Supporting Information

ChemFlow_py: A Flexible Toolkit for Docking and Rescoring

Luca Monari[1], Katia Galentino[1] and Marco Cecchini[1]*

1. Institut de Chimie de Strasbourg, UMR7177, CNRS, Université de Strasbourg, Strasbourg Cedex 67083, France

*Corresponding author : mcecchini@unistra.fr

Ligand Preparation integrated in ChemFlow_py:

The ligand structures obtained by the DUD-E database are selected randomly in order to keep 100 active drugs and 3000 decoys. The resulting mol2 file is used directly to dock the ligands with Plants.

To prepare the structure for autodock, smina and vina we use a script provided by the MGLtools:

· python2 prepare_ligand4.py -l ligand.mol2 -o ligand.pdbqt -U "lps"

This script converts the mol2 structures to pdbqt adding gasteiger charges.

Protein Preparation integrated in ChemFlow_py:

The proteins are provided in pdb format; thus they are converted to mol2 to perform docking:

· obabel -ipdb receptor.pdb -omol2 -Oreceptor.mol2 --title receptor

Then the receptor is prepared for autodock, smina and vina thanks to the MGLtools script:

· python2 prepare_receptor4.py -r receptor.mol2 -o receptor.pdbqt

Results

- ADRB1

PROTOCOL	AUC	TYPE	EF_1	EF_5	EF_10
plants-chemplp	0.772742	dock	10	5	3.6
smina-vinardo	0.752473	dock	14	6	4.2
vina-vina	0.729572	dock	6	4.4	2.9
autodock-autodock	0.626085	dock	2	1.6	2
vina-vinardo	0.755077	rescore	9	6.2	3.9
vina-autodock	0.743723	rescore	11	4.4	3.6
vina-plp	0.737542	rescore	5	3.2	3
plants-plp	0.736533	rescore	9	5.4	3.7
smina-plp	0.726215	rescore	6	3.4	3.2
smina-autodock	0.721428	rescore	7	3	2.8
vina-plp95	0.718187	rescore	6	4	2.9
plants-plp95	0.705245	rescore	7	3.2	2.7
smina-plp95	0.703677	rescore	8	4.2	3.1
vina-dkoes_scoring	0.701127	rescore	6	4	3.1
smina-chemplp	0.695932	rescore	6	3	2.4
smina-vina	0.690863	rescore	4	2.4	2.1
smina-dkoes_scoring	0.684123	rescore	5	4.2	2.9
vina-chemplp	0.679168	rescore	4	2.4	2.3
autodock-plp95	0.653263	rescore	2	2	2.1
plants-vinardo	0.638117	rescore	4	4.2	2.7
autodock-plp	0.6349	rescore	2	2.2	1.6
autodock-dkoes_scoring	0.614273	rescore	2	2.2	1.5
plants-dkoes_scoring	0.609857	rescore	1	1.6	2.1
autodock-chemplp	0.595635	rescore	0	1.2	1.4
plants-vina	0.57849	rescore	2	1.6	1.5

plants-autodock	0.573058	rescore	3	3.2	2.3
autodock-vinardo	0.570363	rescore	1	0.8	1.1
autodock-vina	0.55246	rescore	1	0.8	0.9
plants-dkoes_fast	0.544523	rescore	0	0.4	1.3
autodock-dkoes_fast	0.4925	rescore	1	1.2	0.7
vina-dkoes_fast	0.469643	rescore	0	1.2	1
smina-dkoes_fast	0.4561	rescore	1	0.2	0.6
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.795243	rbn	9	5.6	4.2
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.78355	rbr	7	4.4	3.7
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.781833	z_score	8	5.2	3.8
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.758293	rbv	6	4.8	4.2
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.746353	ans	6	4.6	3.1
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.746173	ass	6	4.8	3.1
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.733915	ecr	6	4.4	3

