

myPresto 5.0

USER'S MANUAL

26th Sep 2018

Program license agreement condition

See the attachment.

About copyright

The copyright notice is as follows, unless otherwise noted.

Copyright indication as follows.

myPresto version 5.0 :

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

ACKNOWLEDGEMENTS:

This work was supported by grants from New Energy and Industrial Technology Development Organization (NEDO) and Ministry of Economy, Trade and Industry (METI), JAPAN. Part of the Version 5.0 has also been supported by Japan Agency for Medical Research and Development (AMED). In addition, the program development was partly supported by Fujifilm Holding Corporation, Japan Tobacco Inc., and Takeda Science Foundation.

This software was originally developed as part of a research project advanced by late Dr. Yoshimasa Kyogoku.

The Particle Mesh Ewald (PME) routines were originally developed by Dr. Tom Darden [28]. National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709 US.

The Accessible Surface Area(ASA) routines were originally developed Dr. Akira R. Kinjo[40]. Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, Mishima, Shizuoka, 411-8540, Japan. He has been working at Institute for Protein Research, Osaka University, 3-2 Yamadaoka, Suita, Osaka 565-0871, Japan

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

Software authors

- **myPresto version 5**

Nakamura, Haruki
Fukunishi, Yoshifumi
Mashimo, Tadaaki
Kurosawa, Takashi
Sugihara, Takanori
Wada, Mitsuhiro
Higo, Junichi
Fukuda, Ikuo
Kawabata, Takeshi
Kamiya, Narutoshi
Kasahara, Kota
Bhaskar Dasgupta
Kanai, Chisato
Yoshikawa, Tatsuya
Sugihara, Yusuke
Nakamura, Hironori

- **myPresto version 4**

Nakamura, Haruki
Fukunishi, Yoshifumi
Eiji, Kanamori
Kubota, Satoru
Omagari, Katsumi
Fukuda, Ikuo
Wada, Mitsuhiro
Mashimo, Tadaaki
Mitomo, Daisuke
Sakamoto, Kuniko
Kurosawa, Takashi
Hashi, Yuichi
Mikami, Yoshiaki
Ueda, Hiroshi

- **myPresto version 3**

Nakamura, Haruki
Fukunishi, Yoshifumi
Jae Gil Kim
Watanabe, Yukihisa S.
Mikami, Yoshiaki
Kubota, Satoru
Omagari, Katsumi
Tatsumi, Rie
Horie, Masaru
Fukuda, Ikuo

- **myPresto version 2**

Nakamura, Haruki
Fukunishi, Yoshifumi
Jae Gil Kim
Mikami, Yoshiaki
Watanabe, Yukihisa S.
Ina, Yasuo
Horie, Masaru
Takahashi, Makoto

Fukuda, Ikuo

- **myPresto version 1**

cosgene version 1 :

Nakamura, Haruki
Fukunishi, Yoshifumi
Hashi, Yuichi
Mikami, Yoshiaki
Kidera, Akinori
Terada, Toru

tplgene version 1 :

Nakamura, Haruki
Fukunishi, Yoshifumi
Kuroda, Masataka
Fukuda, Ikuo

Table of Contents

| | |
|---|-----------|
| 1. Overview and Characteristics of myPresto program package | 1 |
| 1.1. Overview | |
| 1.2. Characteristics | |
| 2. Programs and Data Libraries for myPresto | 3 |
| 3. Access to obtain myPresto program package | 5 |
| 3.1. Download of myPresto program package | 5 |
| 3.2. Deployment of myPresto programs and data | 5 |
| 4. Computations available by myPresto program package | 6 |
| 4.1. Molecular Dynamics Simulation: cosgene | 6 |
| 4.2. Rapid Molecular Dynamics Simulation: psygene/psygene-G | 8 |
| 4.3. Docking Simulation: sievgene/sievgene_M | 9 |
| 4.4. in-silico Screening | 11 |
| 4.5. Search Similar Chemical Compounds : TGS | 13 |
| 4.6. Precise Docking: sievgene_NMR and MVO | 13 |
| 4.7. Estimate of Binding Free Energy: sievgene_DIAV | 13 |
| 4.8. Estimate of Precise Binding Free Energy: SRPG | 14 |
| 4.9. Prediction of Solubility and Aggregator: LogS_predictor | 14 |
| 4.10. Search pockets from a protein surface: MolSite | 14 |
| 4.11. 3D structure construction from 2D information of Chemical Compounds: | |
| 3DdataConstruction..... | 14 |
| 4.12. Editing Tool of Chemical Compounds: EditMol | 15 |
| 5. Tool Programs | 16 |
| 6. Configuration of Environment Variables | 17 |
| 7. List of Files to be Disseminated | 19 |
| 8. List of Documents | 20 |
| 9. References | 21 |

