

jh230055

## 環境循環型社会の実現に向けた ポリマーインフォマティクスのデータ基盤構築

佐藤 正寛（東京大学大学院工学系研究科）

本研究では、計算科学データをもとにしたポリマーインフォマティクスのデータ基盤を構築する。データベースの利活用により、環境循環型社会の実現に必要な高機能ポリマー材料の創出を促すことを目的とする。ポリマーの多階層性を考慮した網羅的な計算結果に対してデータベース化を行い、同時に、各スケールでのポリマー物理量を推算する AI モデルを実装する。また、データの共有と利活用を行うための基盤を整備することでポリマー材料設計を推進する。今年度は量子化学計算の遂行、分子動力学計算で得られるメゾ物理量と力場パラメータの関係の解析、外挿可能な物性予測モデルの構築と評価に注力した。

### 1. 共同研究に関する情報

#### (1) 共同利用・共同研究を実施している拠点名

東京大学 情報基盤センター

大阪大学 サイバーメディアセンター

mdx

- 十二 綱輝：分子動力学計算の実行・評価
- 片瀬 大祐：量子化学計算の実行・評価
- Ruan Haoou：分子動力学計算の実行・評価

### 2. 研究の目的と意義

#### 研究の目的

計算科学データをもとにしたポリマーインフォマティクスのデータ基盤を構築する。データベースの利活用により、環境循環型社会の実現に必要な高機能ポリマー材料の創出を促すことを目的とする。図 1 に基盤システムの概略図を示す。ポリマーの多階層性を考慮し、量子化学計算および分子動力学計算を行い、網羅的な計算結果に対してデータベース化を行う。また同時に、図 2 に示すように、各スケールでのポリマー物理量を推算する AI モデルを実装する。ポリマー情報と物性との関係を単純な深層学習でブラックボックス的に扱うのではなく、ポリマーの多階層性を考慮し、各スケールでの計算科学データを支配因子とする多段型の AI モデルを構築する。各スケールの物理量の間にある相関

#### (2) 課題分野

データ科学・データ利活用課題分野

#### (3) 共同研究分野(HPCI 資源利用課題のみ)

超大規模数値計算系応用分野

超大規模データ処理系応用分野

#### (4) 参加研究者の役割分担

- 佐藤正寛（代表者）：研究統括
- 熊田亜紀子：研究副統括
- 梅本貴弘：研究副統括
- 鈴木豊太郎：データ基盤構築支援
- 田浦健次朗：データ基盤構築支援
- 河村光晶：データ基盤構築支援
- 嶋川肇：コード開発・データ基盤構築
- 森 勇貴：コード開発・データ基盤構築
- Weihao Wang：コード開発・データ基盤構築

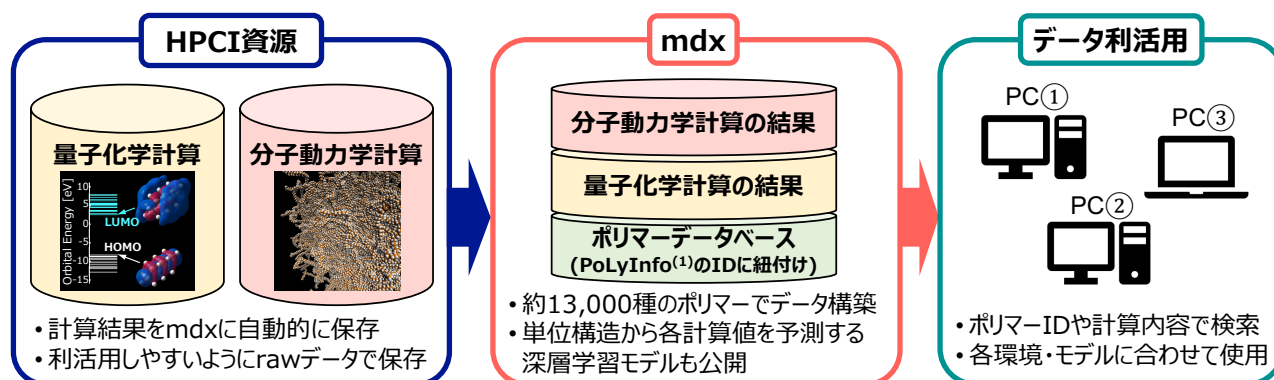


図 1 基盤システムの概略図

関係を明示的に抽出することで、ブラックボックス化することなく、“外挿可能な（説明可能な）”AI モデルを実現する。さらに、データの共有と利活用を行うための基盤を整備することでポリマー材料設計を推進する。

