Convert to 3D | |
| ✓ 2D> 3D | make conformer |
| set substitute property name to identify molecules, if " <nscode>" or "<supplierid_*>" or "<idnumber>" or "<idnumber>" does not exists in sdf file</idnumber></idnumber></supplierid_*></nscode> | SID |
| set supplier name if " <suppliername_*>" does not exists in sdf files (option) :</suppliername_*> | SUPPLIER |
| Tag of moleculer names (option) : | MOLECULE |
| SOURCE of input files (option) : | SOURCE |
| Example of output mol2 file @ <tripos>COMMENT IDNUMBER = Value of <nscode> or <supplierid_*> or <idnumber> or <idnumber> or (Property value user inputs above) SUPPLIER = Value of <suppliername_*> or (User inputs above) LIGANDBOX_ID = MOLECULE-*** (MOLECULE : User inputs above) SOURCE = SOURCE (SOURCE : User inputs above) SOURCE_ID = Value of <namikl_id> (if exists in sdf files.)</namikl_id></suppliername_*></idnumber></idnumber></supplierid_*></nscode></tripos> | |
| Filtering 1 | |
| by moleculer 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 | チェックすると3次元化を行う。 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。 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。 |
| make conformer | 3 次元化する際に、分子の Conformer 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。 |

| | 入力 sdf ファイルの中に、 <nscode> または</nscode> |
|--|--|
| | <supplierid_*> または <idnumber> または</idnumber></supplierid_*> |
| | <idnumber> の property 名が存在しない時に、別の</idnumber> |
| set substitute property | property 名の property 値を、IDNUMBER=として出力 |
| name to identify molecules, | mol2 ファイルに記述する。 |
| if " <nscode>" or</nscode> | IDNUMBER として認識させたい別の property 名を記入す |
| " <supplierid_*>" or</supplierid_*> | る。記入がない場合は、上記3つの prpperty 名を自動判定し |
| " <idnumber>" or</idnumber> | て、IDNUMBER とする。 |
| " <idnumber>" does tag does</idnumber> | 上記3つの property 名が存在せず、ユーザ入力の別名 |
| not exists in sdf file : | 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)。 |

[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</suppliername_*> | 入力 sdf ファイルの中に、<suppliername_*> 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)。</suppliername_*> 入力 sdf ファイルの中に、<suppliername_*> 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。</suppliername_*> (下記の SUPPLIER の部分)。 |
| Tag of moleculer name | 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 @<tripos>MOLECULE の次の行に</tripos> 記載される文字列。 (下記の MOLECULE の部分)。 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。 |
| SOURCE of input files | 入力ファイルの供給源を指定する。 出力 Mol2 ファイルの COMMENT 行に、SOURCE=として 記載される。 (下記の SOURCE の部分)。 |

(mol2 ファイル記述例)

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

@<TRIPOS>MOLECULE MOLECULE-00000001-01

[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 moleculer 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、ファイルセレクターが出るので、入力ファイルを選択して [開く] をクリックします。

| Delect sdf files | | | | | × |
|--|--------|--|------------------|------------------|--------|
| $\leftarrow \rightarrow \checkmark \uparrow$ | << デスク | パップ > MolDesk Screening > sample > sdf | √ Ū | sdfの検索 | م |
| 整理 ▼ 新しいフォ | tルダー | | | | |
| 🖈 クイック アクセス | ^ | 名前 ~ | 更新日時 | 種類 | サイズ |
| 🖊 ダウンロード 🗦 | * | 🖺 multi01.sdf | 2016/06/22 14:33 | SQL Server Compa | 2 KB |
| Creative Cloud | * | 🖺 multi3.sdf | 2016/07/06 16:46 | SQL Server Compa | 7 KB |
| PC 5 | * | 😭 multi28.sdf | 2014/10/23 13:28 | SQL Server Compa | 52 KB |
| ニー デスクトップ | * ~ < | | | | > |
| | ファイル名 | K(N): "multi28.sdf" "multi01.sdf" "multi3.sdf" | ~ | *.sdf | \sim |
| | | | | 開<(O) | キャンセル |

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。

- ※ ver.1.1.95 から入力ファイルが、mol2 ファイルの場合も DB 作成できるようになりました。mol2 ファイル入力の場合は以下の条件があります。
 - 1. すでに化合物の3次元化と電荷付加を実施済であること。
 - 2. mol2 ファイルの COMMENT 行に、各化合物の特定が可能な以下の IDNUMBER の記述があること。

(入力 mol2 ファイル記述例)

@<TRIPOS>COMMENT IDNUMBER = *** @<TRIPOS>MOLECULE

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。

| MolDesk | | | | | | | | | - | |
|----------------------------------|--|------------------|--|-------|-------|----|--------|-------|------|--|
| File Select Display Color Option | Expert Simple Screening Preparation | Window Undo Help | | | | | | | | |
| + + 0 0 | 🛚 🔧 😂 💉 💋 👂 🧿 | 🛃 🔲 📕 | | | | | | | | |
| Source 🛛 🗖 🗆 | 22 Command View 23 C text117: 1 Initial State 2 C Console Docking Info 2 | | | | | | | | | |
| | Convertion 20 Very Andrew Convertion 20 Very Andrew Convertion 20 Mole | | | | image | SA | deltaG | score | RMSD | |
| Ligand Info 🛛 🗖 | | | | | | | | | | |
| inage name SA | | M Consile IS | | • • • | | | | | | |

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 8 並列で実行 | Xeon E5-2697 v2 @ 2.70GHz x 2 (24 コア 48 プロセッサ) / 64GB メモリ / Linux CentOS6 48 並列で実行 |
|------|--|---|
| 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。

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 |
| Ilganu | 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 |
| 10.41 | 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_files | 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 |
| | 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 |
| protein | パク質毎のフォルダに保存 |
| 11 10 | mol2_files 以下のすべての mol2 ファイルをマージした multi |
| all.mol2 | mol2 ファイル。フィルタリングする前。 |
| | 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 |
| all_exclude.mol2 | mol2 ファイル。 |
| | 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 |
| error.log | できなかった分子を確認できる。 |
| | 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、 |
| exclude.info | 各分子の部分構造存在の情報ファイル |
| | 101 ないパタ所のリフト |
| pro.11st | 101クマハク貝のリスト |
| version | データベースのバージョンファイル |
| version | |

「7.3.2 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。

| m Preferences | | | | – o × |
|--|---|----|------------------|-------------------------------|
| type filter text | 2. Screening | | | <p 8<="" td="" ⇒="" ≈=""></p> |
| 1. Molecular dynamics 2. Screening 4. H bond 5. 3D view | Thread number Database directory for screening | 20 | [| Browse |
| 6. Molecule 7. Internet 8. Other | ChEMBL sdfs directory for regression | | | Browse |
| ANSI Support | | | | |
| | | | | |
| | | | | |
| | | | Restore Defaults | Apply |
| | | A | opply and Close | Cancel |

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。

| フォルダーの参照 | × |
|-------------------------|----------|
| | |
| | |
| ✓ MKDB011 | ^ |
| original | |
| ✓ work | |
| 1 | |
| | |
| ✓ database | |
| ilgand | |
| ligandImage | |
| > mol2_files | |
| mts_data | |
| > protein | v |
| | |
| フォルダー(F): database | |
| | |
| 新しいフォルダーの作成(N) OK キャンヤリ | , |
| | |

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。

Prepa ration [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。

| m Remake DB for Screening > | < |
|---------------------------------------|---|
| Select database Division Number: 8 | |
| OK Cancel |] |

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。

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

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 |
|-------------|--|
| 1. 1 | 分子数 10 万個ごとに、c***というフォルダに 3 次元化した |
| liganu | 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 |
| | 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 |
| protein | パク質毎のフォルダに保存 |
| and list | 101 ないパカ町のリフト |
| pro.fist | |
| | データベーフのバージョンファイル |
| version | $\frac{1}{2} - \frac{1}{2} - \frac{1}$ |

「7.3.2 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。

| D Preferences | | | | – o x |
|--|--|-----------|----------|------------------------------------|
| type filter text | 2. Screening | | | <p *="" 8<="" td="" •="" ⇒=""></p> |
| Molecular dynamics Screening H bond 3D view | Thread number Database directory for screening Max number of screening result images | 20 | | Browse |
| 6. Molecule 7. Internet 8. Other ANSI Support | ChEMBL sdfs directory for regression | | | Browse |
| | | Restore | Defaults | Apply |
| | | Apply and | Close | Cancel |

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。

| フォルダーの参照 × |
|--|
| MKDB011 |
| 1 2 database iigand iigandlmage mol2_files mts_data protein |
| フォルダー(F): database 新しいフォルダーの作成(N) OK キャンセル |

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 既知活性化合物 化合物データベース 候補化合物リスト 入力: 複数の mol2 ファイル ファイルセレクターの限界まで選択可 数百万化合物まで 出力:各種物性値の CSV ファイル 既知タンパク質 ML-DSI 法では不要 2D構造図を含む html ファイル ポケット情報 MTS法、ML-MTS法で必要 (いずれも、エクセル読み込み可) 入力: MolDesk でモデリング

※ 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 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。

| | ターゲット | 既知活性 | ユーザ追加による |
|-----------|---------|--------|----------|
| 計算法 | タンパク質 | リガンド | 検索対象化合物 |
| | (PDB) | (mol2) | (mol2) |
| ドッキングスコア順 | \odot | | 0 |
| MTS | Ø | | 0 |
| ML-MTS | Ô | 0 | 0 |
| ML-DSI | | 0 | 0 |

各スクリーニング計算法の入力は以下の通りです。

◎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 8 並列で実行 | Xeon E5-2697 v2@2.70GHz x 2 (24 コア 48 プロセッサ) / 64GB メモリ / Linux CentOS6 48 並列で実行 |
|------------------|--|---|
| ドッキング スコア順 | 35 時間 31 分 | 10時間 3分 |
| または MTS | | 19 吐胆 19 八 |
| ML-MTS ML-DSI | 45 時间 7 分 8 時間 26 分 | 13 時間 12 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 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ブ点をモデリングしたプロジェクトを 作成します。ポケットのプローブ点の作成方法は 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、ポケットのプローブ点の 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 ²² [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 | 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 |
| および | グから入力します。 |
| 検索対象化合物 | |

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、 (M) [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 Screenng -> sample -> screening -> cox2 -> Pro.pdb

| m | Open pdb/mol2/sdf/ | mol | × |
|----------------------|----------------------|-------------------------|------------|
| 🛞 🌛 🔻 🕇 퉬 « sample | e → screening → cox2 | ✓ C cox2の検索 | Q |
| 整理 ▼ 新しいフォルダー | | III 🔻 🔳 | |
| ☆ お気に入り | 名前 | 更新日時 種類 | |
| 🚺 ダウンロード | 🐏 point.pdb | 2015/05/11 0:24 Program | m Debug |
| 三 デスクトップ | 🐏 Pro.pdb | 2015/05/11 0:24 Program | m Debug |
| 週 最近表示した場所 | | | |
| Creative Cloud Files | | | |
| ¥ | < | | > |
| ファイル名(N | I): Pro.pdb | *.pdb;*.ent;*.mol2;*.sd | lf;*.r ∨ |
| | | 開<(O) キャン | /セル .:: |