1. Overview and Characteristics of myPresto program package

1.1. Overview

myPresto is a molecular simulation software system for drug development, supported by Ministry of Economy, Trade and Industry, Japan (METI), New Energy and Industrial Technology Development Organization (NEDO), and Japan Agency for Medical Research and Development (AMED). By using this software, modeling of proteins with small chemical compounds, molecular dynamics simulations for proteins with ligands in aqueous solutions or in membranes, and drug screening with docking simulations are available much more effectively and quickly than before. The name of myPresto comes from Medicinally Yielding Protein Engineering SimulaTor.

myPresto is a program suite, composed of many programs, libraries, and databases. The major programs and database included in myPresto version 5.0 are as follows.

1) Creation of detailed information on molecules necessary for molecular simulation:

“tplgeneX” for peptide chains and nucleic acids, and “tplgeneL” for small chemical compounds.

2) Molecular dynamics simulation programs:

“cosgene” and “psygene”

3) Docking simulation:

“sievgene”

4) Several tools:

“Molsite” for pocket search, EditMol for molecular design, TGS for similar compounds search, LogS_predictor for solubility prediction, SyntheticAccessibility for prediction of synthetic accessibility.

5) Ligand database:

“LigandBOX”

1.2. Characteristics

The characteristic features of myPresto program suite are as follows.

- Coordinated usage of molecular dynamics and docking simulations.

Both molecular dynamics (MD) and docking simulation programs have been developed simultaneously, coordinated usages are designed. For example, results of an MD simulation (snapshot protein structures taken from trajectories of MD) can be easily used for drug docking. As the inversed way, the complexed protein structure with the docked ligand molecule can be used as the initial conformation for MD simulation. Many effective applications are available in myPresto, using the coordinated usages of MD and docking simulations.

- Large-scale in-silico ligand screening

When the ligand database, “LigandBOX”, is used, large-scale in-silico ligand screening is easily performed. Its purpose is to screen only the putative chemical compounds, the activities of which are then examined experimentally, from huge number of the chemical compounds in the library.

There are two ways for in-silico screening. (1) Structure-Based Drug Screening (SBDS), bases on the 3-dimensional (3D) structure of the target protein receptor, and (2) Ligand-Based Drug Screening (LBDS), based on the known ligand structures. myPresto can use both SBDS an LBDS.

- Tools for treating small chemical compounds

In myPresto program suite, many tools are prepared for actual works in drug development, such as “EditMol” for editing compounds, “LogS_predictor” for prediction of solubility of compounds, and “TGS” for similar compounds search.

2. Programs and Data Libraries for myPresto

The following programs are included in myPresto suite:

- myPresto Portal : Graphic user interface program for myPresto5
- tplgeneX : Generation of molecular topology of a peptide and nucleic acids chains, and set the force-field parameters of AMBER and CHARMM
- tplgeneL : Generation of molecular topology of a chemical compound, and set force-field parameters of general AMBER force field (GAFF)
- cosgene : Energy minimization and molecular dynamics (MD) simulation, with NVE/NVT/NPT ensembles, SHAKE, rigid molecular mode, multicanonical MD (McMD), umbrella sampling, GBSA, SRPG for binding free energy computation
- psygene : Rapid MD simulation with non-Ewald method and space decomposition algorithms. GPU version is also available
- Hgene : Addition of hydrogen atoms and atomic partial charges to chemical compounds
- sievgene : Docking program of chemical compounds to a receptor protein
- sievgene_NMR : Docking program of chemical compounds to a receptor protein, with NMR experimental information
- sievgene-MVO : Docking program of chemical compounds to a receptor protein, so that the compounds overlap to the known ligand in the complexed structure
- sievgene_DIAV : Docking program of chemical compounds to a receptor protein, where binding free energy is estimated by a short MD trajectory of an apo-receptor protein.
- PDBcheck : Check a PDB file, process the N- and C-termini of the peptide chain, and identify the ligands clearly.
- UAP : Selection of drug screening by Universal Active Probe (UAP) method
- VCOL : Combinatorial generation of molecules
- confgene/ confgeneC : A tool to generate molecular configurations
- Matrix : in silico screening : Multiple Target Screening (MTS) method as SBDS, and Docking Score Index (DSI) method as LBDS
- 3DdataConstruction : A tool to generate chemical compounds data for LigandBOX
- MVO : Refinement of the docked structure by Maximum Volume Overlap (MBO) method
- MVO_screening : Similar compounds search by MD-MVO method as LBDS
- substructure-search : Similar compounds search by molecular framework as LBDS
- TGS : Similar compounds search by TGS method as LBDS
- LogS_predictor: Prediction of solubility and aggregator by molecular descriptor and physicochemical properties
- MolSite : Search of ligand binding sites on the receptor protein by docking simulation
- SyntheticAccessibility : Prediction of synthetic accessibility of chemical compounds

