

jh240061

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）

概要

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、分子動力学計算による物理量や物性値の計算とその関係の解明、外挿可能な物性予測モデルの構築、ポリマーのプロセス情報と物性値の関係の評価方法の開発に注力した。

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 D3 センター
mdx

(2) 課題分野

大規模計算科学課題分野
データ科学・データ利活用課題分野

(3) 参加研究者一覧と役割分担

佐藤 正寛	研究統括
熊田 亜紀子	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
梅本 貴弘	コード開発・データ基盤構築
片瀬 大祐	コード開発・データ基盤構築
十二 綱輝	コード開発・データ基盤構築
WANG WEIHAO	コード開発・データ基盤構築
RUAN HAOOU	コード開発・データ基盤構築
横山 尋斗	コード開発・データ基盤構築

2. 研究の目的と意義

〈研究計画全体の目的〉

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。

〈今年度の目的〉

具体的に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を行うこと、ポリマーの多階層性を考慮して計算結果の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 モデルを実装することを目的とする。また、ポリマーのプロセス情報とポリマー物性の関係を明らかにする方法を検討する。

〈研究の意義〉

本研究で開発される“外挿可能な”AI モデルは、ポリマーの分子構造からミクロ・メソ

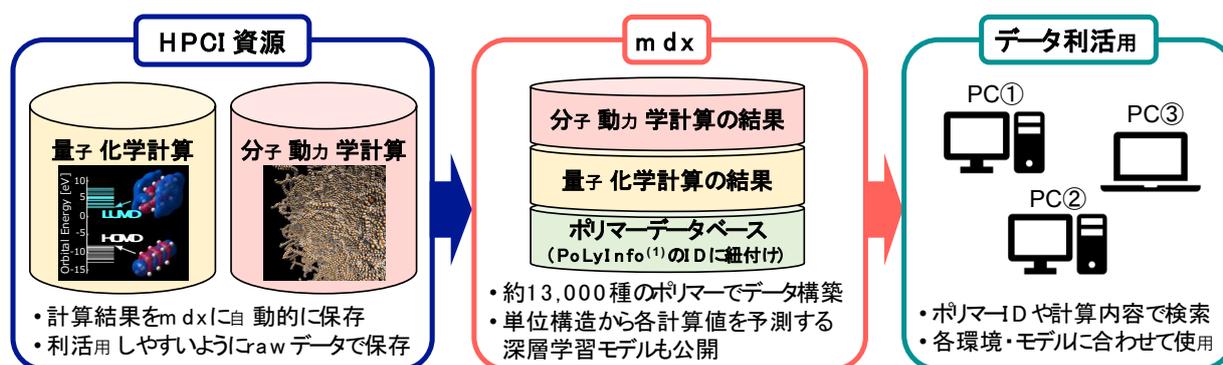


図 1 基盤システムの概略図

スケールを経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すれば、多分野のポリマー材料設計に活かされる。

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、世界に先駆けたデータ利活用基盤構築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。

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あたっては、以下の課題を解決する必要がある。まず、計算科学に基づく新たなデータベースの構築と、それを利活用できる基盤の整備が求められる。加えて、ポリマーの多階層構造を考慮した網羅的な計算を実施するためには、大規模な計算資源が不可欠である。

本提案では、図 1 に示す概略図に従い、公募型共同研究を通じてシステム構築を進める。計算負荷の大きい量子化学計算および分子動力学計算については、HPCI 資源を活用した大規模並列計算により実施する。さらに、データ利活用基盤として mdx を用い、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にとどまらず、材料科学や情報科学分野にも波及効果を持つ学際的な取り組みである。従って、本提案では、工学系研究科・情報理工

学系研究科および情報基盤センターと共同でデータベース構築を行うことができる。また、将来的には材料科学と情報学の融合による、AI モデル開発から材料設計に至る学際的な共同研究へと発展させることを目指している。

4. 前年度までに得られた研究成果の概要

2023 年度は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。また、物性を予測するための機械学習モデルの高精度化とその評価を低分子材料を対象に行った

5. 今年度の研究成果の詳細

本研究のワークフローを図 2 に示す。今年度は、PoLy-Info データベースからホモポリマー約 10,000 種類の繰り返し単位の Simplified Molecular Input Line entrySystem (SMILES) を取得した。次に、得られたポリマーの一部に対して、密度汎関数法 (DFT : Density Functional Theory) に基づく電子状態計算および MD シミュレーションによる物性計算を行った。計算結果をもとに、ポリマーの物性 MD 計算値を予測する線形、非線形回帰モデルを構築し、予測性能の評価を行った。

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

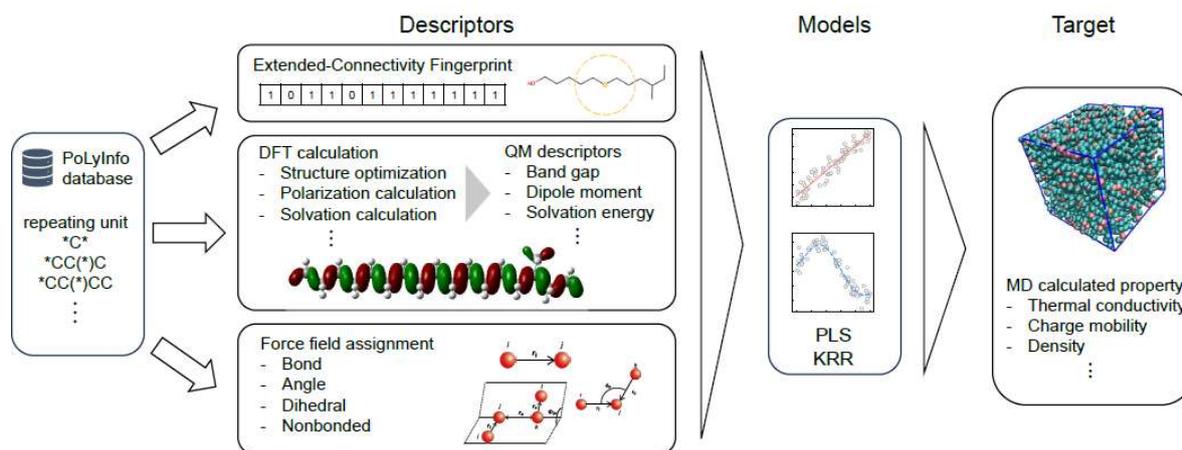


図 2 MD 物性予測モデル構築の手順

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ダイ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。

