

シンポジウム | イノベーション共創プログラム (CIP) : AI×自動計算×自動実験による化学研究のデジタル革新

■ 2025年3月26日(水) 13:00 ~ 15:40 ■ [E]F401(第2学舎 4号館 [4階] F401)

[[E]F401-1pm] AI×自動計算×自動実験による化学研究のデジタル革新

座長、シンポジウム関係者：宮崎 然、高谷 光、酒谷 能彰、松本 和弘

『AI×自動計算×自動実験』を主題として、AI・ハイパフォーマンス・ロボット・自動・自律実験分野の最前線で活躍する研究者を講師に迎え、最新の研究成果について解説して頂きます。2022年から2024年のわずか1年間に、Google DeepMind社をはじめとする巨大企業や卓越研究者をリーダーとする国際コンソーシアムが巨額の資金援助・調達を受けて、革新的なデータ駆動型材料研究手法を開発しており、この分野でのトップの座を巡る競争が激化しています。本企画では、各講演者からの問題提起を通じて、聴講者一人一人が「化学研究のデジタル変革」や「計算科学と自動実験の融合」がもたらすデータ駆動型化学研究のあり方とその可能性について深く考えようとする機会を提供し、本分野での後継者が培養されている我が国において、今何をすべきかを考える気付きの機会としたいと思います。

本セッションは**午前**、**午後**、**夕刻**に実施されます。聴講後の[アンケート](#)へのご協力をお願いいたします。

◆ 日本語 ◆ 依頼講演

13:00 ~ 13:30

[[E]F401-1pm-01]

実験化学の「身体性」に迫るためのAI・自動実験の活用

○ 畠山 歓¹ (1. 東京科学大学)

◆ 日本語 ◆ 依頼講演

13:30 ~ 14:00

[[E]F401-1pm-02]

研究・事業化の競争とリアル・デジタルの共進化

○ 川上 登福^{1,2} (1. 株式会社Material Infinity、2. 株式会社先端技術共創機構)

◆ 日本語 ◆ 依頼講演

14:00 ~ 14:30

[[E]F401-1pm-03]

自動シミュレーションによるデータ駆動型高分子材料研究の最前線

○ 林 慶浩¹ (1. 統数研)

◆ 日本語 ◆ 依頼講演

14:30 ~ 15:00

[[E]F401-1pm-04]

マルチスケールシミュレーションとデータサイエンスの活用による材料設計

○ 小沢 拓¹ (1. 株式会社SOL)

15:00 ~ 15:10

[1E_F40104-07-5add]

インキュベーションタイム

15:10 ~ 15:40

[1E_F40104-07-6add]

パネルディスカッション

実験化学の「身体性」に迫るための AI・自動実験の活用

(東京科学大学 物質理工学院) ○畠山 歓

Utilizing Artificial Intelligence and Automated Experimentation to Explore the Embodiment in Experimental Chemistry (*Materials Science and Engineering, School of Materials and Chemical Technology, Institute of Science Tokyo*) ○Kan Hatakeyama-Sato

In experimental chemistry, embodiment refers to the influence of a researcher's intuition, experience, and physical operations on experimental outcomes. For AI to truly approach human experimental chemists, it is essential to replicate this comprehensive capability, including embodiment, and to develop an understanding of the experimental process. This presentation introduces recent research examples that address this challenge through the utilization of foundational models and automated experimentation technologies.

Keywords : Materials informatics, generative AI, robotics, laboratory automation

実験化学における「身体性」とは、研究者の直観や経験的判断、さらには手技的操作が実験結果に反映される諸事象を指す。近年、分子構造や物性データに基づくケモインフォマティクスやマテリアルズ・インフォマティクスといった分野は大きく発展してきた^[1]。しかし現行の手法は化学構造と物性のみを解析対象とする場合が多く、化学的に不適切な予測提案がなされてしまう事例があるほか、少数データへの適合性が低いといった課題がある。^[2]近年登場した GPT-4 に代表される大規模言語モデルは科学理論に裏打ちされた推論を可能とする^[3]一方で、実験室内の情報を十分には学習していないため、実験現場で蓄積される暗黙知的な事象群までは踏み込めていない^[2]。

一連の問題に対処するにあたっては、実験化学者が単に分子構造の特徴や教科書的な記述といった「記号的」な知識を有するだけでなく、実験ノートや日々のディスカッションから得られる知見、さらには実験時に受け取る五感刺激や筋肉の運動感覚といった「身体性」を基盤とする暗黙知を有している点に注目する必要がある (Fig. 1)。

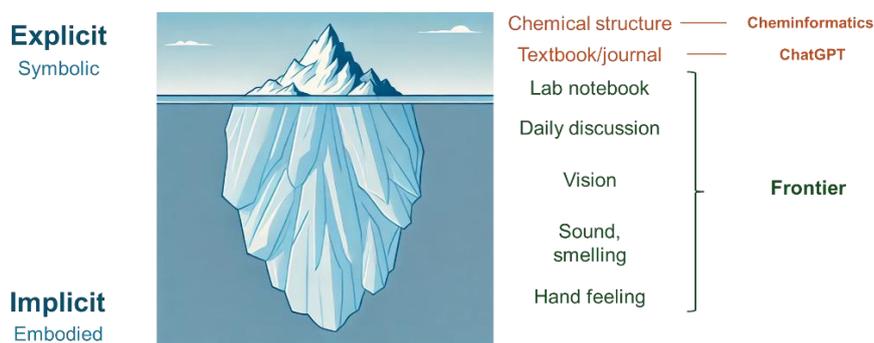


Fig. 1 Symbolic and embodied information.