- `mdmbgene` : Modiling tool for a system including a transmembrane protein and lipid bilayer

3. Access to obtain myPresto program package

3.1. Download of myPresto program package

myPresto programs and data are available from the following web site.

URL of myPresto download site:

[http:// myPresto5.jp](http://myPresto5.jp)

3.2. Deployment of myPresto programs and data

myPresto programs and data are provided as compressed files, which should be uncompressed by the next command. Here, only LINUX system is assumed. Some programs work on the UNIX system on Windows, using cygwin or MinGW, but they are not guaranteed

```
% tar -xzf (program)YYMMDD.tar.gz      (YYMMDD means Year/Month/Date)
```

“(program)” depends on the files to be downloaded.

Many compressed files of myPresto version 5.0 are prepared to be automatically deployed with the following directory structure, after uncompressing the compressed files.

```
(program)YYMMDD/  
|-doc/  
|  |- (program)_manual_v5.000_ja_YYMMDD.pdf  
|  ...  
|-src/  
|-bin/  
... |- install.sh  
...
```

Manual and/or Tutorial files in PDF are installed in the “doc/”directory. Users can refer those for usage of the programs.

When bin/install.sh is executed, programs can be installed. In this case, compilers for programs should be installed beforehand.

Some packages include scripts “test_*.sh” for test computations for programs.

4. Computations available by myPresto program package

4.1. Molecular Dynamics Simulation: cosgene

A molecular dynamics simulation program for protein, DNA, chemical compound, and protein-ligand complex. with AMBER and CHARMM force fields are available. Cosgene can calculate NVE, NVT and NPT ensemble with various boundary conditions. In addition, cosgene can calculate the generalized ensemble like multicanonical ensemble by our original algorithm: Multicanonical MD (McMD). Cosgene can calculate the protein-compound binding free energy by the filling potential method. The generalized ensemble method is useful to calculate the free energy surface (SPRG).

(Fukunishi, Y., Mikami, Y., Nakamura, H. (2003) J. Phys. Chem. B. 107, 13201-13210)

An Example of the protein system is shown in Fig. 1, which is composed of a protein, a small chemical compounds, water and ions. Main functions of the cosgene program are listed in Table 1.

In myPresto 5.0, a package for MD simulations is prepared as cosgene_packYYMMDD.tar.gz, where “YYMMDD” indicates Year/Month/Day.

Parallel usage is also available by using MPI, as cosgene_MPI. In that case, the following “make” command should be used:

```
% make cosgene_MPI
```

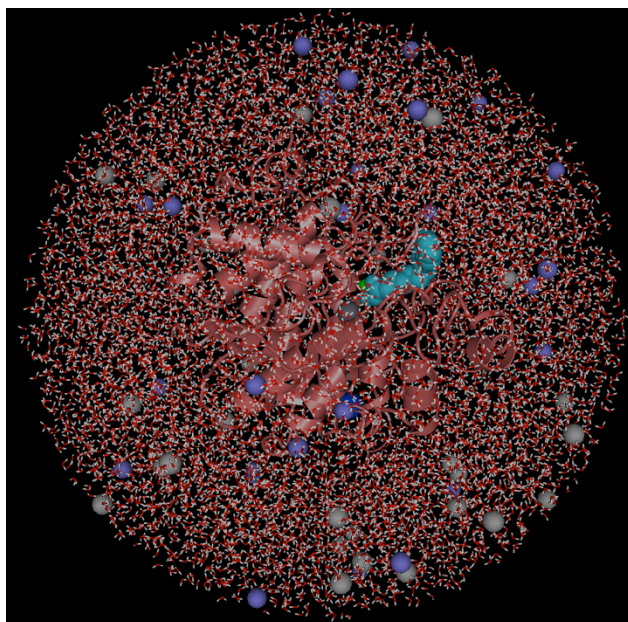


Fig.1. Example of the protein system, a target of cosgene, including a protein, a small chemical compounds, water and ions.