- CDK2

PROTOCOL	AUC	TYPE	EF_1	EF_5	EF_10
smina-vinardo	0.782347	dock	13	5.6	4.2
autodock-autodock	0.775175	dock	14	5.2	4
vina-vina	0.723157	dock	9	4.6	3.4
plants-chemplp	0.602347	dock	5	2	1.8
smina-autodock	0.778405	rescore	16	7	4.3
autodock-vinardo	0.75802	rescore	8	4.8	3.6
smina-dkoes_scoring	0.750007	rescore	10	6	3.7
vina-vinardo	0.738167	rescore	9	4.6	3.9

autodock-dkoes_scoring	0.73312	rescore	10	3.4	2.4
plants-autodock	0.717087	rescore	10	4.4	3.3
smina-vina	0.71556	rescore	9	4	3.5
autodock-vina	0.71382	rescore	8	3.8	3.3
vina-autodock	0.711022	rescore	11	5	3.3
smina-dkoes_fast	0.705437	rescore	3	4.2	3
autodock-plp95	0.691875	rescore	9	4.2	3.2
smina-plp95	0.671755	rescore	8	4.4	2.8
plants-vinardo	0.663197	rescore	3	3.6	2.8
plants-plp95	0.658872	rescore	4	2.6	2.3
autodock-dkoes_fast	0.65593	rescore	3	2.2	1.7
autodock-plp	0.651923	rescore	6	2.8	2.4
vina-dkoes_scoring	0.636807	rescore	5	2.8	2
plants-dkoes_scoring	0.632917	rescore	2	1.4	2.5
vina-dkoes_fast	0.629827	rescore	4	2.6	2.1
vina-plp95	0.628913	rescore	4	2.2	2
plants-dkoes_fast	0.627482	rescore	2	1.4	1.8
plants-vina	0.62052	rescore	4	3.4	2.1
smina-plp	0.608867	rescore	5	2.8	2.1
autodock-chemplp	0.605478	rescore	7	3.2	2.3
plants-plp	0.587662	rescore	3	2	1.8
smina-chemplp	0.567913	rescore	5	2.2	1.8
vina-plp	0.562873	rescore	3	2	1.8
vina-chemplp	0.52044	rescore	4	2	1.6
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.793643	z_score	15	6.2	3.7
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.78447	rbr	9	4.6	3.4

vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.75021	ass	10	5.6	3.6
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.746413	ans	10	5.2	3.6
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.746187	rbv	11	5.8	4.1
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.728312	ecr	10	4.4	3.4
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.676278	rbn	8	3.8	2.9

- HIVPR

PROTOCOL	AUC	TYPE	EF_1	EF_5	EF_10
plants-chemplp	0.73861	dock	6	4	3.3
smina-vinardo	0.730283	dock	3	4.8	3.3
vina-vina	0.710038	dock	4	3.6	2.4
autodock-autodock	0.438518	dock	0	0.6	0.4
vina-dkoes_scoring	0.787287	rescore	8	5.6	4.7
smina-dkoes_scoring	0.786573	rescore	7	4.6	4.2
vina-plp	0.764303	rescore	10	5.4	4
smina-vina	0.760747	rescore	6	3.8	2.6
vina-plp95	0.747667	rescore	11	5	3.7
vina-vinardo	0.73957	rescore	6	4.2	3.2
plants-plp	0.7278	rescore	6	3.2	3.4
vina-chemplp	0.723945	rescore	7	4.8	3.2
smina-plp95	0.720583	rescore	10	4.8	3.2
smina-plp	0.71836	rescore	9	4	3.8
autodock-dkoes_scoring	0.714697	rescore	5	3.6	2.9
plants-plp95	0.693123	rescore	6	4.2	3
smina-chemplp	0.681427	rescore	9	4	3.5

