Automated workflow Production Free Energy Simulation Setup and Analysis (*ProFESSA*)

November 1, 2021

Table 1. Keywords associated with the *ProFESSA* workflow

Keyword	Value	Description	Example
path_to_input	string	Path to directory that contains input files. It should contain a subdirectory *system*.	path_to_input =/home/username/afe/initial
system	string	Name of the system. A folder named system should be present in *path_to_input/*, and should contain the initial structure and parameter files.	system=CDK2
translist	list of strings	A list of desired transformations or edges. In the case of RBFE or RSFE calculations, *translist* should be a list in which each entry consists of two molnames separated by the character "~", while in the case of ASFE calculations, *translist* should be a list of molnames. Initial structure/parameter files of these molnames should be provided in *path_to_input/system* and should be named as follows: For RBFE calculations, each molname present in *translist* should have an associated *molname.pdb*, representing the receptor-ligand complex structure and [*molname_0.mol2 molname_0.lib molname_0.frcmod]*, representing the ligand parameters. Parameters of additional non-standard residues, if present, can be provided as [*molname_1.mol2 molname_1.lib molname_1.frcmod]*, [*molname_2.mol2 molname_2.lib molname_2.frcmod]*, etc. For RSFE and ASFE calculations, each molname present in *translist* should have associated [*molname_0.mol2 molname_0.lib molname_0.frcmod]* files.	translist=(1h1q~1h1r 1h1q~1h1s)
ticalc	string	Specifies nature of calculation. Acceptable values - *rbfe*, *rsfe*, *asfe*.	ticalc=rbfe
nlambda	integer	Number of lambda windows in TI calculation. Acceptable values - positive integers.	nlambda=21
protocol	string	Protocol for running TI simulations. Acceptable value - *unified*.	protocol=unified
mapmethod	integer	Specifies the algorithm using which TS/TC regions are going to be determined. Acceptable values - *0*, *1*, *2*. *0* specifies MCS algorithm. *1* specifies MCS-E algorithm. *2* specifies MCS-E2 algorithm. For a given transformation/edge *molname1~molname2* in *translist*, an atom map file, *molname1~molname2.map.txt*, is generated in the folder *system/setup*.	mapmethod=1
mapinspect	integer	Allows manual inspection of TS/TC regions. Acceptable values - *0, 1, 2.* *0* specifies no manual inspection. *1* specifies manual inspection. *2* specifies resume setup assuming manual inspection has been completed. If *mapinspect* is set to *2*, the necessary atom map files should be present in *system/setup* folder.	mapinspect=true
mapnetwork	string	Specifies if network-wide consistent TS/TC regions of ligands will be generated. Acceptable values - *true*,*false*.	mapnetwork=false

Continued on next page

Table 1 – continued from previous page

Keyword	Value	Description	Example
boxbuild	string / integer	Specifies if and how MD boxes will be built. Acceptable values - *0*, *1*, *2*, *skip*. *0* specifies to build boxes only for "aqueous" state and not for "complex" state. *1* specifies to build boxes for both "aqueous" and "complex" states. *2* specifies to build boxes for both "aqueous" and "complex" states with identical number of water and ions. *skip* specifies to skip box building altogether. For RSFE and ASFE calculations, *boxbuild 0* and *1* are identical.	boxbuild=1
boxbufcom	integer	Specifies MD box buffer for "complex" states. Relevant only for RBFE calculations. Acceptable values - positive integers.	boxbufcom=16
boxbufaq	integer	Specifies MD box buffer for "aqueous" states. Relevant only for RBFE calculations. Acceptable values - positive integers.	boxbufaq=21
ionconc	float	Specifies the MD box ion concentration in units of mol/L (M). Acceptable values - positive real number.	ionconc=0.15
pff	string	Specifies protein forcefield. Acceptable values - *ff14SB*	pff=ff14SB
lff	string	Specifies ligand forcefield. Acceptable values - *gaff*, *gaff2*	lff=gaff2
wm	string	Specifies water model. Acceptable values - *tip4pew*, *tip3p*	wm=tip4pew
mdboxshape	string	Specifies shape of MD box. Acceptable values - *cubic*	mdboxshape=cubic
ntrials	integer	Specifies the number of independent trials of calculation. Acceptable values - positive integers.	ntrials=10
cutoff	integer	Specifies non-bonded cutoff in TI simulations. Acceptable values - positive integers.	cutoff=10
repex	string	Specifies if Hamiltonian Replica Exchange will be employed.	repex=true
nstlimti	integer	Specifies the length of production TI simulations in units of fs. Acceptable values - positive integers.	nstlimti=5000
numexchgti	integer	Specifies the number of exchanges in replica exchange TI simulations. *numexchgti* is ignored is *repex* is set to false. Acceptable values - positive integers.	numexchgti=1000
hmr	string	Specifies if Hydrogen Mass Repartitioning will be used. Acceptable values - *true*, *false*.	hmr=false
notrajectory	string	Specifies if production trajectories will be saved during TI simulations. Acceptable values - *true*, *false*.	notrajectory=true
scalpha	float	Specifies the value of *AMBER DD BOOST* keyword *scalpha* in TI simulations. Acceptable values - positive real numbers.	scalpha=0.5
scbeta	float	Specifies the value of *AMBER DD BOOST* keyword *scbeta* in TI simulations. Acceptable values - positive real numbers.	scbeta=0.5
gti_add_sc	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_add_sc* in TI simulations. Acceptable values - positive integers.	gti_add_sc=5
gti_scale_beta	float	Specifies the value of *AMBER DD BOOST* keyword *gti_scale_beta* in TI simulations. Acceptable values - positive real number.	gti_scale_beta=1
gti_cut	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_cut* in TI simulations. Acceptable values - positive integers.	gti_cut=1
gti_cut_sc_on	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_cut_sc_on* in TI simulations. Acceptable values - positive integers.	gti_cut_sc_on=8