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

| m | Select molecule | file | × |
|----------------------|----------------------|--|----|
| 🔄 🏵 🔻 🕇 🚺 « scre | eening > cox2 | cox2の検索 | P |
| 整理 ▼ 新しいフォルダー | | :== ▼ □ | 0 |
| 🚖 お気に入り | ^ 名前 ^ | 更新日時 | 種 |
| 🗼 ダウンロード | 🐏 point.pdb | 2015/05/11 0:24 | Pr |
| 📰 デスクトップ | 🐏 Pro.pdb | 2015/05/11 0:24 | Pr |
| 3 最近表示した場所 | = Pro.tpl | 2015/05/11 0:24 | TF |
| Oreative Cloud Files | | | |
| 🔲 デスクトップ | ✓ < | | 2 |
| ファイルイ | 더(N): point.pdb | *.*< 開く(0) キャンセル | ¥ |



[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、正確な計算ができません。

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大きい(~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。

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

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。

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

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。

| m | Warning | × |
|---|---|---|
| ▲ | Please set LigandBox at [Help]-[Preference]-[LigandBox]. Or LigandBox database is broken. | |
| | ОК | |
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。

| m Screening : MTS (Mu □ Database for Screening F:¥namiki_medi170313 | ltiple Target Screening metho g: [Help] - [Preference] - [S | d) / Docking score ranking Screening] - [LigandBox] : |) | | × |
|---|--|--|--------------------------|--|---|
| Previous input Proteins Added to 181 | Target-protein | Compounds Added (test AUC) | Active ligands | New input Target-protein Target-protein name for output (in ASCII without blank) | Compounds Add to DB for screening (or for test AUC) Select mol2 files Compounds |
| | Delete selected Reset Method of docking Flexible | | | Method of docking | |
| Delete selected Reset | | Delete selected Reset | Delete selected Reset | Size of output list max No. of lines : 10500 💮 | Delete selected |

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。

| m Screening : MTS (Multiple Target Screening method | I) / Docking score ranking | | × |
|--|---|--|---|
| - Database for Screening : [Help] - [Preference] - [S F:¥namiki_medi170313 | rreening] - [LigandBox] : | | |
| Previous input Proteins Added to 181 Delete selected Reset Delete selected Reset Delete selected Reset | Compounds Added (test AUC) Active ligands Delete selected Reset Delete selected Reset | New input Target-protein Target-protein name for output To-267 (midden to blank) cox2 Method of docking Image: Protein in a protein Size of output list max No. of lines: 10500 Test by AUC | Compounds Active ligands as teacher (or for test AUC) Select mol2 files Compounds Active ligands Compounds Active ligands Delete selected Delete selected |
| | | | OK Cancel |

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

| Select mol2 files to add to DB for screening X | | | | | | | |
|--|--------------------------------|-----------------|-----------------|-------|--|--|--|
| ← → • ↑ <mark> </mark> « | sample > screening > cox2_ligs | 1 v Ö | cox2_ligs1の検索 | Ą | | | |
| 整理 ▼ 新しいフォル | Ø- | | | - 🔳 🕜 | | | |
| 📙 Creative Clou 🖈 | ^ 名前 [^] | 更新日時 | 種類 | サイズ ^ | | | |
| 💻 PC 🛛 🖈 | 上 cox2L1.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | | | |
| 📃 デスクトップ 🖈 | 🛓 cox2L2.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| 📙 USBメモリ 🛛 🖈 | 上 cox2L5.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 3 KB | | | |
| 🛗 ドキュメント 🛛 🖈 | 上 cox2L7.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| | <u>)</u> cox2L12.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| cox2 | 上 cox2L14.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | | | |
| doc | 👱 cox2L18.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| - doc | 🚊 cox2L19.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| img2 | cox2L23.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | | | |
| mol2 | cox2L26.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | | | |
| 🗾 デスクトップ | cox2L31.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| ConeDrive | cox2L34.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| | 👱 cox2L36.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| 10,4147 | cox2L38.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| 画像 | cox2L41.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |
| _ 公開 | cox2L51.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | | |

| m Screening : MTS (Multiple Target Screening metho | d) / Docking score ranking | | × |
|---|---|---|---|
| – Database for Screening : [Help] - [Preference] - [S F:¥namiki_medi170313 | creening] - [LigandBox] : | | |
| Previous input | | - New input | |
| Proteins | Compounds | | Compounds |
| Added to 181 Target-protein Delete selected Reset Method of docking Flexible | - Added (test AUC) - Active ligands | Target-protein Target-protein name for output (in ASCII without blank) cox2 Method of docking Image: the state of the state | Add to DB for screening Active ligands as teacher (or for test AUC) Select mol2 files Componds Cox2L1.mol2 cox2L3.mol2 cox2L14.mol2 cox2L14.mol2 cox2L14.mol2 cox2L14.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 |
| Delete selected Reset | Delete selected Delete selected Reset Reset | Test by AUC | Delete selected |
| | | | OK Cancel |

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。

| | | New input | |
|------------------------|--|--|---|
| Compounds | | | Compounds |
| rotein Added (test AUC |) Active ligands | Target-protein Target-protein name for output (in ASCII without blank) cox2 | Add to DB for screening (or for test AUC) Select mol2 files Compounds cox2l.1.mol2 |
| d of docking | | Method of docking | cox2L2.mol2 cox2L5.mol2 cox2L7.mol2 cox2L12.mol2 cox2L14.mol2 cox2L14.mol2 cox2L18.mol2 cox2L19.mol2 |
| | | Flexible Rigid remake grid files of protein | cox2L26.mol2 cox2L31.mol2 cox2L34.mol2 cox2L36.mol2 |
| | | Size of output list | |
| Delete selected | Delete selected | | |
| Reset | Reset | Test by AUC | Delete selected Delete selected |
| | elected d of docking ? Delete selected Reset | elected d of docking 2 Delete selected Reset | elected d of docking : Delete selected Reset Delete selected Reset |

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のウインドウを広げたり、[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。

| m | | | | | | | | | | | | - 🗆 | × |
|-------------------|--|----------------|---------------------|--------|-------------|---------------------|--------------|---------|--------|--------|---------|-------------|----------|
| 🗖 Scr | Screening Info 🛙 | | | | | | | | | | | | |
| MTS | | | Score | | | Export tabl | e | | | | AUC | | |
| | | | 10 | 6 | 6 ID | | | | | | 6 F | | |
| imag | S C C C C C C C C C C C C C C C C C C C | 1 | HTS1508-03579436-02 | 4.8174 | NS-07213367 | HTS1508-03579436-02 | C21H25N4O | 349.458 | 1.388 | 4.0763 | ENAMINE | Z219349550 | |
| 8 | ¹ 220° | 2 | HTS1508-01970909-02 | 4.4553 | NS-02926541 | HTS1508-01970909-02 | C20H20N2O3 | 336.391 | 1.4704 | 2.9269 | ENAMINE | Z108660526 | |
| o | กา กรุ่ม มรุ่ม มนุ | 3 | HTS1508-02313392-01 | 4.7091 | NS-03806336 | HTS1508-02313392-01 | C17H26N7 | 328.444 | 2.7014 | 3.2778 | ENAMINE | Z195555268 | |
| G | | 4 | HTS1508-01297838-01 | 4.3179 | NS-01905160 | HTS1508-01297838-01 | C22H22N3O | 344.438 | 1.2771 | 3.4332 | ENAMINE | Z90501126 | |
| ţ× <mark>,</mark> | |) ₅ | HTS1508-03818160-01 | 4.088 | NS-08577078 | HTS1508-03818160-01 | C17H17N2O2F3 | 338.329 | 2.3427 | 3.215 | ENAMINE | Z393480320 | |
| _0. V | | 6 | HTS1508-04518302-02 | 4.7931 | NS-10566855 | HTS1508-04518302-02 | C19H26N3O3 | 344.435 | 2.3974 | 2.4917 | ENAMINE | Z1538838180 | |
| Ŷ | ~ <u>\.</u> \ | 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。

[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。

| Screening : ML-MTS | (Machine Learning - Multiple | e Target Screening method) | | × |
|--|-------------------------------|---------------------------------|------------------------------|----------------------------------|
| - Database for Screening F:¥namiki_medi170313 |): [Help] - [Preference] - [S | Screening] - [LigandBox] : | | |
| Descionation | | | Neutinet | |
| Previous input | | Companyada | New input | Concernate |
| Proteins | Townet wanteda | Compounds | Transferration | Compounds |
| Added to 161 | larget-protein | Added (test AOC) Active ligands | Target-protein | (or for test AUC) |
| | | | (in ASCII without blank) | Collect mol2 files |
| | | | | Select mol2 files Active ligands |
| | | | | Compounds |
| | | | | |
| | Delete selected | | | |
| | Reset | | | |
| | Method of docking | | | |
| | Flexible | | | |
| | | | Method of docking | |
| | | | | |
| | | | Flexible O Rigid | |
| | | | | |
| | | | remake grid files of protein | |
| | | | Size of output list | |
| | | | Size of output list | |
| | | | max No. of lines : 10500 🚔 | |
| Delete selected | | Delete selected Delete selected | | |
| belete selected | | belete selected | Test by AUC | Delete selected |
| Reset | | Reset | screening with small DB | |
| | | | | |
| | | | | |
| | | | | OK Cancel |
| | | | | |

