

myPresto 5.0

ユーザー マニュアル

2018 年 9 月 26 日

プログラム使用許諾条件

別紙参照のこと。

コピーライトについて

著作権表示、及び、コピーライトは、特に断りのないものについては下記のとおりです。

myPresto version 5 .0:

次世代天然物化学技術研究組合

Copyright (C) 2006-2018 Next Generation Natural Product Chemistry (N²PC)

謝辞

本ソフトウェアの研究開発は、国立研究開発法人日本医療研究開発機構(AMED)の援助によって行われました。ここに感謝の意を記します。

本ソフトウェアは、故・京極好正博士の始められた研究の中で開発されました。

cosgene で使用している Particle Mesh Ewald (PME) のサブルーチンは、Tom Darden 博士の御厚意により提供されたコードを参考に作成されました。ここに感謝の意を記します。

The Particle Mesh Ewald (PME) routines were originally developed by Dr. Tom Darden [6].
National Institute of Environmental Health Sciences,
Research Triangle Park,
North Carolina 27709 US.

Accessible Surface Area (ASA) のサブルーチンは、金城玲博士の御厚意により提供されたコードを参考に作成されました。ここに感謝の意を記します。

The Accessible Surface Area (ASA) routines were originally developed by Dr. Akira Kinjo [7].
Center for Information Biology and DNA Data Bank of Japan,
National Institute of Genetics,
Mishima, Shizuoka, 411-8540, Japan.
(現在の所属) Institute for Protein Research,
Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

cosgene で使用している Fast Multipole Method (FMM) のサブルーチンは、William T. Rankin 博士の御厚意により提供されたコードを参考に作成されました。ここに感謝の意を記します。

The Fast Multipole Method (FMM) routines were originally developed by Dr. William T. Rankin [8-10].
Center for Computational Science and Engineering
Duke University
Dept. of Electrical and Computer Engineering
Box 90291
Durham, NC 27708-0291 US.

ソフトウェアの著者

◆ myPresto version 5

中村 春木 (Nakamura, Haruki)
福西 快文 (Fukunishi, Yoshifumi)
真下 忠彰 (Mashimo, Tadaaki)
黒澤 隆 (Kurosawa, Takashi)
杉原 崇憲 (Sugihara, Takanori)
和田 光人 (Wada, Mitsuhiro)
肥後 順一 (Higo, Junichi)
福田 育夫 (Fukuda, Ikuo)
川端 猛 (Kawabata, Takeshi)
神谷 成敏 (Kamiya, Narutoshi)
笠原 浩太 (Kasahara, Kota)
Bhaskar Dasgupta
金井 千里 (Kanai, Chisato)
吉川 達也 (Yoshikawa, Tatsuya)
杉原 裕介 (Sugihara, Yusuke)
中村 寛則 (Nakamura, Hironori)

◆ myPresto version 4

中村 春木 (Nakamura, Haruki)
福西 快文 (Fukunishi, Yoshifumi)
金森 英司 (Eiji, Kanamori)
窪田 聡 (Kubota, Satoru)
尾曲 克己 (Omigari, Katsumi)
福田 育夫 (Fukuda, Ikuo)
和田 光人 (Wada, Mitsuhiro)
真下 忠彰 (Mashimo, Tadaaki)
三友 大輔 (Mitomo, Daisuke)
坂本 晋子 (Sakamoto, Kuniko)
黒澤 隆 (Kurosawa, Takashi)
橋 祐一 (Hashi, Yuichi)
三上 義明 (Mikami, Yoshiaki)
上田 寛 (Ueda, Hiroshi)

◆ myPresto version 3

中村 春木 (Nakamura, Haruki)
福西 快文 (Fukunishi, Yoshifumi)
金 在吉 (Jae Gil Kim)
渡辺 幸久 (Watanabe, YS)
三上 義明 (Mikami, Yoshiaki)
窪田 聡 (Kubota, Satoru)
尾曲 克己 (Omigari, Katsumi)
巽 理恵 (Tatsumi, Rie)
堀江 将 (Horie, Masaru)
福田 育夫 (Fukuda, Ikuo)

◆ myPresto version 2

中村 春木 (Nakamura, Haruki)
福西 快文 (Fukunishi, Yoshifumi)
金 在吉 (Jae Gil Kim)
三上 義明 (Mikami, Yoshiaki)
渡辺 幸久 (Watanabe, YS)
伊奈 康夫 (Ina, Yasuo)
堀江 将 (Horie, Masaru)
高橋 誠 (Takahashi, Makoto)
福田 育夫 (Fukuda, Ikuo)

◆ myPresto version 1
 cosgene version 1 :
 中村 春木 (Nakamura, Haruki)
 福西 快文 (Fukunishi, Yoshifumi)
 橋 祐一 (Hashi, Yuichi)
 三上 義明 (Mikami, Yoshiaki)
 木寺 詔紀 (Kidera, Akinori)
 寺田 透 (Terada, Toru)
 tplgene version 1 :
 中村 春木 (Nakamura, Haruki)
 福西 快文 (Fukunishi, Yoshifumi)
 黒田 正孝 (Kuroda, Masataka)
 福田 育夫 (Fukuda, Ikuo)

目次

1. myPresto の概要と特長	1
1.1. 概要	1
1.2. 特長	1
2. myPresto を構成するプログラムとデータ	3
3. ファイルの入手先と利用方法	5
3.1. ファイルの入手先	5
3.2. 利用方法	5
4. myPresto で実行できる計算	7
4.1. 分子動力学シミュレーション (cosgene)	7
4.2. 高速分子動力学計算 (psygene/psygene-G)	9
4.3. ドッキングシミュレーション (sievgene/sievgene_M)	10
4.4. インシリコ・スクリーニング	12
4.5. 類似化合物探索 (substructure_search, TGS)	16
4.6. ドッキング構造の精密計算 (sievgene_NMR, sievgene_MV0)	16
4.7. 結合自由エネルギーの計算 (sievgene_DIAV)	16
4.8. 精密な結合自由エネルギー計算 (SPRG)	16
4.9. 化合物溶解度/アグリゲーター予測 (LogS_predictor)	17
4.10. タンパク質表面のポケット探索 (MolSite)	17
4.11. 化合物の 2D→3D 変換・化合物ライブラリ作成 (3DdataConstruction)	18
4.12. 低分子化合物構造の編集 (EditMol)	18
5. ツールプログラム	19
6. 環境変数の設定	20
7. 配布ファイルリスト	22
8. ドキュメントリスト	23
9. 参考文献	24

