Session 4 Recitation

1) Log into the HPCC (and a development node) and go to your home directory. In the directory that you created in Session 2 (CMSE890Sec303), create a subdirectory Session4 (no spaces).
2) Go to the Protein Databank (https://www.rcsb.org) and search for 1ASY. This should take you to the crystal structure for Aspartyl-tRNA Synthetase complexed with tRNA-Asp. Copy the pdb file to the HPCC. Use more/less to look through this file. Column 1 is the atom type; column 2 is atom number; column 3 is atom type; column 4 is residue name; column 5 is chain ID; column 6 is residue number or ID; column 7 is x coordinate; column 8 is y coordinate; column 9 is z coordinate; column 11 is beta factor or amount of movement during crystallization; column 12 is a basic atom type.
3) Using sed/grep/awk, copy only the atom information to a new file. (Hint: look for more than just ATOM.)
4) Now use sed/grep/awk to copy only the backbone atom info to a new file. Backbone atoms for nucleic acids are “P” and for proteins are “CA” (3rd column).
5) Look at the residue numbers in the 6th column in the backbone file. What do you notice?
6) Look at the residue names in the 4th column of the nucleic acids. Are there any nucleotides that aren’t ATCG? (Hint: look for HETATM)
7) Look at the chain IDs in the 5th column. How many different chains are there and what do they correspond to?
8) Search for SEQRES in the original file. If you wanted to get the sequences of each chain from these lines in the file and reformat them so all the residues are on one line without spaces, what would you do? Don’t spend too much time on this.
9) Turn your commands from #8 into a bash script that you could run on any pdb file.