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

| Select mol2 files to add to DB for screening | | | | | | | |
|---|-------|------------------------------|-----------------|-----------------|------|---|--|
| \leftrightarrow \rightarrow \checkmark \uparrow | « sam | nple > screening > cox2_ligs | 1 v Ö | cox2_ligs1の検索 | Ą | | |
| 整理 ▼ 新しい | フォルダー | | | •== • | | | |
| Creative Clo | u 🖈 🔨 | ~ 名前 | 更新日時 | 種類 サイ | X | ^ | |
| PC | * | 上 cox2L1.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | | |
| 📃 デスクトップ | * | 上 cox2L2.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| USBXモリ | * | 上 cox2L5.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 3 KB | | |
| 🔮 ドキュメント | * | 🛓 cox2L7.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| ■ ピクチャ | * | 上 cox2L12.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| cox2 | | 🛓 cox2L14.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | | |
| - COAL | | <u>)</u> cox2L18.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| doc | | 🚹 cox2L19.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| img2 | | 🚹 cox2L23.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | | |
| mol2 | | 上 cox2L26.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | | |
| デスクトップ | | 🚹 cox2L31.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| | | 🛓 cox2L34.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| Chebrive | | 🚹 cox2L36.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| F#1X7F | | 🛓 cox2L38.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| 画像 | | 🛓 cox2L41.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |
| 公開 | | 🛓 cox2L51.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | | |

| m Screening : ML-MTS (Machine Learning - Multiple Target Screening method) X | | | | | | |
|--|--|---|----------------|--|---|--|
| Database for Screening : [F:¥namiki_medi170313 | Help] - [Preference] - [Scre | ening] - [LigandBox] : | | | | |
| Previous input | | | | - New input | | |
| Proteins Added to 181 D R R FI Delete selected Reset | rget-protein elete selected eset Aethod of docking lexible | Compounds Added (test AUC) Delete selected Reset | Active ligands | Target-protein Target-protein name for output (in ASCII without blank) cox2 Method of docking Image: Second Strategy Strateg | Compounds Add to DB for screening (or for test AUC) Select mol2 files Compounds Cox21.mol2 cox21.mol2 cox21.mol2 cox21.mol2 cox21.mol2 cox21.tamol2 | Active ligands as teacher Select mol2 files Active ligands |
| | | | | | | OK Cancel |

| m Screening : ML-MTS (Machine Learning - Multiple | Target Screening method) | | × |
|--|---------------------------------|--|--|
| - Database for Screening : [Help] - [Preference] - [Sc F:¥namiki_medi170313 | rreening] - [LigandBox] : | | |
| Previous input | | -New input | |
| Proteins | Compounds | | Compounds |
| Added to 181 Target-protein | Added (test AUC) Active ligands | - Target-protein | Add to DB for screening |
| Delete selected Reset | Delete selected Reset | Target-protein name for output (in ASCI without blank) cox2 Method of docking | (or for test AUC) Select mol2 files Compounds Cox21.mol2 cox21.smol2 cox21.smol2 cox21 |
| | | | OK Cancel |

[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

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

| Select mol2 files of active ligand as teacher | | | | | | | |
|---|-------------------------------|------------------------------|-----------------|--|--|--|--|
| ← → × ↑ 📙 « sa | mple > screening > cox2_ligs2 | ✓ Č cox2_ligs2Ø ³ | _{検索} ク | | | | |
| 整理 ▼ 新しいフォルダー | | | :== (?) | | | | |
| Creative Clou 🖈 🔦 | 名前 | 更新日時 | 種類 ^ | | | | |
| PC 🖈 | 1 cox2L0.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| 📃 デスクトップ 🖈 | 上 cox2L3.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| USBXEU 🖈 | 1 cox2L4.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| 🔮 ドキュメント 🛛 🖈 | 1 cox2L6.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| | 1 cox2L8.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| cox2 | 📩 cox2L9.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| dos | 1 cox2L10.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| doc | 1 cox2L11.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| img2 | 📩 cox2L13.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| mol2 | 1 cox2L15.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| デスクトップ | 1 cox2L16.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| | 1 cox2L17.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| | 1 cox2L22.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| 10,000 | <u>1</u> cox2L24.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| 画像 | <u>1</u> cox2L25.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |
| - 公開 | 🛓 cox2L29.mol2 | 2015/05/11 0:24 | ArgusLab Docume | | | | |

| w Scleaning with the to (Wald time Leanin Database for Screening : [Help] - [Pr F:¥namiki_medi170313 | ng - woldgie Target Screening Hendol) eference] - [Screening] - [LigandBox] : | |
|--|--|---|
| Previous input Proteins Added to 181 Delete selec Reset Method of Flexible | ted docking | New input Compounds Target-protein Add to DB for screening Target-protein name for output (or for test AUC) Active ligands as teacher (or for test AUC) Select mol2 files Active ligands cox2 Compounds Cox2L1.mol2 cox2L3.mol2 |
| Delete selected Reset | Delete selected Delete selecte Reset Reset | Test by AUC Delete selected Delete selected |

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。

| m Screening : ML-MTS (Machine Learning - Multiple | e Target Screening method) | | × |
|---|---|---|--|
| - Database for Screening : [Help] - [Preference] - [S F:¥namiki_medi170313 | Screening] - [LigandBox] : | | |
| Previous input | | New input | |
| Proteins | Compounds | | Compounds |
| Added to 181 Target-protein | Added (test AUC) Active ligands | - Target-protein | Add to DB for screening Active ligands as teacher |
| | | Target-protein name for output (in ASCII without blank) | (or for test AUC) Select mol2 files Active ligands |
| Delete selected Reset Method of docking Flexible | | Method of docking | Compounds cox21.0.mol2 cox21.1.mol2 cox21.3.mol2 cox21.2.mol2 cox21.4.mol2 cox21.5.mol2 cox21.6.mol2 cox21.7.mol2 cox21.6.mol2 cox21.12.mol2 cox21.4.mol2 cox21.12.mol2 cox21.10mol2 cox21.13.mol2 cox21.1.mol2 cox21.13.mol2 cox21.10mol2 cox21.13.mol2 cox21.13.mol2 cox21.3.mol2 cox2.13.mol2 |
| | | Flexible Rigid Iremake grid files of protein Size of output list | cox2126.mol2 cox2116.mol2 cox2131.mol2 cox217.mol2 cox2134.mol2 cox2122.mol2 cox2136.mol2 |
| Delete selected Reset | Delete selected Delete selected Reset Reset | max No. of lines : 10500 🔹 Test by AUC Screening with small DB | Delete selected |
| | | | OK Cancel |

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。

| m Screening : ML-DSI (Machine Learning - Docking Sco | ore Index method) | | × |
|--|---------------------------------|--|--|
| – Database for Screening : [Help] - [Preference] - [Scre F:¥namiki_medi170313 | eening] - [LigandBox] : | | |
| Previous input | | New input | |
| Proteins | Compounds | | Compounds |
| Added to 181 Target-protein Delete selected Reset | Added (test AUC) Active ligands | Target-protein Target-protein name for output (in ASCII without blank) | Add to DB for screening (or for test AUC) Select mol2 files Compounds |
| Method of docking Flexible | | Method of docking | |
| | | ☐ remake grid files of protein Size of output list | |
| Delete selected | Delete selected Delete selected | max No. of lines : 10500 | |
| Reset | Reset | screening with small DB | Delete selected |

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。

| m Screening : ML-DSI (Machine | Learning - Docking Score Index method) | | | | × |
|--|---|-----------------|--|---|---|
| – Database for Screening : [He F:¥namiki_medi170313 | p] - [Preference] - [Screening] - [LigandBox] : | | | | |
| Previous input | | Ne | ew input | | |
| Added to 181 | t-protein Added (test AUC) | Active ligands | Target-protein Target-protein name for output (in ASCII without blank) | Compounds Add to DB for screening or for test AUC Select mol2 files Compounds | ctive ligands as teacher – Select mol2 files Active ligands |
| Dele Rese Met Flex | te selected t hod of docking ible | | Method of docking | | |
| | | | Flexible Rigid remake grid files of protein Size of output list | | |
| Delete selected Reset | Delete selected Reset | Delete selected | max No. of lines : 10500 🔹 Test by AUC Screening with small DB | Delete selected | Delete selected |
| | | | | | OK Cancel |

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

| Delect mol2 files to a select mol2 files t | to add to | DB for screening | | | > | × |
|---|-----------|----------------------------|-----------------|-----------------|-------|---|
| $\leftarrow \rightarrow \cdot \uparrow$ | « sam | ple > screening > cox2_lig | gs1 ∨ Ĉ |) cox2_ligs1の検索 | م | |
| 整理 ▼ 新しい | フォルダー | | | 8== | - 🔳 ? |) |
| Creative Clou | u# ^ | 名前 ^ | 更新日時 | 種類 | サイズ | ^ |
| PC | * | 1 cox2L1.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | |
| 📃 デスクトップ | * | 🛓 cox2L2.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| USBXモリ | * | 1 cox2L5.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 3 KB | |
| 🔮 ドキュメント | * | 🛓 cox2L7.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| ■ ピクチャ | * | 🛓 cox2L12.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| cox2 | | <u> cox2L14.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | |
| dec | | <u> cox2L18.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| doc | | <u>1</u> cox2L19.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| img2 | | 👱 cox2L23.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | |
| mol2 | | <u> cox2L26.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume | 5 KB | |
| デスクトップ | | <u> cox2L31.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| ConeDrive | | <u> c</u> ox2L34.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| | | 👱 cox2L36.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| | | <u> cox2L38.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| 画像 | | <u> cox2L41.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |
| - 公開 | | 🛓 cox2L51.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB | |

| m Screening : ML-DSI (| Machine Learning - Docking | Score Index method) | | | | × |
|--|---|----------------------------|----------------|-------------------------|---|----------------------------|
| Database for Screening F:¥namiki_medi170313 | g : [Help] - [Preference] - [S | Screening] - [LigandBox] : | | | | |
| Previous input | | | | New input | | |
| - Proteins | | Compounds | | New input | Compounds | |
| - Added to 181 | Target-protein | Added (test AUC) | Active ligands | Target-protein | -Add to DB for screening - | Active ligands as teacher- |
| | Delete selected Reset Method of docking Flexible | | | Method of docking | (or for test AUC) Select mol2 files Cox21.mol2 cox21.mol2 cox21.mol2 cox21.mol2 cox21.mol2 cox21.4mol2 cox21.4mol2 cox21.18mol2 cox21.9mol2 cox21.9mol2 cox21.9mol2 cox21.34mol2 cox21.34mol2 cox21.34mol2 cox21.34mol2 cox21.34mol2 cox21.36mol2 | Select mol2 files |
| Parat | | Parat | Poset | Test by AUC | Delete selected | Delete selected |
| iveset | | Neset | | screening with small DB | | |
| | | | | | | OK Cancel |

| m Screening : ML-DSI (M | lachine Learning - Docking | Score Index method) | | × |
|--|--|---------------------------------|---|--|
| – Database for Screening : F:¥namiki_medi170313 | [Help] - [Preference] - [S | creening] - [LigandBox] : | | |
| Previous input | | | - New input | |
| Proteins | | Compounds | | Compounds |
| Added to 181 | Delete selected Reset Method of docking Flexible | Added (test AUC) Active ligands | Target-protein Target-protein name for output (in ASCI without blank) Method of docking Image: protein Image: protein Size of output list max No. of lines : 10500 | Active ligands as teacher (or for test AUC) Select mol2 files Compounds Cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L1.mol2 cox2L3. |
| Reset | | Reset | Test by AUC | Delete selected |
| | | | | OK Cancel |