1. myPresto の概要と特長

1.1. 概要

myPresto (マイ・プレスト) は、経済産業省、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) および国立研究開発法人日本医療研究開発機構 (AMED) からの委託で実施されたプロジェクトにおいて開発された医薬品開発支援のための分子シミュレーションシステムです。従来に比べて短時間で高精度・高能率に、タンパク質や低分子化合物等のモデリング、水溶液中や脂質二重膜中でのタンパク質と低分子化合物との複合体の分子動力学計算、薬物ドッキング・薬物スクリーニング等を行うことができます。myPresto の名称は、Medicinally Yielding Protein Engineering SimulaTor に由来しています。

myPresto は、多数のプログラム群、データセットの総称です。myPresto version 5.0 に含まれる代表的なプログラムおよびデータには以下にものがあります。

- ① 分子シミュレーションに必要な分子の詳細情報の作成:

tplgeneX(ペプチド鎖等高分子用), tplgeneL(低分子化合物用)

- ② 分子動力学(molecular dynamics(MD))計算プログラム: cosgene, psygene(GPU 対応版)

- ③ ドッキングプログラム: sievgene

- ④ 各種ツール群: MolSite(ポケット探索), EditMol(分子編集), TGS(類似化合物探索), LogS_predictor(溶解度予測), SyntheticAccessibility(合成容易性予測)

- ⑤ 化合物データベース: LigandBOX

1.2. 特長

[分子動力学計算とドッキング計算との連携が容易]

myPresto には、MD 計算プログラムとドッキングプログラムの両方が含まれているため、MD 計算の結果 (例えば、MD 計算のトラジェクトリーにおけるタンパク質立体構造のスナップショット) を使ったドッキングシミュレーションや相互作用解析を簡単に行うことができます。また逆に、ドッキング計算で得られたタンパク質・低分子化合物の複合体構造を初期構造とした MD シミュレーションを容易に実行することができます。myPresto では、MD 計算とドッキング計算を連携して行う高度な計算方法が多数採用されています。

[インシリコ・スクリーニング計算実施が容易]

myPresto の一部として提供されている化合物ライブラリ LigandBOX を使用すれば、200 万化合物を対象としたインシリコ・スクリーニングを実行することができます。インシリコ・スクリーニングは、膨大な数の低分子化合物を含む化合物ライブラリから、実験的に効能を検証する化合物を絞り込む計算のことを言います。インシリコ・スクリーニングの

計算方法は、大きく分けて2つあります。ターゲットタンパク質の立体構造に基づくスクリーニング方法 (Structure-based drug screening (SBDS)) と、既知活性化合物の構造に基づくスクリーニング方法 (Ligand-based drug screening (LBDS)) です。myPresto では、SBDS と LBDS の両方のスクリーニング法を使うことができます。通常は、インシリコ・スクリーニング計算で選出された化合物の多くは、実験的には不活性です。実験で効能を評価した化合物の中に含まれる活性のある化合物 (ヒット化合物) の割合は、ターゲットタンパク質に大きく依存しますが、一般のインシリコ・スクリーニングでは 1%程度だと言われています。myPresto を使用したインシリコ・スクリーニングでは、11 種類のターゲットタンパク質に対して、200 以上の新規活性化合物が 3~50%のヒット率で発見されています。

[低分子化合物関連ツールが充実]

myPresto では、既存の分子の一部を編集するツール (EditMol) や、化合物の溶解度予測ツール (LogS_predictor)、類似化合物探索ツール (TGS) など、創薬の現場で役に立つツール群が多数用意されています。

2. myPresto5 を構成するプログラムとデータ

myPresto に含まれるプログラムとデータには以下のものがあります。

- myPresto Portal: myPresto5 用のグラフィック・ユーザー・インターフェイス・プログラム
- tplgeneX: タンパク質・核酸等の高分子鎖に対して AMBER/CHARMm の力場を割り当てる。
- tplgeneL: 化合物に対して general AMBER force field (GAFF) を割り当てる。
- cosgene: 分子動力学シミュレーションを行う。NVE/NVT/NPT アンサンブル、SHAKE, 剛体モデル、マルチカノニカルMD, 各種アンブレラサンプリング、GBSA を行える。SRPG 法による自由エネルギー計算が可能。
- psygene/psygene-G: GPU も用いた高速の分子動力学シミュレーションを行う。
- Hgene: 化合物への H 原子付加、原子電荷計算など。
- sievgene/sievgene_M: タンパク質—化合物ドッキングソフト
- sievgene_NMR: タンパク質—化合物溶液の NMR 実験による化合物のシグナルを再現するようなタンパク質—化合物複合体構造を予測するタンパク質—化合物ドッキングソフト
- sievgene-MVO: ドッキングスクリーニングにおいて、既知薬物とタンパク質の複合体を参考にして、既知薬物への分子の重ね合わせとドッキングを同時に行う薬物スクリーニング
- sievgene_DIAV: タンパク質—化合物複合体において、MD トラジェクトリーから結合自由エネルギーを推算する。
- PDBcheck: PDB ファイルの中身をチェックし、ペプチド鎖 N, C 末端の処理や、化合物を分子単位に TER で区切ることで、PDB を構成する分子を明確にする。
- UAP: Universal Active Probe 法による薬物スクリーニング結果の選択
- VCOL: コンビナトリアルな分子の生成ツール
- confgene/ confgeneC: 分子の配座生成ツール
- Matrix: in silico screening (Multiple Target Screening method, Docking Score Index method): MTS 法、DSI 法での薬物スクリーニング (MTS 法は SBDS, DSI 法は LBDS)
- 3DdataConstruction: 化合物データベース生成ツール
- MVO: Maximum Volume Overlap 法によるドッキング構造の精密化
- MVO_screening: MD-MVO 法による類似化合物探索 (LBDS)
- substructure-search: 分子骨格による類似化合物探索 (LBDS)
- TGS: TGS 法による類似化合物探索 (LBDS)
- LogS_predictor/Property_predictor: 分子記述子などによる溶解度推算及びアグリゲーターの予測
- MolSite: 化合物ドッキングによるタンパク質ポケット及び活性の予測
- SyntheticAccessibility: 化合物合成容易性の予測

- membgene: 膜タンパク質のモデルに脂質二重膜と水分子を配置するプログラム

3. ファイルの入手先と利用方法

3.1. ファイルの入手先

myPresto のプログラムやデータは、以下の URL のサーバーからダウンロード可能です。

myPresto ダウンロードサイトの URL:

<http://mypresto5.jp/>

3.2. 利用方法

myPresto の配布用プログラムやデータは、圧縮ファイルで提供されています。圧縮されたファイルは、以下のコマンドで解凍してください。(Linux での使用を想定しています。cygwin や MinGW 等の Windows 上で設定した UNIX 環境では、不具合が発生する場合があります。一部のプログラムは Windows 上でも利用可能です。)

```
% tar -xzf (program)YYMMDD.tar.gz      (YYMMDD には年月日を示す数字が入ります。)
```

