ZIBAffinity Workflow

1. Application

MADF	Pilot	Test	Benchmark
-	ZiBAffinity	-	-

2. Requirements

ID	Type	Location	Description
1	Bash script	/mnt/lustre/scratch/home/otras/zib/vdu/data_4_javier/za_main.sh	Gets input, executes container (=ID 2) tools, writes output files to “./sim_dir”
2	Container	/mnt/lustre/scratch/home/otras/zib/vdu/data_4_javier/ubuntu-16.04.img	Contains all the tools and libraries required by ZIBaffinity
3	Path to target molecule files	Input argument to ID 1	Path to ligand input file given as user input (selected from database). **For testing**: “./input/erAlpha-1gwr”
4	Path to ligand molecule files	Input argument to ID 1	Path to ligand input file given as user input. **For testing**: “./input/17betaEstradiol.pdb”
5	Signed integer	Input argument to ID 1	Formal charge of ligand molecule (ID 4) **For testing**: “0”
6	Path to simulation directory and output files	Input argument to ID 1	Path for simulations and analysis. **For testing**: /mnt/lustre/scratch/home/otras/zib/vdu/data_4_javier (= “.”)
7	User email address	Input argument to ID 1	User Email address used to name some directory

3. Steps

Step ID	Command
Main script (Req-ID 1)	za_main.sh <proteinPath> <ligandFile> <ligandCharge> <workPath> <userEmail> (bash script performing/executing the next four steps) **For testing:** cd /mnt/lustre/scratch/home/otras/zib/vdu/data_4_javier ./za_main.sh ./input/17betaEstradiol.pdb ./input/erAlpha-1gwr 0 . durmaz@zib.de
Preliminary	Several commands executed by main script. Preliminary checking/ creation of input files and directories needed for the simulation.
Preprocess	run_singularity.sh $IMG_NAME_PREP /bin/bash $SCRIPTPATH/za_prep_sim.sh $INFILE_LIG_FULL $INFILE_TAR_FULL $CHARGE $WORKPATH (bash script “za_prep_sim.sh” inside container executed by main script) Force field parameterization of the ligand molecule and composition of initial binding modes of protein-ligand systems in explicit water.
Simulation	OUT=$(sbatch -N 61 -n 24 -t 01:30:00 -p cola-corta --array=0-60 --export=IMG_NAME_PREP,EMAILADRS,JOBDIR,COMPLEXDIR,zaLogFile,TARCONF,SING_MPI_MOD,SCRIPTPATH,za_cfg_file temp_mdrun) (batch job script submitted by main script) 61 independent molecular dynamics simulations using SLURM job arrays. Batch script is copied from container to local file on host called “temp_mdrun” using singularity and submitted to job queue by host.
Postprocess	OUT=$(sbatch -N 1 -n 1 -t 00:20:00 -p cola-corta --dependency afterok:$DEPENDENCY --export=orientSep,IMG_NAME_PREP,EMAILADRS,JOBDIR,COMPLEXDIR,zaLogFile,TARCONF,SING_MOD,SCRIPTPATH,za_cfg_file,MOL temp_postsim) (batch job script submitted by main script and depending on the output of previous step) Derives thermodynamic quantities from Gromacs output files, determines favorable binding mode and estimates its binding energy. Batch script is copied from container to local file on host called “temp_postsim” using singularity and submitted to job queue by host.

4. Workflow

Step ID	Depends on	Required by	Success definition	Conditional jump
			Main log file: ./${EMAILADRS}___ ${UNIXTIME}/sim___ ${TARGETNAME}/za .log	
Preliminary		Preprocess	Log search: “Preliminary checks of input files or directories successful”	If found, no action required (next script, preprocessing, executed by main script). Otherwise (“unsuccessful”) program abortion.
Preprocess	Preliminary	Simulation	Log search: “Preprocessing terminated successfully”	If found, no action required (next script, simulation batch job, submitted by main script). Otherwise (“unsuccessfully”) program abortion.
Simulation	Preprocess	Postprocess	Log search: “Simulations terminated properly” [${integer}|”None”] out of 60 runs with different initial binding modes completed successfully”	If found, no action required (next script, postprocessing batch job, submitted by main script). Otherwise (“improperly”) program abortion. Also an unsigned ${integer} should be provided to user.
Postprocess	Simulation		Log search: “Binding affinity calculation successful”	If found, done. Send results (two output files) to user. Otherwise (“unsuccessful”) program abortion. Inform user.

5. Resources

Step ID	Cores	Nodes	GPU	RAM	Time	# runs
Preliminary	1	1	-	-	1 second	1
Preprocess	1	1	-	-	3-5 minutes	1
Simulation	24	1	-	< 1GB	1.5-2 hours	61 (SLURM job array. Up to 20 simultaneous runs observed.)
Postprocess	1	1	-	< 0.2 GB	20 minutes	1 (Slurm job, depending on “Simulation”)

6. Appendix