

# 統計数理研究所におけるデータ駆動型材料研究の実践

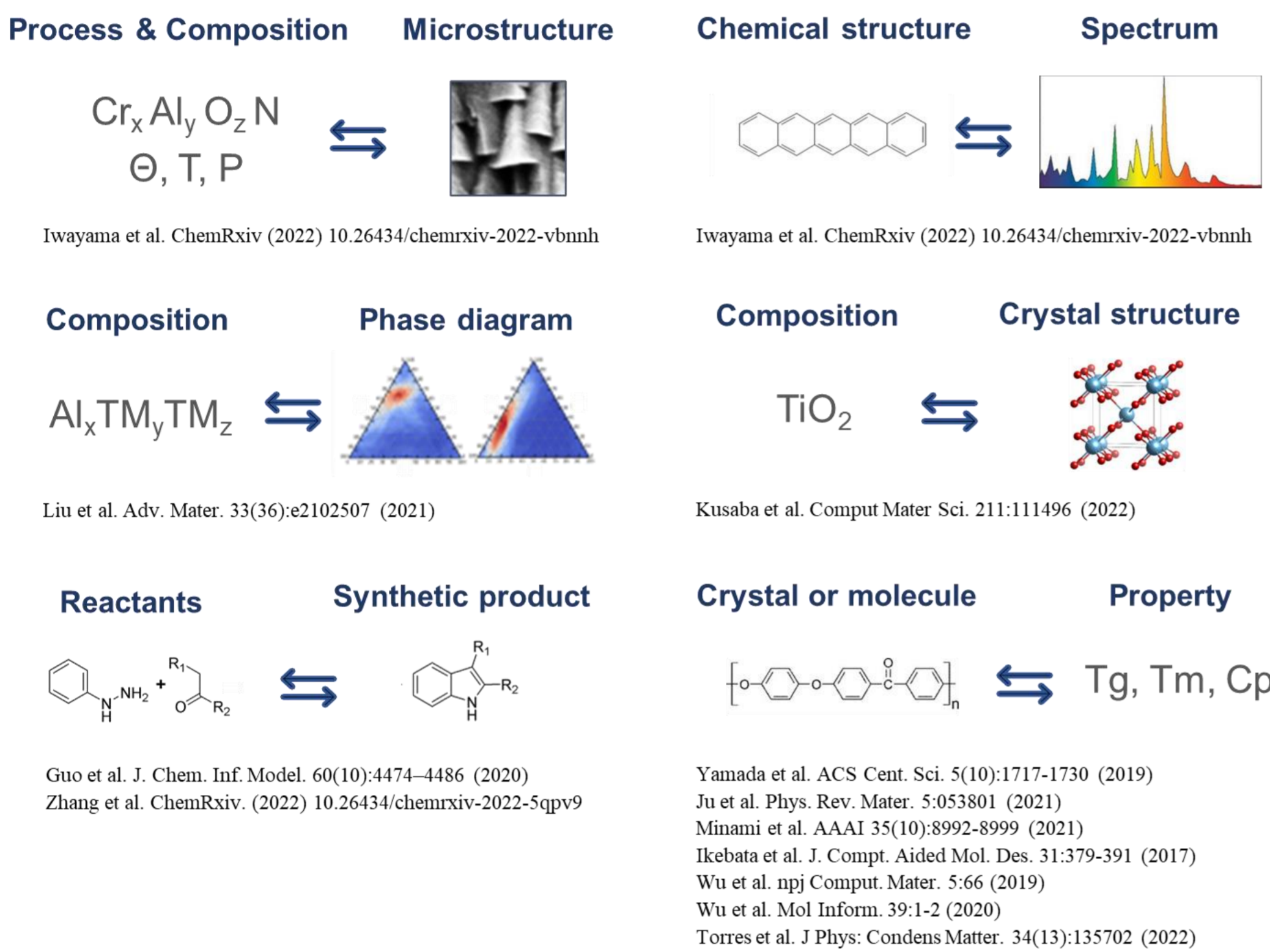
吉田 亮 データ科学研究系 教授

材料設計のパラメータ空間は極めて広大です。例えば、有機化合物のケミカルスペースには $10^{60}$ を超える候補物質が存在すると言われています。マテリアルズインフォマティクスの問題は、このような広大な未踏空間から所望の特性を有する埋蔵物質を発掘することです。我々はデータ科学のユニークな視点から新しい問題を発掘し、マテリアルズインフォマティクスの基盤技術の開発と実践・実証を推進しています。