[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

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

| Select mol2 files of act | ive ligand as teacher | | × |
|--------------------------|---------------------------------|-----------------|-----------------|
| ← → • ↑ <mark>.</mark> « | sample > screening > cox2_ligs2 | ✓ Cox2_ligs2の | _{食索} ク |
| 整理 マ 新しいフォルタ | <i>i</i> – | | ::: • 🔟 ? |
| 🔓 Creative Clou 🖈 🖞 | * 名前 * | 更新日時 | 種類 ^ |
| 💻 PC 🛛 🖈 | 上 cox2L0.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| 📃 デスクトップ 🖈 | 1 cox2L3.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| USBXEU 🖈 | 上 cox2L4.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| 🗎 ドキュメント 🛛 🖈 | <u> cox2L6.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume |
| | 1 cox2L8.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| cox2 | <u>1</u> cox2L9.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| des | 1 cox2L10.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| doc | <u>1</u> cox2L11.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| img2 | 1 cox2L13.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| mol2 | <u>1</u> cox2L15.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| デスクトップ | 1 cox2L16.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| | cox2L17.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| | 1 cox2L22.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| | 🛓 cox2L24.mol2 | 2015/05/11 0:24 | ArgusLab Docume |
| 画像 | <u> cox2L25.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume |
| 公開 | 上 cox2L29.mol2 | 2015/05/11 0:24 | ArgusLab Docume |

| Screening : ML-DSI (| (Machine Learning - Docking | Score Index method) | | × |
|---|---|---|---|---|
| Database for Screenin F:¥namiki_medi170313 | g : [Help] - [Preference] - [} | Screening] - [LigandBox] : | | |
| Previous input | | | New input | |
| Proteins Added to 181 | Target-protein Delete selected Reset Method of docking Flexible | Compounds Added (test AUC) Active ligands | Target-protein Target-protein name for output (in ASCII without blank) Method of docking Image: Plexible Rigid Termake grid files of protein Size of output list max No. of lines: Test by AUC | Compounds Active ligands as teacher (or for test AUC) Select mol2 files Compounds Cox21.4mol2 cox21.4mol2 cox21.4mol2 cox21.5mol2 cox21.4mol2 cox21.7mol2 cox21.4mol2 cox21.8mol2 cox21.4mol2 cox21.8mol2 cox21.8mol2 cox21.8mol2 cox21.8mol2 cox21.8mol2 cox21.8mol2 cox21.8mol2 cox21.1mol2 cox21.8mol2 cox21.1mol2 cox21.8mol2 cox21.1mol2 cox21.18mol2 cox21.1mol2 cox21.18mol2 cox21.1mol2 cox21.18mol2 cox21.1mol2 cox21.18mol2 cox21.1mol2 cox21.18mol2 cox21.17mol2 cox21.34mol2 cox21.17mol2 cox21.34mol2 cox21.22mol2 |
| | | | Screening with small Db | 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。

| n Screening : ML-DSI (Machine Le | earning - Docking Score Index method) | | | | × |
|---|--|--------------------|---|---|---|
| – Database for Screening : [Help] F:¥namiki_medi170313 |] - [Preference] - [Screening] - [LigandBox] : | | | | |
| Previous input | | - New i | input | | |
| Proteins | Compounds | | C | ompounds | |
| Added to 181 Target- | selected of docking le | Active ligands Tai | ethod of docking ethod | Add to DB for screening or for test AUC) Select mol2 files Compounds cox2L1.mol2 cox2L2.mol2 cox2L3.mol2 cox2L1.mol2 cox2L1.mol2 cox2L14.mol2 cox2L14.mol2 cox2L14.mol2 cox2L14.mol2 cox2L14.mol2 cox2L14.mol2 cox2L13.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 cox2L3.mol2 | Active ligands as teacher Select mol2 files Active ligands cox2L0.mol2 cox2L3.mol2 cox2L4.mol2 cox2L4.mol2 cox2L6.mol2 cox2L10.mol2 cox2L10.mol2 cox2L10.mol2 cox2L110.mol2 cox2L113.mol2 cox2L13.mol2 cox2L13.mol2 cox2L13.mol2 cox2L15.mol2 cox2L15.mol2 cox2L15.mol2 cox2L22.mol2 |
| Reset | Reset | Reset | screening with small DB | Delete selected | Delete selected |
| | | | | | OK Cancel |

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



既知活性化合物 / 既知タンパク質 / 検索対象化合物 を追加して、繰り返しスクリーニング

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いてもか 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。

ッリー表示画面で □ 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。

| | | | | New input | | |
|--------------------------|---|--|--------------------------|---|--|-------------------------|
| oteins | | Compounds | | | Compounds | |
| Added to 181 | Target-protein cox2 Delete selected Reset Method of docking Flexible | Added (test AUC) cox21.mol2 cox212.mol2 cox215.mol2 cox215.mol2 cox2112.mol2 cox2114.mol2 cox2114.mol2 cox2118.mol2 cox2119.mol2 cox213.mol2 cox2125.mol2 | Active ligands | Target-protein Target-protein mame for output (in ASCII without blank) Method of docking Flexible Rigid remake grid files of protein Size of output list | Add to DB for screening (or for test AUC) Select mol2 files Compounds | Active ligands as teach |
| Delete selected Reset | | Delete selected Reset | Delete selected Reset | Test by AUC | Delete selected | Delete selected |

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。



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 -> cox2L62.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L62.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L66.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 -> cox2L71.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L72.mol2 MolDesk Screening -> sample -> screening -> cox2_ligs1 -> cox2L73.mol2

| → × ↑ 📙 < | < sam | ple > screening > cox2_ | ligs1 v | ට cox2_ligs1の検索 | م |
|------------------|-------|--------------------------|-------------------------|-----------------|-------|
| 経理 ▼ 新しいフォ. | ルダー | | | | - 💷 🤇 |
| Creative Clou 🖈 | ^ | 名前 ^ | 更新日時 | 種類 | サイズ |
| 💻 PC 🛛 🛪 | | 👖 cox2L34.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| | | 🛓 cox2L36.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| | | 🛓 cox2L38.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| | | <u> cox2L41.mol2</u> | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| E 141X21 x | | 🛓 cox2L51.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| 📰 ピクチャ 🛛 🖈 | • | 🛓 cox2L52.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| cox2 | | 🛓 cox2L53.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| doc | | 1 cox2L54.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| ima2 | | 🛓 cox2L55.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| mol2 | | 🛓 cox2L62.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| Inoi2 | | 1 cox2L66.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| 🔜 デスクトップ | | 1 cox2L67.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 6 KB |
| 🕋 OneDrive | | 1 cox2L68.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| ドキュメント | | 1 cox2L71.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| 画曲 | | 1 cox2L72.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| | | 1 cox2L73.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| 公開 | | 1 cox2L74.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| 👗 Kiyotaka Misoo | | 上 cox2L79.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 4 KB |
| PC PC | ~ | 上 cox2L80.mol2 | 2015/05/11 0:24 | ArgusLab Docume | 5 KB |
| 5 | ファイル・ | 名(N): "cox2L74.mol2" "co | x2L38.mol2" "cox2L41.mc | *.mol2 | ~ |

| evious input | | | | - New input | | |
|--------------------------|---|---|--------------------------|---|--|--|
| oteins | | Compounds | | | Compounds | |
| Added to 181 | Target-protein cox2 Delete selected Reset Method of docking Flexible | Added (test AUC) cxx21.mol2 cxx21.mol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 cxx21.smol2 | Active ligands | Target-protein Target-protein name for output (in ASCII without blank) cox2 Method of docking Flexible Rigid remake grid files of protein Size of output list max No. of lines: 10500 10500 | Add to DB for screening (or for test AUC) Select mol2 files Compounds | Active ligands as teach Select mol2 files Cox2L38.mol2 cox2L31.mol2 cox2L51.mol2 cox2L51.mol2 cox2L53.mol2 cox2L53.mol2 cox2L53.mol2 cox2L53.mol2 cox2L62.mol2 cox2L62.mol2 cox2L67.mol2 cox2L67.mol2 cox2L71.mol2 cox2L73.mol2 cox2L73.mol2 cox2L73.mol2 cox2L74.mol2 |
| Delete selected Reset | | Delete selected Reset | Delete selected Reset | Test by AUC | Delete selected | Delete selected |

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。

| evious input | | | | - New input | | |
|--------------------------|---|--|-----------------|--|--|---|
| roteins | | Compounds | | | Compounds | |
| Added to 181 | Target-protein cox2 Delete selected Reset Method of docking Flexible | Added (test AUC) cox21.mol2 cox21.mol2 cox21.smol2 cox21.smol2 cox21.stmol2 cox21.stmol2 cox21.18.mol2 cox21.98.mol2 cox21.98.mol2 cox22.smol2 | Active ligands | Target-protein Target-protein name for output (in ASCI without blank) cox2 Method of docking Flexible Rigid remake grid files of protein Size of output list max No. of lines : 10500 | Add to DB for screening (or for test AUC) Select mol2 files Compounds | Active ligands as teache Select mol2 files Active ligands cox2138.mol2 cox2141.mol2 cox2151.mol2 cox2153.mol2 cox2153.mol2 cox2154.mol2 cox2154.mol2 cox2167.mol2 cox2167.mol2 cox2167.mol2 cox2167.mol2 cox2171.mol2 cox2173.mol2 cox2173.mol2 cox2174.mol2 |
| Parat | | Parat | Parat | Test by AUC | Delete selected | Delete selected |
| Delete selected Reset | | Delete selected | Delete selected | Test by AUC ✓ screening with small DB | Delete selected | De |

1回目の計算結果は再計算せずに再利用す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



既知活性化合物 / 既知タンパク質 / 検索対象化合物 を追加して、繰り返しスクリーニング

「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。

| Compounds | | |
|---|--|---|
| Target-protein Added (test AUC) cox2 cox2.1.mol2 Delete selected cox2.1.mol2 Reset cox2.1.mol2 Method of docking cox2.14.mol2 Flexible cox2.13.mol2 cox2.13.mol2 cox2.14.mol2 cox2.14.mol2 cox2.13.mol2 cox2.14.mol2 cox2.13.mol2 cox2.19.mol2 cox2.12.mol2 cox2.126.mol2 cox2.126.mol2 | Active ligands Cox2138.mol2 Cox214.mol2 Cox214.mol2 Cox2152.mol2 Cox2152.mol2 Cox2152.mol2 Cox2152.mol2 Cox2153.mol2 Cox2162.mol2 Cox2167.mol2 Cox2167.mol2 Cox2167.mol2 Cox2167.mol2 Cox2174.mol2 Cox21 | Compounds Add to DB for screening (or for test AUC) Select mol2 files Compounds |
| Reset | Test by AUC | Delete selected Delete selected |
| | Delete selected Reset | Delete selected Delete selected Reset Reset |