(program) の部分は、ダウンロードするファイルによって異なります。

myPresto version 5.0 では、多くのダウンロードファイルは、解凍後に、以下のディレクトリ構造を持つように準備されています。

```
(program)YYMMDD/  
├── doc/  
│   ├── (program)_manual_v5.000_ja_YYMMDD.pdf  
│   └── ...  
├── src/  
├── bin/  
└── ...  
    ├── install.sh  
    └── ...
```

doc/の下に、マニュアル、もしくは、チュートリアル の PDF ファイルが置かれています。プログラムの使い方は、これらを参照してください。

また、bin/install.sh を実行すると、プログラムをインストールできるようになっています。(必要なコンパイラがインストールされている必要があります。)

その他、プログラムによっては、プログラムの動作確認用に、テスト計算実行スクリプト (test_*.sh) が用意されています。

4. myPresto で実行できる計算

本章では、myPresto で実行できる計算について説明します。

4.1. 分子動力学シミュレーション(cosgene)

myPresto では、水中もしくは脂質二重膜中でのタンパク質と低分子化合物の複合体構造の分子動力学シミュレーションを実行できます。myPresto において分子動力学計算を行うプログラムはcosgeneです。cosgeneでの計算を行うためには、tplgeneX、tplgeneL、setwater、add_ion 等、複数のプログラムを用いて準備を行う必要があります。myPresto ver. 5.0 では、分子動力学計算を行うために必要なプログラムを集めたパッケージをcosgene_packYYMMDD.tar.gzとして提供しています。(YYMMDDは年月日を示す数字です。)

myPresto では、拡張アンサンブル法的一种であるマルチカノニカル法を使った MD 計算 (McMD)を行うための支援ツールプログラムも提供されています。cosgene の実行サンプルファイルを含むセットは、cosgene_sample_packYYMMDD.tar.gzとして配布しています。

複数のコアを同時に利用して計算を実行する cosgene_MPI (MPI を利用) も用意されています。cosgene_MPI は、cosgene_pack に含まれる cosgene のソースファイルから作成可能ですが、MPI が利用な環境を事前に準備されていることが必要です。Makefile を編集して、次のコマンドを実行することにより cosgene_MPI の実行ファイルが作成されます。編集する内容は、環境によって異なります。MPI 用のコンパイラ、ライブラリ等を適切に設定する必要があります。

```
% make cosgene_MPI
```

(Makefile を編集する際には、現在の Makefile をコピーしてから行ってください。)

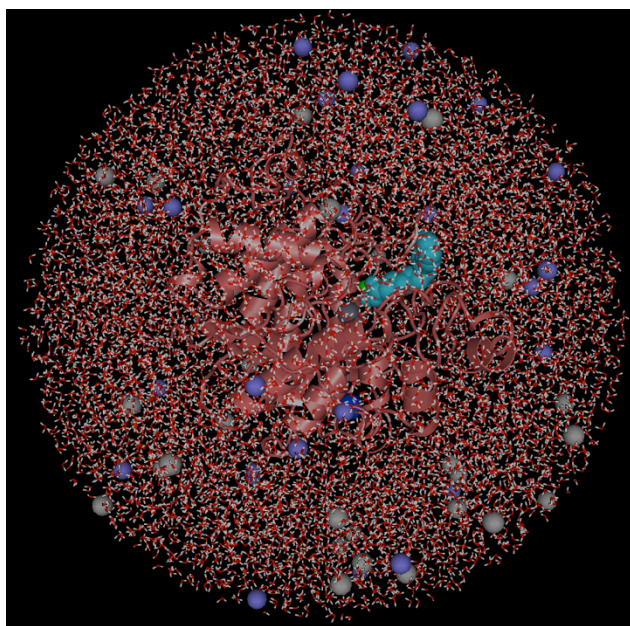


図 1. **cosgene** で計算可能な系の例
(タンパク質・低分子化合物・水・イオンを含む系)

cosgene の主な機能について、表 1 に挙げます。

表 1. **cosgene** の主な機能

機能分類	内容
エネルギー最小化	Steepest decent method, Steepest descent method with SHAKE, Conjugate gradient method
シミュレーション可能なアンサンブルおよび拡張アンサンブル	Micro-Canonical, Canonical, Force-biased Multi-Canonical, Tsallis Dynamics
積分器	Leap-frog (Verlet), Velocity Verlet, RESPA
Thermostat	Hoover-Evans Gaussian constraint, Nose-Hoover
Barostat	Andersen, Parrinello-Rahman
長距離相互作用	Direct summation, Direct summation & Cutoff, Ewald, Particle Mesh Ewald, Fast Multipole Method
拘束法	SHAKE, RATTLE, Rigid-body, Position restraint, Distance restraint
境界条件	球・楕円体、周期境界条件

4. 2. 高速分子動力学計算 (psygene/psygene-G)

myPresto では、空間分割のアルゴリズムとオリジナルな非エバルト法 (Zero-Multipole summation method: ZMM) のアルゴリズムを用いて高速に計算を行う分子動力学シミュレーションプログラム psygene (Protein dYnamics SimulatinG ENginE) が別途開発し、利用することができます。psygene-G では、MPI によって複数の GPU による並列化計算が容易に実行できます。 (Mashimo, T., Fukunishi, Y., Kamiya, N., Takano, Y., Fukuda, I., Nakamura, H. (2013) J. Chem. Theory Comput. 9, 5599-5609)

psygene-G の主な機能について、表 1 に挙げます。

表 1. Psygene-G の主な機能

機能分類	内容
エネルギー最小化	Steepest descent, Conjugate gradient
MD 計算 (アンサンブル)	Micro-Canonical, Canonical, NPT
積分器	Leap-frog (Verlet), Velocity Verlet
Thermostat	Berendsen, Hoover-Evans Gaussian constraint, Nose-Hoover
Barostat	Berendsen
長距離相互作用	Direct summation, Direct summation & Cutoff, Zero Dipole method
拘束法	SHAKE, Position restraint, Distance restraint
境界条件	周期境界条件

4.3. ドッキングシミュレーション(sievgene/sievgene_M)

sievgene は、タンパク質と低分子化合物とのドッキング計算を行うプログラムです。sievgene には、ドッキングポーズの精度を重視した sievgene_for_dockingpose と、計算速度を重視した sievgene_for_screening の2つのバージョンがあります。

sievgene でドッキング計算を実行するためには、ドッキングに適した PDB ファイルや、計算に使用するトポロジーファイル、ドッキングサイトを指定するファイルを事前に作成する必要があります。これらの事前準備には、pdbcheck、tplgeneX、make_point 等を使用します。sievgene を使う際に必要となるプログラムは、sievgene_packYYMMDD.tar.gz の中に含まれています。

sievgene_M は、計算効率を改善した sievgene の改良版です。sievgene_M の使い方は、sievgene とほとんど同じで、同じ入力ファイルを使用できます。sievgene_M には、並列処理版と逐次処理版があります。

sievgene_M の並列処理版は、1 つのドッキング計算に対して同時に多数の CPU コアを使用した並列処理で高速な計算が可能です。これは、1 つのドッキング計算に長い時間がかかる計算対象、例えば、多くのコンフォメーションを取りうる比較的大きな化合物の計算に適しています。

sievgene_M の逐次処理版(特に、こちらを sievgene_M_serial と呼ぶ)は、1 つのドッキング計算に1つの CPU コアのみを使用します。こちらは、特に、多数の低分子化合物(100 万件以上)を対象とした薬剤候補化合物スクリーニング計算に適しています。