Table 1. Main functions of cosgene

| Function type | Description |
|---------------------------|--|
| Energy minimization | Steepest decent method, Steepest decent method with SHAKE, Conjugate gradient method |
| MD calculation | Micro-Canonical, Canonical, Force-biased Multi-Canonical |
| | Tsallis Dynamics |
| Integrator | Leap-frog (Verlet), Velocity Verlet, RESPA |
| Thermostat | Hoover-Evans Gaussian constraint, Nose-Hoover |
| Barostat | Andersen, Parrinello-Rahman |
| Long distance interaction | Direct summation, Direct summation & Cutoff, Ewald, Particle Mesh Ewald, Fast Multipole Method |
| Restraints methods | SHAKE, RATTLE, Rigid-body, Position restraint, Distance restraint |
| Boundary conditions | Sphere, ellipsoid, periodic boundary conditions |

4.2. Rapid Molecular Dynamics Simulation: psygene/psygene-G

The program psygene (Protein dYnamics SimulatinG ENginE), was developed for rapid MD simulations. psygene uses our original non-Ewald algorithm (Zero-Dipole summation method) for electrostatic interactions with the space decomposition algorithm. The GPU (Graphics Processing Units) version is also available as psygene-G for a single and multiple GPUs.

(Mashimo, T., Fukunishi, Y., Kamiya, N., Takano, Y., Fukuda, I., Nakamura, H. (2013) J. Chem. Theory Comput. 9, 5599-5609)

Table 2. Main functions of psygene/psygene-G

| Function type | Description |
|------------------------------|--|
| Energy minimization | Steepest decent method, Conjugate gradient method |
| MD calculation (ensemble) | Micro-Canonical, Canonical, NPT |
| Integrator | Leap-frog (Verlet), Velocity Verlet |
| Thermostat | Berendsen, Hoover-Evans Gaussian constraint, Nose-Hoover |
| Barostat | Berendsen |
| Long distance interaction | Direct summation, Direct summation & Cutoff, Zero Dipole method |
| Restraints methods | SHAKE, Position restraint, Distance restraint |
| Boundary conditions | periodic boundary conditions |

4.3. Docking Simulation: sievgene/sievgene_M

Pprograms, sievgene (SIEVinG ENginE) and sievgene_M are protein-compound docking programs for in-silico (virtual) drug screening and prediction of protein-ligand complex structure. Sievgene is a fast flexible docking program. On average, it can dock one compound to a target in 1 sec with high accuracy (56% complex structures satisfy $\text{RMSD} < 2\text{\AA}$ in a self-docking test) on usual PC. It is designed for PC grid computing.

(Fukunishi, Y., Mikami, Y., Nakamura, H. (2005) J. Mol. Graph. Model. 24, 34-45)

There are two versions in sievgene: 1) sievgene_for_dockingpose, which can provide accurate docking poses, and 2) sievgene_for_screening, which can make rapid computations.

sievgene_M is an improved version of sievgene_for_screening that improved computational efficiency. The usage of sievgene_M is almost the same as sievgene, you can use the same input file. sievgene_M has two versions: parallel processing version and serial processing version.

The parallel processing version of sievgene_M enables high-speed calculation by parallel processing using many CPU cores simultaneously for one docking calculation. This is suitable for calculation of computation objects that take a long time for one docking calculation, for example, relatively large compounds that can take huge numbers of conformations.