- 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。

| Screening : ML-MT | S (Machine Learning - Multip | le Target Screening metho | od) | | | × |
|----------------------------------|--|--|--|---|---|-------------------------------------|
| Previous input | | | | - New input | | |
| Proteins | | Compounds | | | Compounds | |
| Added to 181 | Target-protein | Added (test AUC) | Active ligands | Target-protein | Add to DB for screening | - Active ligands as teacher- |
| Cox2 Delete selected Reset | Cox2a Delete selected Reset Method of docking Flexible | cox2L1.mol2 cox2L2.mol2 cox2L5.mol2 cox2L7.mol2 cox2L12.mol2 cox2L14.mol2 cox2L14.mol2 cox2L14.mol2 cox2L23.mol2 cox2L23.mol2 cox2L23.mol2 cox2L26.mol2 | cox2138.mol2 cox2134.mol2 cox2151.mol2 cox2151.mol2 cox2155.mol2 cox2155.mol2 cox2155.mol2 cox2155.mol2 cox2166.mol2 cox2166.mol2 cox2167.mol2 cox2177.mol2 cox2177.mol2 cox2174.mol2 cox2174.mol2 Delete selected Reset | Target-protein name for output (in ASCII without blank) Method of docking Image: State of Plexible Rigid remake grid files of protein Size of output list max No. of lines : 10500 Test by AUC Screening with small DB | (or for test AUC) Select mol2 files Compounds Delete selected | Select mol2 files Active ligands |
| | | | | | | OK Cancel |

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. 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ΔG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(IC50 値,%阻害値,活性値など)は、全て結合自由エネルギ ーΔG 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(Kd=Ki 等)。

2.1. 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(IC50 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。 <u>https://www.ebi.ac.uk/chembl/</u> (ChEMBLE トップページ) にアクセスします。

| 🖨 EMBL-EBI | Services | 🕅 Resear | rch 📥 Training | About us | | | | | | EMBL-EBI |
|------------|------------------------------|-------------|--------------------|------------------------------|------------------|-----------------|---------------|-------------------------------|-----------------|-----------------------|
| 62 | Chr | - К Л Г | וכ | | | Search in Ch | hEMBL | | | Q |
| | UNE | | JL | | | Examples: Imati | nib erbB2 bra | ain MDCK c | Draw a Structur | re Enter a Sequence |
| UniChem | ChEMB | L-NTD | SureChEMBL | Downloads | We | b Services | More | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | C | hEMBL is | a manually cura | ited database of | bioact to aid | ive molecules | s with drug | g-like propertion | es. It brings | |
| | logen | nor chomi | cui, biouctivity u | r | new dr | ugs. | on or gene | | | |
| | | | | | | | | | | |
| | | | | | | | | | | |
| | | | | | | Evolore | | MRI | | |
| | | | | | | | | VIDL | | |
| | | | | | | and quantit | ties of dat | s a summary ta for each of | them. | L entities |
| | | | | | | Instructio | ons: Click | on a bubb | le to explore a | a specific |
| | | 26K | 11K | 1.9M | | ChEMBL en | ntity in mo | ore detail. | | |
| | | Indications | Drugs | Compound | s | | | | | |
| | | 4. Mech | 7K 12K Targets | | > | | | | | |
| 1 | - 1 N/ | | angus | | | | | | | |

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。

| | | | | EMBL-EBI 🌘 |
|----------------------|--------------------|---------|------------------------|---------------|
| Tyrosine-protein | kinase ABL | | ĺ | ۹ |
| Examples: Imatinib e | rbB2 brain MDCK c1 | ccccc1N | Draw a Structure Ent | er a Sequence |
| s Old Interface | More | | | |

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 > Da | EBL > Databases > Chemical Biology > ChEMBL Database > Taroets Search Results > Tvrosine-protein kinase ABL | | | | | | |
|----------|---|---------------------------|--------------------------------------|--|------------------------------------|----------------------|-----------------------|
| Sea | rch Results | | | | | | |
| All R | esults 299 Compounds 245 T | Fargets 7 Assays 1 | 3 Documents 23 | Cells 11 Tissues | 0 | | |
| Tai | rgets | | | | | | |
| Show | Full Query | | | | | | |
| Table H | leatmap | | 0 Selecto Browse | Targets ed - Select All Activities 🕢 | | ≛ C | SV 🛓 TSV |
| ₩ | Filters 😋 | Records per page: 20 • | | Show/Hide Columns | | | ٩ |
| ~ | Organism Taxonomy L1 | Showing 1-7 out of | 7 records | | | < | 1 > |
| | Eukaryotes 6 | . e | | | | | • |
| | Viruses 1 • Organism Taxonomy L2 | ChEMBL \$ | Search Name 🗘 Hit | UniProt Type Accessions | | Compounds 🗘 | Activities |
| | Mammalia 6 retro-transcribing 1 | CHEMBL5166 | Tyrosine- protein kinase V-ABL | P00521 SINGLE PROTEIN | Abelson murine N leukemia virus | 52 By Mol. Wt.: | 68 By Std. Type |
| | Primates 5 - N/A - 1 Rodentia 1 | CHEMBL1862 | Tyrosine- protein kinase ABL | P00519 SINGLE PROTEI | N Homo sapiens | 4669 By Mol. Wt.: | 12638 By Std. Type |

以下のページが表示されます。

| T | EBI > Databases > Chemical Biology > ChEMBL Database > CHEMBL1862 | | | | | | |
|---|---|---|------------|--|--|--|--|
| ٩ | Jame And C | Classification | - | | | | |
| | ID: | CHEMBL1862 | CHEMBL1862 | | | | |
| | Туре: | 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 > Da | atabases > Chemical Biology | > ChEMBL Databa | se > Activ | ities > Query | | | | | | | | |
|---------------|-----------------------------------|-----------------|------------|---------------|----------------|----------------------------------|------------------|----------|----------|---------|-----------|----------|
| Bro | owse Act | tivities | | | | | | | | | | |
| Edit O | | | | | | | | | | | | |
| <u>Eult Q</u> | | | | | | | | | | | | |
| Snow | Full Query Ø | | | | | | | | | | | |
| | | | | | | 12,638 Activ 0 Selected - Sel | ities ect All | | | | | |
| Table | | | | | E | Browse Compo | unds 🔞 | | | | SV SV TSV | |
| | | | | | | | | | | | | |
| 筆 | Filters | | Record | s per page: | | | | | * | | ٩ | 1 |
| * | There | ~ | 20 | • | | Show/Hi | de Columns | | | | | , |
| | Standard Type | | Sho | wing 1-20 out | of 12,638 reco | rds | | | < 1 | 2 3 4 | 5 ··· > | |
| | Inhibition | 6131 | 4 | | | | | | | | | ×. |
| | IC50 | 2306 | | | | | | | | | | |
| | Ki | 1626 | П | ChEMBL \$ | Compound | Standard | Standard | Standard | Standard | pChEMBL | Comment ≑ | As Ch |
| | Kd | 1524 | | ID | Key | туре | Relation | value | Units | value | | ID |
| | Residual Activity | 316 | | | | | | | | | | |
| | Kd apparent | 243 | | | | | | | | | | |
| | Km | 29 | | | | | | | | | | |
| | Vmax | 17 | | | 18 | IC50 | > | 100000 | nM | No Data | No Data | CHE |
| | Vmax(app) | 12 | | | | | | | | | | |
| | Other Categories | 32 | | CHEMBL538507 | | | | | | | | |
| | Target Type | | | | | | | | | | | |
| | SINGLE PROTEIN | 12638 | | | | | | | | | | |
| | Organism Taxo | nomy L1 | | ma. | Genistein | IC50 | = | 10000 | nM | 5 | No Data | CHE |
| | | | | CHEMBL44 | | | | | | | | |

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。

| III Table | 12,638 Activities 0 Selected - Select All Browse Compounds 🚱 | ± csv ± tsv |
|--------------|--|-------------|
| Clic | t here to download your file. The download will expire on 2020-01-01T00:56:13.965907+00:00. Learn More | |

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

 \downarrow

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

2.2. 回帰パラメターの計算

2.2.1. 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。



2.2.2. 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。



2.2.3. 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、ファイルセレクターで前節で 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。

| D Make DB to predict Activity | | | × |
|---|----------|------------------|--------|
| ChEMBL download data file (*.tsv) | | | |
| 1. | Browse | Name of protein: | |
| 2 | Browse N | Name of protein: | |
| 3. | Browse N | Name of protein: | |
| 4. | Browse N | Name of protein: | |
| 5. | Browse N | Name of protein: | |
| 6. | Browse N | Name of protein: | |
| 7. | Browse N | Name of protein: | |
| 8. | Browse N | Name of protein: | |
| 9. | Browse N | Name of protein: | |
| 10. | Browse N | Name of protein: | |
| 11. | Browse N | Name of protein: | |
| 12 | Browse N | Name of protein: | |
| 13. | Browse N | Name of protein: | |
| 14. | Browse N | Name of protein: | |
| 15. | Browse N | Name of protein: | |
| 16 | Browse N | Name of protein: | |
| >>> Name of protein: in ASCII without blank | | | |
| Experiments used in calculation | | | |
| | | | |
| ritering | | | |
| by moleculer 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。

| MolDesk Screening | | - 🗆 × |
|--|--------------------------------------|--|
| File Select Display Color Option Expert Simple Sci | reening Preparation Window Undo Help | |
| 🔹 🗈 🗢 🖉 💿 🚹 🔣 🛃 | 💋 💰 P 💿 🛃 🗔 📕 | |
| P2-L2-p1 🛛 🗖 Command View | 🛛 🗖 🗖 project001 : 1 Initial State 🖾 | " 🗆 📄 Console 📄 Docki 📄 MD An 📄 Scree 📄 QSPR L 💥 👘 🗖 |
| De Dyna Do | Scree Prepa | |
| | | Learn Predicate |
| (1) Convert to | 3D Mol2 | Protein/Target Q |
| | | |
| Make DB fo | / Screening | |
| S Remake DB | for Screening | |
| | | |
| S Make DB to | predict Activity | |
| S. Make Rem | arrian model | |
| S make negr | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| Ligand Info 🛛 | | |
| image name SA | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |
| | 📄 jV Console 🕴 📄 Warning Console | |
| | | |
| | | A |
| | | |
| | | |
| < > | | |
| | | START: Convert to 3D Mols 31/148 |
| | | |
| | | |

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。

2.2.4. 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変わり 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。

| Consc | ole 🗖 D | ocki | ۸D An | S | cree | [🗖 🕻 | QSPR I | X | |
|---------|---------|--------|-------|----------|------|-------|--------|---|---|
| | | | | | | | | | |
| Learn | Predict | | | | | | | | |
| 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 ΔG と、回帰 パラメタ計算時の ΔG 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。



2.2.5. 回帰パラメターファイルの確認

計算結果の回帰パラメ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 |

2.3. 活性値の予測計算

前節で作成した回帰パラメ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。

2.3.1. 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。

| Predict Activity Input | × |
|------------------------|-----------|
| PCA parameter | select |
| Input Mol2 Files | |
| | |
| | |
| | |
| select | |
| | OK Cancel |

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ターファイルを下図 で選択します。

| Select PCA parameter file. | | | × |
|---|----------------------------|------------------|------------|
| \leftarrow \rightarrow \checkmark \uparrow \square \ll work \Rightarrow dates | tabase_qsar > 09.param ∨ Č | 09.paramの検索 | <i>م</i> |
| 整理 ▼ 新しいフォルダー | | | • 🔳 🕜 |
| omp ^ 名前 | | 更新日時 | 種類 |
| hread 🗋 AE | BL.param | 2018/09/03 15:33 | PARAM ファイル |
| pleiades | | | |
| project-jsch-test | | | |
| projects | | | |
| protein 🗸 < | | | > |
| ファイル名(N): 【 | ABL.param 🗸 | *.param | ~ |
| | | 開<(0) ▼ | キャンセル |

