

# Clusterizer–DockAccessor v1.1 Tutorial

Written by Flavio Ballante, Chase Wehrle and Garland R. Marshall

Department of Biochemistry and Molecular Biophysics,  
Washington University School of Medicine, St. Louis, Missouri 63110

February 27, 2016  
(revised September 24, 2017)

```
*****  
****  
**  
*****  
*****  
*****  
**  
*  
*
```

Clusterizer–DockAccessor v1.1 Tutorial  
February 27, 2016 (revised September 24, 2017)

Clusterizer–DockAccessor software and  
documentation are distributed free of charge at:

<http://cmd.wustl.edu>

<http://www.cheminf.com>

Department of Biochemistry and Molecular Biophysics,  
Washington University School of Medicine  
St. Louis, Missouri 63110

# Contents

1. Introduction.....	4
2. Before Starting .....	4
3. Decompress the Sample Data .....	4
4. Sample Data Composition .....	5
5. Re-Docking.....	6
Perform Cluster Analysis and Binding Poses Extrapolation .....	6
Compute the Predictive Capability of a Docking Program: Docking Accuracy .....	7
6. Cross-Docking .....	8
Compute the Predictive Capability of a Docking Program: Docking Accuracy .....	10
7. Novel Compounds: Binding Pose Extrapolation .....	11
8. Citation.....	11
9. License .....	11
References.....	12

# 1. Introduction

This tutorial explains how to use Clusterizer-DockAccessor v1.1 to extrapolate best-docked, best-cluster and best-fit docked poses, perform cluster analysis, as well as compute the docking accuracy.

Along with this document, a sample data is also provided and herein referred.

**Note:** purpose of this tutorial is strictly to show the use of Clusterizer and DockAccessor.

**Note:** for a comprehensive background theory and methodology procedure refer to DOI: 10.1021/acs.jcim.5b00603.

Software, documentation and sample data can be downloaded free of charge from:

<http://cmd.wustl.edu>

<http://www.cheminf.com>

## 2. Before Starting

Before starting this tutorial it is recommended to read the Clusterizer-DockAccessor v1.1 user manual.

It is necessary first to install 1) **DOCK 6.7 or 6.8**;<sup>1</sup> 2) the **Open Babel** suite;<sup>2</sup> 3) set the “**DOCKPATH**” environment variable (as specified in the user manual, section 2.1); 4) have the **write\_conformation\_from\_dlg.py** file as well as the **vina\_split** program in your bin folder (see the user manual section 2.1); and 5) install **Clusterizer-DockAccessor** v1.1 software (see the user manual section 2.2).

The sample data is characterized by full docking calculations, thus to strictly perform this tutorial it is not necessary to have installed the docking programs as those specified in the user manual Table 2.

## 3. Decompress the Sample Data

1. Download Clusterizer-DockAccessor\_SAMPLE\_DATA.tar.gz in a preferred folder;
2. Unpack tar.gz file:

**tar xvzf Clusterizer-DockAccessor\_SAMPLE\_DATA.tar.gz**

A folder named SAMPLE\_DATA will be created.

3. Enter in the SAMPLE\_DATA folder:

**cd SAMPLE\_DATA**

## 4. Sample Data Composition

The SAMPLE\_DATA directory comprises the RE-DOCKING and CROSS-DOCKING calculations (mimicking the templates as those depicted in Figures 1 and 2 from the user manual, respectively). RE-DOCKING and CROSS-DOCKING folders are characterized by 9 DOCKING PROGRAM sub-directories (one for each docking-scoring function combination):

AUTODOCK4
AUTODOCK4_Zn
DOCK6
MPSDOCKZn
PLANTS_chemlp
PLANTS_plp
PLANTS_plp95
SURFLEX_DOCK
VINA

Each of them, have docking calculation sub-folders:

**RE-DOCKING:**

ECRD (Experimental Conformation Re-Docking)  
RCRD (Random Conformation Re-Docking)

**CROSS-DOCKING:**

ECCD (Experimental Conformation Cross-Docking)  
RCCD (Random Conformation Cross-Docking)

The hKDAC2 3MAX<sup>3</sup> chain A PDB complex has been included as RE-DOCKING calculations; The hKDAC8 inhibitor (code 1DI)<sup>4</sup> from 3SFH<sup>5</sup> chain A PDB complex and the hKDAC8 3SFF<sup>5</sup> chain A PDB protein have been included as CROSS-DOCKING calculations.

hKDAC stands for human Lysine Deacetylase.