Serial processing version of sievgene_M (in particular, this is called sievgene_M_serial) uses only one CPU core for one docking calculation. This is particularly suitable for drug candidate compound screening calculations targeting a large number of low molecular compounds (more than 1 million)

For example, if you are using a 20 core system, you can calculate a docking between a proteins and a compounds using 20 cores simultaneously in one job. In the serial processing version, it is possible to efficiently calculate 20 dockings of 20 pairs of proteins and compounds with 20 jobs using one core for each job.

sievgene_M also works on a system with a many-core CPU (Intel Xeon Phi) where more than 50 cores are mounted on one CPU, and it can fully exploit its performance. For example, in a model with 60 cores in one CPU, it is possible to efficiently calculate one docking using 60 parallel processes at the same time. sievgene_M is not a program dedicated to Intel Xeon Phi, it also works well with many CPUs. Currently, compilation of sievgene_M requires a Intel compiler.

In order to execute a docking calculation, users should prepare the protein atomic coordinates file (PDB file) with the molecular topology file and a file identifying the docking site on the protein surface. For those preparations, several tools, pdbcheck, tplgeneX, and make_point, are available. All those programs are included in a program package: sievgene_packYYMMDD.tar.gz, sievgene_M_packYYMMDD.tar.gz, where YYMMDD indicates Year/Month/Date.

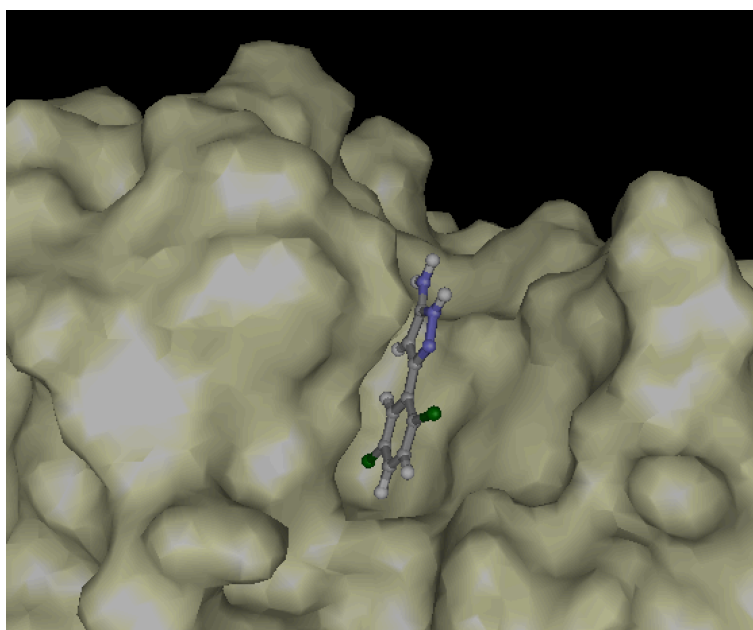


Fig. 2. Example of ligand docking to a protein by sievgene

4.4. in-silico Screening

myPresto performs docking calculations of combinations of proteins and low-molecular compounds, and stores the data. Considering the number of combinations that are possible using 181 types of proteins and 2 million low molecules, docking must be performed more than $181 \times 2 \text{ million} = 360 \text{ million}$ times. For this reason, we improved work efficiency by automating the majority of the data storage tasks. We also automated the screening process to allow users to perform screening using any desired protein and low molecule. Available screening methods are MTS (Multiple Target Screening) and DSI (Docking Score Index).

- MTS: For a target protein, this method examines the order of docking scores of low-molecular compounds and selects compounds that rank high with that target protein.
- DSI: For each low-molecular compound, this method examines the patterns of the protein docking scores and selects low molecules whose patterns resemble the docking score pattern of the specified low-molecular compound.

MTS method with machine-learning, ML-MTS, is also prepared for high accuracy. For these program are used depending whether the 3D structure of the target protein receptor is known or not, and whether the known ligands exist or not.

myPresto system provides the following programs and data for in-silico screening from 2 million small chemical compounds.

- (1) Scripts for screening calculations
- (2) 3D database for 2 million chemical compounds, LigandBOX
- (3) Interaction matrix for the above (2) database

Scripts at the above (1) make it possible to perform effective screening from 2 million chemical compounds, using many possible CPU cores. For example, when 200 CPU cores are available, complete parallel computation using 200 CPU cores can be done, where each core computes screening from 10,000 chemical compounds. Some parts of the scripts use Job scheduling program, which is LSF in the default scripts. Users should change these parts, using bsub or qsub commands.