autodock-vina	0.672577	rescore	1	2.4	2.5
autodock-plp95	0.6725	rescore	3	3.2	2.5
autodock-plp	0.668558	rescore	3	3.8	2.8
autodock-dkoes_fast	0.65679	rescore	6	3.6	2.8
vina-dkoes_fast	0.63862	rescore	5	3.4	2.5
smina-dkoes_fast	0.629997	rescore	4	2.2	2
plants-vinardo	0.626277	rescore	4	3.2	2.3
autodock-chemplp	0.62273	rescore	1	2.6	2.3
plants-dkoes_scoring	0.618	rescore	8	3.4	2.5
plants-vina	0.606547	rescore	4	3.4	2.1
vina-autodock	0.599735	rescore	0	0.8	1.3
autodock-vinardo	0.59919	rescore	3	1.4	1.5
smina-autodock	0.571685	rescore	2	1.2	1.4
plants-dkoes_fast	0.544677	rescore	4	2.6	1.7
plants-autodock	0.488792	rescore	1	1.2	1.2
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.744613	rbn	8	4.2	3.5
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.72398	rbv	7	3.8	2.8
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.712722	ecr	4	3.4	2.4
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.711367	ans	5	3.4	2.6
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.709807	ass	5	3.4	2.6
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.707177	z_score	5	3.8	2.5
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.694857	rbr	4	2.6	2.3

- HMDH

PROTOCOL	AUC	TYPE	EF_1	EF_5	EF_10

smina-vinardo	0.776877	dock	11	7	4.6
vina-vina	0.742818	dock	3	2.2	2
autodock-autodock	0.602253	dock	2	1.6	1.8
plants-chemplp	0.565998	dock	0	1.6	1.5
vina-dkoes_scoring	0.816027	rescore	10	6.4	4.5
smina-dkoes_scoring	0.808623	rescore	13	7	4.8
vina-autodock	0.789198	rescore	10	5.2	4.3
smina-vina	0.78862	rescore	7	5	4
vina-vinardo	0.78099	rescore	14	6.4	4.5
vina-plp	0.775767	rescore	10	5.4	4
smina-plp	0.740537	rescore	10	5.2	3.3
vina-plp95	0.738632	rescore	8	4.4	3.2
smina-autodock	0.728297	rescore	8	4.6	4.1
vina-chemplp	0.715463	rescore	4	3	2.8
smina-plp95	0.699188	rescore	8	4.4	3
autodock-vina	0.698597	rescore	1	3.2	2.3
smina-chemplp	0.67882	rescore	6	2	2.1
autodock-vinardo	0.651737	rescore	2	1.8	1.9
autodock-plp	0.640707	rescore	3	2.4	2.3
autodock-dkoes_scoring	0.63686	rescore	1	1.6	2.2
autodock-plp95	0.605592	rescore	3	2.6	2.1
autodock-chemplp	0.592782	rescore	3	1.8	1.4
plants-vinardo	0.575233	rescore	0	1	1.2
plants-plp	0.546407	rescore	0	1	1.1
plants-autodock	0.545868	rescore	3	1	1.1
plants-vina	0.529853	rescore	1	1	1.1
plants-plp95	0.504132	rescore	0	0.2	0.6
plants-dkoes_scoring	0.4734	rescore	0	0.4	0.7

plants-dkoes_fast	0.37624	rescore	0	0.4	0.4
autodock-dkoes_fast	0.3524	rescore	1	0.4	0.3
smina-dkoes_fast	0.329877	rescore	1	0.6	0.5
vina-dkoes_fast	0.319997	rescore	1	0.4	0.4
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.77159	rbv	4	3.6	3.7
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.762887	ass	5	2.4	2.6
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.758013	z_score	4	3	3.3
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.75428	ans	4	2.2	2.6
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.746895	ecr	3	2.6	2
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.719005	rbr	3	3	2.3
vina-vina--autodock-autodock--plants-chemplp--smina-vinardo	0.623003	rbn	0	1.8	2

