California Polytechnic State University

Department of Chemistry and Biochemistry
Dr. McDonald Computational Research Laboratory

Protein Structure and Design Project: MEK1

Previous Work:

Sabsay, K. R.; Lee, R. T.; Ravatt, L. M.; Oza, J. P.; McDonald, A. R. Computational Models for Activated Human MEK1: Identification of Key Active Site Residues and Interactions. *J. Chem. Inf. Model.* **2019**, acs.jcim.8b00989. https://doi.org/10.1021/acs.jcim.8b00989.

Proposed Next Steps:

1. **Focus on the 20ns simulation of the MEK_{1IR3g6} Complex (MEK_{1IR3} : $ERK2_{model}$: ATP). Extract coordinate files from the simulation trajectory and run interaction energy calculations (e.g., SAPT) for each frame.**

Potential New Directions:

- Re-run simulation with the addition of Magnesium ions.
- Consider potential mutations of the ERK2 activation loop (i.e., the model peptide) that would strengthen the interaction energy.
- Consider the MEK1-drug interactions and any potential drug variations to enhance interaction energy.

Helpful Tools:

1. Terminal (access and submit jobs on skylight)
2. VMD (view pdb files, calculate RMSD, alignments)
3. PyMol (view pdb files, sequence specific alignments)
4. TextWrangler (view and manipulate text files)

Questions? Contact: +1 (408) 712-3165 | kimberlyrsabsay@gmail.com

California Polytechnic State University

Department of Chemistry and Biochemistry

Dr. McDonald Computational Research Laboratory

Interaction Energy Calculations: MEK_{1IR3g6} Complex (MEK_{1IR3}:$ERK2_{model}$:ATP)

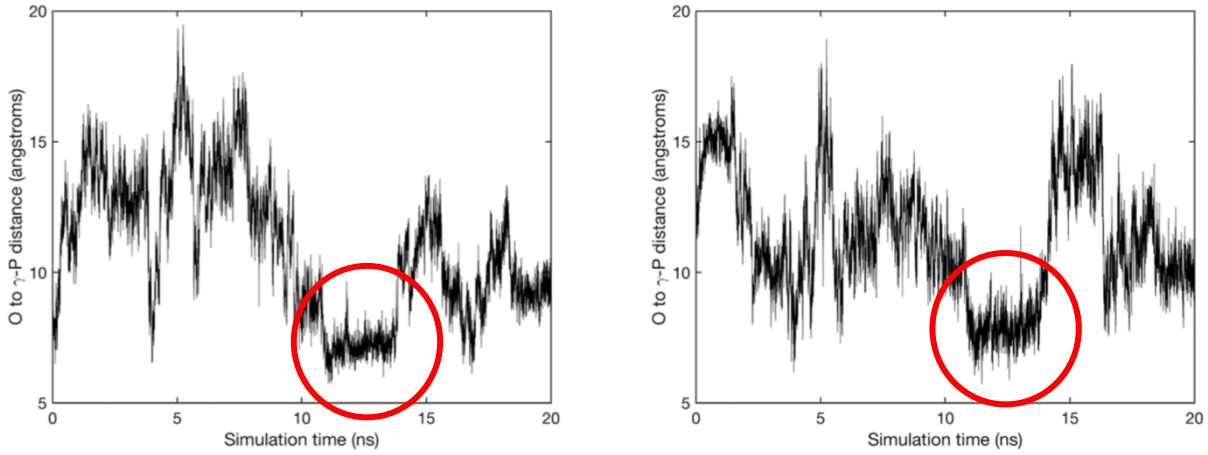

Figure 6. Time traces of the distances between the hydroxyl oxygens of (left) Thr-4 and (right) Tyr-6 and the phosphorus atom of the γ-phosphate of ATP.

We are focusing on the region (**red circle**) where Thr-4 and Tyr-6 were within 10 angstroms of key catalytic residues, between 11-13 ns. There are **11 coordinate snapshots** within this region, taken every 0.2 ns from the trajectory file.

Initially, three snapshots were analyzed in calculating SAPT interaction energies: at 11 ns (frame 5500), at 12 ns (frame 6000), and at 13 ns (frame 6500).