AI が今後、人間の実験化学者の能力に追従するためには、これら記号的知識と身体

性を包含した多面的な能力を統合し、実験プロセスを包括的に理解する過程が不可欠となる。

本発表では、基盤モデルの活用と自動実験技術の融合により、この身体性を取り込んだ総合的な実験プロセス理解を目指す最近の研究例を紹介する。これにより、究極的には、従来のインフォマティクス手法では到達困難であった実験空間の理解と、それに基づく新しい知識の創出を目指す道筋を考える。

例として、実験者の手技をウェアラブルデバイスで観察し、その様子をマルチモーダルな生成 AI で解析する手法について報告する。一般常識を含む基礎知識を既に有した基盤モデルを活用することで、特別なモデルチューニングをすることなく、実験作業の様子を一定精度で記述できる。実験者の手技をリアルタイムで観察し、試薬を秤量する様子を自動で読み取って記録したり、フローチャートの形でまとめ直すといった作業が実現しつつある。

カスタムした液体ハンドリング装置とロボットアームを用いた、ポリアミック酸粒子の半自動合成システムについても報告する^[4]。このシステムにカメラおよびマルチモーダルな大規模言語モデルを統合することにより、合成実験を継続的にモニタリングし、その過程を自動的かつ詳細に記録できた。

実験研究に適合した基盤モデルの研究や開発を通し、実験手続き全体の客観性が強化されるとともに、実験条件の再現性や、化学構造—物性—プロセスの相関をつなぐ、新たな実験科学の手法や学理が切り開かれるものと期待される。

[1] A. Mirza, N. Alampara, S. Kunchapu, B. Emoekabu, A. Krishnan, M. Wilhelmi, M. Okereke, J. Eberhardt, A. M. Elahi, M. Greiner, C. T. Holick, T. Gupta, M. Asgari, C. Glaubitz, L. C. Klepsch, Y. Köster, J. Meyer, S. Miret, T. Hoffmann, F. A. Kreth, M. Ringleb, N. Roesner, U. S. Schubert, L. M. Stafast, D. Wonanke, M. Pieler, P. Schwaller, K. M. Jablonka, 2024, arXiv:2404.01475.

[2] K. Hatakeyama-Sato, N. Yamane, Y. Igarashi, Y. Nabae, T. Hayakawa, *Sci. Technol. Adv. Mater.: Methods* 2023, 3, 2260300.

[3] K. Hatakeyama-Sato, S. Watanabe, N. Yamane, Y. Igarashi, K. Oyaizu, *Digital Discovery* 2023, 2, 1548.

[4] K. Hatakeyama-Sato, H. Ishikawa, S. Takaishi, Y. Igarashi, Y. Nabae, T. Hayakawa, *Polym. J.* 2024, 56, 977.

研究・事業化の競争とリアル・デジタルの共進化

(Material Infinity¹, 先端技術共創機構 (ATAC)²) ○川上 登福^{1,2}

Competition in research and commercialization and co-evolution of real and digital (¹Material Infinity Inc., ²Advanced Technology Acceleration Corporation) ○Takayoshi Kawakami^{1,2}

Amid intensifying global research competition, maximizing limited resources and improving productivity are essential to enhancing research competitiveness. Sharing research facilities enables broader access to experimentation, while consolidation enhances resource efficiency and supports the establishment of specialized organizations for experimentation, measurement, and evaluation. These organizations allow for effective use of diverse expertise and increase investment efficiency in automation and autonomy. Automating and autonomizing these processes accelerates research speed and generates valuable data. When combined with computational science, this data enables further speed and efficiency improvements. Accelerated and optimized research processes directly strengthen international competitiveness and foster greater innovation capacity.

Keywords : Sharing; Consolidation; Professional Operators; Automation and Autonomy;

国際的な研究競争が激化する中で、限られたリソースを最大限活用し、生産性を向上させるとともに、研究競争力を向上させる取り組みが求められている。研究設備の共用化は、研究を広く行うことを可能にし、集約化は更に設備効率を高めると共に、実験・計測・評価を担う専門組織化を可能にする。専門組織化によって広く知見を活用することを可能にし、自動自律化への投資効率を向上させる。実験・計測・評価の自動化・自律化は、研究のスピードを向上させ、そこから生まれるデータの活用と計算科学は、更なるスピードアップ・効率化を可能にする。研究のスピードアップ・効率化は、研究の国際的な競争力を高め、イノベーション創出力の強化に直結する。



自動シミュレーションによるデータ駆動型高分子材料研究の最前線

(統数研¹) ○林 慶浩¹

Frontiers of Data-Driven Polymer Materials Research by Automated Molecular Simulation
(¹*The Institute of Statistical Mathematics, Research Organization of Information and Systems*)
○Yoshihiro Hayashi¹

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,^{1,2} 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 10^5 polymer skeletons.