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

| m Predict Activity Input | × |
|---|-----------------------------|
| PCA parameter C:¥data¥QSAR003¥work¥database_qsa | r¥09.param¥ABL.param select |
| Input Mol2 Files | |
| | |
| | |
| | |
| select | |
| | OK Cancel |

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。

| Select Input mol2 files. | | × |
|---|------------------|----------------|
| \leftarrow \rightarrow \checkmark \uparrow \square \ll sample \Rightarrow TGS \Rightarrow ligand \checkmark \circlearrowright | ligandの検索 | م |
| 整理 ▼ 新しいフォルダー | · = = = = = | - 🔳 🕜 |
| ligand ^ 名前 | 更新日時 | 種類 ^ |
| MolDesk Basic 0c_3.mol2 | 2009/07/01 17:12 | ArgusLab Doc |
| MolDesk Screeni | 2009/07/01 17:12 | ArgusLab Doc |
| nolmil-gh-page <u>1</u> 0c_8.mol2 | 2009/07/01 17:12 | ArgusLab Doc |
| mypresto <u><u></u> 0c_13.mol2</u> | 2009/07/01 17:12 | ArgusLab Doc |
| myPresto Portal v C 1 1.mol2 | 2009/07/01 17:12 | ArgusLab Doc ¥ |
| ファイル名(N): "0c_4.mol2" "0c_3.mol2" ~ | *.mol2 | ~ |
| | 開<(O) ▼ | キャンセル |

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。

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

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。

2.3.2. 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変わります。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。

| Console Dock | i 🗖 MD A | An 📑 Scree. | 🗖 QS | PR I ⊠ | |
|------------------|-----------|--------------|--------|----------|-------|
| Learn Predict | | | | | |
| Principal Compo. | 0 | × 1 | ~ | show gra | aph |
| image | name | deltaG/Prop. | PC 0 | PC 1 | PC 2 |
| | 0c_3.mol2 | -6.319 | 0.022 | 0.403 | -0.57 |
| | 0c_4.mol2 | -6.009 | -0.007 | 0.374 | -0.60 |
| | | | | | |

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) のカラムに表示します。

3. 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。

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

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

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 > Databas | es > Chemical Biology > ChEMBL Data | ibase > | Assays Se | arch Results > | aqueo | ous solubility | | | | | | | | |
|--|--------------------------------------|-----------------|---------------|-----------------------------------|-------|----------------|---------------|--------|--|----------|---------------------|----------------------------|---|--------------------------|
| Search | Results | | | | | | | | | | | | | |
| All Result | cs 2656 Compounds 19 | Targ | gets 67 | Assays | 205 | 4 Docur | ments 5 | 15 | Cells 0 T | issues 1 | | | | |
| Assa | ays | | | | | | | | | | | | | |
| Show Full | Query 🕑 | | | | | | | | | | | | | |
| III Table | | | | | | | | С В | 2,054 Assays D Selected - Select rowse Activities | All O | | | ≛ CSV | ± TSV |
| Ŧ | Filters | o; | Record: 20 | s per page: | • | | | | Show/Hide 0 | Columns | | * | | ٩ |
| . | ▲ Type Label | | | Showing 1-20 out of 2,054 records | | | | | | | < 1 2 | 345 | · > | |
| P - P A - A B - E | Physicochemical 2 ADME Binding | 2031 19 4 | | ChEMBL ID | \$ | Search Hit | Assay Type | \$ | Description \Rightarrow | Organism | Compounds 👻 | Document ChEMBL 🗘 ID | BAO \$ Format | Source 🌩 |
| Classifications L1 | | | | | | | | | | | 148 |) | | |
| - N// | A - 2 Classifications L2 | 2054 | | CHEMBL6319 | 62 |] | Р | | Aqueous solubility | No Data | By Mol. Wt.: | CHEMBL1132890 | small-molecule physicochemical format | Scientific Literature |
| - N// | A - 2 | 2054 | | CHEMBL1034 | 1535 | | Р | | Aqueous solubility in phosphate buffered saline by | No Data | 147 By Mol. Wt.: | CHEMBL1151935 | small-molecule physicochemical | Scientific |
| ^ | Classifications L3 | _ | | | | | | | multi-screen solubility assay | | | | format | 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

| | | C |
|--------------------------|--|---|
| Assay ID: | CHEMBL631962 | |
| Туре: | 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 > D | atabases > Chemical Biology > ChEMBL Database | > Activiti | es > Query | | | | | | | | |
|----------------|---|------------|----------------|-------------------|--|--|------------|------------|-----------|--------------|-----------------|
| Br | owse Activities | | | | | | | | | | |
| Edit C Show | uerystring 🛛 | | | | | | | | | | |
| III Table | | | | | 148 Activ 0 Selected - S Browse Comp | r ities Select All Dounds 🚱 | | | | ≛ CSV | ± TSV |
| ₽ ≈ | Filters | Reco 20 | rds per page: | • | Show | /Hide Columns | | | | | ٩ |
| | Standard Type | Sh | nowing 1-20 ou | t of 148 records | | | | 4 | (1 2 | 3 4 5 | ·· > |
| | Log S 148 | • | | | | | | | | | ÷ |
| | Target Type | Г | Molecule | Compound Key 🍦 | Standard 🖕 | Standard 👙 | Standard 🌲 | Standard 🚖 | pChEMBL 🔶 | Comment ≑ | Assay ChEMBL |
| | NO TARGET 148 | | ID | | Туре | Relation | value | Units | Value | | ID |
| | Organism Taxonomy L1 | | + | | | | | | | | |
| | - N/A - 148 | | 1.010. | DDT | Log S | = | -7.15 | No Data | No Data | No Data | CHEMBL6319 |
| | Organism Taxonomy L2 | | CHEMBL41689 | 3 | | | | | | | |
| | - N/A - 148 | | 」→ <u>→</u> | Hexamethylbenzene | Log S | = | -5.23 | No Data | No Data | No Data | CHEMBL6319 |
| | Organism Taxonomy L3 | | CHEMBL16347 | | | | | | | | |

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。

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

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

 \downarrow

DOWNLOAD-XXXX.tsv

3.2. 入力実験データファイルの作成

3.2.1. 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。

| m Warning | | | | |
|-----------|---|--|--|--|
| | Make DB to predict Activity : Please execute "Save As". This command is available after saving the project by [File] - [Save As]. | | | |
| | OK | | | |

3.2.2. 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。

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

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。

| Make experiment data file Input | × |
|---|--------|
| Input ChEMBL Download File | select |
| Target: | |
| Output Experiment Data File [F:¥project¥work¥regression¥experiment.db | select |
| OK | incel |

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。)

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

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 |
| • • • | | | | | |

| 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) | 0 |
| 7 | STANDARD_TYPE(データ種別) | 0 |
| 8 | STANDARD_VALUE(データ値) | 0 |
| 9 | STANDARD_UNITS(単位) | × |
| 10 | 重み(デフォルト 1.0) | 0 |

各カラム (スペース区切り)の意味は以下の通りです。

※ 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) |
| Р | |
| 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 | 11 (sc .ge. 30.07 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。

3.2.3. 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。

3.2.4. 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 | Ра | VER | CHEMBL3430218 | COMP | CHEMBL421362 Papp 0.93 10'-6_cm/s 1.0 |
| TARGET | Ра | VER | CHEMBL3430218 | COMP | CHEMBL3425620 logPapp 0.98 - 1.0 |
| TARGET | Ра | VER | CHEMBL3430218 | COMP | CHEMBL3425623 logPapp 1.24 - 1.0 |
| TARGET | Ра | VER | CHEMBL3430218 | COMP | CHEMBL3425624 Papp 5.95 10'-6_cm/s 1.0 |
| TARGET | Ра | 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。)

105

3.2.5. 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。

| 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) | 0 |
| 7 | STANDARD_TYPE (データ種別) | 0 |
| 8 | STANDARD_VALUE(データ値) | 0 |
| 9 | STANDARD_UNITS(単位) | × |
| 10 | 重み(デフォルト 1.0) | 0 |

各カラム(スペース区切り)の意味は以下の通りです。

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。

3.3. 回帰パラメターの計算

3.3.1. 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。

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

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) | 入力する mol2 ファイル (複数選択可)。 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。 (入力は必須でない) |
| Input Regression Parameter Files (Already learned) (Option) | すでに過去に計算して得た回帰パラメターファイル (複数選択可)(入力は必須でない) |
| Regression method | 回帰計算の方法 ・SIMP: 全データ同じ重み ・RBST: ロバスト推定で重みを自動調整(デフォルト) ・WEIT: 実験データファイルに記載した重みを使用 |
| Descriptor | 計算で使用する記述子の種類 (デフォルトはすべて ON) ・Fluctuation: 物理量のゆらぎ・分散 ・Physical: 物理的なもの ・ASA: ASA ・Charge polarization: 水素結合数や水素結合しうる原子の電荷分極 ・MACCSKey: MACCSKey ・Radius: 分子の半径(平均半径、3 極での Rgyr) |
| Output Regression Parameter File | 出力する回帰パラメターファイル(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、ファイル入力が必須の項目は、

[Input Experiment Data Files] [Output Regression Paramater 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'-6_cm/s 1.0

 TARGET Pa VER CHEMBL3430218 COMP CHEMBL421362 Papp 0.93 10'-6_cm/s 1.0

 TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425620 logPapp 0.98 - 2.0

 TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425623 logPapp 1.24 - 2.0

 TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425624 Papp 5.95 10'-6_cm/s 1.0

 TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425624 Papp 5.95 10'-6_cm/s 1.0

 TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425629 logPapp -1.05 - 2.0

 TARGET Pa VER CHEMBL3430218 COMP CHEMBL3425629 logPapp -1.05 - 2.0

[Despriptor] 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。

3.3.2. 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。)

| Select Input experiment data | ata files. | | | × |
|---|---------------------------------------|------------------|-----------------|--------|
| $\leftarrow \rightarrow \cdot \uparrow \square \ll \text{proj}$ | ject > project001 > work > regression | ~ č |) regressionの検索 | ٩ |
| 整理 ▼ 新しいフォルダー | | | | = |
| MolDesk_sup ^ | 名前 ^ | 更新日時 | 種類 | サイズ |
| MolDesk_wor | 💿 experiment.db | 2018/12/06 18:42 | Data Base File | 10 KB |
| myPrestoPort | experiment2.db | 2018/12/06 20:08 | Data Base File | 11 KB |
| namiki_medi1 | experiment3.db | 2018/12/06 20:09 | Data Base File | 11 KB |
| project | experiment4.db | 2018/12/06 20:09 | Data Base File | 8 KB |
| project001 | experiment5.db | 2018/12/06 20:10 | Data Base File | 267 KB |
| System Volun 🗸 | | | | |
| ファイル | 名(N): experiment.db | | ✓ *.db | ~ |
| | | | 開く(O) | キャンセル |