For **each** snapshot, **six SAPT calculations** need to be performed in order to assess the strength of the interactions (ATP or the peptide with the six $MEK1_{1IR3}$ residues determined to be the most critical in the binding complex):

1. ATP:Glu144
2. ATP:Met146
3. ATP:Leu197
4. Peptide:Gln110
5. Peptide:Arg234
6. Peptide:Asn221

When running the initial 18 SAPT calculations, 5 failed: frame 5500 (all three peptide interactions) and frame 6500 (two peptide interactions). The remaining calculations ran successfully. The source of the error in the 5 failed jobs remains unclear.

California Polytechnic State University

Department of Chemistry and Biochemistry
Dr. McDonald Computational Research Laboratory

Interaction Energy Calculations | *Outline of Main Goals*:

1. Find and fix the error causing the jobs to fail
2. Re-run the 5 failed jobs until successful
3. Complete the 3-snapshot analysis report
4. Repeat the procedure with the remaining 8 coordinate snapshots obtained in the interaction region
5. Compile a larger data set that represents the entirety of the 11-13ns timeframe with the collected data.

Note: If the entire trajectory file can be found (still searching…) there will be even more snapshots to run calculations on.

Performing Interaction Energy Analysis on a Coordinate-Snapshot

Starting from the .pdb coordinate file extracted from the trajectory:

Part I: Extract coordinates (PDB files) of interest (8 files total)

1. 1IR3frame#_ATPcoord.pdb
2. 1IR3frame#_PEPcoord.pdb
3. 1IR3frame#_Glu144coord.pdb
4. 1IR3frame#_Met146coord.pdb
5. 1IR3frame#_Leu197coord.pdb
6. 1IR3frame#_Gln110coord.pdb
7. 1IR3frame#_Arg234coord.pdb
8. 1IR3frame#_Asn221coord.pdb

*Note: Files in **red** are of MEK_{1IR3} residues that interact with ATP. Files in **blue** are of MEK_{1IR3} residues that interact with the peptide.*

Example PDB File:

```
#MEK1 1IR3 Model g6 COMPLEX frame 5500 (11 ns) PDB Coordinates
ATOM      1  N   GLU   62      28.469  32.823  62.163  1.00  0.00           N
ATOM      2  HT1 GLU   62      29.168  32.954  62.923  1.00  0.00           H
ATOM      3  HT2 GLU   62      28.199  31.827  62.290  1.00  0.00           H
ATOM      4  HT3 GLU   62      27.707  33.531  62.180  1.00  0.00           H
ATOM      5  CA  GLU   62      29.101  32.985  60.813  1.00  0.00           C
ATOM      6  HA  GLU   62      29.985  33.589  60.959  1.00  0.00           H
ATOM      7  CB  GLU   62      29.860  31.661  60.367  1.00  0.00           C
```

Atom # | Atom type | Amino Acid type | Residue # | xyz coordinates

Note: Residue numbering does not start at 1

California Polytechnic State University

Department of Chemistry and Biochemistry
Dr. McDonald Computational Research Laboratory

Organization is incredibly important because files accumulate rapidly. Be sure to include a header **# comment line** for each file to describe what the file contains.

e.g. #MEK1 1IR3 Model g6 {*insert residue, small molecule, etc.*} frame {#} PDB Coordinates

Part II: Convert PDB files to .xyz files

This step essentially removes extraneous information that is not needed in SAPT calculations. All that is needed is each atom's chemical element and coordinates in 3D space.

Like almost everything in the computational world, there are many ways to do this. Below is only one way, others can also work.

Use **Open Babel** (free software that converts molecular modeling data files)

To install: https://openbabel.org/wiki/Category:Installation

Standard conversion in terminal:

```
>> obabel -ipdb {name}.pdb -oxyz -m
```

Helpful Tip: using the -m command allows babel to perform **batch operations**, thus allowing you to complete multiple conversions at once:

```
>> obabel -ipdb *.pdb -oxyz -m
```

The command above will convert every file ending with '.pdb' in the current directory into a '.xyz' file. *The resulting files should have the same names as the inputs but differ only in their extensions.*

Part III: Stitch together appropriate .xyz files to create the SAPT input files

Recall there are six interactions that need to be calculated, thus six input files need to be generated.

