

産学連携コンソーシアムによるデータ駆動型高分子材料研究を変革する 高分子物性データ基盤の創出

研究の背景と概要

本研究は、全原子分子動力学(MD)シミュレーションに基づく高分子物性計算を全自動化するソフトウェア RadonPyを用いて、データ駆動型高分子材料研究(高分子インフォマティクス)の学術基盤となる体系的なデータベースを創出する。データ駆動型研究の最も重要な資源はデータである。しかしながら、現時点においてデータ駆動型研究に資する高分子物性データベースは存在しない(表1)。その原因として、次の3点が考えられる。

- ① 他の材料系に比べて、高分子材料の実験やシミュレーションはコストが非常に高い。
- ② 研究者の興味や設計変数(試料の作製方法、プロセスパラメータなど)が多様であるため、コミュニティ全体でコモンデータを創出しようという動きが起きにくい。
- ③ 基礎研究と応用研究の垣根が低いと、競合相手に対する情報秘匿の意識が高くなり、データを公開するインセンティブが研究者に働きにくい。

このような背景から高分子材料のオープンデータベースの開発はほとんど進んでいない。本研究では、複数の大学・企業からなるコンソーシアムを形成し、統計数理研究所のものづくりデータ科学研究センターの研究者が開発した高分子物性自動計算手法と我が国のフラグシップスパコン「富岳」を融合し、世界最大の高分子物性データベースを創出する。

ミッションを達成するための駆動力

① 高分子物性自動計算ソフトウェア RadonPyの開発

RadonPyは高分子材料のMDシミュレーションを全自動化できる現時点における唯一のソフトウェアである。高分子材料のMDシミュレーションは全自動化が技術的に難しく、さらに膨大な計算資源を必要とするため、包括的な計算物性データベースを創出しようという動向は国内外ともに皆無となっている。我々は、RadonPyの開発に成功したことで、世界初となる包括的な高分子物性データベースの創出の実現に向けた道筋を見出すことができた。