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。

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

[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。

| m Warı | ning | × |
|--------|--|---|
| | Please set correct ChEMBL sdfs database for regression at [Help]-[Preference]-[Screening]-[ChEMBL sdfs directory for regression]. | |
| | OK | |

[Input Mol2 Files (Option)] を入力してないので、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。

3.3.3. 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変わり 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。

| Console | 📩 Docki | MD An | Scree | QSPR | I 🖾 | |
|-------------|---------|-------|-------|------|-----|---|
| | | | | | | |
| Learn Pre | edict | | | | | |
| Protein/Tar | rget 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。



3.3.4. 回帰パラメターファイルの確認

計算結果の回帰パラメ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) なども保存される。 |
| 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ターファイル |

3.4. 特性値の予測計算

前節で作成した回帰パラメ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。

3.4.1. 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。

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

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

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。

| Select Input mol2 files. | | × |
|--|---------------------|---------------|
| \leftarrow \rightarrow \checkmark \Uparrow \checkmark sample \rightarrow TGS \Rightarrow ligand \checkmark | ご ligandの検索 | ρ |
| 整理 ▼ 新しいフォルダー | | |
| ligand 个 名前 | 更新日時 種 | 類 ^ |
| MolDesk Basic 0c_3.mol2 | 2009/07/01 17:12 Ar | gusLab Doc |
| MolDesk Screeni 0c_4.mol2 | 2009/07/01 17:12 Ar | rgusLab Doc |
| molmil-gh-page 1_0c_8.mol2 | 2009/07/01 17:12 Ar | rgusLab Doc |
| mypresto <u>1</u> 0c_13.mol2 | 2009/07/01 17:12 Ar | rgusLab Doc |
| myPresto Portal V < 01 1.mol2 | 2009/07/01 17:12 Ar | rausLab Doc ¥ |
| ファイル名(N): "0c_4.mol2" "0c_3.mol2" | ∽ *.mol2 | ~ |
| | 開く(O) ► = | キャンセル: |

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。

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

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。

3.4.2. 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変わります。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。

| 🗖 Console 📄 Dock | i 🗖 MD A | An 📩 Scree. | 🗖 QSP | RI ⊠ | |
|------------------|----------|--------------|----------|----------|-------|
| Learn Predict | | | | | |
| Principal Compo. | 0 | × 1 | ~ | show gra | ph |
| image | name | deltaG/Prop. | PC 0 | PC 1 | PC 2 |
| | 0c_4 | -36.27 | 1794.614 | -329.625 | -146. |
| | 0c_3 | -31.956 | 1695.912 | -255.335 | 56.29 |
| | | | | | |

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。

4. 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、重なりの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の%の値です。

4.1. 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



4.2. 検索対象分子の選択

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

| m MVO Screening | × |
|--|-----|
| Select property of COMMENT line in Mol2 files to identify the molecule | 5.1 |
| OK Cancel | |

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。

| m MVO Screening | × |
|---|-----|
| | |
| Select property of COMMENT line in Mol2 files to identify the molecul | e.: |
| | |
| OK Correct | |
| OK Cancel | |

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

| m Select mol2 files | | | × |
|---|------------------|-----------|----------|
| $\leftarrow \rightarrow \checkmark \uparrow$ _ < sample > MVO_screening > target | ٽ ~ | targetの検索 | م |
| 整理 ▼ 新しいフォルダー | | | |
| 素 クイック アクセス ^^ 名前 ^^ | 更新日時 | 種類 | サイズ |
| ↓ ダウンロード ★ interval inter | 2017/01/18 15:46 | MOL2 ファイル | 1,211 KB |
| 🔜 デスクトップ 🖈 | | | |
| ドキュメント オ | | | |
| 📰 ピクチャ 🔹 🗸 | | | |
| ファイル名(N): all.mol2 | ~ | *.mol2 | ~ |
| | | 開く(O) | キャンセル |

※ 例では、1ファイルですが、複数ファイルも選択できます。

4.3. 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



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。

5. 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。

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] - [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。



5.2. 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。

| m Select mol2 files | | | | | × |
|---------------------|--|----------------------|----------------|---------|-------|
| 📀 🎯 🔹 🕇 🐌 « MolDes | | | | | |
| 整理 ▼ 新しいフォルダー | | | | | |
| projects ^ | 名前 | 更新日時 | 種類 | サイズ | ^ |
| Jacsy Anna | 0c_3.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | |
| 4prq | 0c_4.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | |
| | 0c_8.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | |
| | 0c_13.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 4 KB | |
| proj007 | 01_1.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | |
| proj007 | 01ca7.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | |
| proj009 | 01gcz.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | |
| test001 | 01ljt.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 4 KB | |
| test002 | a4g.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 5 KB | |
| test003 | 1a4q.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 6 KB | |
| test004 | 1a28.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 6 KB | |
| test005 | 1a42.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 5 KB | |
| test006 | 1abe1.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | |
| test124 ACE | 1abe2.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | |
| test124_ACE_NME ¥ | 1abf1.mol2 | 2009/07/01 17:12 | MOL2 ファイル | 3 KB | ~ |
| ファイル名(N |): "supr.mol2" "0c_3.mol2" "0c_4.mol2" "0c_4 | 8.mol2" "0c_13.mol2" | "01_1.mo v *.I | mol2 | ~ |
| | | | | 開<(0) 4 | Fヤンセル |

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。

| 💼 *untitled : 5 Docking 📄 *untitled : 2 Topology Graph Simi 🛛 🖓 | 🗖 🖳 Console 🕮 📩 Screening Info | |
|--|--|-------|
| | Standard Output | |
| | calculate DB enrichment? (y or n) | ~ |
| | readmol3= | |
| | \1\inp0.mol2 | |
| | | |
| | numatom 36 36 | |
| | === predicted compounds === | |
| | 1 .\ligand\abclen.mol2 0.0000000 | 000 |
| | 2 .\ligand\absalb.mol2 0.0000261 | 141 |
| | 3 .\ligand\0303_1.mol2 0.00001611 | 131 |
| | 4 .\ligand\0303 2.mol2 0.00001611 | 131 |
| | 5 .\ligand\lckp.mol2 0.00001893 | 301 |
| | 6 .\ligand\01 1.mol2 0.00002332 | 246 |
| | 7 .\ligand\s 8ohdpa.mol2 0.00002356 | 670 |
| | 8 .\ligand\3cpa.mol2 0.00002363 | 342 |
| | 9 .\ligand\1phg.mol2 0.00002392 | 249 |
| | 10 .\ligand\01gcz.mol2 0.00002675 | 501 |
| | 11 .\ligand\1mif1.mol2 0.00002686 | 643 |
| | 12 .\ligand\ldr1.mol2 0.00002720 | 029 |
| | 13 .\ligand\lc1e.mol2 0.00002811 | 120 |
| | 14 .\ligand\abpropr.mol2 0.00003180 | 040 |
| | 15 .\ligand\0c 3.mol2 0.00003478 | 880 |
| | 16 .\ligand\2ada.mol2 0.00003560 | 047 |
| | 17 .\ligand\0c 4.mol2 0.00003624 | 421 |
| | 18 .\ligand\2ifb.mol2 0.00003689 | 928 |
| | 19 .\ligand\d_quinpi.mol2 0.00003703 | 373 |
| | 20 .\ligand\0c_8.mol2 0.00003745 | 566 |
| | 21 .\ligand\abmetop.mol2 0.00003872 | 219 |
| | 22 .\ligand\abateno.mol2 0.00004214 | 454 |
| | 23 .\ligand\keto.mol2 0.00004349 | 973 |
| | 24 .\ligand\livb.mol2 0.00004442 | 297 |
| | 25 .\ligand\abterb.mol2 0.00004502 | 286 |
| | 26 .\ligand\abalpre.mol2 0.00004509 | 969 |
| | 27 .\ligand\d_apomor.mol2 0.00004587 | 711 |
| | 28 .\ligand\abpindo.mol2 0.00004845 | 559 |
| | 29 .\ligand\1d3h.mol2 0.00004852 | 247 |
| | 30 .\ligand\lokl.mol2 0.00004861 | 159 |
| IV Console 1 | 31 .\ligand\d_skf383.mol2 0.00004896 | 636 |
| | 32 .\ligand\Diflunisal.mol2 0.00004963 | 398 |
| | 33 .\ligand\d_dobuta.mol2 0.00005486 | 639 |
| | 34 .\ligand\1dala.mol2 0.00005558 | 880 |
| Distance #1-#2; U.U Atom #1: CA 259 Pacidua: DHE 699 Chain(auth/labol): A/2 Eiler 1 | 35 .\ligand\llala.mol2 0.00005558 | 880 |
| Coordinates: (16.854.15.448.189.056.) | 36 .\ligand\s_sumatr.mol2 0.00005678 | 821 |
| Atom #2: CA 258 Residue: PHE 688 Chain(auth/label): A/? File: 1 | 37 .\ligand\1f3d.mol2 0.00005731 | 121 |
| Coordinates: (16.854, 15.448, 189.056) | 38 .\ligand\01ca7.mol2 0.00005806 | 679 |
| Distance #1-#2: 0.0 | 39 .\ligand\3cla.mol2 0.00005952 | 256 |
| Atom #1: CA 258 Residue: PHE 688 Chain(auth/label): A/? File: 1 | 40 .\ligand\lfen.mol2 0.00005954 | 422 |
| Coordinates: (16.854, 15.448, 189.056) | 41 .\ligand\DHICA.mol2 0.00006276 | 675 |
| | 42 .\ligand\D_dopachrome.mol2 0.00006356 | 609 Y |
| | | |
| | | |
| | | |

6. 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。

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 -> 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。



6.2. 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。

| | rch → ligand | u e lina | | |
|---|-------------------------|-----------------|-----------------|---|
| | | ♥ 0 liga | indの検索 | Q |
| 整理 ▼ 新しいフォルダー | | | •== • | |
| Deprojects 个 名前 | 更新日時 | 種類 | サイズ | |
| 3csy 001.mol2 | 2010/03/08 16:29 | MOL2 ファイル | 3 KB | |
| 4prq 002.mol2 | 2010/03/08 16:29 | MOL2 ファイル | 3 KB | |
| o03.mol2 | 2010/03/08 16:29 | MOL2 ファイル | 3 KB | |
| metan001 | 2010/03/08 16:29 | MOL2 ファイル | 3 KB | |
| 005.mol2 | 2010/03/08 16:29 | MOL2 ファイル | 3 KB | |
| proj007 006.mol2 | 2010/03/08 16:29 | MOL2 ファイル | 3 KB | |
| 007.mol2 | 2010/03/08 16:29 | MOL2 ファイル | 3 KB | |
| test001 | 2010/03/08 16:29 | MOL2 ファイル | 3 KB | |
| test002 | 2010/03/08 16:29 | MOL2 ファイル | 4 KB | |
| test003 | 2010/03/08 16:31 | MOL2 ファイル | 3 KB | |
| test004 | | | | |
| test005 | | | | |
| Lest006 | | | | |
| 🔰 test124_ACE | | | | |
| 🚡 test124_ACE_NME 🖌 | | | | |
| ファイル名(N): ["010.mol2" "001.mol2" "002.mol2" "003. | mol2" "004.mol2" "005.r | mol2" "00 ∨ *.r | nol2 開〈(O) : | ▼ |