## 順問題と逆問題

マテリアルズインフォマティクスの問題の多くは、順問題と逆問題の形式に帰着します。順問題の目的は、系の入力に対する出力の予測です。これに対し逆問題では、モデルの逆写像を求めて所望の物性を有する材料を予測します。解析対象の変数は、分子、組成、結晶、混合物、プロセス、合成経路など問題に応じて多様な形式をとります。我々は、材料研究の多岐に渡る分野で、データ科学の先進技術を駆使し、新材料創製に取り組んでいます。

材料研究における様々な順問題と逆問題と機械学習アルゴリズム

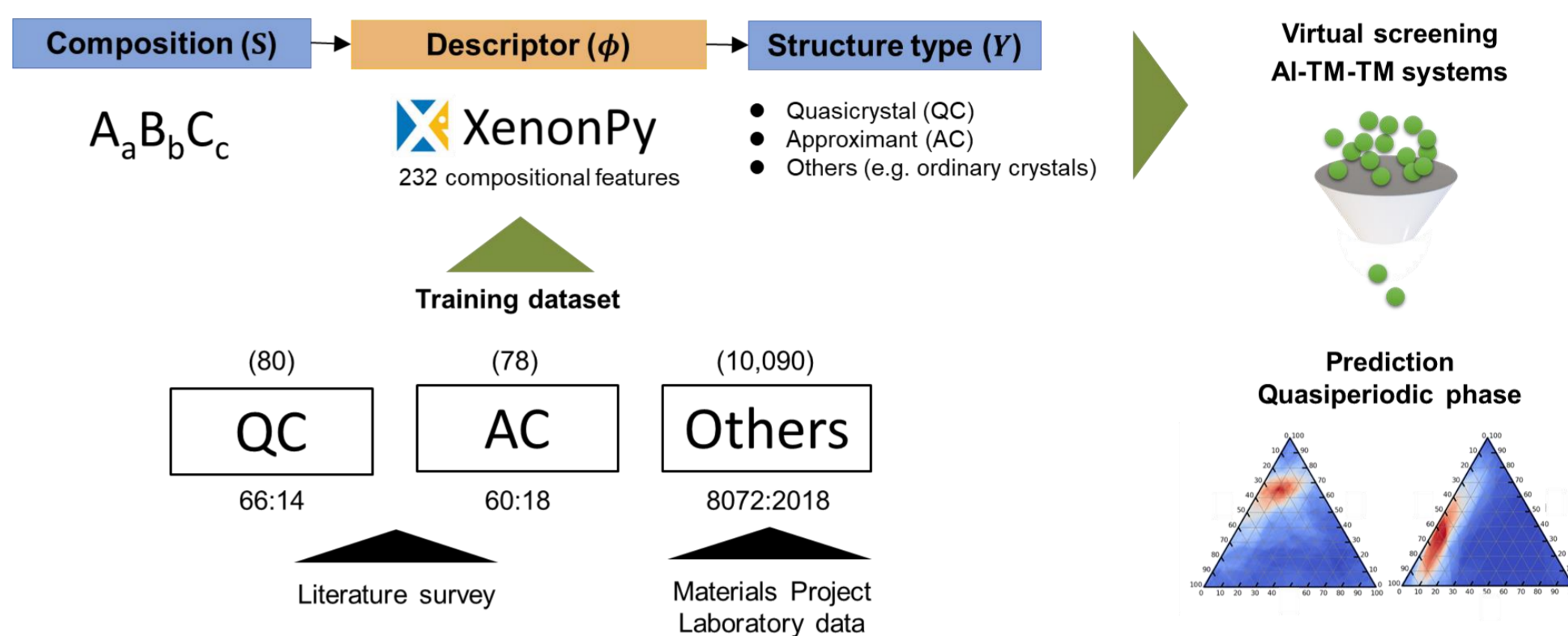