LigandBOX mentioned at the above (2) is an open database, containing the three-dimensional (3D) molecular structures and atomic charges of chemical compounds, which are available from industries. Some are collected from the KEGG DRUG database. A package including 2 million chemical compounds in mol2 format is also provided.

The interaction matrix at the above (3) is necessary for DSI, MTS and ML-MTS methods. Here, docking scores for 181 referenced protein vs 2 million chemical compounds are included. (181 x 2,000,000).

The packages, `sievgene_packYYMMDD.tar.gz`, `screening_packYYMMDD.tar.gz`, and `screening_dataYYMMDD.tar.gz` are available for in-silico screening.

myPresto also provides tools for ensemble docking, where docking simulations are made for a lot of structures generated by MD trajectories. When a few putative template receptor structures are selected, universal active probes (UAPs) are used. A program package for the ensemble docking is available as `UAPyymmdd.tar.gz`

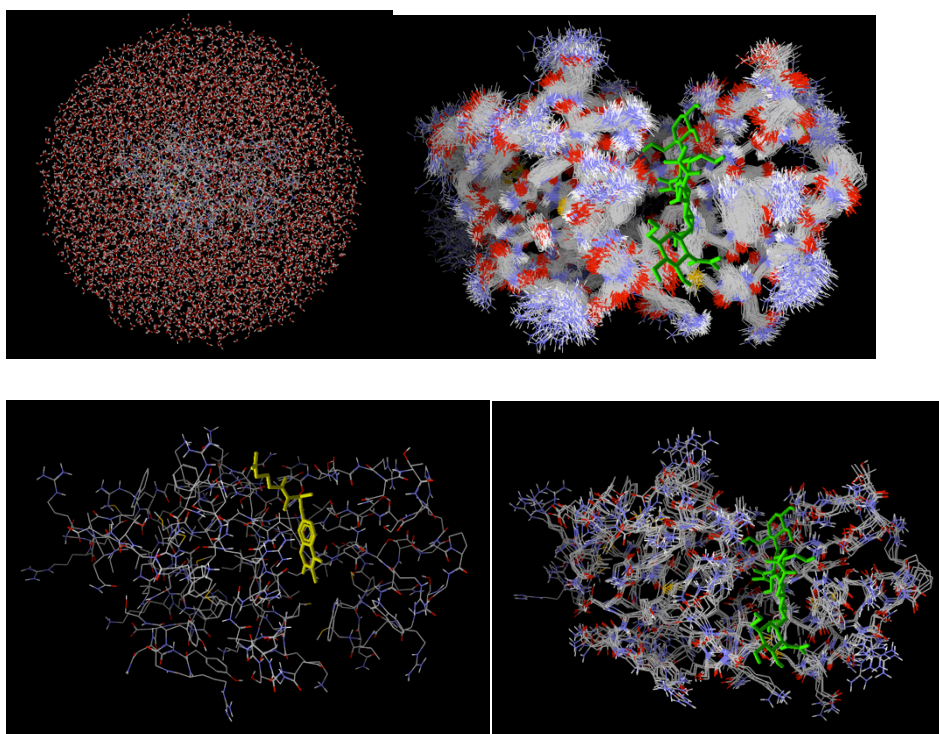


Fig. 3. Example of ensemble docking. First, MD simulation of a receptor protein in water is performed (Upper left). Then, about 100 template receptor structures are randomly selected (Upper right). By using UAPs, about 5 templates are selected (Lower right). Finally, best ligand candidates are given by docking calculations (Lower left).

4.5. Search Similar Chemical Compounds : TGS

(1) Substructure_search

A ligand-based drug screening program based on chemical compound structure comparison. A substructure search method based on the molecular structure by using the Ullman's method.

A package program is available as substructure_searchYYMMDD.tar.gz

(2) TGS: Topological Graph Search (A Similarity Search Using Molecular Topological Graphs)

A similarity search method based on eigen values of molecular edge matrix. (Y. Fukunishi, H. Nakamura. "A similarity search using molecular topological graphs", Journal of Biomedicine and Biotechnology, Volume 2009 (2009), Article ID 231780)

A package program is available as TGSyymmdd.tar.gz

4.6. Precise Docking: sievgene_NMR and sievgene_MVO

(1) sievgene_NMR: SIEVinG ENginE for NMR

A protein-compound docking program for prediction of protein-ligand complex structure with using NMR signal of ligand. The prediction accuracy is improved approximately twice comparing to the original sievgene by using the DIRECTION NMR signal of ligand. (Fukunishi, Y., Mizukoshi, Y., Takeuchi, K., Shimada, I., Takahashi, H., Nakamura, H. (2012) J. Mol. Graph. Model. 31, 20-27)
Last update on May 11,2012

