Composition-Independent Prediction of Self-Assembled Structures in Multicomponent Systems Using Machine Learning

Yuuki Ishiwatari [a], Takahiro Yokoyama [a], Tomoya Kojima [b], Taisuke Banno [b] and Noriyoshi Arai [a]

[a] *Department of Mechanical Engineering, Keio University*
[b] *Department of Applied Chemistry, Keio University*

ishiwatari-527.uki@keio.jp

Predicting self-assembled structures from molecular structures is a central challenge in the design of functional soft materials, especially in multicomponent systems, where nonlinear mixing effects make rational design difficult. In this study, we developed a machine-learning framework to predict self-assembled structures in multicomponent surfactant systems directly from molecular information.

To generate training data, dissipative particle dynamics (DPD) simulations were performed using amphiphilic molecular models, and surfactant systems containing one to five components were constructed. The critical packing parameter (CPP)[1] was used as a descriptor of the resulting self-assembled structures because it showed a clear correspondence with the observed morphologies.

Several model architectures were compared to identify a suitable approach for multicomponent prediction. When the training and test data contained the same number of components, models that treated each component separately showed higher accuracy than those that handled all components together. Among them, graph-based models, which represent molecules as networks of atoms and bonds, performed particularly well, indicating that such representations are effective for capturing structural features relevant to self-assembly.

We then evaluated generalization across different numbers of components. Models trained on one- to four-component systems were successfully extended to five-component systems. Among the tested approaches, models that use graph-based learning both to extract structural features of each molecule and to explicitly represent interactions between components showed the strongest generalization ability. These models maintained high accuracy for five-component systems even when they contained molecular species not included in the training data, and achieved an R^2 of approximately 0.8 even when trained only on two-component systems. These results suggest that interaction patterns learned from simple mixtures can be transferred to more complex multicomponent systems.

Overall, this study demonstrates that machine learning combined with coarse-grained simulation enables accurate prediction of self-assembled structures in multicomponent systems with substantial generalization across composition space. The proposed framework provides a promising strategy for the data-efficient design of complex soft materials.

Keywords:
Machine learning, Self-assembly, Molecular simulation, Multicomponent systems

References:

[1] N. Israelachvili, *Intermolecular and Surface Forces* , Academic Press, 3rd edn, 2011.