## 新しい準結晶の予測と発見

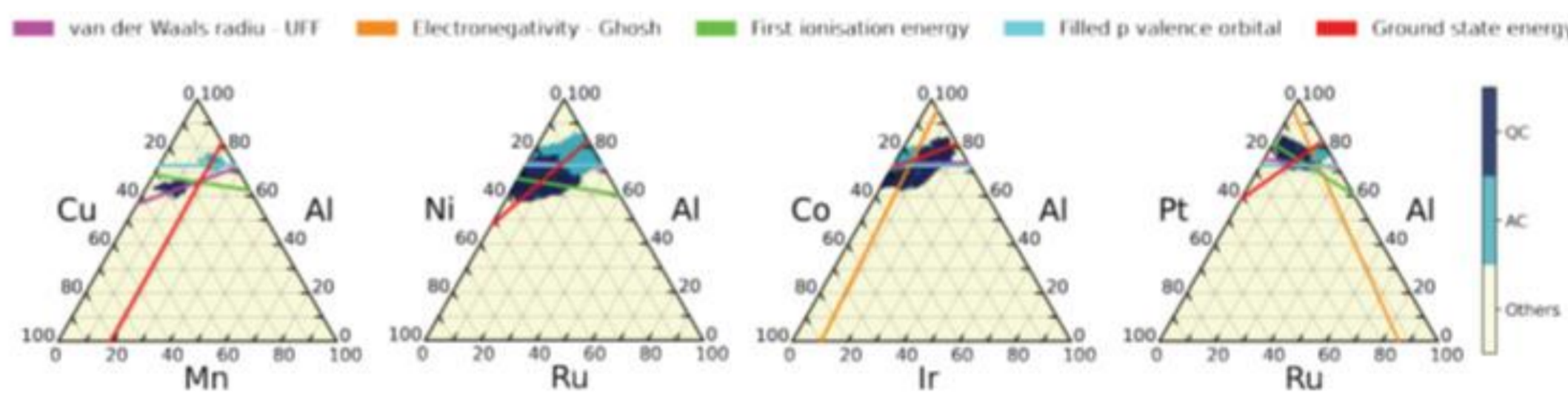
準結晶は通常の結晶のような並進対称性を持たないが、原子配列に高度な秩序がある物質群です。最初の準結晶からおおよそ35年間で100種類以上の安定な準結晶が見つかってきましたが、近年は準結晶の発見のペースが著しく低下しています。我々は、準結晶を形成する化学組成を予測する機械学習アルゴリズムを開発し、複数の新しい準結晶を発見することに成功しました。

Reference: Liu et al. Adv. Mater. 33(36):e2102507 (2021)