(2) sievgene_MVO: SIEVinG ENginE for Maximum Volume Overlap

A protein-compound docking program for prediction of protein-ligand complex structure / drug screening with using a given known-ligand docking pose. This software performs a ligand docking and overlapping the ligand onto the given known-ligand structure given by the user at the same time. It is a combination of docking and similarity search. (Fukunishi, Y., Nakamura, H. (2012) Pharmaceuticals 5, 1332-1345)

4.7. Estimate of Binding Free Energy: sievgene_DIAV

sievgene_DIAV, SIEVinG ENginE with DIAV method, is a protein-compound docking program, sievgene, with DIAV method to estimate protein-compound binding free energy more correctly than the original sievgene does (less than about 1.3 kcal/mol). (Fukunishi, Y., Nakamura, H. (2013) Pharmaceuticals 6, 604-622). This uses short MD simulation for an apo-receptor protein.

A program package is available as sievgene_DIAVyymmdd.tat.gz

4.8. Estimate of Precise Binding Free Energy: SRPG

SRPG (Smooth Reaction Path Generation) method is available to estimate very precise binding free energy, ΔG , of ligand, by first preparing a smoothed path, on which a ligand leaves the target receptor protein. The path was first computed by a filling potential method (), which was then smoothed by approximated with a Legendre polynomial. The PMF (Potential of Mean Force) at each point on the smoothed path was computed by umbrella sampling method. (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)

This program is included in a cosgene package (cosgene_packYYMMDD.tar.gz) and a cosgene sample package (cosgene_sample_packYYMMDD.tar.gz). Sample files are included in SPRG_sampleYYMMDD.tar.gz.

4.9. Prediction of Solubility and Aggregator: LogS_predictor

A solubility (LogS) predictor based on molecular descriptor and salvation free energy calculation. Aggregators (frequent hitters) can be predicted based on the calculated LogS values (Mashimo, T., Fukunishi, Y., Orita, M., Katayama, N., Fujita, S., Nakamura, H. (2010) International Journal of High Throughput Screening, 1, 99-107)

A program package is available as LogS_predictorYYMMDD.tar.gz

4.10. Search pockets from a protein surface: MolSite

MolSite predicts ligand-binding site of target protein and its affinity by a compound docking simulation (Fukunishi, Y., Nakamura, H. (2011) Protein Science 20, 95-106.)

A program package is available as MolSiteYYMMDD.tar.gz

4.11. 3D structure construction from 2D information of Chemical Compounds:

3DdataConstruction

A tool box for 3D mol2 file generation from 2D mol2 file. It can generate 3D compound database from a 2D chemical compound catalog.

A program package is available as 3DdataConstructionYYMMDD.tar.gz

4.12. Editing Tool of Chemical Compounds: EditMol

A part of a chemical compounds in mol2 format can be edited by a tool, EditMol.

A program package is available as EditMol_YYMMDD.tar.gz.

5. Tool Programs

There are many tool programs included in myPresto as shown below. They are available as `toolsYYMMDD.tar.gz`, or included in the relating program packages.

The details of these tools are described in `tools_manual_v5.000_en_YYMMDD.tar.gz`

- `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. Configuration of Environment Variables

The directories are designed so that users can work under the directories, which are generated by uncompressing the downloaded compressed files of myPresto version 5.0. Therefore, in many cases, the environment variables such as paths are not particularly set, but only some relative paths are assigned at executing the program.

When a path to the myPresto program is included as PATH, users are not required to make any configurations by inputting the path information.

Followings are the example cases for execution of myPresto programs.

In the case, where the path information is not assigned:

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

Here, “../bin/” is the relative path for sievgene program.

In the case, where the path information is assigned:

```
% sievgene < control
```

The followings indicate the environment variables and how to assign the path information.