ポリマーの量子力学 (QM : Quantum Mechanics) 記述子のデータセットを作成するにあたり、繰り返し単位の SMILES を CH3 で終端し、得られた座標を初期構造として構造最適化を行った。構造最適化計算における理論レベルは B3LYP/6-31G(d) とした。得られた最適化構造をもとに物理量の計算を行った。取得した物理量の例を以下に示す。

- 1 原子当たりの電子数
- 電子空間の範囲
- 分子体積
- 分極率
- 分子体積当たりの分極率
- 双極子モーメント
- 最高被占有分子軌道 (HOMO) のエネルギー
- 最低空軌道 (LUMO) のエネルギー
- 垂直イオン化エネルギー
- 垂直電子親和力
- 標準自由エネルギー
- 分子分配関数
- 回転定数
- ESP 電荷の最小値

- ESP 電荷の最大値
- 水中溶媒和エネルギー
- オクタノール中溶媒和エネルギー

これらの物理量は昨年度から今年度にかけて QMex 記述子として提案したものの一部である。本研究ではポリマーにおいても QMex 記述子が物性の予測に有効であるのか評価を行う。

C. 分子動力学計算

A で構造を取得した約 13,000 種類のポリマーのうち、約 5,500 種類のポリマーに対して分子動力学計算を実施した。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メソスケールの物理量を取得した。一連の分子動力学計算は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 LAMMPS を用いて行った。分子動力学計算の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、各種物性の計算までの直列実行を可能にした。

計算対象としたポリマー物性は、RadonPy に実装されている 12 の物性 (密度, 回転半径 R_g , 定圧比熱 C_p , 定積比熱 C_v , 等温および等エントロピー圧縮率, 等温および等エントロピー体積弾性率, 体積膨張係数, 線膨張係数, 自己拡散係数, 熱伝導率) と, そのポリマーマトリックス中における Na^+ および Cl^- のイオン移動度である。ポリマーの力場は RadonPy に組み込まれている GAFF2 を使用した。 Na^+ イオンおよび Cl^- イオンの力場は文献から取得した。ポリマーの各原子電荷は, 50 原子程度のオリゴマー構造を用いた DFT 計算により取得した。シミュレーションボックスには 1 分子当たり 600 原子程度の CH_3 終端のオリゴマーを 6 つ, および Cl^- イオンを 1 つ挿入した。平衡化計算の過程は文

献を参考にし, Larsen らによって提案された 21 段階の圧縮・減圧平衡化プロトコルを実施した。 Na^+ 含有の構造は Cl^- 含有の平衡化構造からイオン原子の交換を行い, 追加の平衡化計算を行うことで作成した。イオン移動度は温度が 360K, 圧力が 1 atm のシミュレーションボックスで計算し, その他は温度が 300K, 圧力が 1 atm のシミュレーションボックスで計算した。熱伝導率は Langevin サーモスタットを用いた非平衡分子動力学法 (NEMD: Non-Equilibrium Molecular Dynamics) により計算した。イオン移動度は, 平衡化構造に対して x 方向に電界を印加し, NPT 計算を行うことでシミュレートした。イオン移動度 μ は, イオン原子の x 方向の移動距離を d_x とし $\mu = d_x/(Et)$ のように計算した。

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, 手始めに分子動力学計算で得られた物性量を予測する手法を開発した。入力記述子として A で得られた量子化学計算のミクロな物理量や分子動力学の力場パラメータを用いることで, 既存の手法に比べてより高い外挿予測性能が得られることを示した。また, B や C でポリマーの計算物性データベースを拡張したことにより, 内挿予測のみならず外挿予測においても予測精度が向上することを実証した。さらに, 計算された物理量を用いることで, 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た。B で取得した量子化学計算のミクロな物理量や分子動力学の力場パラメータに加え, C で取得した分子動力学のメソスケール物理量を入力記述子としたマルチスケールなモデルにより, ポ

リマーの実験値であっても外挿的な予測が可能であることを示した。ポリマー分子設計を目指し、外挿的な物性予測法の開発に加えて、新たに分子生成手法の開発にも取り組み始めた。初期段階として、原子数 13 以下の低分子を対象に、存在可能なすべての分子構造を網羅的に生成する手法の検討を進めた。

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、今年度は分子動力学計算において、計算実行によるデータ蓄積を 5,500 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 10,000 点程度まで拡張したいと考えている。また、前章のポリマー物性のための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・分子動力学計算の実行…100%
(計画通りであるが、さらなる拡張が必要)
- ・ メゾ物理量の AI モデル構築…70%
(現在は数種類の物性量の予測ができておらず、さらなる拡張が必要)
- ・ 計算結果および AI モデルのデータベース化…60%
(データベース化を行っているところである)

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、様々な物性料に関して可能にする(現在は

数種類の物性量の予測にとどまっている)。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 10,000 点程度まで拡張する予定である。また、引き続き、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。

より将来的には、分子構造生成やプロセス情報とあわせたデータ基盤、ポリマー基盤モデル構築も行いたい。

※7. 研究業績はウェブ入力です