1. Frame#_ATP_Glu144.in
2. Frame#_ATP_Met146.in
3. Frame#_ATP_Leu197.in
4. Frame#_Peptide_Gln110.in
5. Frame#_Peptide_Arg234.in
6. Frame#_Peptide_Asn221.in

A simple script could stitch together files and thus generate the input files. GitHub may have helpful scripts, but exercise caution when testing the scripts as they may need modification before they are successful for the task at hand (i.e., make duplicate files to test the new script on). Below is an alternate working method.

California Polytechnic State University

Department of Chemistry and Biochemistry

Dr. McDonald Computational Research Laboratory

Brute-force method of creating input files:

1. Create a new text file in a text editing software.
2. Paste the two .xyz files together according to the example below. Two files include: (1) **ATP.xyz** or **Peptide.xyz** and (2) **MEK1_residue.xyz**.
3. Save the file with the extension '.in'.

4. Once all six SAPT input files are created, save them in the shared **OneDrive**.

Note: input files are very sensitive to extra spaces or unintentional blank lines, thus if there is an immediate error the first trouble-shooting tactic is to make sure you didn't hit enter after the last line or that there isn't an extra space somewhere.

California Polytechnic State University

Department of Chemistry and Biochemistry

Dr. McDonald Computational Research Laboratory

Part IV: Generate .slurm scripts for each input file/calculation

In order to run these calculations (or "jobs"), the input files have to be submitted on some computer center. The **'.slurm' script** stands for Simple Linux Utility for Resource Management which is a job scheduler that decides where resources go on the queue of submitted jobs. The Mercury supercomputer utilizes this open-source scheduler.

1. Create a new text file in a text editing software.
2. Set up the '.slurm' script according to the example below. The values that will differ for each script are boxed in **red**.
3. Save the file with the extension '.slurm'.

ExampleSAPT.slurm

~/Desktop/ExampleSAPT.slurm

```
1  #!/bin/tcsh
2  #SBATCH -p stdmem
3  #SBATCH --nodes=1
4  #SBATCH --ntasks-per-node=2
5  #SBATCH --mem-per-cpu=30GB
6  #SBATCH --export=ALL
7  #SBATCH -t 96:00:00
8  #SBATCH -J 1IR3_ATPGlu144_SAPT3000
9  #SBATCH -o 1IR3_ATPGlu144_SAPT3000.o%j
10
11 source /usr/share/Modules/init/tcsh
12 set echo
13 hostname
14
15 module load psi4conda/1.1
16 setenv PSI_SCRATCH /scratch/$GROUP/$USER
17
18 psi4 -i 1IR3_ATPGlu144_SAPT3000.in -o 1IR3_ATPGlu144_SAPT3000.out -n 2
```

4. Once all six '.slurm' scripts are created, make a new folder that contains all six input files and all six '.slurm' scripts.

Part V: Move files to the supercomputer and run jobs

1. First copy the folder containing the files into marcy:
   >> scp -r folder_name username@marcy.furman.edu:
   Note: if only copying a single file, omit '-r'
2. Log into marcy on another terminal:
   >> ssh username@marcy.furman.edu
   Enter password when prompted
3. Ensure the folder was copied into marcy:
   >> ls
   This command will list everything in marcy. You should see the folder_name.
4. Copy the folder to skylight from marcy:
   >> scp -r folder_name skylight:

5. Navigate to skylight:
   >> ssh skylight
6. Ensure the folder was copied into skylight using 'ls'
7. Navigate into the folder:
   >> cd folder_name
8. Run job in skylight:
   >> sbatch ExampleSAPT.slurm
   All six '.slurm' files need to be submitted!
9. Check the status of running or pending jobs:
   >> squeue -u username

One the jobs are complete; output files will appear in the folder on skylight. You can view them using the VI editor prior to moving them to your computer.