• Environment Variables used in myPresto

| Environment variable: | meaning of the variable |
|-----------------------|--|
| PATH | : Path to the program |
| TPL_INPUT_PATH | : input file directory for tplgeneX |
| TPL_OUTPUT_PATH | : output file directory for tplgeneX |
| TPL_DB_PATH | : force-field libraries directory for tplgeneX |
| TPLL_INPUT_PATH | : input file directory for tplgeneL |
| TPLL_OUTPUT_PATH | : output file directory for tplgeneL |
| TPLL_DB_PATH | : force-field libraries directory for tplgeneL |

In particular, TPL_DB_PATH and TPLL_DB_PATH are important for the usage of tplgeneX and tplgeneL.

[Example]

In the case when the force-field parameters for `tplgeneL` is set at “`/home/user01/myPresto/tplgeneL/DB`”:

Configuration of the environment variables depends on shell (the underlines are user’s inputs):

(1) In case of bash:

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

(2) In case of csh:

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

When the paths to the directories configured by the environment variables are fixed, it is useful to describe them in the LOGIN scripts like “`.bashrc`” or “`.cshrc`”.

The current shell is confirmed by the command below.

```
% ps
```

For the programs, `tplgeneX` and `tplgeneL`, `bin/exec_tplgeneX.sh` and `bin/exec_tplgeneL.sh` are prepared in `cosgene_pack` and `sievgene_pack`, respectively. These shells configure `TPL_DB_PAH` and `TPLL_DB_PATH`, respectively, and so `tplgeneX` and `tplgeneL` can run without any configurations by users themselves.

7. List of Files to be Disseminated

The myPresto version 5.0 files disseminated are as follows:

tplgeneXYMMDD.tar.gz
tplgeneLYMMDD.tar.gz
pdbcheckXYMMDD.tar.gz
HgeneXYMMDD.tar.gz

sievgene_packXYMMDD.tar.gz
MolSiteXYMMDD.tar.gz
sceening_packXYMMDD.tar.gz
UAPXYMMDD.tar.gz
screening_dataXYMMDD.tar.gz (Only this file is disseminated from Next Generation Natural Product Chemistry (N²PC), not from Web site. Please contact N²PC.)

cosgene_packXYMMDD.tar.gz
cosgene_sample_packXYMMDD.tar.gz
SRPG_sampleXYMMDD.tar.gz

psygene-GXYMMDD.tar.gz

3DdataConstructionXYMMDD.tar.gz
EditMol_XYMMDD.tar.gz
LogS_predictorXYMMDD.tar.gz
TGSXYMMDD.tar.gz
substructure_searchXYMMDD.tar.gz
SyntheticAccessibilityXYMMDD.tar.gz

sievgene_DIAVXYMMDD.tar.gz
sievgene_MVOXYMMDD.tar.gz
sievgene_NMRXYMMDD.tar.gz
MD-MVOXYMMDD.tar.gz

toolsXYMMDD.tar.gz
LibMyPrestoXYMMDD.tar.gz
doc_enXYMMDD.tar.gz

8. List of Documents

Documents in myPresto version 5.0 are as follows. They are included in doc_enYYMMDD.tar.gz or in the directory of the relating program, “doc/”.

user_manual_v5.000_en_YYMMDD.pdf (YYMMDD shows Year/Month/Date)

tplgeneX_manual_v5.000_en_YYMMDD.pdf

tplgeneL_manual_v5.000_en_YYMMDD.pdf

Hgene_manual_v5.000_en_YYMMDD.pdf

sievgene_manual_v5.000_en_YYMMDD.pdf

cosgene_manual_v5.000_en_YYMMDD.pdf

cosgene_sample_manual_v5.000_en_YYMMDD.pdf

psygene-G_manual_v5.000_en_YYMMDD.pdf

tools_manual_v5.000_en_YYMMDD.pdf

VCOL_manual_v5.000_en_YYMMDD.pdf

9. References

Please refer the following References when you make any presentation for myPresto.

Entire myPresto system and Filling potential/SRPG method

- 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 with Multicanonical MD

- 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 (2004) 70, 057103.

Particle Mesh Ewald (PME)

- 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)

- 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)

- 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

- 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

- 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

- 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)

- 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)

- 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)

- 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)

- 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

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

LigandBOX

- 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, Yusuke Sugihara, Yoshifumi Fukunishi, Haruki Nakamura, Biophysics (2013) 9, 113-121.

Sptool for solubility (LogS) and aggregator (frequent hitter) prediction

- 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.

Prediction of ligand-binding site, MolSite

- 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)

- 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

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

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