Structures preparation and docking calculations have been performed as previously reported.<sup>4</sup>

**Note:** Clusterizer and DockAccessor results are already included. Either Clusterizer or DockAccessor will overwrite the pre-existent results when used.

## 5. Re-Docking

Begin by operating in the SAMPLE\_DATA directory and enter into the RE-DOCKING folder:

```
cd RE-DOCKING/
```

### Perform Cluster Analysis and Binding Poses Extrapolation

1. Enter the ECRD folder for a <DOCKING PROGRAM> and run Clusterizer.  
<DOCKING PROGRAM> is one of the sub-directories listed above.

This can also be done for RCRD by entering that directory instead and replacing ECRD with RCRD throughout this tutorial. When performing RCRD analysis be sure to insert the correct input key filename (check Example 2 in this tutorial).

```
cd <DOCKING PROGRAM>/ECRD
```

Note that there are three different RMSD Metrics that Clusterizer can use.  
Commands for each of them are listed below.

Standard heavy-atom RMSD:

```
Clusterizer.1.1.<DOCKING PROGRAM>.sh 3MAX.A.key.min 3MAX.A.lock.min  
3MAX.A.key.min 3MAX.A s
```

Hungarian symmetry-corrected heavy-atom RMSD:

```
Clusterizer.1.1.<DOCKING PROGRAM>.sh 3MAX.A.key.min 3MAX.A.lock.min  
3MAX.A.key.min 3MAX.A h
```

Minimum-distance heavy-atom RMSD:

```
Clusterizer.1.1.<DOCKING PROGRAM>.sh 3MAX.A.key.min 3MAX.A.lock.min  
3MAX.A.key.min 3MAX.A m
```

---

#### EXAMPLE 1

If you want to analyze ECRD AUTODOCK4 results using the standard heavy-atom RMSD:

```
cd AUTODOCK4/ECRD
```

```
Clusterizer.1.1.AutoDock.sh 3MAX.A.key.min 3MAX.A.lock.min 3MAX.A.key.min 3MAX.A s
```

---

---

## EXAMPLE 2

If you want to analyze RCRD AUTODOCK4 results using the standard heavy-atom RMSD:

```
cd AUTODOCK4/RCRD
```

---

```
Clusterizer.1.1.AutoDock.sh 3MAX.A.key.min.center.conf 3MAX.A.lock.min 3MAX.A.key.min 3MAX.A s
```

---

2. Each different RMSD Metric creates its own sub-directory comprising the log file RMSD\_and\_CLUSTER.ANALYSIS.log . These directories can be accessed as shown below.

Standard heavy-atom RMSD:

```
cd 3MAX.A/CLUSTER.ANALYSIS_HA_RMSDs
```

Hungarian symmetry-corrected heavy-atom RMSD:

```
cd 3MAX.A/CLUSTER.ANALYSIS_HA_RMSDh
```

Minimum-distance heavy-atom RMSD:

```
cd 3MAX.A/CLUSTER.ANALYSIS_HA_RMSDm
```

3. View the log file from the Clusterizer analysis:

```
vi RMSD_and_CLUSTER.ANALYSIS.log
```

## Compute the Predictive Capability of a Docking Program: Docking Accuracy

1. Navigate to the <DOCKING PROGRAM> directory:

```
/SAMPLE_DATA/RE-DOCKING/<DOCKING PROGRAM>/
```

2. Run DockAccessor:

This can also be done for RCRD by entering that directory instead and replacing ECRD with RCRD throughout this tutorial.