#### 研究の意義

本研究ではポリマーの計算科学データを活用し、ポリマー物性に対して“外挿可能な”AI モデルの構築を目指す。この AI モデルは、ポリマーの分子構造からマイクロ・メソスケールを經由してマクロ物性に至るまでの過程を定量的に評価できるものであり、ポリマー物性発現の学理構築にも貢献する。また、基盤データベースを活用した自身の研究事例を発信し、多分野のポリマー材料設計に本データベースの利活用を推進することで、最終的には広範なポリマー学理構築につながると考えられる。ポリマー材料は未来の環境循環型社会の実現に必要不可欠であり、インフォマティクス技術はその設計開発を加速度的に促進するポテンシャルを秘めている。その推進に最も重要なものはデータ利活用の基盤を構築することであり、本提案はその先駆けとしての役割を果たす。

- 計算科学の新たなデータベースを構築し、データ利活用を促す基盤が必要
  - ポリマーの多階層性を考慮した計算を網羅的に行うためには大規模な計算資源が必要
- そこで図 1 に示すような概略図にしたがい、公募型共同研究を通じてシステム構築を実施する。計算負荷が大きい量子化学計算および分子動力学計算は HPCI 資源を用いた大規模並列計算により実施する。データ利活用のための基盤には mdx を使用し、検索可能な計算結果のデータベース化を行う。

本提案は申請者が所属する電気電子分野だけでなく、材料科学分野と情報科学分野にまで影響を与える学際性を持つ。すでに関連テーマの検討事項に関して、東大情報理工学系研究科の先生方や化学・材料メーカーの연구원の方々と議論を実施している。後述する補足資料の研究課題のように、我々はデータベースの構築だけでなく、実際にデータを利活用したポリマー材料の創出を進めており、その過程で共同研究先の材料メーカーとともに新材料の合成・性能評価までを行う予定である。本提案では東大の情報理工学研究科と情報基盤センターの先生方と共同でデータベース構築を実施する予定であり、最終的

