

# BISC/QBIO 481, Fall 2020

## Topics

Third Assignment for 1<sup>st</sup> Section of the Course (Rohs) **Total 60+30 pts.**

**Due:** Thursday, October 15, 2020 11:59pm PST

**Submission Instructions:** All answers to the questions must be submitted as a typed PDF file via blackboard before the due date. Any mathematical equations must be correctly formatted, or you will receive a penalty in points.

### Question 1: High-throughput methods

- A. Briefly describe the *in vitro* experiments SELEX-seq and PBM and the *in vivo* experiment ChIP-seq. **5pts**
- B. Compare and discuss the advantage and disadvantage of these methods and what sort of biological questions they may be relevant to. **5pts**

### Question 2: Protein-DNA recognition

- A. Briefly explain the physical mechanisms relevant for (a) base readout and (b) shape readout of DNA sequence by DNA binding proteins. **4pts**
- B. Download and install the PyMol plugin PDIviz, which is designed for visualizing protein-DNA interfaces. PDIviz can be found here: <http://melolab.org/mbi/pdiviz> General instructions for installing PyMol plugins can be found here: [https://pymolwiki.org/index.php/Plugins#Installing\\_Plugins](https://pymolwiki.org/index.php/Plugins#Installing_Plugins) Using the plugin, visualize the structures (a) 1T2K and (b) 1KX5. Determine which regions of the DNA are contacted by the protein and explain any differences that you observe. Include the structure visualizations in your report. **5pts**
- C. Access the Proteopedia pages (a) [http://www.proteopedia.org/w/Hox-DNA\\_Recognition](http://www.proteopedia.org/w/Hox-DNA_Recognition) and (b) [http://www.proteopedia.org/w/P53-DNA\\_Recognition](http://www.proteopedia.org/w/P53-DNA_Recognition). Briefly explain for (a) Hox proteins and (b) p53 which structural features of the respective DNA binding sites are important for achieving binding specificity. **6pts.**

For any questions related to this homework, please post on the discussion boards located on blackboard (<https://blackboard.usc.edu>).

### Question 3: Molecular Dynamics

In this question we're going to implement a teeny tiny molecular dynamics simulation for a single particle in the potential  $V(x) = x^2$ . Molecular Dynamics is based on solving Newton's equations of motion which relates the force on a particle to its acceleration:

$$F(x) = ma$$

This equation says that the force acting on a particle (which is a function of its position  $x$ ) is equal to the mass of the particle times its acceleration. Remember that acceleration is defined as:

$$a = \frac{d^2x}{dt^2}$$

Therefore, we can write:

$$F(x) = m \frac{d^2x}{dt^2}$$

Our goal is, given a force, to find  $x$  as a function of time:  $x = x(t)$ . To solve this equation numerically (i.e. with the aid of a computer), we can use what is known as Verlet integration. A full discussion of this is beyond the scope of this course, but just know that Verlet integration is a way to accurately solve Newton's equation numerically (I encourage you to learn more on your own).

First, we have to discretize our position and time variables. This means that rather than treating them as continuous variables, we only consider them at particular fixed intervals. For example, we can have that  $t = 0.0, 0.2, 0.4, 0.6, \dots$  (which corresponds to a time step  $\Delta t = 0.2$ ) and we would then want to figure out what the corresponding values of  $x$  are at each time step, based on Newton's equation.

Verlet integration tells us that, given a time step  $\Delta t$ , (which means how far apart our time intervals are spaced), and a function which defines a force  $F$ , that

$$x_i = 2x_{i-1} - x_{i-2} + \frac{F(x_{i-1})}{m} \Delta t^2 \quad (1)$$

Here,  $x_i$  simply means the position of the particle at the  $i^{\text{th}}$  time step. Please answer the following questions and include your answers in your report.

For any questions related to this homework, please post on the discussion boards located on blackboard (<https://blackboard.usc.edu>).

- A.** Using the potential  $V(x) = x^2$ , what is the corresponding force? Be sure you get the sign correct, or the later part of this question will not work! **2pts**
- B.** Let's do a quick practice run with Verlet integration. Using a constant force  $F = -9.8$ , a mass  $m = 1.0$ , and a time step  $\Delta t = 0.1$  complete the following table. Use equation (1) to update the position ( $x_i$ ) at each time step and fill out the table below. Include the completed table in your report. **6pts**

$i$	$x_i$	$x_{i-1}$	$x_{i-2}$	$F$	$t_i$
2		9.951	10	-9.8	0
3				-9.8	
4				-9.8	
5				-9.8	

- C.** Once you've figured out how to do part **B**, it's trivial to pop in a different force, time step and initial conditions and solve Newton's equation for any system, including an entire protein structure! Of course, how good we can do depends on how accurate our *force field* is in the case of a protein, but let's do a very simple example using a single particle in one dimension.

