

## 自動分子シミュレーションによる高分子物性データプラットフォームの産学共創

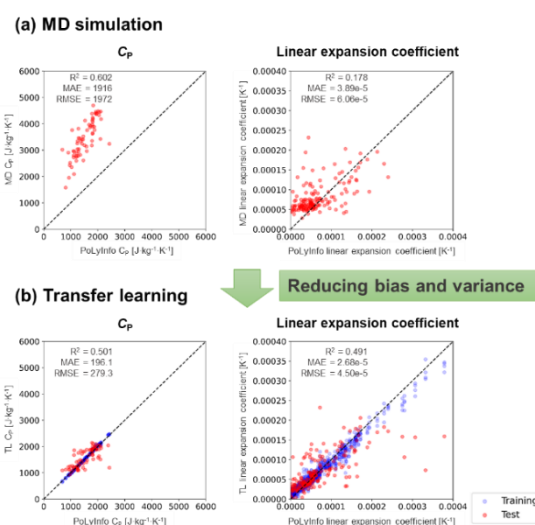
(統数研<sup>1</sup>) ○林 慶浩<sup>1</sup>

Co-creation of polymer properties data platform by automated molecular simulation through industry-academia collaboration (<sup>1</sup>*The Institute of Statistical Mathematics, Research Organization of Information and Systems*) ○Yoshihiro Hayashi<sup>1</sup>

In recent years, data-driven materials design techniques, known as materials informatics, have been rapidly introduced into the field of materials research. Needless to say, the most important resource of data-driven research is data. However, the amount of data on polymeric materials is overwhelmingly small, and at present there is practically no polymer property database that contributes to data-driven research. In order to develop a database of polymer properties based on molecular simulations, we have developed RadonPy,<sup>1,2</sup> a Python library that supports the automation of polymer property calculations based on molecular dynamics (MD) simulations. RadonPy fully automates a series of processes required to perform MD calculations, including initial structure generation, charge and force field assignment, equilibration and nonequilibrium MD calculations, and calculation of physical properties. The vast computational resources of Fugaku and the industry-academia collaboration of 5 universities and 30 companies, led by the Institute of Statistical Mathematics, are producing data with the aim of constructing a database that includes more than 10<sup>5</sup> polymer skeletons.

Equilibration calculations for about 77,000 polymers were completed by the end of 2023 using RadonPy, of which thermal conductivity calculations for about 71,000 polymers and T<sub>g</sub> calculations for about 40,000 polymers were completed. Comparison of the calculated and experimental values showed that there were large systematic biases and variations in specific heat and coefficient of linear expansion. The gap between the calculated and experimental values can be corrected by a machine learning method called transition learning, as shown in Fig. 1. To quantify the value of the database from simulations, we observed a scaling law for the prediction accuracy of transition learning with respect to

the number of simulated data. This scaling law has been shown to theoretically follow a power law, and we observed that the prediction accuracy improves in accordance with the power law for the RadonPy simulation data, as shown in Fig. 2. This scaling curve enables us to estimate the number of data required for the database and the achievable prediction performance, which is an important guideline for constructing the database.

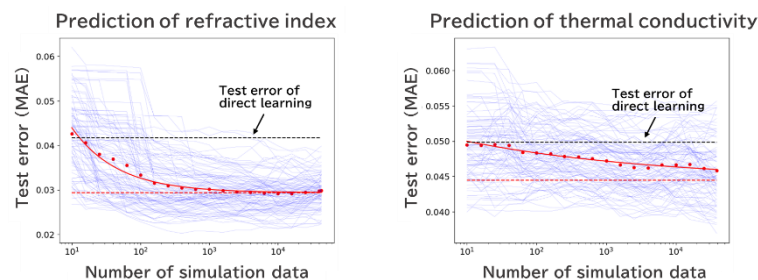


**Figure 1.** Calibration of MD calculated values by the transfer learning.

In addition, Bayesian optimization and automatic simulation using RadonPy are combined to design polymers with desired properties. An example of molecular design of polymers with both a high refractive index and a high

Abbe number, which are required for optical polymers, will be presented.

**Keywords :** Materials Informatics; Polymer Informatics; Molecular Dynamics Simulation; Transfer Learning; Industry-Academia Collaboration



**Figure 2.** Scaling law of transfer learning using simulation data.

近年、マテリアルズインフォマティクスと呼ばれるデータ駆動型材料設計の技術が材料研究の分野に急速に導入されている。データ駆動型研究の源泉は、言うまでもなくデータである。しかしながら、高分子材料のデータ量は圧倒的に少なく、データ駆動型研究に資するレベルの高分子物性データベースは現時点において実質的に存在しない。そこで、分子シミュレーションによる高分子物性データベースを開発するために、分子動力学 (MD) シミュレーションによる高分子物性計算の自動化を支援する Python ライブラリである RadonPy を開発した。<sup>1,2)</sup> RadonPy は MD 計算の実行に必要な初期構造の生成、電荷・力場の割り当て、平衡化計算、非平衡 MD 計算、物性値算出の一連の工程を完全自動化する。そして、富岳の膨大な計算資源と、統計数理研究所が中心となり 5 大学・30 企業の産学連合体により、 $10^5$  個を超える高分子骨格を包含するデータベースの構築を目指しデータ生産を行っている。

RadonPy を用いて 2023 年末までに約 77,000 ポリマーの平衡化計算が完了し、そのうち約 71,000 ポリマーの熱伝導率計算と約 40,000 ポリマーの Tg 計算が完了した。また、計算値と実験値を比較すると、比熱や線膨張係数において大きな系統バイアスやばらつきが存在した。この計算値と実験値の間のギャップは、Fig. 1 に示すように転移学習とよばれる機械学習手法により補正できることが示された。計算値によるデータベースの価値を定量的に示すべく、計算データ数に対する転移学習の予測精度のスケールリング則を観測した。このスケールリング則は、理論的にはべき乗則に従うことが示されており、RadonPy の計算データにおいてもべき乗則に従い予測精度が向上することが観測された。このスケールリングカーブにより、必要なデータ数や到達可能な予測性能の見積もりが可能となり、データベースを構築するための重要な指針となる。

また、ベイズ最適化と RadonPy による自動シミュレーションを融合し、所望の特性を有する分子設計を行っている。当日は光学用高分子に要求される、高屈折率と高アッベ数を両立する高分子の分子設計の事例を紹介する。

1) Y. Hayashi, J. Shiomi, J. Morikawa, R. Yoshida, *npj Comput. Mater.* **2022**, 8, 222.

2) <https://github.com/RadonPy/RadonPy>