Questions/clarifications and comments about tools used by VR,PL

* Check when closed loop coproduct simulation with scoring will be finished. Adjust/revise schedule to account for delay.

We will judge which of the following pairs to present after checking these results

-4FVT/4BVG approach

-4BVH/34GLS approach

RSK: I hope to have the MD simulation results by the end of Monday and the MM/GBSA scores by Tuesday.

* Ligand placement for INT/NAM complex (not in native structure) – PL has planned to try various approaches (see wiki/task list) to reduce energy of INT/NAM complex (vdw scaling, ala mutagenesis, etc). Was this applied or not?

RSK: In fact PL has tried side chain repacking and Induced fit docking approach for NAM placement. He has reported on this. However, going by the documents it looks like the model was chosen based on remodeling of side chain, over IFD docking. Vdw scaling and alanine mutations are an integral part of the IFD protocol. I remember he has mentioned in some document that IFD didn’t produce satisfactory results.

* MM-GBSA/PBSA scoring automation script (need for this had been previously been mentioned by RC): frame by frame scoring not finished?; need clarification, what does this do that mm-pbsa.py does not? [Note from software dev page: we had been developing scripts for mm-gb(pb)sa scoring of docked poses using amber for ifd or design applications. Answer the above questions on that page.]

RSK: The frame by frame scoring script is a part of the schedule listed under miscellaneous items. The MMPSA.py script provided by Amber developers computes the binding energy of each frame internally and reports the ensemble average binding energy over all the frames. However, if one needs the binding energy of each frame, a script has to be written to mine the energy of each frame which it has stored internally in a log file created during the MM/PBSA run. This is required if one needs to plot a distribution of the binding/conformational energies. This functionality needs to be scripted.

RSK: The shell script which I wrote for automating the MMPSA calculation was to compute the MMPBSA binding energies/binding free energies for a specified time interval. The does the flowing functionality

1. It reads the entire trajectory and creates multiple MMPBSA input scrips t based on user defined time interval (ie it creates a MMPBSA input script for parsing frames between 1-2 ns, 2-3 ns, 3-4ns etc ) and splits the trajectory based on the time interval.
2. Once it creates all the required input MMPBSA scripts, it submits the job to each available thread on the CPU and thereby parallelizes it into multiple jobs.

Advantages the script offers are

 The need to create multiple MMMPBSA input script is automated

 In one MMPBSA run one can get the energies for the entire defined time interval

 period. Else, one will have to create multiple scripts manually and run them serially.

 But here all the jobs are run in parallel in one go.

 It can be used to check for convergence of free energies and to identify the converged

 time period for computing free energies ( ie say the last 10 ns or last 5 ns etc.. of the

 simulation)

* Software used to prepare systems for MD simulations – in case of any differences wrt Ping. To facilitate methods, confirm:

PC\_GAMESS for Gaussian calculations rather than AM1-BCC

Antechamber for partial charge fitting?

specify software used for ligand interaction diagrams

adding hydrogens – may indicate that either protein prep wizard or other software package was used

RSK: There is no difference between the method which I used and Ping had used for calculating the RESP charges. There is only a slight difference in the way hydrogen was added to the protein earlier. However, this will not be an issue because we are going to use Schrodinger for preparing the open loop simulation of 4BVH and for the closed loop ( the simulation which is running presently, I had a 4BVH protein structure prepared by Ping using Schrodinger) So there is absolute consistency in the methods. In fact I have adapted the same RESP for AADPR based on what Ping has calculated so there is no difference here.

Software’s used

RED programs using Firefly (PC\_GAMESS) for deriving RESP fitted partial charges

Amber14 tools (Antechamber program) are employed for parametrizing the non-standard residues and tleap program for creating topology and force field parameters for the standard residues.

 NAMD for MD simulation employing Amber ff and TIP3P solvent model

* You should not do a hard shutdown of nodes unless absolutely required. You will note in my email to you when you inquired about this that I assumed that prior to shutdown you had

verified this was not a networking issue. I believe you only checked that afterwards.

RSK: Okay, point taken.

 -Plan to reinstall mpi? Need to reinstall Amber in that case pointing to new paths?

 I think that there will need to reinstall Amber, however the MPI /MPICH issue needs to be fixed

 in the long run. Will look into it later after the completion of the paper work together with the

 slave003 issue.

-Issues with outdated version of PC-GAMESS. If you were not able to get the job completed on the webserver, why not reinstall / clean install PC gamess?

Yes the complete RED package for RESP charge derivation needs to be reinstalled along with the latest version of PC\_GAMESS/Firefly. This will be done later ( less priority)

- Do not reinstall schrodinger unless required – use same force field

I have installed Schrodinger on the windows machine now. There is an issue with the license key they send us( Sherry has been copied on all the email communications with Schrodinger pertaining to the licensing issue). They have acknowledged the issue and they will be sending out a new license file today. This shouldn’t be an issue, but I anticipated such trivial issues with license file and the possible need to reach out to their support team for assistance. That’s the reason I had kept aside 2 days for installation in my first schedule.

I talked with Schrodinger support team yesterday, and they told me that their current license can be used to invoke any Schrodinger version post 2014 release and I am not sure if the current license will help us to run the 2014 release which Ping had previously installed on the windows machine too.

In any case that shouldn’t be a problem because we will have the option to turn on the old OPLS version too.

