

User Guide for LeDock



Hongtao Zhao, PhD

Email: htzhao@lephar.com

Website: www.lephar.com

Introduction

LeDock is flexible small-molecule docking software, which performs an exhaustive search of position, orientation and conformation of a ligand in the active site of a protein.

Citation

Zhang, N.; Zhao, H. Enriching screening libraries with bioactive fragment space. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 3594-7.

Theory

Scoring Function. From the perspective of molecular recognition, a ligand binds to its receptor via shape complementarity using hydrogen bonds as anchors. The binding affinity ΔG_{bind} is determined by (Eq. 1).

$$\Delta G_{\text{bind}} = \alpha \sum_{i \in \text{lig}} (E_i^{\text{vdw}} + E_i^{\text{hb}}) \Theta(E_{\text{co}} - E_i^{\text{vdw}} - E_i^{\text{hb}}) + \beta(r) \sum_{i \in \text{lig}} \sum_{j \in \text{pro}} \frac{q_i q_j}{r_{ij}} + \gamma E_{\text{lig}}^{\text{strain}} \quad (1)$$

The first term is the sum of vdW interaction E^{vdw} (Eq. 2) and hydrogen bonding energy E^{hb} (Eq. 3), where Θ is the Heaviside step function and E_{co} is the cutoff energy to allow for soft docking. The second term is the electrostatic interaction energy with a distance function $\beta(r)$ accounting for both electrostatic screening and desolvation effects, where q is the partial atomic charge and r is the distance between a pair of atoms. The last term is the ligand conformational strain upon binding, and is contributed by the intra-molecular clash and/or torsion strain. Coefficients α , β and γ were empirically determined by referring to the previous scoring function.¹

$$E_i^{\text{vdw}} = \sum_{j \in \text{pro}} \varepsilon_{ij} \left[\left(\frac{r_{ij}^{\text{min}}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij}^{\text{min}}}{r_{ij}} \right)^6 \right] \quad (2)$$

$$E_i^{\text{hb}} = \sum_{j \in \text{pro}} w_{ij} (r_{ij} - r_{\text{co}}) \Theta(r_{\text{co}} - r_{ij}) \quad (3)$$

where ε and r^{min} are LJ parameters. E^{hb} runs over hydrogen bonding pairs where r_{co} is the cutoff distance of forming a H-bond, and w is the H-bonding weight to account for a variety of H-bonding strength. Although H-bonding weights may be derived in accordance with H-bond basicity,^{2, 3} they are taken from those previously used to calculate H-bonding penalty upon ligand binding.¹

Search Algorithm. The evolutionary algorithm is adopted in combination with simulated annealing search, which is used to generate the first generation of docking poses. At the beginning of each simulated annealing search, the input ligand conformation, including its position, orientation and torsions, is randomly changed, so that each search starts with a different ligand conformation.

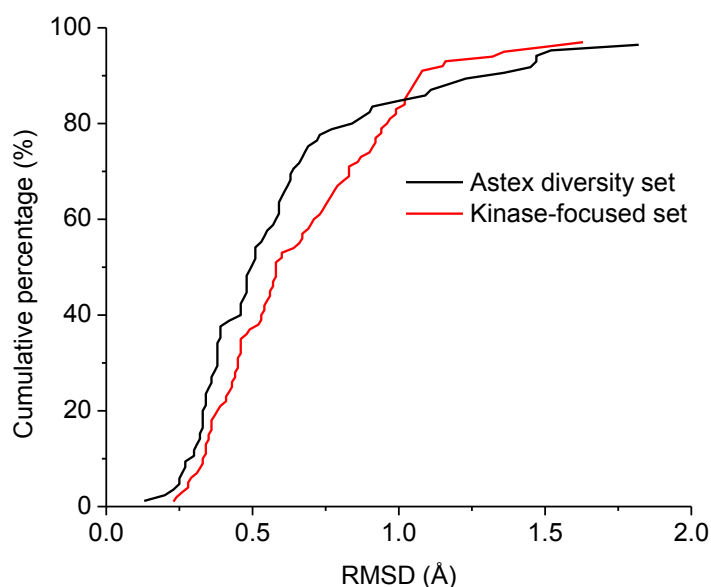
Implementation. LeDock was coded in C++ language, with a simplified input interface: definition of the binding site, number of runs, root mean square deviation (RMSD) cutoff for pose clustering, and a ligand list for sequential docking of a library. LeDock reads small molecules in a Sybyl Mol2 format, where both Sybyl atom types and bond order shall be correctly assigned.

Performance

Docking performance on the Astex diversity set and kinase-focused set

	$f_{\text{RMSD} \leq 2.0 \text{Å}}^{\text{a}}$ (%)	$f_{\text{RMSD} \leq 2.0 \text{Å}}^{\text{b}}$ (%)	Time ^c (minutes)
Astex diversity set ⁴	94.1 ± 2.0 ^d	82.4 ± 0.0 ^d	1.2
kinase-focused set ¹	97	91	1.4

^aBy the closest pose among 20 runs. ^bBy the best-scored pose among 20 runs. ^cAverage time per 20 runs. ^dStandard deviation based on three replicates each with 20 runs.



Cumulative distribution of RMSD values within the cutoff of 2 Å by the closest poses relative to the crystallographic binding modes for the Astex diversity set (black) and kinase-focused set (red).

Usage

`$LeDock`

Get “Help” information.

`$LeDock [input file]`

Dock small molecules into the protein binding site one by one. The output for each small molecule is written into a file named *[Small-molecule-name].dok*.

