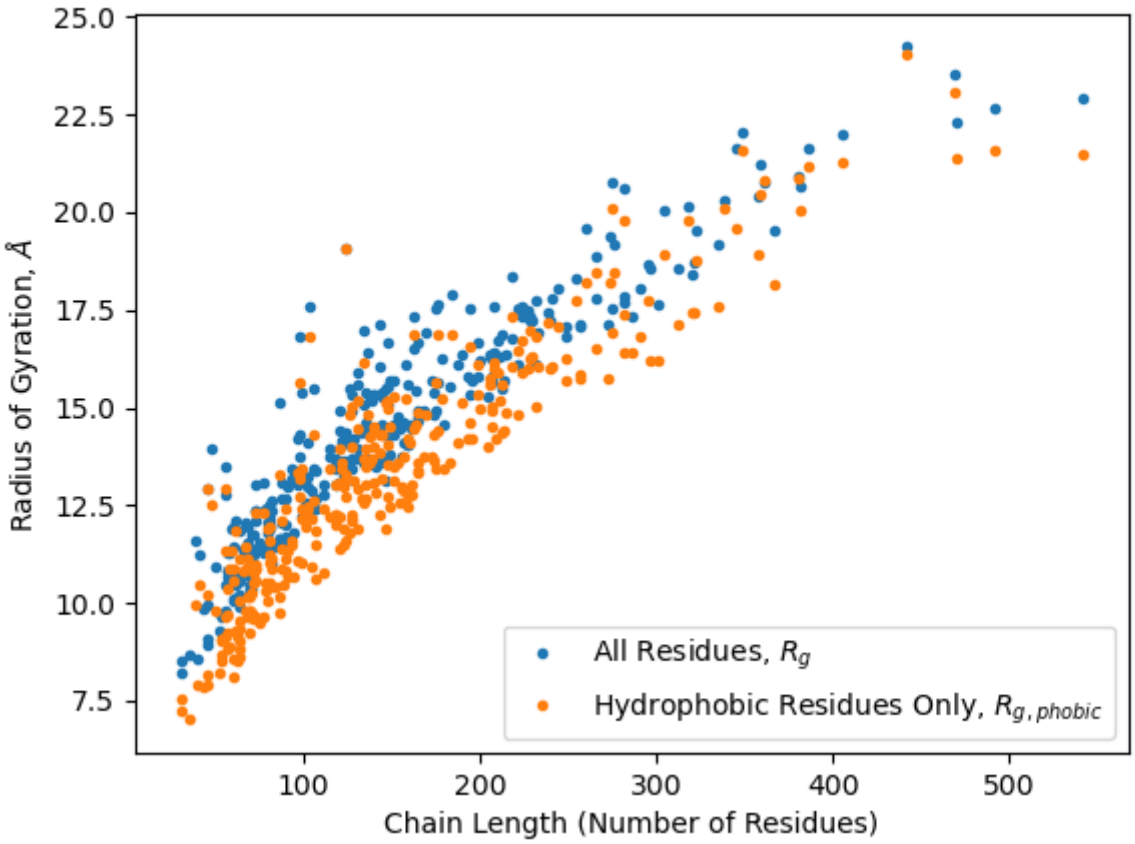
pdb_analysis

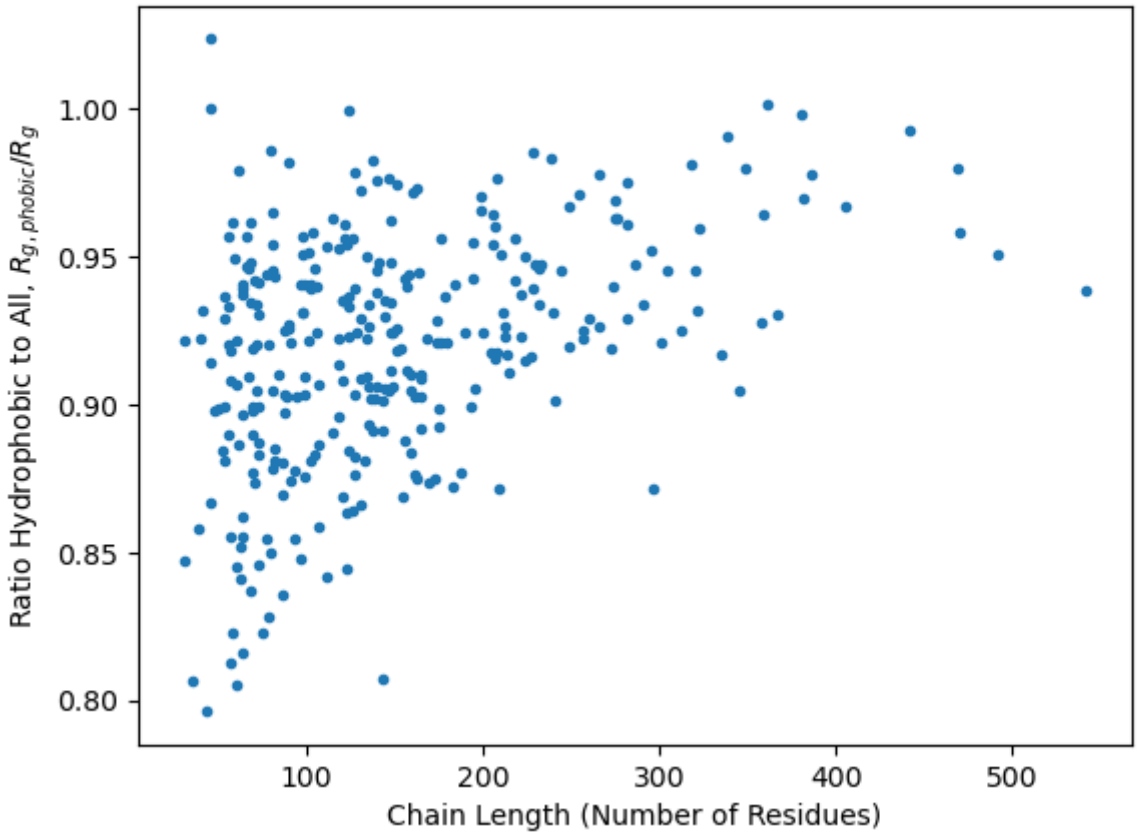
Jackson Sheppard
CH E 210D, Exercise 1
10/05/22\