* See attached annotated “side chain prediction + refinement RC comments.doc” for comments and questions regarding these methods

Revised list of miscellaneous tasks

* The list below specifies those which can be done concurrently with simulations/scoring (“ready” indicates they can be done at any time – indicate when you will do them)

-Fig ------Simulated B factor values for Sirt3/OAADPr complexes modeled based on native closed conformation (4BVH) and an open loop conformation as seen in apo enzyme (3GLS). Related, where did the “SIRT3/aADPR” plot come from in the current SIRT3 simulated B factor SI Fig?

I see that Ping has done some simulations using Sirt3/AADPr and Sirt3/AADPR/Ex423 complex (between May 12- 20 2014). But the details and the contents of the folders are not that clear like those documented for the Sirt3/INT/NAM complex. Probably, he felt that this simulation was of less priority over the Sirt3/INT/NAM simulations at that point in time. He has calculated the B factors from these simulations I suppose.

 Although it may sound that the simulation we are running is currently redundant it’s certainly not considering the very limited documentation we have. Also, I feel that we compare the energies between a pair (open loop vs closed loop) of simulation, it’s better that we used the same protocol for setting up the MD system or be absolutely clear of the method used for setting up the earlier simulation, as it may reflect on the MM/PBSA and MM/GBSA energies.

Secondly, although I see those trajectories are available, but the naming convention used for the files makes it hard to decipher.

For example under the folder

/home2/plin/work/project01/MD\_4BVH\_AADPR\_bin which has the trajectories for sirt3/AADPr there are two topologies for the same receptor and I am not sure what that means (two topologies for the same receptor???)

 4BVH\_SIRT3\_AADPR\_rec1.prmtop

 4BVH\_SIRT3\_rec3.prmtop

-Per-residue RMSD values for the cofactor binding loop region calculated with respect to

MD averaged structure of Sirt3/OAADPr complex modeled based on an open and closed loop conformation.

RSK: Once the MD run is over this will be completed. Listed in the revised schedule.

-Fig MD averages for Sirt3/OAADPr complexes modeled based on native closed conformation (4BVH) and an open loop conformation as see in apo enzyme (3GLS).

-Fig Ligand interaction diagrams for the product interacting with open and closed loop conformation for Sirt3/INT/NAM complex some ready

RSK: The MD average structure of the last 10 ps will be used to create this interaction diagram for Sirt3/AADPr in open and closed loop conformation. The other two for Sirt3/INT/NAM complex also needs to be done. For consistency purpose all interaction diagram will be done using the Schrodinger ligand interaction tool.

-Table …..MM/GBSA and MM/PBSA conformational energies and binding affinity calculation based on the new simulation results. Do not report NAM binding affinities: insufficient sampling

RSK: Okay the table will be revised when we add the values for the Sirt3/ AADPR conformational energies.

-Time series plot of MM/GBSA and MM/PBSA energies for Sirt3/OAADPr with open and closed loop conformation. Some ready

RSK: for Sirt3/INT/NAM we do have it. We need to plot a similar one for Sirt3/AADPr. However, based on your above comment that “. Do not report NAM binding affinities: insufficient sampling” do we need this plot if you feel that the binding energies are not converged due to finite time scale sampling issues..

-Creating probability density distribution plots based on the energies of each frame in the MD simulation. Includes conformational energy distribution fns for all simulations including 4BVG with loop replacement from 4FVT for purpose of next paper. VR should prepare pdfs at successive times; RC will use for convergence analysis.

RSK: Okay.

Make another version of each of these distribution figures wherein the x axis is RMSD with respect to starting structure. Annotate the location (RMSD) of the MD average structure in each case

 some ready

 RSK: I don’t think that we have an RMSD vs time (plot showing the evolution of RMSD with

 time). I feel this is a definite requirement to show converge of the simulation. Ideally, checking

 energies for convergence is quite tricky because even a small change in conformation reflects a

 large charge in energies. A RMSD plot will be created ( Time vs RMSD and a distribution plot as

 suggested by you)

-SI Fig ------: Plot showing crystallographic B-factor values of the Cα atoms belonging to the co-factor binding loop region of Sir2 in different states. Ready

RSK:Okay

-B factors for any sir2 simulations available from PL’s work (see RC’s prior comment on this)? If so make the corresponding plot analogous to that fro SIRT3

RSK: Okay.

-Sir2 structure alignments (for SI) ready

 -Starting structures for simulations (for SI) some ready

-Method for Ligand/NAM placement needs to be worked out for the supplementary section

-MD simulation method in particular treatment of non-standard residues has to be worked out

- Computational method references from prev paper to be added vis a vis methods. if any references are no longer relevant or new ones are relevant to the tools used, indicate that

--Side chain prediction validation: Need to summarize side chain prediction validation method and provide associated plots.

-one of new required methods subsections

-two validation datasets: 4FVT and 4BVG

-present the data in scatterplot format

-consider whether to use subsets of residues in each as separate datasets

-specify how the residue sets were chosen, along w/ answers to any other questions from RC’s notes

-do not subdivide into energy and sampling errors at this time(?). do not correct energy errors

-**further details to be provided shortly**

-time series of energies: Note that showing frame at t=0 will depict immediate product energies for both open and closed loop conformations. Consider showing 3 traces for INT/NAM: complex, receptor+INT, receptor+NAM

RSK: Okay. I am not sure if INT with NAM together as a ligand will work in MM/PBSA, because it may have been hardcoded that MMPBSA.py script may not treat both INT and NAM as a ligand together because MMPBSA approach basically follows a 1:1 equilibrium ratio.. Anyways, let me try it. It’s just an heads up that we may run into issues when we attempt INT-NAM