Equilibration calculations for about 137,000 data and about 87,000 unique polymers were completed by the end of 2024 using RadonPy. 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. In addition, Bayesian optimization and automatic simulation using RadonPy are combined to design polymers with desired properties.

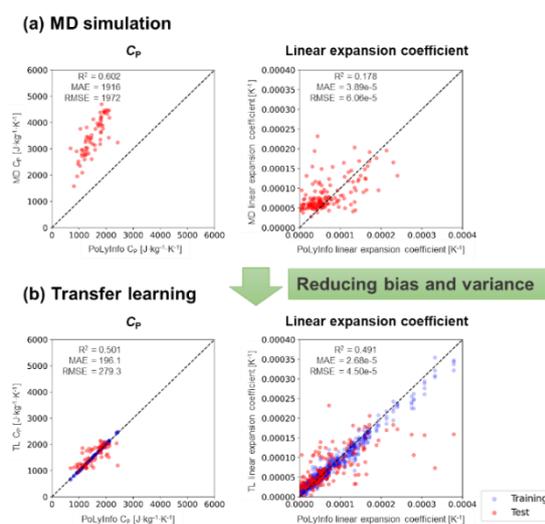


Figure 1. Calibration of MD calculated values by the transfer learning.

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. 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

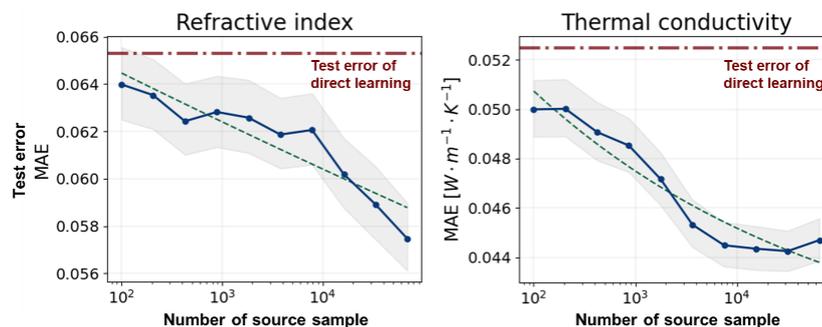


Figure 2. Scaling law of transfer learning using simulation data.

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。^{1,2)} 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、10⁶ 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 2024 年末までに約 137,000 件のデータ (ユニークポリマー数約 87,000) の平衡化計算が完了した。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リング則を観測した。このスケーリング則は、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>

マルチスケールシミュレーションとデータサイエンスの活用による材料設計

(JSOL¹) ○小沢 拓¹

Materials design using multiscale simulation and data science

(¹JSOL Corporation) ○Taku Ozawa¹

In materials design, it is necessary to pay attention to the multiscale structures inside materials, i.e., to which scale the target properties are caused. The same care must be taken when applying computer science (simulation and data science) methods.

In the case of simulations, it is necessary to apply the appropriate theory and methods for each scale. For example, quantum chemistry-based methods such as molecular orbital and density functional theory are applicable to properties related to electronic states, molecular dynamics based on classical mechanics and its coarse-grained models are applicable to dynamics of molecular aggregates, and phase-field models are applicable to meso-scale phenomena such as phase separation.

In addition, the linkage between data science and simulation includes the use of the results of a large number of simulation data runs as training data and the use of surrogate models to accelerate simulations.

This presentation will provide an overview along with recent software trends.

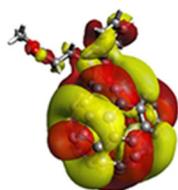
Keywords : Multiscale simulation, molecular simulation, molecular modeling, data science, material design

材料設計においては多くの場合、材料内部のマルチスケール特性、つまりターゲットとなる物性がどのスケールに起因するかに注意する必要がある。計算機科学（シミュレーションやデータサイエンス）の方法を適用するにあたっては同様に、シミュレーションであれば各スケールに適した理論や手法の適用が必要となる。

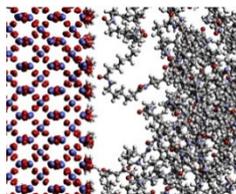
電子状態に関連する物性であれば分子軌道法や密度汎関数法など量子化学に基づく手法、分子の集合体のダイナミクスであれば古典力学に基づく分子動力学やその粗視化モデル、さらに相分離などのメソスケールの現象であればフェーズフィールドモデルなどが該当する。

また、データサイエンスとシミュレーションの連携については、シミュレーションデータを大量に実行した結果を教師データに用いたり、シミュレーションを高速化するためのサロゲートモデルなどの活用がある。

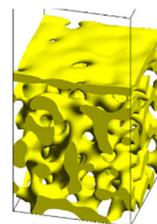
本講演では最近のソフトウェアの動向と併せて概観する。



量子化学計算



全原子分子動力学



フェーズフィールドモデル