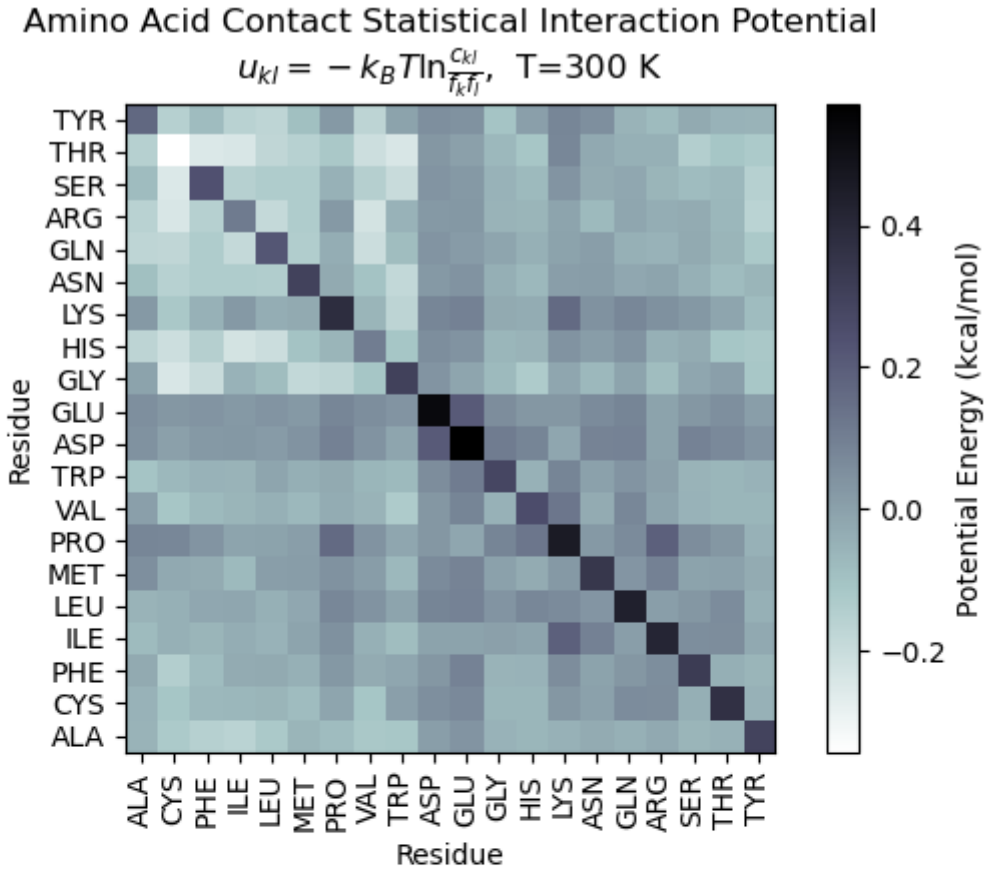
Here we present an analysis of 309 protein sequences, computing structural quantities such as the radius of gyration and statistical interaction potentials. The input sequences are stored in The Protein Data Bank `*.pdb` format and reside in the `proteins/` directory at the root of this repository. Each file consists of a protein X-ray crystal structure and thus includes atomic X, Y, Z coordinates of each amino acid (residue) comprising the protein structure. In this analysis, we read the sequences of our 309 proteins and compute the radius of gyration, R_g, defined as follows:

$$ R_g^2 = \frac{1}{N}\sum_{i=1}^N |\vec{r}*i - |*langle |*vec{r} |*rangle|^2, |*quad |*langle |*vec{r} |*rangle = |*frac{1}{N}|*sum{i=1}^N \vec{r}_i $$

and hydrophobic residues only, $R_{g,phobic}$. We then plot both of these quantities along with their ratio against the total number of residues in the structure. Finally, we compute and visualize the "statistical" interation potential for the 20 amino acid types by considering amino acid contacts present in this data set.




1.00
0.95
0.90
0.85
0.80
Ratio Hydrophobic to All, $R_{g,phobic}/R_g$
100
200
300
400
500
Chain Length (Number of Residues)



Lowest Energy Interactions (RES-RES)	Lowest Energies (kcal/mol)	Highest Energy Interactions (RES-RES)	Highest Energies (kcal/mol)
CYS-CYS	-0.343109	GLU-GLU	0.571927
CYS-PHE	-0.247388	ASP-ASP	0.530072
CYS-ILE	-0.242653	LYS-LYS	0.458285
CYS-TRP	-0.240138	GLN-GLN	0.438962
ILE-VAL	-0.225613	ARG-ARG	0.417976