`$LeDock -spli [name].dok`

Split the docking poses of a docking output into individual PDB coordinates.

Input

An input file contains the following information:

```
//  
  
    Receptor  
    pro.pdb  
  
    RMSD  
    1.0  
  
    Binding pocket  
     $x_{\min}$     $x_{\max}$   
     $y_{\min}$     $y_{\max}$   
     $z_{\min}$     $z_{\max}$   
  
    Number of binding poses  
    20  
  
    Ligands list  
    ligands.list
```

```
//
```

“Receptor”, “Binding pocket”, “Number of binding poses”, and “Ligands list” are keywords, which shall not be changed.

● Receptor

Explicit hydrogen atoms should be included in the PDB file. The atom will be read in only when a line starts with “ATOM”, so if water molecules, metal ions, cofactors, substrates need to be kept during docking, **please ensure the line starting with “ATOM” instead of “HETATM”**.

The residue name and atom name of a protein shall follow the nomenclature of CHARMM. The original heavy atom name and residue name of a protein in a PDB file directly downloaded from Protein Data Bank shall also be fine. **However, please use residue name “HOH” for water molecules, instead of “TIP3”**.

Section 4 will introduce the program LePro, which can be used to automatically process a PDB file downloaded from Protein Data Bank by removing all water molecules and cofactors, and adding hydrogen atoms to the protein. LePro will also generate a docking input file automatically. Please take a look at the generated protein file and get a flavor of PDB format required for docking.

- **RMSD**

When multiple docking runs are carried out, one might find some docking poses are very similar. In order to reduce redundancy, a clustering by RMSD can be performed. The recommended value is 1.0 Å.

- **Binding pocket**

The binding pocket is set as a rectangular box. The size of the binding pocket is recommended to include any protein atom within 4 to 6 Å of any heavy atom of a drug-like molecule.

- **Number of binding poses**

20 by default.

- **Ligands list**

A file such as “ligands.list” contains all small molecules to be docked. Please prepare your small molecules in SYBYL Mol2 format. A Mol2 file of a small molecule can be directly downloaded from ZINC database (<http://zinc.docking.org/>). One can use the program LeFrag to split a batch of Mol2 ligands into separate Mol2 files:

```
$lefrag -spli batch.mol2
```

Then to generate a ligand list, one can simply use the following command:

```
$ls *mol2 > ligands.list
```

Using the program LePro to automatically process the protein and generate an input file for docking

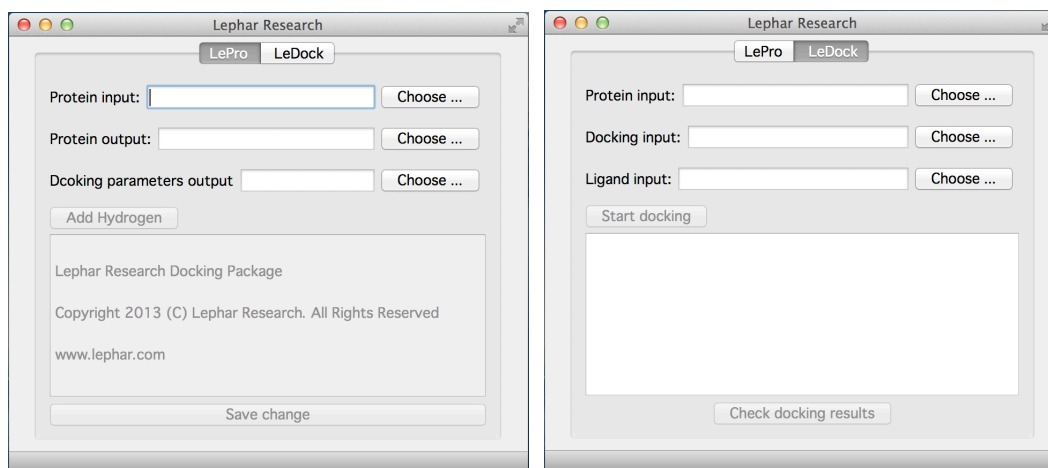
Please download your target protein from Protein Data Bank, and then use the following command:

```
$LePro [PDB code].pdb
```

Two files will be generated. One is “pro.pdb” which can be used for docking. All water molecules, cofactors, and ligands are removed. Since the hydrogen atom of OH group in residues such Tyr, Ser and Thr may have different orientations, please check if the orientation of Hydrogen is correct and if not make changes accordingly.

The other is “dock.in”. The size of the binding site is determined as to include any protein atom within 4 Å of any heavy atom of the first ligand. **However, this is not necessarily the binding site one might expect. Please ensure a correct binding site.** For more information about LePro, please refer to its manual.

Graphic interface



Tutorial

Please refer to the tutorial, which includes test cases with detailed explanation and troubleshooting.

References

1. Zhao, H.; Huang, D. Hydrogen bonding penalty upon ligand binding. *PLoS One* **2011**, *6*, e19923.
2. Laurence, C.; Brameld, K. A.; Graton, J.; Le Questel, J. Y.; Renault, E. The pK(BHX) database: toward a better understanding of hydrogen-bond basicity for medicinal chemists. *J. Med. Chem.* **2009**, *52*, 4073-86.
3. Green, A. J.; Popelier, P. L. Theoretical prediction of hydrogen-bond basicity pKBHX using quantum chemical topology descriptors. *J. Chem. Inf. Model.* **2014**, *54*, 553-61.
4. Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T.; Mortenson, P. N.; Murray, C. W. Diverse, high-quality test set for the validation of protein-ligand docking performance. *J Med Chem* **2007**, *50*, 726-41.