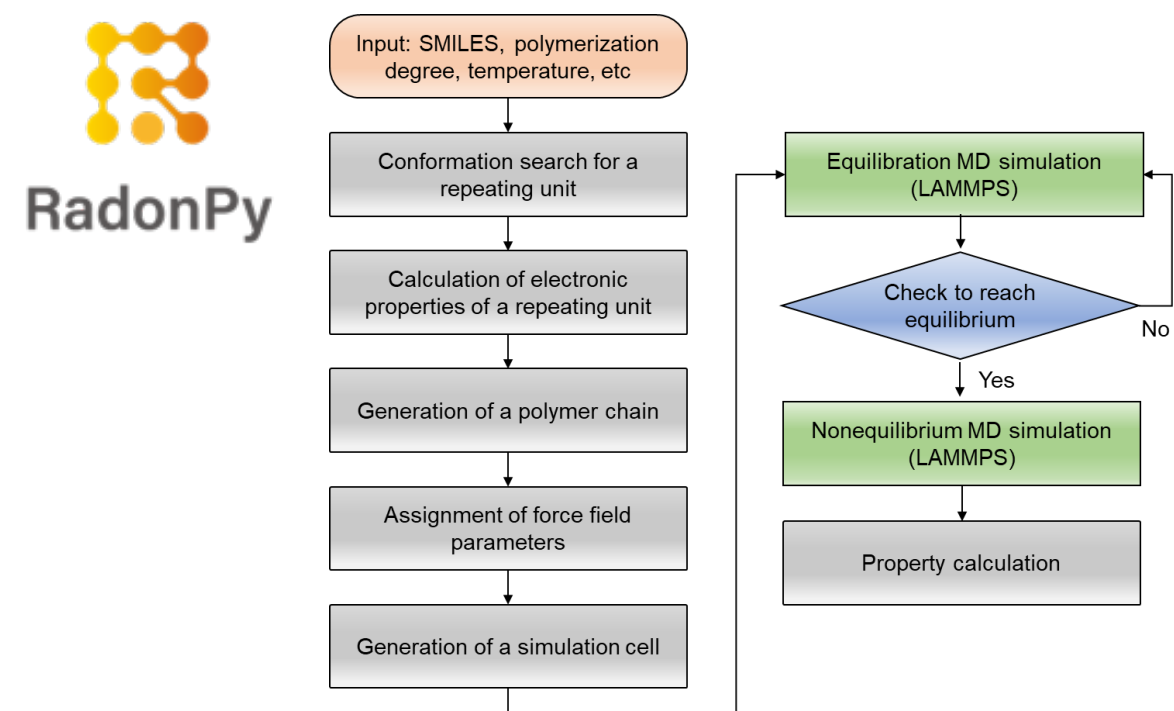


図1: RadonPyのワークフロー

② 産学の垣根を超えたデータの共同生産

高分子物性のMD計算は膨大な計算コストを伴うため、小中規模の研究グループでは、データ駆動型研究に資するレベルのデータを生産できない。そこで、多数の大学・企業からなるコンソーシアムを形成し、データベースを共同開発する仕組みを構築した。現在、3大学・19企業に属する82名がコンソーシアムに参画し、RadonPy及びデータベースの共同開発事業を推進している。最終的には、10万種類以上の分子骨格を包含する高分子物性データベースを創出する。