Continued on next page

Table 1 – continued from previous page

Keyword	Value	Description	Example
gti_cut_sc_off	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_cut_sc_off* in TI simulations. Acceptable values - positive integers.	gti_cut_sc_off=10
gti_lam_sch	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_lam_sch* in TI simulations. Acceptable values - positive integers.	gti_lam_sch=1
gti_ele_sc	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_ele_sc* in TI simulations. Acceptable values - positive integers.	gti_ele_sc=1
gti_vdw_sc	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_vdw_sc* in TI simulations. Acceptable values - positive integers.	gti_vdw_sc=1
gti_cut_sc	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_cut_sc* in TI simulations. Acceptable values - positive integers.	gti_cut_sc=1
gti_ele_exp	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_ele_exp* in TI simulations. Acceptable values - positive integers.	gti_ele_exp=2
gti_vdw_exp	integer	Specifies the value of *AMBER DD BOOST* keyword *gti_vdw_exp* in TI simulations. Acceptable values - positive integers.	gti_vdw_exp=2
twostate	string	Specifies if twostate setup will be employed in TI simulations. Acceptable values - *true*, *false*.	twostate=true
bidirection_aq	string	Specifies if bidirectional setup will be used for "aqueous" state TI simulations. Applicable when *twostate* is set to *false*. Acceptable values - *true*, *false*.	bidirection_aq=false
bidirection_com	string	Specifies if bidirectional setup will be used for "complex" state TI simulations. Applicable when *twostate* is set to *false*. Acceptable values - *true*, *false*.	bidirection_aq=false
stage	string	Specifies the action of the script. Acceptable values - *setup*, *analysis*. *setup* specifies script to set up TI simulations. *analysis* specifies script to perform analysis.	stage=setup
setupmode	integer	Specifies the mode of simulation setup. Acceptable values - *0* *0* sets up regular TI simulations.	setupmode=0
partition	string	Specifies the HPC partition on which TI runs will be performed. Acceptable values - *null*, *name of HPC partition*.	partition=gpu
nnodes	integer	Specifies the number of nodes to be requested for a single set of TI simulations. Acceptable values - positive integer.	nnodes=1
ngpus	integer	Specifies the number of gpus per node to be requested for a single set of TI simulations. Acceptable values - positive integer.	ngpus=8
wallclock	string	Specifies the wallclock on TI jobs. Acceptable values - formatted time in hours:minutes:days.	wallclock=3-00:00:00
path_to_data	string	Specifies the path to production runs. Default path is set to *system/protocol/run*	path_to_data=CDK2/unified/run
exptdatafile	string	Specifies the name of a text file containing experimental ligand binding free energies. Acceptable values - *skip*, *filename*. The text file should have two columns corresponding to *molname* (column 1) and relative ligand binding free energy (column 2).	exptdatafile=Expt.dat
bar	string	Specifies if BAR is going to be used for analysis instead of MBAR. Acceptable values - *true*, *false*.	bar=false
ccc	string	Specifies if cycle closure corrections are to be applied during analysis. Acceptable values - *true*, *false*.	ccc=true

Continued on next page

Table 1 – continued from previous page

Keyword	Value	Description	Example
ccc_ddG	string	Specifies if cycle closures will be applied to *"complex"* - *"aqueous"* delta delta Gs instead of *"complex"* and *"aqueous"* delta Gs, independently. Acceptable values - *true*, *false*.	ccc_ddG=true
start	float	Specifies the percentage of data to ignore from the beginning of TI production runs. Acceptable values - float numbers ranging from 0 to 100, and less than *stop*.	start=20.0
stop	float	Specifies the percentage of data to read from the start of TI production runs. Acceptable values - float numbers ranging from 0 to 100, and greater than *start*.	stop=100.00
check_convergence	string	Specifies if check of data convergence will be carried out during analysis. Acceptable values - *true*, *false*. If *check_convergence* is set to *true*, the analysis is carried out multiple times, for a range of *start* and *stop* values.	check_convergence=true
showallcycles	string	Specifies if the output should show information on all possible cycles within the given transformation network. Acceptable values - *true*, *false*.	showallcycles=true