

Homework 13
Polymer Physics 2024
Due Tuesday April 16 at noon
(Please list the contributors to the HW at the top of the document)

Tamasi MJ, Patel RA, Borca CH, Kosuri S, Mugnier H, Upadhya R, Murthy NS, Webb MA, Gormley AJ *Machine Learning on a Robotic Platform for the Design of Polymer-Protein Hybrids* Adv. Mat. **34** 2201809 (2022) describe automated random copolymer synthesis for polymers that stabilize proteins. The automated synthesis is coupled with machine learning to find new polymers that can preserve and/or enhance protein biofunction after thermal treatment. Figure 1 shows the three study proteins and the eight monomer types that are considered, as well as the flow chart for synthesis and machine learning. Tamasi M, Kosuri S, DiStefano J, Chapman R, Gormley AJ *Automation of Controlled/Living Radical Polymerization* Adv. Intel. Sys. **2** 1900126 (2020) describes the automatic polymer synthesis protocol and instrumentation.

- a) Define the following terms: Gaussian process regression; Bayesian optimization; PET-RAFT polymerization; PPH; REA; active learning; Mann–Whitney U test; Shapley additive explanations; power transform; Box-Cox transform; Yeo-Johnson transform; squared exponential kernel basis function; Bayesian optimization; Tree-structured Parzen Estimator; kriging; density-based spatial clustering of applications with noise (DBSCAN); k-Means clustering; one-shot learning (computer vision); latin hypercube sampling; Bayesian inference; active learning; mined data; collected data; seed data; design of experiments (DOE); well-plate format; supervised learning; unsupervised learning; support vector machine (SVM); ridge regression; lasso regression; classification and regression trees (CARTs); Random Forests; Gradient Boosted Trees; rectified linear aggregation (ReLU).
- b) Explain the experiment done by Tamasi and the protocol shown in Figure 1.
- c) Figure 3 gives an analysis of the ML process. Explain each of the parts of figure 3.
- d) Meyer TA, Ramirez C, Tamasi MJ, Gormley AJ *A User's Guide to Machine Learning for Polymeric Biomaterials* ACS Polym. **3** 141-157 (2023) is a paper intended as a tutorial for AI/ML. Run the code in Google Colab from the Gormley link and describe each of the steps of the AI/ML process. <http://www.gormleylab.com/MLcolab>
- e) Rerun the analysis changing the number of decision trees, *n-estimators*, to 50 evenly spaced values in the range 10-3000. Compare the SHAP analysis with that from part d.