Supplementary files for docking

- Autodock grid:

```
npts {grid_points[0]} {grid_points[2]} {grid_points[2]}          # num.grid points in xyz
gridfld receptor.maps.fld                       # grid_data_file
spacing 0.375                                   # spacing(A)
receptor_types A C H HD N NA OA S SA                    # receptor atom types
ligand_types A C H HD N NA OA S SA                      # ligand atom types
receptor receptor.pdbqt                         # macromolecule
gridcenter {center[0]} {center[1]} {center[2]}                  # xyz-coordinates or auto
smooth 0.5                                              # store minimum energy w/in rad(A)
map receptor.A.map                                      # atom-specific affinity map
map receptor.C.map                                      # atom-specific affinity map
map receptor.H.map                                      # atom-specific affinity map
map receptor.HD.map                                     # atom-specific affinity map
map receptor.N.map                                      # atom-specific affinity map
map receptor.NA.map                                     # atom-specific affinity map
map receptor.OA.map                                     # atom-specific affinity map
map receptor.S.map                                      # atom-specific affinity map
map receptor.SA.map                                     # atom-specific affinity map
elecmap receptor.e.map                          # electrostatic potential map
dsolvmap receptor.d.map                         # desolvation potential map
dielectric -0.1465                      # <0, AD4 distance-dep.diel;>0,constant
```

- Autodock docking:

```
## GENERIC SECTION
autodock_parameter_version 4.2 # used by autodock to validate parameter set
outlev adt                                      # diagnostic output level
intelec                                 # calculate internal electrostatics
seed pid time                                   # seeds for random generator

## LIGAND-SPECIFIC SECTION
ligand_types A C H HD N NA OA S SA                      # atoms types in ligand
fld receptor.maps.fld                           # grid_data_file
map receptor.A.map                                      # atom-specific affinity map
map receptor.C.map                                      # atom-specific affinity map
map receptor.H.map                                      # atom-specific affinity map
map receptor.HD.map                                     # atom-specific affinity map
map receptor.N.map                                      # atom-specific affinity map
map receptor.NA.map                             # atom-specific affinity map
map receptor.OA.map                             # atom-specific affinity map
map receptor.S.map                                      # atom-specific affinity map
map receptor.SA.map                             # atom-specific affinity map
elecmap receptor.e.map                          # electrostatics map
desolvmap receptor.d.map                        # desolvation map
move ligand.pdbqt                               # small molecule

## INITIAL SEARCH STATE SECTION
tran0 random                            # initial coordinates/A or random
quaternion0 random                      # initial quaternion
dihe0 random                            # initial dihedrals (relative) or random

## SEARCH-SPECIFIC SECTION
```

```
ga_pop_size 150                 # number of individuals in population
ga_num_evals 2500000                    # maximum number of energy evaluations
ga_num_generations 27000        # maximum number of generations
ga_elitism 1                            # number of top individuals to survive to next generation
ga_mutation_rate 0.02                   # rate of gene mutation
ga_crossover_rate 0.8                   # rate of crossover
set_ga                          # set the above parameters for GA or LGA

## LOCAL SEARCH PARAMETERS SECTION
sw_max_its 300          # iterations of Solis & Wets local search
sw_max_succ 4                   # consecutive successes before changing rho
sw_max_fail 4                   # consecutive failures before changing rho
sw_rho 1.0                      # size of local search space to sample
sw_lb_rho 0.01          # lower bound on rho
ls_search_freq 0.06             # probability of performing local search on individual
set_psw1                # set the above pseudo Solis & Wets parameters

## PERFORM SEARCH SECTION
ga_run 10               # do this many hybrid GA-LS runs

## ANALYSIS SECTION
rmstol 0.5                      # cluster_tolerance/A
analysis                # perform a ranked cluster analysis
```

- Smina docking:

```
#CENTER DEFINITION
receptor=../receptor.pdbqt
ligand=ligand.pdbqt
center_x={center[0]}
center_y={center[1]}
center_z={center[2]}

#SIZE BOX
size_x={size[0]}
size_y={size[1]}
size_z={size[2]}

#SCORING FUNCTION
scoring=vinardo

#OTHER PARAMETERS
energy_range=3.0
exhaustiveness=8
num_modes=10

#USE ACCURATE LINE SEARCH
accurate_line=yes
out=output.pdbqt
log=output.log
```