例えば、20 コアのシステムを使用する場合には、並列処理版では1つのジョブで20 コアを同時に使って1組のタンパク質と化合物とのドッキング計算を行うことができます。逐次処理版では、1つのジョブが1コアずつを使用して、同時に20個のジョブで20組のタンパク質と化合物のドッキング計算を効率良く行うことができます。

sievgene_M は、50 個以上のコアが1つの CPU に搭載されているメニーコア CPU (インテル社の Xeon Phi) でも動作し、その性能を十分に引き出すことができます。例えば、1 つの CPU の中に60個のコアを搭載したモデルでは、60 並列の処理を同時に実行して効率良く計算を行うことが可能です。sievgene_M は、Intel Xeon Phi 専用のプログラムという訳ではなく、多くの CPU で問題なく動作します。現在は、sievgene_M のコンパイルには、Intel 社のコンパイラが必要です。

sievgene の配布ファイル名: sievgene_packYYMMDD.tar.gz (YYMMDD は年月日を示す数字)

sievgene_M の配布ファイル名: sievgene_packYYMMDD.tar.gz (YYMMDD は年月日を示す数字)

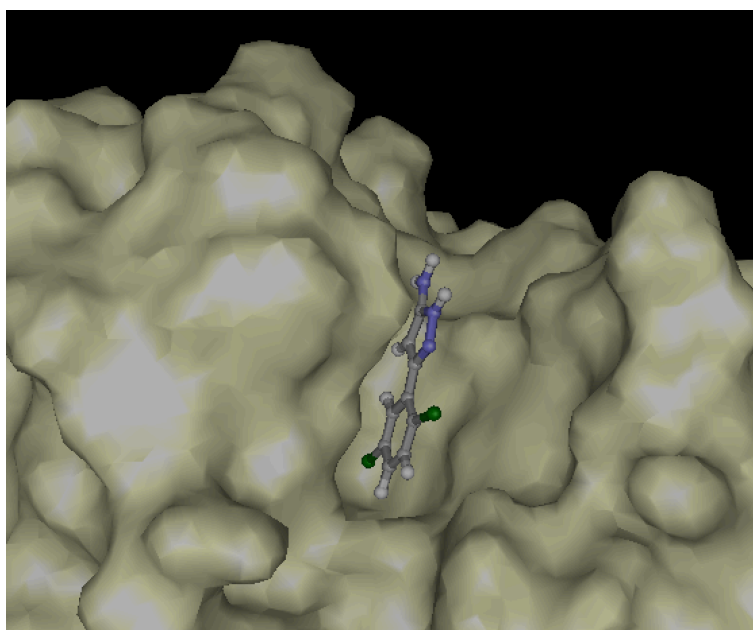


図 2. sievgene を使って低分子化合物をタンパク質にドッキングさせた例

4.4. インシリコ・スクリーニング

myPresto で実行できるインシリコ・スクリーニングには、multiple target screening(MTS)法、docking score index(DSI)法、機械学習 MTS(ML-MTS)法等、複数の異なる方法があります。ターゲットタンパク質の立体構造が利用可能であるか、既知活性化化合物が存在しているか等によって、適切な方法を選択します。

myPresto では、200 万化合物を対象としたインシリコ・スクリーニング計算を容易に実行できるように、以下のプログラム及びデータを提供しています。

- (1) スクリーニング計算用スクリプトプログラム群
 - (2) スクリーニング計算用化合物ライブラリ LigandBOX(200 万化合物)
 - (3) 上記(2)に対応した相互作用行列

上記(1)のスクリプトプログラムは、200 万化合物の計算を効率的に実行するためのプログラム群です。多くの CPU コアを利用して分散した計算ができる仕組みになっています。例えば、200 個のコアが利用可能な場合は、200 個のコアを使用して、各コアで 1 万化合物ずつ分散して同時に計算することが可能です。スクリプトの一部では、ジョブスケジューリングプログラムを使用しています。これは、環境に依存します。現在の設定では、LSF を使用しています。ジョブスケジューリングプログラムを利用するプログラムの一部は、お使いの環境に応じて、変更する必要があります。具体的には、bsub、qsub 等、ジョブ投入のコマンドを使用して、利用可能なキューの設定をします。これは、ソースコードを書き換えて行います。

上記(2)の LigandBOX は、市販されている低分子化合物の 3 次元分子構造と分子電荷や物理化学的パラメータを含むオープンなデータベースです。また KEGG DRUG に集められた医薬品についても同様の情報を含んでいます。3D 化・水素原子付加・部分電荷付加等の処理が施されていて、ドッキングにすぐに利用可能となっています。スクリーニング用のセット用に、200 万個ちょうどで mol2 ファイル形式で提供されている化合物ライブラリを package 化したものも用意しています。

上記(3)の相互作用行列は、MTS 法、ML-MTS 法、DSI 法で使用するデータです。リファレンスタンパク質 181 個と、(2)の 200 万化合物とのドッキングスコア(181x200 万)を含んでいます。相互作用データがなければ、MTS 法、ML-MTS 法、DSI 法を用いた計算は実行できません。

MTS 法を用いたスクリーニング方法

MTS 法は、SBDS に分類される計算手法で、ターゲットタンパク質に対して、化合物ライブラリに含まれる化合物とのドッキング計算を行います。その後に、リファレンスタンパク質と化合物とのドッキング結果を参考にして、化合物のランキングリストを作成します。MTS 法では、ターゲットタンパク質だけでなく多数のリファレンスタンパク質とのドッキングスコア (sievgene スコア) も参考にして、ランキングを決定します。

この計算を実施するためには、まず、ターゲットタンパク質の準備をします。これは、`sievgene_packYYMMDD.tar.gz` に含まれるプログラムを使用します。`pdbcheck`、`tplgeneX`、`make_point` 等を用います。ドッキング用に処理した PDB ファイルを `Pro_md.pdb`、トポロジーファイルを `Pro.tpl`、ドッキングサイトを指定するプローブ点情報ファイル `point.pdb` を作成します。ターゲットタンパク質の目印となる 4 文字程度の ID をディレクトリ名に与えて、その下に、`Pro_md.pdb`、`Pro.tpl`、`point.pdb` を置きます。スクリーニング計算用スク립トの中で、これらのファイル名を使っていますので、`Pro_md.pdb`、`Pro.tpl`、`point.pdb` は、大文字・小文字を含めて同じ名前を使用してください。

