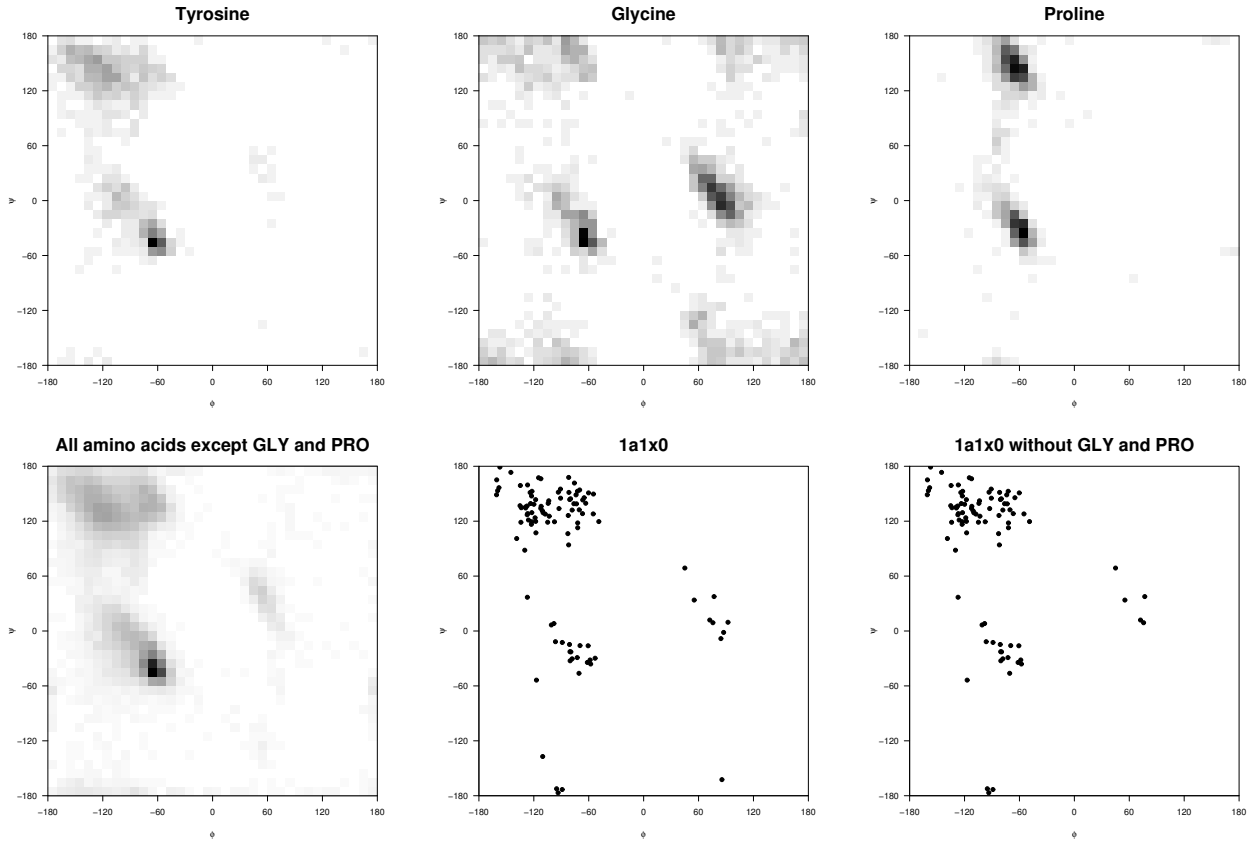
Statistical Topics in Genetics and Genomics (140.668)
Winter 2004

Homework Assignment #3
(Due Wednesday, December 22, 2004)

Please hand in a hard copy of your `R` code, and send an electronic version of it to me (ingo@jhu.edu).

Get `http://www.biostat.jhsph.edu/~iruczins/pdbs.tar.gz` which contains files with the atomic coordinates of 100 proteins from the Protein Data Bank. Write an `R` function to generate the ϕ and ψ angles for the Ramachandran plot. Your function should have the option to produce a scatterplot for the ϕ and ψ angles as well as an image for "binning" the angles, with an option for the width of the bins. You should also be able to specify or exclude certain amino acids from the plots. In particular, generate the following plots:



You should pay attention to the following details:

- To calculate ϕ and ψ around a C_α atom in an amino acid, we need the carbon from the previous and the nitrogen from the following amino acid. Therefore, for a protein with n amino acids, the ϕ and ψ angles are only defined for residues $2, \ldots, n-1$.
- Not every amino acid in a protein has coordinates in its PDB file. For example, if something went wrong in the crystallization procedure, there might be a chunk of the protein that was not solved. Hence ϕ and ψ angles are also not defined for amino acids just before the start or just after the end of such a missing chunk.