- Vina docking:

```
#CENTER DEFINITION
receptor=../receptor.pdbqt
ligand=ligand.pdbqt
center_x={center[0]}
```

```
center_y={center[1]}
center_z={center[2]}

#SIZE BOX
size_x={size[0]}
size_y={size[1]}
size_z={size[2]}

#OTHER PARAMETERS
energy_range=3.0
exhaustiveness=8
num_modes=10
out=output.pdbqt
log=output.log
```

- Plants docking:

```
# input files
protein_file ../receptor.mol2
ligand_file  ligand.mol2

# scoring function and search settings
scoring_function chemplp
search_speed speed1
aco_ants 20
aco_evap 2
aco_sigma 1.0

# binding site definition
bindingsite_center {center[0]} {center[1]} {center[2]}
bindingsite_radius {radius}

# write mol2 files as a single (1) or multiple (0) mol2 files
write_multi_mol2 1

# cluster algorithm, save the best DOCK_POSES.
cluster_structures 10
cluster_rmsd 1

# write
write_ranking_links 0
write_protein_bindingsite 1
write_protein_conformations 0

#parameters for output
output_dir  .
```

Supplementary files for rescoring

- Smina rescoring (scoring function: dkoes_fast, dkoes_scoring, vina, vinardo):

```
#CENTER DEFINITION
```

```
receptor=../receptor.pdbqt
ligand=ligand.pdbqt
center_x={center[0]}
center_y={center[1]}
center_z={center[2]}

#SIZE BOX
size_x={size[0]}
size_y={size[1]}
size_z={size[2]}

#SCORING FUNCTION
scoring=vina

#RESCORING
score_only=yes

#USE ACCURATE LINE SEARCH
minimize=yes
accurate_line=yes
out=output.pdbqt
log=output.log
```

- Plants rescoring (scoring function: plp, plp95, chemplp):

```
# input files
protein_file ../receptor.mol2
ligand_file  ligand.mol2

# scoring function and search settings
scoring_function plp
rescore_mode simplex


# binding site definition
bindingsite_center {center[0]} {center[1]} {center[2]}
bindingsite_radius {radius}

# write mol2 files as a single (1) or multiple (0) mol2 files
write_multi_mol2 1

# write
write_rescored_structures 1
write_ranking_links 0
write_protein_bindingsite 1
write_protein_conformations 0

#parameters for output
output_dir  .
```

- Autodock rescoring (scoring function: autodock):

```
## GENERIC SECTION
autodock_parameter_version 4.2 # used by autodock to validate parameter set
outlev adt                                  # diagnostic output level
intelec                                 # calculate internal electrostatics
```

```
## LIGAND-SPECIFIC SECTION
unbound_model bound                     # state of unbound ligand
ligand_types A C H HD N NA OA S SA              # atoms types in ligand
fld receptor.maps.fld                   # grid_data_file
map receptor.A.map                              # atom-specific affinity map
map receptor.C.map                              # atom-specific affinity map
map receptor.H.map                              # atom-specific affinity map
map receptor.HD.map                             # atom-specific affinity map
map receptor.N.map                              # atom-specific affinity map
map receptor.NA.map                     # atom-specific affinity map
map receptor.OA.map                     # atom-specific affinity map
map receptor.S.map                              # atom-specific affinity map
map receptor.SA.map                     # atom-specific affinity map
elecmap receptor.e.map                  # electrostatics map
desolvmap receptor.d.map                # desolvation map
move ligand.pdbqt                       # small molecule

## LOCAL SEARCH PARAMETERS SECTION
sw_max_its 300                  # iterations of Solis & Wets local search
sw_max_succ 4                           # consecutive successes before changing rho
sw_max_fail 4                           # consecutive failures before changing rho
sw_rho 1.0                              # size of local search space to sample
sw_lb_rho 0.01                  # lower bound on rho
ls_search_freq 1.00                     # probability of performing local search on individual
set_psw1                        # set the above psw Solis & Wets parameters

## PERFORM SEARCH SECTION
do_local_only 50                # do this many hybrid GA-LS runs

## ANALYSIS SECTION
rmstol 0.5                              # cluster_tolerance/A
analysis                        # perform a ranked cluster analysis
```