### 3. 当拠点の公募型共同研究として実施した意義

ポリマーインフォマティクスの基盤構築には以下の課題を解決する必要がある。

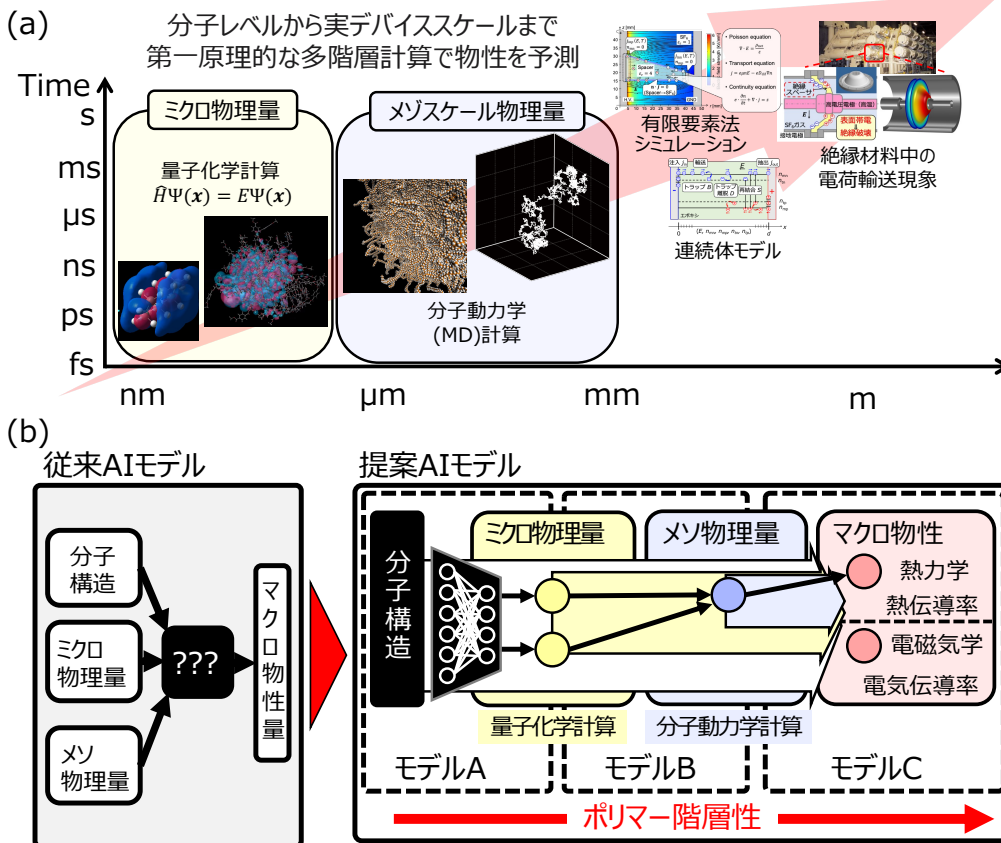


図 2 (a) ポリマー材料のマルチスケールモデリングの一例

(b) ポリマーの多階層性を考慮した提案 AI モデル

な発展として材料学と情報学の融合により AI モデルの検討から材料設計までを実施する学際的な共同研究に位置する。

4. 前年度までに得られた研究成果の概要

該当なし

5. 今年度の研究成果の詳細

A. ポリマー構造の取得

データベース化する約 13,000 種類のポリマー構造を大規模ポリマーデータベースである PoLyInfo (<https://polymer.nims.go.jp/>) から取得した。また、量子化学計算や分子動力学計算においてポリマーの繰り返し構造を操作しやすいように、モノマー構造の SMILES (構造の文字列表記) からオリゴマー構造の SMILES を生成するためのコードを作

成した。

B. 量子化学計算

A で構造を取得した約 13,000 種類のポリマーに対し、モノマー構造および原子数 50 前後のオリゴマー構造を用いた量子化学計算を実施した。実験で得られるポリマー物性を予測するために、説明変数となる種々のミクロ物理量を取得した。一連の量子化学計算は大阪大学サイバーメディアセンターの SQUID に搭載される Gaussian16 を用いて行った。各構造に対して、構造最適化、各種エネルギー、分極特性、電子密度分布、振動特性、溶媒効果の計算を実施した。また、1 万点を超える大量の構造に対して連鎖的な量子化学計算を遂行するため、並列計算および後処理を効率化するためのコードを作成した。個々の量子化学計算は計算時間が小さく

なるように、原子数が 50 程度以下の小さい構造を用い、理論レベルを軽めに設定して行った。各ジョブは同じ条件の計算を大量の構造に対して直列に実行していく仕様にした。また、計算負荷に合わせて 1 ノード内で最大 4 並列の計算が実行されるようにバックグラウンド処理を施して各ジョブを投下した。計算結果は raw データ、および物性値を抽出したテーブルデータとして mdx 上に保存した。

### C. 分子動力学計算

ポリマー物性の高精度予測に向けて、今年度は分子動力学計算で用いる力場パラメータと各物性との関係を検討した。分子の構造情報ではなく力場パラメータに着目することで、定量的かつ外挿的な物性予測、および所望物性を起点にした分子設計が可能になると考えた。今回対象にしたポリマー物性は熱伝導率およびイオン性電気伝導率である。まずは HPC 上で分子動力学計算によってポリマーの結晶構造およびアモルファス構造の熱伝導率およびイオン移動度を推定するシステムを構築した。分子動力学計算の全自動化ライブラリである RadonPy を活用し、モノマー構造の SMILES からオリゴマー構造の作成、原子電荷と力場の作成、収束判定を含めた平衡化計算、熱伝導率およびイオン移動度の計算までの直列実行を可能にした。