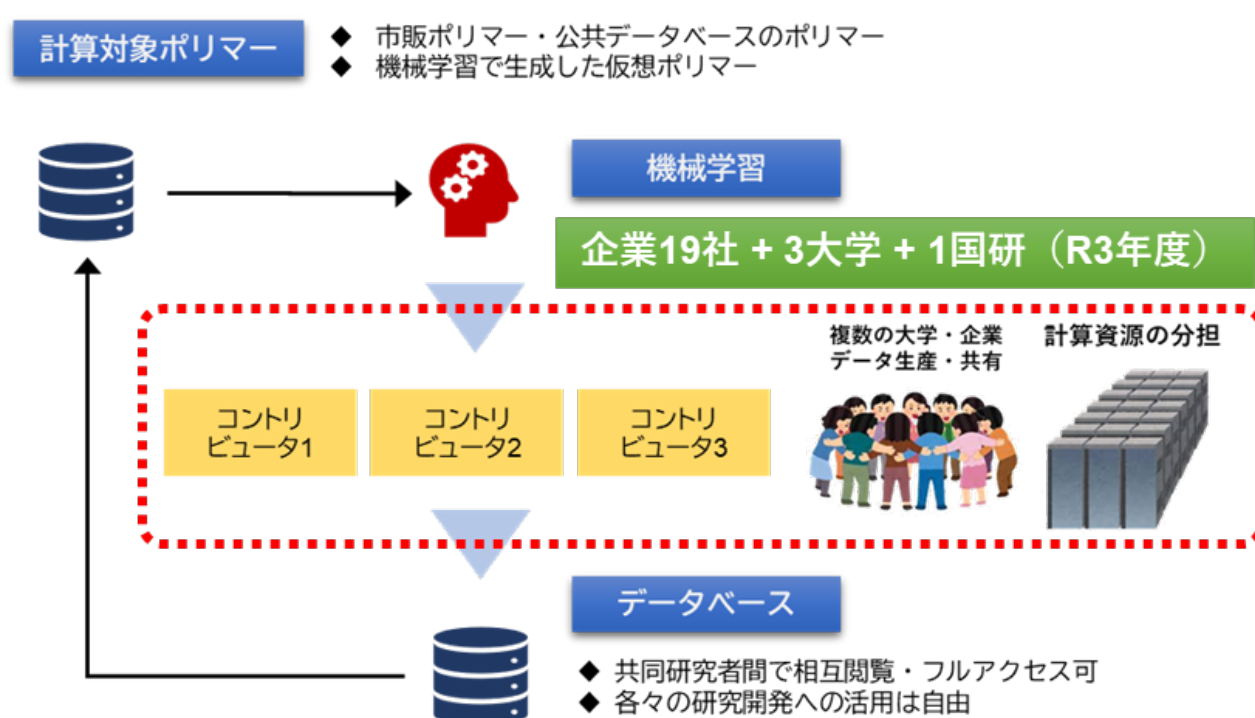


図2: コンソーシアムによるデータベースの共同開発

③ 富岳の利用

本事業は、文部科学省「富岳」成果創出加速プログラムの支援を受けている(2021-課題名「データ駆動型高分子材料研究を変革するデータ基盤創出」)。複数のポリマーの独立な物性計算を富岳の大量ノードに分担させることで、前人未踏の大規模データベースを構築する。プロジェクト初年度には、アモルファスポリマーの15種類の物性の計算が完了した。大量の高分子材料の複数物性の同時分布を観測することで、物性間のトレードオフを表すパレート境界の位置が特定される。また、データ生産の過程でパレート境界を超える特異な分子骨格も同定される。

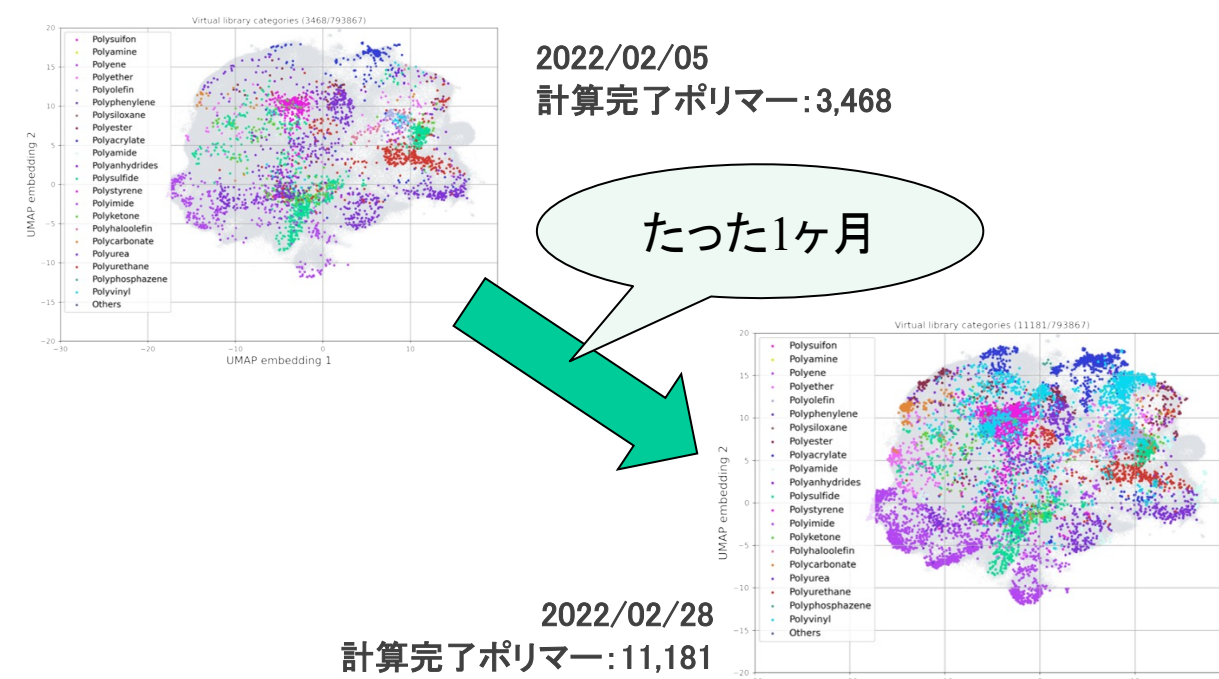


図3: 富岳を活用した大量ポリマーの物性自動計算

現時点における成果

現時点での成果として以下の3点があげられる。

(1) RadonPyの開発

2021年度にアモルファスポリマーの15物性の自動計算を実施し、プレプリントの公開とRadonPyのファーストリリースを行った。
論文: Hayashi, Y., Shiomi, J., Morikawa, J., Yoshida, R., RadonPy: automated physical property calculation using all-atom classical molecular dynamics simulations for polymer informatics. arXiv preprint. arXiv:2203.14090 (2022).
DOI: <https://doi.org/10.48550/arXiv.2203.14090>
コード(GitHub): <https://github.com/RadonPy/RadonPy>