これらのファイル(ディレクトリ)を用意した後は、`screening_packYYMMDD.tar.gz` に含まれるプログラムを使用して、多数の化合物に対するドッキング計算を実行します。ドッキング計算後には、データ集計を行い、その後に MTS 法によるランキングを決定します。最終的に、上位化合物のリストを出力します。ランキングを決定するまでの計算では、ドッキングポーズを記録していませんので、上位の化合物に対してドッキングポーズを取得する場合には、インプットファイルの設定を変更して、上位化合物に対して再計算を実行します。スクリーニング用 `LigandBOX`(200 万化合物)を使う場合には、`screenig_dataYYMMDD.tar.gz` を入手して使用します。その中に、化合物の `mol2` ファイル、リファレンスタンパク質と化合物との相互作用行列が含まれています。`screening_dataYYMMDD.tar.gz` は、適当な時期に更新される場合がありますので、スクリーニング計算を実行する前に、新しいバージョンのものがリリースされていないかを確認するとよいでしょう。

機械学習 MTS 法を用いたスクリーニング方法

既知活性化合物についての情報がある場合には、それらの情報を用いてヒット率の改善を試みることができます。myPresto では、MTS 法を機能拡張して、既知活性化合物の情報に基づいて機械学習を実行して、既知活性化合物がランキングの上位にシフトするように、スコアを変更することによって、未知の活性化合物の順位を上位に来ることを期待することができます。これは、機械学習 MTS 法 (ML-MTS 法) と呼んでいます。ML-MTS 法を実行するためには、MTS 法で計算したターゲットタンパク質とライブラリ化合物とのドッキングに加えて、既知活性化合物とターゲットタンパク質、および、既知活性化合物とリファレンスタンパク質とのドッキングを行います。機械学習 MTS 法で使用するファイルセットは、MTS 法と同じで、`sievgene_packYYMMDD.tar.gz`、`screening_packYYMMDD.tar.gz`、`screening_dataYYMMDD.tar.gz` です。

DSI 法を用いたスクリーニング方法

DSI 法では、ターゲットタンパク質の立体構造情報を使用しません。既知活性化合物とリファレンスタンパク質とのドッキング計算を行い、既知活性化合物とリファレンスタンパク質とのドッキングスコア (どのリファレンスタンパク質に結合しやすいか) のパターンが近い化合物を、スクリーニング用 LigandBOX の中から選出します。DSI 法についても、MTS 法、ML-MTS 法と同じファイルセット (`sievgene_packYYMMDD.tar.gz`, `screening_packYYMMDD.tar.gz`, `screening_dataYYMMDD.tar.gz`) を使用します。

アンサンブル・ドッキング法を用いたスクリーニング方法

myPresto では、分子動力学シミュレーションで発生させたタンパク質の構造を利用したインシリコ・スクリーニング計算を実施できます。分子動力学シミュレーションで得られたトラジェクトリーからタンパク質の立体構造を多数取り出して、その中から、構造を絞り込み、絞り込んだ構造に対して行って得られたドッキング結果を使って化合物のランキングを行うことが比較的簡単にできます。タンパク質の多数の立体構造から、少数を選出する際には、universal active probe(UAP)を使います。UAP は、上市された薬剤、もしくは、臨床試験に使われた活性のある化合物群です。UAP が結合しやすい立体構造を、実際に薬剤になるような性質を持つ化合物が結合しやすいものとして選出して使用します。

アンサンブル・ドッキング法では、まず、MD 計算を実行します。アンサンブル・ドッキング用のプログラムセットは、UAPyymdd. tar. gz として配布しています。

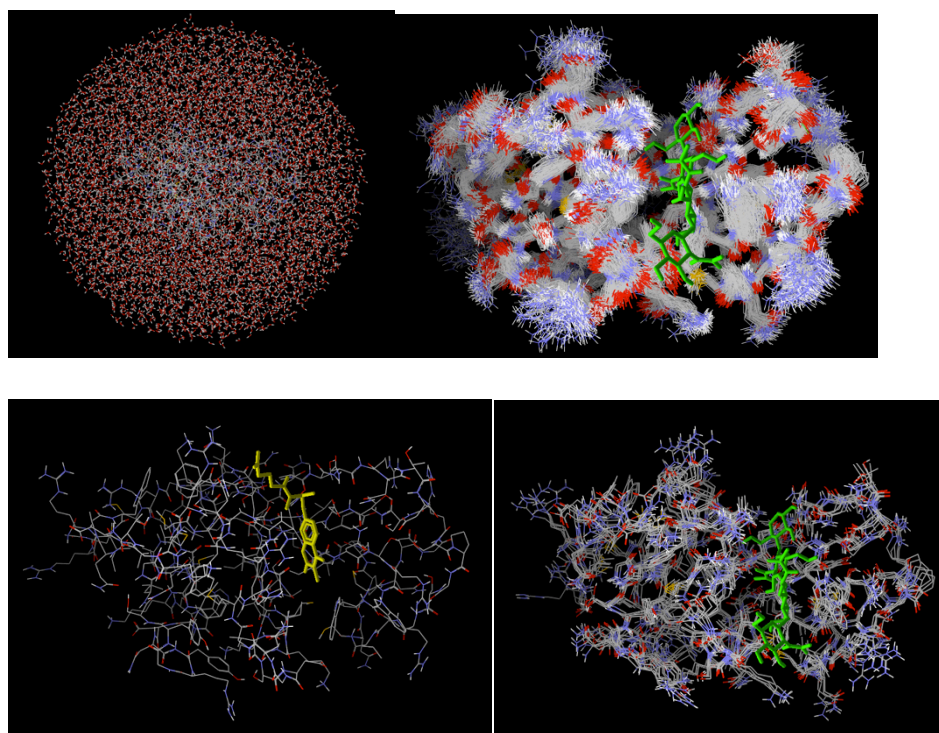


図 3. アンサンブル・ドッキングを用いたスクリーニングのイメージ図

まず、タンパク質を水中に置いた系(左上)で MD 計算を実行して得られたトラジェクトリーから多数のタンパク質構造を 100 個程度取得します(右上)。UAP を用いて評価方法によって、立体構造を 5 個程度に絞り込み(右下)、選出した構造をそれぞれ用いて、ドッキング計算を行って、得られたスコアを使用して最終的に化合物を選出します(左下)。