準結晶を形成する化学組成を予測する機械学習アルゴリズム



機械学習アルゴリズムが発見した準結晶形成の5つのルール

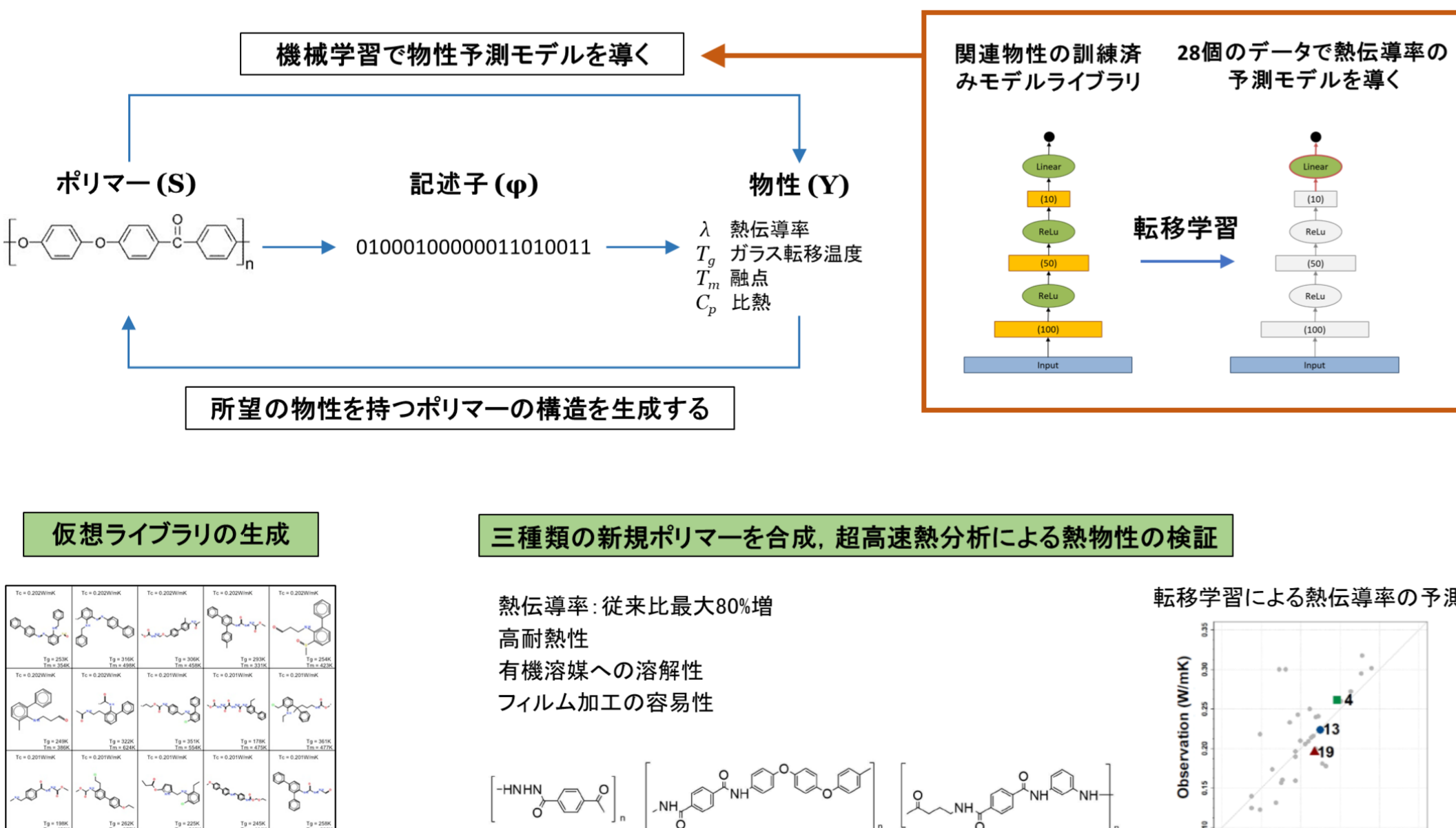


## 新しいポリマーの予測と発見

高分子材料の熱伝導率は、金属やセラミックスに比べると非常に低いことが知られています。しかしながら、近年、特異的に高い熱伝導率を持つ高分子材料が見つかり、放熱性の向上が要求される電子デバイスに高分子材料を応用する研究に注目が集まっています。我々は転移学習とベイズ推論による分子設計技術を用いて、非常に高い熱伝導率を持つアモルファスポリマーの予測と合成に成功しました。

Reference: Wu et al. npj Comput. Mater. 5:66 (2019)

機械学習による高い熱伝導率を持つポリマーの予測と発見



## データベースをつくる

データ駆動型研究の源泉は、言うまでもなくデータですが、現時点において、データ駆動型研究に資する高分子物性のデータベースは存在しません。本グループは、分子動力学シミュレーションによる高分子物性計算を全自動化するソフトウェアRadonPyを開発しています。現在、多数の大学・企業とRadonPyを用いて高分子物性のデータベースを共同開発しています。最終的には10万種類以上の分子骨格を包含する体系的なデータベースを構築します。本事業は「富岳」成果創出加速プログラム「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」の支援の下で推進されています。

Reference: arXiv preprint. arXiv:2203.14090 (2022)

RadonPy (GitHub): <https://github.com/RadonPy/RadonPy>