Standard heavy-atom RMSD:

```
DockAccessor.1.1.sh ECRD cplx.list s
```

Hungarian symmetry-corrected heavy-atom RMSD:

```
DockAccessor.1.1.sh ECRD cplx.list h
```

Minimum-distance heavy-atom RMSD:

**DockAccessor.1.1.sh ECRD cplx.list m**

3. View DockAccessor Output

Standard heavy-atom RMSD:

**vi ECRD.DOCKING.ACCURACY.HA\_RMSDs.log**

Hungarian symmetry-corrected heavy-atom RMSD:

**vi ECRD.DOCKING.ACCURACY.HA\_RMSDh.log**

Minimum-distance heavy-atom RMSD:

**vi ECRD.DOCKING.ACCURACY.HA\_RMSDm.log**

## 6. Cross-Docking

Begin by operating in the SAMPLE\_DATA directory. Enter into the CROSS-DOCKING folder:

**cd CROSS-DOCKING/**

1. Enter the ECCD folder for a <DOCKING PROGRAM> and run Clusterizer. A <DOCKING PROGRAM> is one of the sub-directories listed above.

This can also be done for RCCD by entering that directory instead and replacing ECCD with RCCD throughout this tutorial. When performing RCRD analysis be sure to insert the correct input key filename (check Example 4 in this tutorial).

**cd <DOCKING PROGRAM>/ECCD**

Note that there are three different RMSD Metrics that Clusterizer can use. Commands for each of them are listed below.

Standard heavy-atom RMSD:

**Clusterizer.1.1.<DOCKING PROGRAM>.sh 3SFH.A.key.min 3SFF.A.lock.min  
3SFH.A.key.min 3SFH.A\_3SFF.A s**

Hungarian symmetry-corrected heavy-atom RMSD:



**Clusterizer.1.1.<DOCKING PROGRAM>.sh 3SFH.A.key.min 3SFF.A.lock.min  
3SFH.A.key.min 3SFH.A\_3SFF.A h**

Minimum-distance heavy-atom RMSD:

**Clusterizer.1.1.<DOCKING PROGRAM>.sh 3SFH.A.key.min 3SFF.A.lock.min  
3SFH.A.key.min 3SFH.A\_3SFF.A m**

---

**EXAMPLE 3**

*If you want to analyze ECCD PLANTSplp results using the standard heavy-atom RMSD:*

**cd PLANTS\_plp/ECCD**

**Clusterizer.1.1.PLANTS.sh 3SFH.A.key.min 3SFF.A.lock.min 3SFH.A.key.min 3SFH.A\_3SFF.A s**

---

**EXAMPLE 4**

*If you want to analyze RCCD PLANTSplp results using the standard heavy-atom RMSD:*

**cd PLANTS\_plp/RCCD**

**Clusterizer.1.1.PLANTS.sh 3SFH.A.key.min.center.conf 3SFF.A.lock.min 3SFH.A.key.min 3SFH.A\_3SFF.A s**

---

2. Each different RMSD Metric creates its own sub-directory comprising the log file RMSD\_and\_CLUSTER.ANALYSIS.log . These directories can be accessed as shown below.

Standard heavy-atom RMSD:

**cd 3SFH.A\_3SFF.A/CLUSTER.ANALYSIS\_HA\_RMSDs**

Hungarian symmetry-corrected heavy-atom RMSD:

**cd 3SFH.A\_3SFF.A/CLUSTER.ANALYSIS\_HA\_RMSDh**

Minimum-distance heavy-atom RMSD:

**cd 3SFH.A\_3SFF.A/CLUSTER.ANALYSIS\_HA\_RMSDm**

3. View the log file from the Clusterizer analysis :

**vi RMSD\_and\_CLUSTER.ANALYSIS.log**

## Compute the Predictive Capability of a Docking Program: Docking Accuracy

1. Navigate to the <DOCKING PROGRAM> directory:

```
/SAMPLE_DATA/CROSS-DOCKING/<DOCKING PROGRAM>/
```

2. Run DockAccessor:

This can also be done for RCCD by entering that directory instead and replacing ECCD with RCCD throughout this tutorial.