4.5. 類似化合物探索(substructure_search, TGS)

myPresto に含まれている類似化合物探索法には、いくつか異なるものがあります。

- (1) substructure_search: 部分構造探索

配布ファイル: substructure_searchYYMDD.tar.gz

- (2) TGS: Topological Graph Search(TGS)法を用いた類似化合物探索

配布ファイル: TGSyymmdd.tar.gz

4.6. ドッキング構造の精密計算(sievgene_NMR, sievgene_MV0)

myPresto では、sievgene よりもドッキング精度を高めたプログラムも用意されています。

- (1) sievgene_NMR: NMR のシグナルデータを利用してドッキング精度を高めた計算を行うプログラム。sievgene_NMRyymmdd.tar.gz で配布。

- (2) sievgene_MV0: 既知の低分子化合物・タンパク質の複合体構造を利用して、既知の複合体構造の低分子化合物に新規低分子化合物をできるだけ重ね合わせる計算を行うプログラム。

4.7. 結合自由エネルギーの計算(sievgene_DIAV)

sievgene_DIAV は、DIA(direct interaction approximation)法を使った結合自由エネルギー推定方法で、タンパク質 (アポ体) の MD 計算を行って得られたトラジェクトリーファイルを利用して、トラジェクトリーファイルに含まれる複数のタンパク質構造に対してドッキングを行って、それらの結果から結合自由エネルギーを推定します。このプログラムの実行では、事前に分子動力学シミュレーションを実行してトラジェクトリーファイルが得られていることを前提としています。

配布ファイル名: sievgene_DIAVyymmdd.tat.gz

4.8. 精密な結合自由エネルギー計算(SPRG)

SRPG(Smooth reaction path generation) 法は、滑らかな解離経路を用意して、それに沿った各点で自由エネルギー計算を行う事で、タンパク質とリガンドの結合自由エネルギー (ΔG) を精密に計算する方法です。解離経路の生成には Filling Potential 法を使用し、経路そのものはリガンドのランダムマーク原子 (例えばリガンドの重心に近い重原子) の平均座標で表現します。その際、ルジャンドル多項式を使って経路を滑らかにしておきます。解離過程の PMF は、経路に沿った各点で最近接の構造を選び、それらに対応する計算系を

用意した後、経路に沿ってアンブレラ・サンプリング法で計算します。各計算系は、選択した構造に水分子とイオンを加える事により生成します。補正項は、初期座標周辺でリガンド配置を微量ずらす事により複数構造を生成し、これらを使って計算した PMF から計算します。最後に経路の PMF と補正項から結合自由エネルギーを求めます。(Fukunishi, Y., Mikami, Y., Nakamura, H. (2003) J. Phys. Chem. B. 107, 13201-13210; Fukunishi, Y., Mashimo, T., Orita, M., Ohno, K., Nakamura, H. (2009) J. Chem. Inform. Model. 49, 925-933; Nguyen, H., Tran, T., Fukunishi, Y., Higo, J., Nakamura, H., Le, L. (2015) J. Chem. Inform. Model. 55, 1936-1943)

SRPG 法による結合自由エネルギー計算のプログラムは、`cosgene` のパッケージ `cosgene_packYYMDD.tar.gz` および `cosgene_sample_packYYMDD.tar.gz` に含まれていますが、利用例については、`cosgene_sample_manual_5.000_ja_YYMDD.pdf` の 18 章 (Sample-16:SRPG-Streptavidin の結合自由エネルギー計算) を参照してください。`cosgene_sample_manual_5.000_ja_YYMDD.pdf` は、`doc_jaYYMDD.tar.gz`、または `cosgene_sample_packYYMDD.tar.gz` に含まれています。サンプルファイルは別途 `SRPG_sampleYYMDD.tar.gz` を使用してください。

4.9. 化合物溶解度/アグリゲーター予測 (LogS_predictor)

分子記述子と溶媒和自由エネルギーに基づいて低分子化合物の溶解度 (LogS) を推定し、アグリゲータも予測します。(T. Mashimo, Y. Fukunishi, M. Orita, N. Katayama, S. Fujita, H. Nakamura. International Journal of High Throughput Screening, 2010:1, 99-107)

化合物溶解度/アグリゲーター予測を行うプログラムセットは、`LogS_predictionYYMDD.tar.gz` に含まれています。

4.10. タンパク質表面のポケット探索 (MolSite)

低分子リガンド・ライブラリーを用いて、それらのリガンドをタンパク質にドッキングさせることによって、リガンド結合ポケットの探索を行います。(Fukunishi, Y., Nakamura, H. (2011) Protein Science 20, 95-106.)

このタンパク質表面のポケット探索を行うプログラムセットは、`MolSiteYYMDD.tar.gz` に含まれています。

4.11. 化合物の 2D → 3D 変換・化合物ライブラリ作成 (3DdataConstruction)

化合物の 2D→3D 変換、および、変換した化合物のライブラリ化をするためのツール群は、3DdataConstructionYYMDD.tar.gz に含まれています。

4.12. 低分子化合物構造の編集(EditMol)

EditMol を使うと、mol2 ファイル形式の低分子化合物の一部の構造を変更することが可能です。EditMol は、EditMol_YYMDD.tar.gz に含まれています。

5. ツールプログラム

myPresto に含まれるツールプログラムには、以下のものがあります。これらは、toolsYYMMDD.tar.gz、もしくは、関連するプログラムセットに含まれています。詳細は、tools_manual_v5.000_ja_YYMMDD.tar.gz を参照してください。

- set_water
- mergetpl
- SHAKEinp
- RIGIDinp
- GBSAinp
- reweightFB
- reweightST
- reweightGST
- selection
- clustering
- Gamess2tplinp
- Gauss2tplinp
- tpl2mol2
- add_ion
- confgene
- confgeneC
- FEP
- Hgene
- MVO
- trans_code
- tpl2capbc

6. 環境変数の設定

myPresto version 5.0 では、ダウンロードした圧縮ファイルを解凍すると作成されるディレクトリの下で容易に作業できるようにディレクトリ構成を設計しています。そのため、多くの場合は、パス等の環境変数を設定しなくても、簡単な相対パスを実行時に指定することによって、プログラムを実行することができます。環境変数 PATH に、myPresto のプログラムへのパスを追加しておく、と、コマンド実行の際に、パスの入力を省くことができます。以下に、パスを設定していない場合と設定している場合のプログラムの実行例を挙げます。

パスを設定していない場合の実行方法の例：

```
% ../bin/sievgene < control
```

ここでは、(../bin/)が sievgene への相対パスになります。相対パスは、カレントディレクトリの場所によって異なります。

パスを設定している場合の実行方法の例：

```
% sievgene < control
```

myPresto version 4.3 までは、特定のディレクトリに実行ファイルをインストールし、そこへのパスを設定することによって、myPresto のプログラムを実行するようにしていました。以下、プログラムへのパス等、システムの環境変数の設定について説明します。

■myPresto で使用することがある環境変数について

環境変数名	説明
-------	----

PATH	： プログラムへのパスの追加
------	----------------

TPL_INPUT_PATH	： tplgeneX 入力ファイル用ディレクトリ（パスを含めて記述する）
----------------	--------------------------------------

