**Supporting Information (SI)**

# Developing Efficient Deep Learning Model for Predicting Copolymer Properties

Himanshu, Kaushik Chakraborty and Tarak K Patra*
Department of Chemical Engineering and Center for Atomistic Modeling and Materials Design, Indian Institute of Technology Madras, Chennai TN 600036, India

## I. Model Polymer and Data Generation

The data sets are generated using MD simulation-based genetic algorithms. Starting from a random sequence, genetic operators – mutation and crossover are applied to generate new polymer sequences. Each time the GA algorithm produces a new polymer sequence, MD simulation is conducted for measuring its property.[1,2] A very generic phenomenological polymer model within the Kremer-Grest (KG)[3] framework is used in this study. In the model, two adjacent coarse-grained monomers of a copolymer are connected by the Finitely Extensible Nonlinear Elastic (FENE) potential of the form $E = -\frac{1}{2}KR_0^2\, ln\left[1 - \left(\frac{r}{R_0}\right)^2\right]$, where $K = 30\epsilon/\sigma^2$ and $R_0 = 1.5\sigma$ for bond length $r \leq R_0$ and $E = \infty$ for $r > R_0$. The pair interaction between any two monomers is modelled by the well-known Lennard-Jones (LJ) potential of the form $V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$. The $\epsilon$ is the unit of pair interaction energy. The size of all the monomers is σ. The LJ interaction is truncated and shifted to zero at a cut-off distance $r_c = 2.5\sigma$ to represent attractive interaction among the monomers. Here, the size of a monomer (σ) is constant and same for both the systems. The two moieties have different interaction strength (ε) to represent their different chemical nature in the Rg data set. For the Rg data set, $\epsilon_{AA} = \epsilon$ and $\epsilon_{BB} = 0.2\epsilon$ and $\epsilon_{AB} = 0.2\epsilon$. For the surface tension data set, $\epsilon_{AA} = \epsilon$, $\epsilon_{BB} = \epsilon$ are considered. The AB interaction is repulsive which is modeled by considering $r_c = 1.12\sigma$ and $\epsilon_{AB} = \epsilon$. The KG polymer model is a standard model for studying generic polymer properties using molecular dynamics simulations. It is very popular for MD study because of its simplicity and computational efficiency, rather than its ability to represent specific polymers species and conditions. Moreover, this model can be mapped to any commodity polymer.[4]

## II. Loss Functions of DNN Models

Few representative loss functions during the layer-by-layer growth of a DNN model is shown in Figure S1. The Figure S1 corresponds to the surface tension model. It clearly suggests that the loss function is systematically decreasing during the training of the model without any significant overfitting.

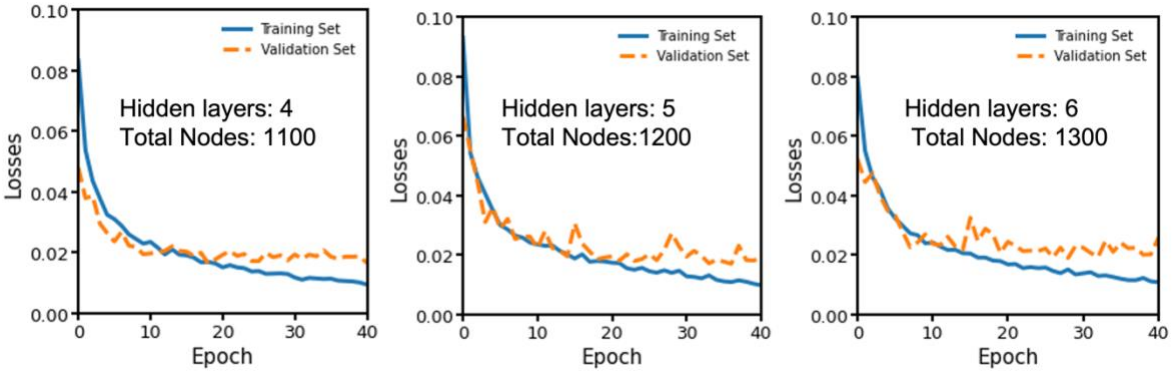


*Figure S1: Few representative Loss function are shown. The number of layers and nodes in a model is as mentioned in the panels.*

**References:**


(1) Bale, A. A.; Gautham, S. M. B.; Patra, T. K. Sequence-Defined Pareto Frontier of a Copolymer Structure. *J. Polym. Sci.* **2022**, *60* (14), 2100–2113. https://doi.org/10.1002/pol.20220088.

(2) Meenakshisundaram, V.; Hung, J.-H.; Patra, T. K.; Simmons, D. S. Designing Sequence-Specific Copolymer Compatibilizers Using a Molecular-Dynamics-Simulation-Based Genetic Algorithm. *Macromolecules* **2017**, *50* (3), 1155–1166. https://doi.org/10.1021/acs.macromol.6b01747.

(3) Kremer, K.; Grest, G. S. Molecular Dynamics (MD) Simulations for Polymers. *J. Phys. Condens. Matter* **1990**, *2* (S), SA295–SA298.

(4) Everaers, R.; Karimi-Varzaneh, H. A.; Fleck, F.; Hojdis, N.; Svaneborg, C. Kremer–Grest Models for Commodity Polymer Melts: Linking Theory, Experiment, and Simulation at the Kuhn Scale. *Macromolecules* **2020**, *53* (6), 1901–1916. https://doi.org/10.1021/acs.macromol.9b02428.