Standard heavy-atom RMSD:

```
DockAccessor.1.1.sh ECCD cplx.list s
```

Hungarian symmetry-corrected heavy-atom RMSD:

```
DockAccessor.1.1.sh ECCD cplx.list h
```

Minimum-distance heavy-atom RMSD:

```
DockAccessor.1.1.sh ECCD cplx.list m
```

3. View DockAccessor Output

Standard heavy-atom RMSD:

```
vi ECCD.DOCKING.ACCURACY.HA_RMSDs.log
```

Hungarian symmetry-corrected heavy-atom RMSD:

```
vi ECCD.DOCKING.ACCURACY.HA_RMSDh.log
```

Minimum-distance heavy-atom RMSD:

```
vi ECCD.DOCKING.ACCURACY.HA_RMSDm.log
```

## 7. Novel Compounds: Binding Pose Extrapolation

If a docking protocol (using a docking program from those in Table 2) has been selected to dock novel/untested compounds in a certain target protein (e.g. during a virtual screening application), then Clusterizer can be employed to extrapolate BD and BC poses from the obtained docking calculations following the same guidelines as those specified for the docking assessment analysis, but specifying as reference pose the file used as input key conformation. Thus, since the absence of any experimental structure of the candidate compound, the BF pose can't be obtained (thus BF results from Clusterizer should be ignored), thus only the BD and BC poses may be considered.

## 8. Citation

If you used Clusterizer-DockAccessor in your work, please cite:

Ballante, F.; Marshall, G. R., An Automated Strategy for Binding-Pose Selection and Docking Assessment in Structure-Based Drug Design. *J. Chem. Inf. Model.* **2016**, *56*, 54-72.

## 9. License

Clusterizer and DockAccessor are free software: you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation, either version 3 of the License, or (at your option) any later version. This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. You should have received a copy of the GNU General Public License along with this program. If not, see <<http://www.gnu.org/licenses/>>.

## References

1. Allen, W. J.; Balias, T. E.; Mukherjee, S.; Brozell, S. R.; Moustakas, D. T.; Lang, P. T.; Case, D. A.; Kuntz, I. D.; Rizzo, R. C., DOCK 6: Impact of new features and current docking performance. *J. Comput. Chem.* **2015**, *36*, 1132-1156.
2. O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R., Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33 1-14.
3. Bressi, J. C.; Jennings, A. J.; Skene, R.; Wu, Y.; Melkus, R.; De Jong, R.; O'Connell, S.; Grimshaw, C. E.; Navre, M.; Gangloff, A. R., Exploration of the HDAC2 foot pocket: Synthesis and SAR of substituted N-(2-aminophenyl)benzamides. *Bioorg. Med. Chem. Lett.* **2010**, *20*, 3142-3145.
4. Ballante, F.; Marshall, G. R., An Automated Strategy for Binding-Pose Selection and Docking Assessment in Structure-Based Drug Design. *J. Chem. Inf. Model.* **2016**, *56*, 54-72.
5. Whitehead, L.; Dobler, M. R.; Radetich, B.; Zhu, Y.; Atadja, P. W.; Claiborne, T.; Grob, J. E.; McRiner, A.; Pancost, M. R.; Patnaik, A.; Shao, W.; Shultz, M.; Tichkule, R.; Tommasi, R. A.; Vash, B.; Wang, P.; Stams, T., Human HDAC isoform selectivity achieved via exploitation of the acetate release channel with structurally unique small molecule inhibitors. *Bioorg. Med. Chem.* **2011**, *19*, 4626-4634.