TPL_OUTPUT_PATH	： tplgeneX 出力ファイル用ディレクトリ（同上）
-----------------	------------------------------

TPL_DB_PATH	： tplgeneX 力場 DB 用ディレクトリ（同上）
-------------	------------------------------

TPLL_INPUT_PATH	： tplgeneL 入力ファイル用ディレクトリ（パスを含めて記述する）
-----------------	--------------------------------------

TPLL_OUTPUT_PATH	： tplgeneL 出力ファイル用ディレクトリ（同上）
------------------	------------------------------

TPLL_DB_PATH	： tplgeneL 力場パラメータ DB 用ディレクトリ（同上）
--------------	-----------------------------------

特に、TPL_DB_PATH と TPLL_DB_PATH の設定は、tplgeneX と tplgeneL を使う場合には重要です。

(設定例)

tplgeneL カ場パラメータ DB 用ディレクトリを"/home/user01/myPresto/tplgeneL/DB"とする場合。

環境変数の設定は、使用しているシェルの設定方法に従います。

(下線部が入力部分です)

(1) bash の場合

```
% export TPLL_DB_PATH=/home/user01/myPresto/tplgeneL/DB
```

(2) csh の場合

```
% setenv TPLL_DB_PATH /home/user01/myPresto/tplgeneL/DB
```

※環境変数で設定するディレクトリへのパスが固定される場合は、ログインスクリプト(.bashrc, .cshrc) や専用スクリプトに書き込んでおく と 便利 です。

なお、現在使用しているシェルは以下のコマンドで確認できます。

```
% ps
```

tplgeneX と tplgeneL については、cosgene_pack, sievgene_pack において、bin/exec_tplgeneX.sh, bin/exec_tplgeneL.sh を用意しています。これらは、そのプログラムの中で、TPLL_DB_PATH、TPLL_DB_PATH をそれぞれ設定していますので、環境変数を自分で設定することなく tplgeneX もしくは tplgeneL を実行することができます。

7. 配布ファイルリスト

myPresto version 5.0における配布プログラムファイルは以下の通りです。

```
tplgeneXYMMDD.tar.gz
tplgeneLYMMDD.tar.gz
pdbcheckXYMMDD.tar.gz
HgeneYMMDD.tar.gz

sievgene_packYMMDD.tar.gz
MolSiteYMMDD.tar.gz
screening_packYMMDD.tar.gz
UAPYMMDD.tar.gz
screening_dataYMMDD.tar.gz（次世代天然物化学技術研究組合から配布）

cosgene_packYMMDD.tar.gz
cosgene_sample_packYMMDD.tar.gz
SRPG_sampleYMMDD.tar.gz

psygene-GYMMDD.tar.gz

3DdataConstructionYMMDD.tar.gz
EditMol_YMMDD.tar.gz
LogS_predictorYMMDD.tar.gz
TGSYMMDD.tar.gz
substructure_searchYMMDD.tar.gz
SyntheticAccessibilityYMMDD.tar.gz

sievgene_DIAVYMMDD.tar.gz
sievgene_MVOYMMDD.tar.gz
sievgene_NMRYMMDD.tar.gz
MD-MVOYMMDD.tar.gz

toolsYMMDD.tar.gz
LibMyPrestoYMMDD.tar.gz
doc_jaYMMDD.tar.gz
```

8. ドキュメントリスト

myPresto version 5.0におけるドキュメントファイルは以下の通りです。

doc_jaYYMMDD.tar.gz、または、関連するプログラムセットのdoc/に含まれています。

user_manual_v5.000_ja_YYMMDD.pdf (YYMMDD には数字が入ります。)

tplgeneX_manual_v5.000_ja_YYMMDD.pdf

tplgeneL_manual_v5.000_ja_YYMMDD.pdf

pdbcheck_manual_v5.000_ja_YYMMDD.pdf

Hgene_manual_v5.000_ja_YYMMDD.pdf

sievgene_manual_v5.000_ja_YYMMDD.pdf

sievgene_tutorial_v5.000_ja_YYMMDD.pdf

MolSite_manual_v5.000_ja_YYMMDD.pdf

screening_tutorial_v5.000_ja_YYMMDD..pdf

UAP_manual_v5.000_ja_YYMMDD.pdf

cosgene_manual_v5.000_ja_YYMMDD.pdf

cosgene_tutorial_v5.000_ja_YYMMDD.pdf

cosgene_sample_manual_v5.000_ja_YYMMDD.pdf

psygene-G_manual_v5.000_ja_YYMMDD.pdf

3DdataConstruction_manual_v5.000_ja_YYMMDD.pdf

EditMol_manual_v5.000_ja_YYMMDD.pdf

LogS_predictor_manual_v5.000_ja_YYMMDD.pdf

TGS_manual_v5.000_ja_YYMMDD.pdf

substructure_search_manual_v5.000_ja_YYMMDD.pdf

SyntheticAccessibility_manual_v5.000_ja_YYMMDD.pdf

sievgene_DIAV_manual_v5.000_ja_YYMMDD.pdf

sievgene_MV0_manual_v5.000_ja_YYMMDD.pdf

sievgene_NMR_manual_v5.000_ja_YYMMDD.pdf

MD-MV0_manual_v5.000_ja_YYMMDD.pdf

tools_manual_v5.000_ja_YYMMDD.pdf

VCOL_manual_v5.000_ja_YYMMDD.pdf

9. 参考文献

引用文献について

本ソフトウェアの使用時には、以下の文献を引用してください。

myPresto 全体及び Filling potential/SRPG 法について

- 1) "The filling potential method: A method for estimating the free energy surface for protein-ligand docking", Yoshifumi Fukunishi, Yoshiaki Mikami, and Haruki Nakamura, J. Phys. Chem. B. (2003) 107, 13201-13210.
- 2) "Protein-ligand binding free energy calculation by the smooth reaction path generation (SRPG) method", Yoshifumi Fukunishi, Daisuke Mitomo, and Haruki Nakamura, J. Chem. Inf. Model., 2009, 49, 1944-1951.
- 3) "Computational Study of Drug Binding Affinity to Influenza A Neuraminidase Using Smooth Reaction Path Generation (SRPG) Method", Nguyen, H., Tran, T., Fukunishi, Y., Higo, J., Nakamura, H., Le, L. Journal of Chemical Information and Modeling, (2015) 55, 1936-1943.

cosgene :Multicanonical MD について

- 4) "Determination of multicanonical weight based on a stochastic model of sampling dynamics", Jae Gill Kim, Yoshifumi Fukunishi, Akinori Kidera and Haruki Nakamura, Physical Review E (2003) 68, 021110.
- 5) "Multicanonical molecular dynamics algorithm employing adaptive force-biased iteration scheme", Jae Gil Kim, Yoshifumi Fukunishi, Haruki Nakamura, Phys. Rev. E 70, 057103 (2004).