手始めに、結晶構造とアモルファス構造のポリエチレンを用いて、力場パラメータと熱伝導率との関係を検討した。その結果、各構造で熱伝導に寄与する力場パラメータおよび伝熱成分の違いが明らかになった。(業績リスト[4,7,10,11]) さらに、ポリマー種を 8 種類に拡張して、力場と熱伝導率、およびイオン移動度の関係を検討した。その中で得られた力場と物性の相関に着目し、熱伝導率とイオン移動度の間に存在するトレードオフの関係を明らかにした。また、力場パラメータを系統的に変化させた熱伝導率とイオン移

動度の計算により、トレードオフを克服し得るポリマー材料の分子設計指針を示した。この方法は第一原理に基づくアプローチであり、多様なポリマーの分子設計において大きな利点を持つと考えられる。(業績リスト [3,12])

### D. 機械学習

これまで報告されているポリマーの物性予測モデルは、学習データと同じ範囲の物性予測(内挿)を扱うものばかりであり、実用的なポリマー材料探索において不可欠である学習データ範囲外の物性予測(外挿)に対する検討が不十分である。そこで、ポリマーの多階層 AI モデル開発に先立ち、図 1(b)においてモデル B を省略した、少数の階層性を考慮した AI モデルを開発し、物性予測精度のベンチマークを取得した。その結果、提案モデルは従来モデルと比べると 2 桁程度小さいデータサイズでも同等の物性推定精度を有し、さらに従来手法では不可能であった外挿的物性予測が可能であることを実証した。(業績リスト[1,9])

この AI モデルを用いたポリマーの物性予測にも取り組んだ。モノマー構造を用いた量子化学計算の結果を元に、ポリマーの基本物性の実験値に対する外挿モデルの構築を進め、速報的に国際学会で報告を行った。(業績リスト[5]) さらに、分子動力学計算により 200 点程度まで蓄積した熱伝導率とイオン移動度の計算データを使用し、力場パラメータを説明変数とする予測モデルの構築を進めた。量子化学計算のミクロ物理量に加えて力場パラメータを説明変数とすることで、予測精度が向上することを実証した。(業績リスト [6,13,14])

また、実験データを用いたポリマー物性予測において機械学習の汎化性能が低下する「リーケージ」の問題に着目し、よりロバストな評価手法を開発した。ここでは特に、フ

イラー材料を含むポリマーコンポジットの物性に対し、加工プロセスや測定手法といった実験条件を特徴量化した場合の問題点を検討した。実験条件を特徴量化する際に提案の評価手法を導入することで、リーケージによる過大評価を防ぎつつ予測精度の向上が得られることを示した。(業績リスト[2,8])

## 6. 進捗状況の自己評価と今後の展望

### 進捗状況の自己評価

全体の進捗としては概ね順調である。分子動力学計算を進めるにあたって、今年度は力場パラメータに着目した解析に注力し、計算実行によるデータ蓄積は200点程度まで完了した。一方、ポリマー構造の多様性を考慮し、分子動力学の計算データを1000点程度まで拡張したいと考えて2024年度の継続課題申請を行った。また、前章のD節にある機械学習モデルの高精度化とその評価にも注力した。当初の研究計画に従い、進捗状況を次のように評価した。

- 量子化学計算の実行…100%
- ミクロ物理量の AI モデル構築…40%  
(モノマー構造に対するモデルは未完)
- 分子動力学計算の実行…100%  
(計画通り 200 種のデータ取得を完了)
- メゾ物理量の AI モデル構築…50%  
(ミクロ物理量および力場パラメータを説明変数とする試験的なモデルを構築、力場導入において改善の余地あり)
- 計算結果および AI モデルのデータベース化…30%  
(計算結果の保存は完了したが、データベース化には至っていない。AI モデルの構築と保存も必要である)

### 今後の展望

ポリマーの多階層性を考慮して量子化学計算と分子動力学計算の結果を段階的に導入した機械学習モデルとすることで、スモールデータのポリマー物性に対する外挿予測を可能にする。その過程でポリマー構造の多様性を考慮し、分子動力学の計算データを1000点程度まで拡張する予定である。あわせて力場パラメータを導入した機械学習モデルの高精度化を進める。また、各種計算結果およびそれらを段階的に予測する機械学習モデルの公開に向けて MLflow (<https://mlflow.org>) を用いたデータベース構築を行う。