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。



7. 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54A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54Å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。



7.1. MPI 動作環境の設定方法

7.1.1. 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。

7.1.2. 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

Ubuntsu 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。

7.2. 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。

7.2.1. 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

7.2.2. 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

7.3. 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。

7.3.1. 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。

| • | no MPI | (cosgene) |
|---|--------|-----------|
|---|--------|-----------|

| Preferences | | | – o x |
|---------------------------------|---|--|-----------|
| | 1. Molecular dynamics | | ⇔ ⇒ ⇒ ₹ 8 |
| 1. Molecular dy 2. Screening | Molecular Dynamics Program : o no MPI (cosgene) MPI (cosg | ene_MPI) MPI (psygene) MPI + GPU (psygene-G) • GROMACS | |
| 3. Docking | Processes of MPI | 10 | |
| 4. H bond 5. 3D view | GPU 0 ID | 0 | |
| 6. Molecule | GPU 1 ID | | |
| 7. Internet | GPU 2 ID | | |
| 8. Other | GPU 3 ID | | |
| ANSI Support | | | |
| | GROMACS directory | C:¥Program Files¥MolDeskScreening¥gromacs¥win64¥gromacs-2021.7 | Browse |
| | Steps to be automatically devided | 1000000 | |
| | Use chiral server : • Yes • No Chiral Server User ID Chiral Server API Key | | |
| | | Restore Defaults | Apply |
| | | Apply and Close | Cancel |

[no 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)

| Preferences | | | – o × |
|-----------------|-----------------------------------|--|-------------|
| | 1. Molecular dynamics | | (- ▼ -> ▼ 8 |
| 1. Molecular dy | Molecular Dynamics Program : | | |
| 2. Screening | ono MPI (cosgene) OMPI (cosg | ene_MPI) O MPI (psygene) O MPI + GPU (psygene-G) O GROMACS | |
| 3. Docking | Processes of MPI | 10 | |
| 4. H bond | GPU 0 ID | 0 | |
| 5. 3D view | | | |
| 6. Molecule | GPU 1 ID | | |
| 7. Internet | GPU 2 ID | | |
| 8. Other | GPU 3 ID | | |
| ANSI Support | | | |
| | GROMACS directory | C:¥Program Files¥MolDeskScreening¥gromacs¥win64¥gromacs-20217 | Browse |
| | | 1000000 | |
| | steps to be automatically devided | 100000 | |
| | | | |
| | Use chiral server : | | |
| | ○Yes ⁰No | | |
| | Chiral Server User ID | | |
| | Chiral Server API Key | | |
| | | | |
| | | Pectore Defaulte | Apply |
| | | Restore Defaults | Арріу |
| | | Apply and Close | Cancel |
| | | | |

[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÷2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 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)

| M Preferences | | | – 🗆 X |
|-----------------|-----------------------------------|---|---|
| | 1. Molecular dynamics | | ⇔ • ⇔ • 8 |
| 1. Molecular dy | Molecular Dynamics Program : | | |
| 2. Screening | ○ no MPI (cosgene) ○ MPI (cosg | ene MPI) OMPI (psygene) OMPI + GPU (psygene-G) GROMACS | |
| 3. Docking | Processes of MPI | 10 | |
| 4. H bond | | 0 | |
| 5. 3D view | GPOOLD | 0 | |
| 6. Molecule | GPU 1 ID | | |
| 7. Internet | GPU 2 ID | | |
| 8. Other | GPU 3 ID | | |
| ANSI Support | | | |
| | GROMACS directory | C:¥Program Files¥MolDeskScreening¥gromacs¥win64¥gromacs-20217 | Browse |
| | | | |
| | steps to be automatically devided | 100000 | |
| | | | |
| | Use chiral server : | | |
| | ○Yes ⁰No | | |
| | Chiral Server User ID | | |
| | Chiral Server API Key | | |
| | | | |
| | | Restore Defaults | Apply |
| | | | , ,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,,, |
| | | Apply and Close | Cancel |
| | | | |

[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÷2です。

[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

| m Preferences | | | – o x |
|---|--|---|-----------|
| | 1. Molecular dynamics | | ⇔ • ⇔ • ≬ |
| 1. Molecular dy 2. Screening 3. Docking 4. H bond 5. 3D view 6. Molecula | Molecular Dynamics Program : ono MPI (cosgene) OMPI (cosg Processes of MPI GPU 0 ID GPU 1 ID | ene_MPI) OMPI (psygene) OMPI + GPU (psygene-G) GROMACS 10 0 | |
| 7. Internet 8. Other ANSI Support | GPU 2 ID GPU 3 ID | | |
| | GROMACS directory Steps to be automatically devided | C:¥Program Files¥MolDeskScreening¥gromacs¥win64¥gromacs-2021.7 1000000 | Browse |
| | Use chiral server : • Yes • No Chiral Server User ID Chiral Server API Key | | |
| | | Restore Defaults | Apply |
| | | Apply and Close | Cancel |

[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。

• Use chiral Server=Yes

| Preferences | | | – 🗆 🗙 |
|--|---|---|-------------|
| | 1. Molecular dynamics | | (- • -> • 8 |
| 1. Molecular dy 2. Screening 3. Docking 4. H bond 5. 3D view 6. Molecule 7. Internet 8. Other ANSI Support | Molecular Dynamics Program : no MPI (cosgene) MPI (cosge Processes of MPI GPU 0 ID GPU 1 ID GPU 2 ID GPU 3 ID | ene_MPI) OMPI (psygene) OMPI + GPU (psygene-G) GROMACS 10 0 | |
| | GROMACS directory Steps to be automatically devided | C:¥Program Files¥MolDeskScreening¥gromacs¥win64¥gromacs-2021.7 1000000 | Browse |
| | Use chiral server : • Yes • No Chiral Server User ID Chiral Server API Key | | |
| | | Restore Defaults | Apply |
| | | Apply and Close | Cancel |

[Use chiral server=Yes] を選択したときは、Chiral 社が提供するクラウドサーバの GROMACS によるエネルギー極小化計算と MD 計算を実行します。別途、ユーザと Chiral 社との契約が必要です。

7.3.2. 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。

| m Preferences | | | - D X |
|---------------------------|---|------|----------------------|
| type filter text | 2. Screening | | ↓ ↓ ↓ ↓ |
| 2. Screening 4. H bond | Thread number Database directory for screening | 8 | Browse |
| 5. 3D view 6. Molecule | Max number of screening result images | 1000 | |
| 7. Internet | ChEMBL sdfs directory for regression | | Browse |
| | | Re | store Defaults Apply |
| | | | OK Cancel |

「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



| m Preferences | | — | |
|-------------------------|---------------------------------------|----------------------|-------------------------------|
| type filter text | 2. Screening | | <p th="" ⇒="" ▼="" ▼<=""></p> |
| 2. Screening | Thread number | 8 | |
| 4. H bond 5. 3D view | Database directory for screening | F:¥namiki_medi170313 | Browse |
| 6. Molecule | Max number of screening result images | 1000 | |
| 7. Internet | ChEMBL sdfs directory for regression | | Browse |
| | | | |
| | | Restore Defaults | Apply |
| | | ОК | Cancel |
| | | | |

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。

| フォルダーの参照 | × |
|-------------------------|---|
| | |
| | |
| | |
| | _ |
| ✓ MKDB011 | ^ |
| original | |
| 🗸 📊 work | |
| 1 | |
| > 2 | |
| V database | |
| > ligand | |
| ligandImage | |
| mol2 files | |
| mtr data | |
| mis_data | |
| > protein | × |
| | _ |
| フォルター(F): database | |
| | |
| 新しいフォルダーの作成(N) OK キャンセル | , |
| | |

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。

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

[Screening Info] 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。

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

[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

| フォルダーの参照 | | × |
|-----------|------------------------|-----|
| | | |
| | | |
| 🗸 🗸 cl | hembl_24_sdfs_moldesk | ^ |
| | 2000 | |
| | c001 | |
| | c002 | |
| | c003 | |
| | c004 | |
| | c005 | ~ |
| フォルダー(F): | chembl_24_sdfs_moldesk | |
| | | |
| 新しいフォルダー | の作成(N) OK キャンセ | 211 |

| m Preferences | | | □ × |
|---------------------------------------|--|-----------------------|-------------------------|
| type filter text | 2. Screening | < | ;→ - ;→ - |
| 1. Molecular dynamics 2. Screening | Thread number 8 | | |
| 4. H bond 5. 3D view | Database directory for screening | amiki_medi170313 | Browse |
| 6. Molecule | Max number of screening result images 1000 |) | |
| 7. Internet | ChEMBL sdfs directory for regression F:¥ch | hembl_24_sdfs_moldesk | Browse |
| | 1 | | |
| | | Restore Defaults | Apply |
| | | ОК | Cancel |
| | | | |

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。

7.3.3. 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。

| Preferences | | - 0 | × | |
|--|---|--|---|--|
| | 6. Molecule | ¢ • ¢ • § | | |
| Molecular dy Screening Docking | Force Field [Protein/RNA/DNA/Metals (TplgeneX)] : AMBER ff99SB | t parm96 OCHARMm22 (for Amino Acid Only) OCHARMm19 (for Amino Acid Only) | | |
| 4. H bond 5. 3D view | Protein Force Field : | ff99_SB_ILDN_aa | ~ | |
| 6. Molecule | RNA Force Field : | RNA_OL3 | ~ | |
| 7. Internet | DNA Force Field : | DNA_OL15 | ~ | |
| 8. Other ANSI Support | Force Field [Water (TplgeneX)]: OTIP3P OOPC3 Force Field [Small molecules (TplgeneL)]: OGAFF ver.2.1 OGAFF ver.1.8 OGAFF ver.1.7 OAMBER parm99 | | | |
| | Protein > Content of standard amino acid (%) | 80 | | |
| | Nucleotide > Content of standard nucleotide residue (%) | 80 | | |
| | Glycan > Content of standard glycan residue (%) | 80 | | |
| | MolSite UAP file path | C:¥workspace¥moldesk¥MolDesk¥mypresto¥template¥uap.mol2 Browse | | |
| | | Restore Defaults Apply | | |
| | | Apply and Close Cancel | | |
7.4. psygene / psygene-Gによる MD 計算

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

において、[MPI (psygene)] または [MPI + GPU (psygene-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 (直 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。