(2) データベース

2021年度は、計11,181アモルファスポリマーの15物性の計算を実施した。図4に示すように、大量のポリマーの複数物性の同時分布を観測することに成功した。これにより、複数物性のパレート面の位置が特定されて、パレート面を構成するポリマー群の構造的特徴に関する体系的な知見を得るに至った。これらのポリマーの構造的特徴を解析した結果、分子骨格の剛直性、水素結合可能なユニットが高密度で存在すること、水素結合やdipole-dipole相互作用を介したメカニズムがアモルファスポリマーの高熱伝導化を実現していることが明らかになった(Hayashi et al. arXiv:2203.14090 (2022))。

(3) コンソーシアムの運営と事業推進

2021年度は、統計数理研究所に加えて3大学19企業に所属する82名が本プロジェクトに参画した(図5)。参画者の一部は、RadonPyの開発に加わり、ガラス転移温度、溶媒和自由エネルギー、レオロジー特性、熱硬化性樹脂の自動計算手法の開発に取り組んだ。また、多数の参画者が富岳を利用してデータ生産を実施した。参画者の多くは、MDシミュレーションやスパコンの専門知識を持たないため、毎月定例で開催されるミーティングやSlackを活用して技術指導や情報交換を行っている。また、2022年3月31日に成果報告会を開催し、プロジェクトの研究紹介、RadonPyファーストリリースのアナウンス、関連コミュニティからの意見収集を行った。参加者数は219名であった。

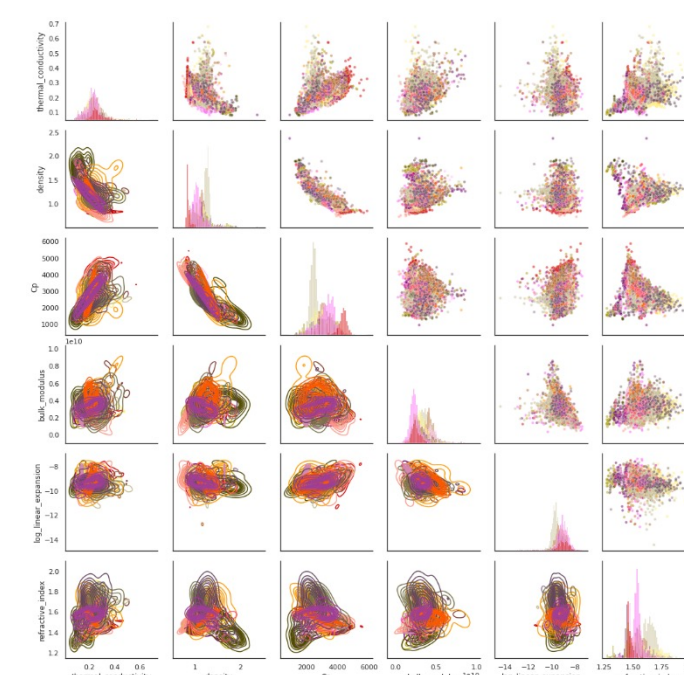


図4 複数物性の同時分布の観測とパレート境界の同定



図5 産学連携の新しいモデルケースを創出

展望

10万種類以上の分子骨格を包含するポリマーの物性の大地図を作成する。さらにデータベースと統計数理研究所の統計的機械学習の先端技術をパッケージ化し、大学共同利用・国際共同研究・産学連携ネットワークを介して関連分野に学術資源を提供していく。また、分野・組織・国境の垣根を超えたデータベースの共創、学融合というモデルケースを社会に発信していく。