Particle Mesh Ewald(PME)について

- 6) "A smooth particle meth Ewald method" U.Essmann, L.Perera, M.L.Berkowitz, T.Darden, H.Lee and L.G.Pedersen. (1995) *J. Chem. Phys.* 103, 8577-8593.

Accessible surface area (ASA)について

- 7) "Physicochemical evaluation of protein folds predicted by threading" Kinjo, A. R., Kidera, A., Nakamura, H., Nishikawa, K. (2001) *Eur Biophys J* 30, 1-10.

Fast Multipole Method (FMM)について

- 8) "Scalable variants of multipole-accelerated algorithms for molecular dynamics applications" In *Proceedings of the Seventh SIAM Conference on Parallel Processing for Scientific Computing*, February 1995, J. A. Board, Z. S. Hakura, W. D. Elliott, W. T. Rankin.

-
- 9) W. T. Rankin, “*Efficient Parallel Implementations of Multipole Based N-Body Algorithms*.” PhD thesis, Duke University, Department of Electrical and Computer Engineering, P.O.Box 90291, Durham, NC 27708-0291, April 1999.
 - 10) W. T. Rankin, DPMTA -Distributed Parallel Multipole Tree Algorithm, Duke University, Durham, NC (2002).

psygene/psygene-G について

- 11) “Molecular Dynamics Simulations Accelerated by GPU for Biological Macromolecules with a Non-Ewald Scheme for Electrostatic Interactions”, Tadaaki Mashimo, Yoshifumi Fukunishi, Narutoshi Kamiya, Yu Takano, Ikuo Fukuda, Haruki Nakamura, J. Chem. Theory Comput. (2013) 9, 5599-5609.

sievgene について

- 12) “Similarity among receptor pockets and among compounds: Analysis and application to in silico ligand screening”, Y. Fukunishi, Y. Mikami, and H. Nakamura, The Journal of Molecular Graphics and Modelling 24 (2005) 34-45.

Sievgene_NMR について

- 13) Yoshifumi Fukunishi, Yumiko Mizukoshi, Koh Takeuchi, Ichio Shimada, Hideo Takahashi, Haruki Nakamura. Protein-ligand docking guided by ligand pharmacophore-mapping experiment by NMR. Journal of Molecular Graphics and Modelling, 2011, 31, 20-27.

Multiple Target Screening (MTS)法について

- 14) “Multiple target screening method for robust and accurate in silico screening”, Y. Fukunishi, Y. Mikami, S. Kubota, H. Nakamura, Journal of Molecular Graphics and Modelling, 25, 61-70 (2005).
- 15) “Noise reduction method for molecular interaction energy: application to in silico drug screening and in silico target protein screening”, Y. Fukunishi, S. Kubota, H. Nakamura, Journal of Chemical Information and Modeling, 46, 2071-2084 (2006).
- 16) “Improvement of protein-compound docking scores by using amino-acid sequence similarities of proteins”, Y. Fukunishi, H. Nakamura, Journal of chemical information and modeling, 48, 148-156 (2008)

Docking Score Index (DSI)法について

- 17) “Classification of chemical compounds by protein-compound docking for use in

designing a focused library”, Y. Fukunishi, Y. Mikami, K. Takedomi, M. Yamanouchi, H. Shima, H. Nakamura, Journal of Medicinal Chemistry, 49, 523-533 (2006).

- 18) “An efficient in silico screening method based on the protein-compound affinity matrix and its application to the design of a focused library for cytochrome P450 (CYP) ligands”, Y. Fukunishi, S. Hojo, H. Nakamura, Journal of chemical information and modeling, 46, 2610-22 (2006).

Maximum Volume Overlap (MVO)法について

- 19) “Prediction of protein-ligand complex by docking software guided by other complex structures”, Y. Fukunishi, H. Nakamura, Journal of Molecular Graphics and Modelling, 26 (2008) 1030-1033.

Molecular-dynamics Maximum Volume Overlap (MD-MVO)法について

- 20) “A new method for in-silico drug screening and similarity search using molecular-dynamics maximum-volume overlap (MD-MVO) method”, Y. Fukunishi, H. Nakamura, Journal of Molecular Graphics and Modelling, 2009, 27, 628-636.

TGS 法について

- 21) “A similarity search using molecular topological graphs”, Y. Fukunishi, H. Nakamura, Journal of Biomedicine and Biotechnology, Volume 2009 (2009), Article ID 231780

LigandBox について

- 22) “Advanced in-silico drug screening to achieve high hit ratio-development of 3D-compound database”, Yoshifumi Fukunishi, Yuusuke Sugihara, Yoshiaki Mikami, Kohta Sakai, Hiroshi Kusudo, Haruki Nakamura, Synthesiology, 2009, 2, 60-68.
- 23) “LigandBox: A database for 3D structures of chemical Compounds”, Takeshi Kawabata, Yuusuke Sugihara, Yoshifumi Fukunishi, Haruki Nakamura, Biophysics (2013) 9, 113-121.

Sptool 溶解度(LogS)及び aggregator (frequent hitter)の予測 について

- 24) Tadaaki Mashimo, Yoshifumi Fukunishi, Masaya Orita, Naoko Katayama, Shigeo Fujita, Haruki Nakamura, Quantitative analysis of aggregation-solubility relationship by in-silico solubility prediction, International Journal of High Throughput Screening, 2010:1, 99-107.

MolSite (タンパク質リガンドポケット予測) について

25) Yoshifumi Fukunishi, Haruki Nakamura. “Prediction of ligand-binding sites of proteins by molecular docking calculation for a random ligand library” , Protein Science. 2011: 20, 95-106.

Universal active probe (UAP) について

26) Fukunishi, Yoshifumi; Ono, Kazuki; Orita, Masaya; Nakamura, Haruki, Selection of in-silico drug screening result by using universal active probes (UAPs), Journal of Chemical Information and Modeling, 2010, 50, 1233-1240.

27) Mitsuhiro Wada, Eiji Kanamori, Haruki Nakamura, Yoshifumi Fukunishi, Selection of in-silico drug screening results for G-protein-coupled receptors by using universal active probe. Journal of Chemical Information and Modeling. 2011, 51 (9) 2398-2407.

BindMol について

28) Yoshifumi Fukunishi. Prediction of positions of active compounds makes it possible to increase activity in fragment-based drug development. Pharmaceuticals. 2011, 4, 758-769.

Sievgene-MVO について

29) Yoshifumi Fukunishi, Haruki Nakamura. “Integration of Ligand-Based Drug Screening with Structure-Based Drug Screening by Combining Maximum Volume Overlapping Score with Ligand Docking” , Pharmaceuticals. 2012: 5, 1332-1345.

myPresto 5.000