Part VI: Move completed output data to your computer

1. Copy finished files/folder back to marcy:
   >> scp -r folder_name username@marcy.furman.edu:
2. Open a new terminal and navigate to the directory you want to copy the folder into (e.g., desktop).
3. Copy folder from marcy to computer in the new terminal window:
   >> scp -r username@marcy.furman.edu:folder_name .

If there is already a folder of the same name in that directory, this process may override that folder. This is OK, the new folder contains the input files and '.slurm' scripts as well as the output files.

California Polytechnic State University

Department of Chemistry and Biochemistry

Dr. McDonald Computational Research Laboratory

Part VII: Accessing the SAPT data from the output files

Take a minute to step back and look at the big picture. There should be six output files, one for each of the interactions. These calculations are all part of one coordinate-snapshot (i.e., one frame). The calculations represent the six different interaction energies of the MEK-ATP-peptide complex at a frozen moment in time within the simulation.

Open an output file. The results will look like the example below:

SAPT Results

Electrostatics	−9.89783305 [mEh]	−6.21098427 [kcal/mol]	−25.98676067 [kJ/mol]
Elst10,r	−9.89783305 [mEh]	−6.21098427 [kcal/mol]	−25.98676067 [kJ/mol]
Exchange	11.13686953 [mEh]	6.98849143 [kcal/mol]	29.23985095 [kJ/mol]
Exch10	11.13686953 [mEh]	6.98849143 [kcal/mol]	29.23985095 [kJ/mol]
Exch10(S^2)	11.05764014 [mEh]	6.93877424 [kcal/mol]	29.03183420 [kJ/mol]
Induction	−5.71222229 [mEh]	−3.58447375 [kcal/mol]	−14.99743963 [kJ/mol]
Ind20,r	−5.83762556 [mEh]	−3.66316549 [kcal/mol]	−15.32668590 [kJ/mol]
Exch-Ind20,r	1.94279767 [mEh]	1.21912400 [kcal/mol]	5.10081529 [kJ/mol]
delta HF,r (2)	−1.81739441 [mEh]	−1.14043226 [kcal/mol]	−4.77156902 [kJ/mol]
Dispersion	−4.10497814 [mEh]	−2.57591278 [kcal/mol]	−10.77762011 [kJ/mol]
Disp20	−4.76039721 [mEh]	−2.98719447 [kcal/mol]	−12.49842286 [kJ/mol]
Exch-Disp20	0.65541906 [mEh]	0.41128169 [kcal/mol]	1.72080275 [kJ/mol]
Disp20 (SS)	−2.38019860 [mEh]	−1.49359724 [kcal/mol]	−6.24921143 [kJ/mol]
Disp20 (OS)	−2.38019860 [mEh]	−1.49359724 [kcal/mol]	−6.24921143 [kJ/mol]
Exch-Disp20 (SS)	0.36161838 [mEh]	0.22691897 [kcal/mol]	0.94942905 [kJ/mol]
Exch-Disp20 (OS)	0.29380068 [mEh]	0.18436272 [kcal/mol]	0.77137370 [kJ/mol]
Total HF	−4.47318581 [mEh]	−2.80696659 [kcal/mol]	−11.74434935 [kJ/mol]
Total SAPT0	−8.57816396 [mEh]	−5.38287937 [kcal/mol]	−22.52196946 [kJ/mol]
Special recipe for scaled SAPT0 (see Manual):			
Electrostatics sSAPT0	−9.89783305 [mEh]	−6.21098427 [kcal/mol]	−25.98676067 [kJ/mol]
Exchange sSAPT0	11.13686953 [mEh]	6.98849143 [kcal/mol]	29.23985095 [kJ/mol]
Induction sSAPT0	−5.67016118 [mEh]	−3.55808001 [kcal/mol]	−14.88700818 [kJ/mol]
Dispersion sSAPT0	−4.09078848 [mEh]	−2.56700863 [kcal/mol]	−10.74036514 [kJ/mol]
Total sSAPT0	−8.52191318 [mEh]	−5.34758148 [kcal/mol]	−22.37428305 [kJ/mol]

The interaction energy is broken down according to the categories listed on the left.

Record the relevant information in a well-organized table. Save these consolidated results in the shared **OneDrive**.