## 7. 研究業績

### (1) 学術論文 (査読あり)

- [1] H. Shimakawa, A. Kumada, and M. Sato, "Extrapolative molecular property prediction using quantum chemistry-assisted machine learning," *npj Computational Materials*, vol. 10, 11, 2024, doi: 10.1038/s41524-023-01194-2.
- [2] H. Shimakawa, A. Kumada, and M. Sato, "Prevention of Leakage in Machine Learning Prediction for Polymer Composite Properties," *Journal of Chemical Information and Modeling*, vol. 64, 9, 3621–3629, 2024, doi: 10.1021/acs.jcim.3c01894.
- [3] H. Yokoyama, H. Shimakawa, A. Kumada and M. Sato, "Modulating thermal and electrical conductivities in polymers: an approach towards extracting molecular design rules through atomistic simulations," *Applied Physics Letters*, 124, 182905, 2024, doi: 10.1063/5.0198445.
- [4] 横山尋斗, 嶋川肇, 熊田亜紀子, 佐藤正寛, 「ポリマー中の熱伝導現象の解明に向けた分子動力学的アプローチ」, *電気学会論文誌 A*, 144 巻, 5 号, p. 183–184, 2024.

### (2) 国際会議プロシーディングス (査読あり)

- [5] H. Shimakawa, A. Kumada, and M. Sato,

“Extensive Evaluation and Advancement of Extrapolation Performance in Polymer Property Prediction,” A Meeting of the Materials Research Society 2023 Fall, DS05.04.02, 2023.

[6] H. Yokoyama, H. Shimakawa, A. Kumada, and M. Sato, "Molecular dynamics simulation of thermal conduction in polymers," 2024 IEEE International Conference on Dielectrics, 2024 (accepted)

(3) 国際会議発表（査読なし）

(4) 国内会議発表（査読なし）

[7] 横山尋斗, 嶋川肇, 熊田亜紀子, 佐藤正寛, 「ポリエチレンにおける熱伝導の分子動力学シミュレーション」, 令和5年電気学会基礎・材料・共通部門大会, 8-B-a1-1, 2023.

[8] 嶋川肇, 熊田亜紀子, 佐藤正寛, 「ポリマーコンジット材料の熱伝導率および電気伝導率予測のための汎用型機械学習」, 電気学会誘電・絶縁材料/ケーブル合同研究会, DEI-23-067, EWC-23-016, 2023.

[9] 嶋川肇, 熊田亜紀子, 佐藤正寛, 「外挿可能な分子探索のための量子化学計算と機械学習を掛け合わせた物性予測モデル」, 第46回ケモインフォマティクス討論会, P08, 2023

[10] 横山尋斗, 嶋川肇, 熊田亜紀子, 佐藤正寛 「結晶/非晶ポリエチレンにおける熱伝導の分子動力学シミュレーション」, 2023年度放電学会若手セミナー, 2023

横山尋斗, 嶋川肇, 熊田亜紀子, 佐藤正寛, [11] 「結晶/非晶ポリエチレンにおける熱伝導の分子動力学シミュレーション」, 2023年度放電学会年次大会, A-5, 2023

[12] 横山尋斗, 嶋川肇, 熊田亜紀子, 佐藤正寛, 「ポリマー物性間トレードオフを打破するための分子設計指針の検討」, 電気学会誘電・絶縁材料/放電・プラズマ・パルスパワー/高電圧合同研究会, DEI-24-017 EPP-24-

017 HV-24-017, 2024

[13] 横山尋斗, 嶋川肇, 熊田亜紀子, 佐藤正寛, 「ポリマーインフォマティクスによるイオン移動度予測の検討」, 令和6年電気学会全国大会, 2-054, 2024

[14] 嶋川肇, 横山尋斗, 熊田亜紀子, 佐藤正寛, 「ポリマーインフォマティクスによるイオン移動度予測の検討」, 令和6年電気学会全国大会, 2-055, 2024

(5) 公開したライブラリなど

(6) その他（特許, プレスリリース, 著書等）