Create a similar table as in part **B** (or arrays/vectors if you're using something like Python or R), but this time use the force you determined in part **A** (which is NOT constant). I have provided initial conditions below. Compute  $x_i$  and  $t_i$  for about 400 time steps, and then plot  $x_i$  vs  $t_i$  ( $t_i$  on the horizontal axis). Use  $m = 1$  and  $\Delta t = 0.01$ . What kind of motion is the particle undergoing? Include your plot with proper labels. Do not include a table in your report. **12pts**

$i$	$x_i$	$x_{i-1}$	$x_{i-2}$	$F$	$t_i$
2		0.002	0	-0.004	0.02
3					
...					
400					

#### Question 4: Structure Determination

For any questions related to this homework, please post on the discussion boards located on blackboard (<https://blackboard.usc.edu>).

Let's look at some X-ray crystallography data. I have provided three solved protein structures, a6.pdb, 06.pdb and 08.pdb and corresponding \*.dsn6 files that contain electron densities for these proteins.

- A. For each PDB file, open it in PyMol and then also open the corresponding \*.dsn6 file. You'll be presented with a dialog window when you load the \*.dsn6 file, make sure to check the isomesh box and leave everything else as-is.



Change the visualization type for the protein structure from “cartoon” to “lines”. Then, using a PyMol selection, select and zoom in on the residue ASN 132 for each structure, showing a clear view of the side chain and electron density. Include these visualizations in your report, making sure to label each one appropriately. Each of these structures was determined from a different electron density resolution. Based on what you observe of the electron density contours, order the structures for highest (best) resolution to lowest (worst) resolution. **6pts**

- B. You may notice that the density contour on the poorer resolution structures could leave some room for interpretation about the orientation of the ASN side-chain. How might a structural biologist resolve this issue? **4pts**
- C. For the a6 structure/electron density, have a look at the residue GLN 126. Zoom in on it and provide a visualization in your report. You may notice that there are two side chains for this residue! What is going on here, and what might be the cause of this? **5pts**

**Question 5: \*\*BONUS\*\* 30pts**

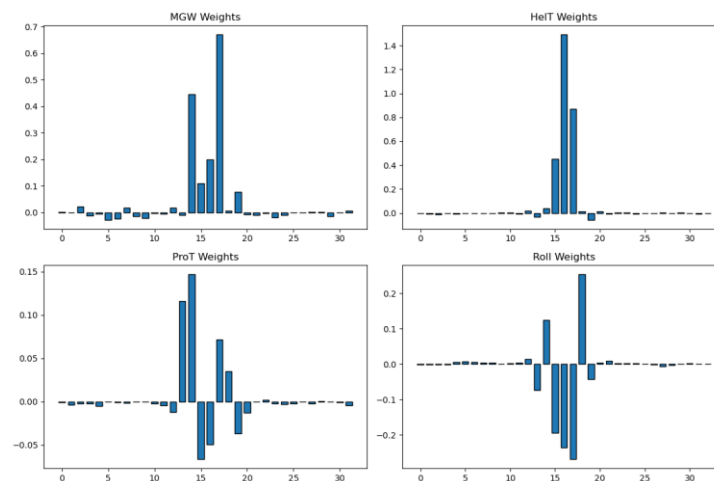
For any questions related to this homework, please post on the discussion boards located on blackboard (<https://blackboard.usc.edu>).

In order to do this question, please install the following python modules in conda if not already installed:

```
conda install sklearn
conda install jupyter
```

This question will use the file `hw3.ipynb` which is a python notebook. To open it, run `jupyter notebook hw3.ipynb` from a terminal window. This is an interactive python session, if you're unfamiliar with how these work take a quick look here: <https://towardsdatascience.com/a-beginners-tutorial-to-jupyter-notebooks-1b2f8705888a>

Follow along with the notebook and write the code as you go! This question is based on this paper: [https://rohslab.usc.edu/Papers/2015\\_Pnas\\_Zhou.pdf](https://rohslab.usc.edu/Papers/2015_Pnas_Zhou.pdf) so you may want to have a look at that for some more background. Your goal is to produce a plot similar to the one below (you'll understand what this means when you go through the notebook). Note that the values here have been intentionally changed, you won't get the same results.



Make sure to turn in your notebook with your report.

For any questions related to this homework, please post on the discussion boards located on blackboard (<https://blackboard.usc.edu>).