



Title	有機・無機化合物を包括する化学のための基盤モデルの構築
Author(s)	水上, 渉
Citation	サイバーメディアHPCジャーナル. 2026, 16, p. 68-71
Version Type	VoR
URL	https://doi.org/10.18910/106238
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

有機・無機化合物を包括する化学のための基盤モデルの構築

水上 渉

大阪大学 量子情報・量子生命研究センター

1. はじめに

近年、情報科学と化学の融合領域において、機械学習ポテンシャル(Machine Learning Interatomic Potential: MLIP)と呼ばれる、物質を構成する原子間に働くエネルギーと力を予測する機械学習モデルの研究が盛んに行われています。

原子・分子のスケールで見ると、化学とは化合物がある構造 X から別の構造 X' へと変化する過程を理解する学問です。化学反応に伴う熱は両構造のポテンシャルエネルギー $E(X)$ と $E(X')$ の差として与えられ、両者を結ぶ最小作用経路上の一次の鞍点(遷移状態)のエネルギーからは反応速度を算出できます。さらに $E(X)$ の微分量は、屈折率を決める分極率(電場による二階微分)をはじめ、種々の物性値と直接結び付いています。

このように、化合物のある構造情報 X をインプットして、その化合物のエネルギーを返すような「ポテンシャル関数」は、分子・原子レベルでの化学の理解を支える最も基礎的な関数の一つと言えます。言い換えれば、任意の化合物 X に対して、 $E(X)$ を高精度かつ高速に予測できるモデルを構築できれば、それは化学分野における「基盤モデル」ということができるでしょう。

2010年代後半までは、機械学習により構築されたポテンシャル関数は、特定の系に特化したものか、精度を大きく欠くものかのいずれかでした。しかし近年、グラフニューラルネットワーク(GNN)の発展と大規模な第一原理計算・量子化学計算データセットの公開を背景に、多様な化合物に適用可能な「汎用」MLIPが登場してきます[1]。汎用MLIPは第一原理計算に匹敵する精度を保ちながら桁違いに高速であり、多くの場合、計算コストが系のサイズに対して線形にスケー

ルします。そのため、従来は扱えなかった大規模系をこれまでにない信頼性をもって検証できる道具として大きな注目を集めています。

一方で、「汎用」と謳いつつ現在の汎用MLIPにはまだまだ課題が山積みです。第一に、「汎用」と称するモデルの大半は、有機系化合物か無機系化合物のいずれか一方のみを対象としてきました。第二に、高性能な汎用MLIPは高価なクラウドモデルであったり、ライセンスに制限があったりするなど、必ずしも使いやすいものではないという問題もありました。我々は、真に化学分野の基盤モデルと呼びうるモデルは、有機・無機の双方に対応し、周期表のできる限り広い範囲をカバーし、かつ誰もが利用できる形で提供されることが求められていると考えています。

こうした問題意識のもと、我々は2024年度までに、計算設定の異なる第一原理計算データセットを再計算なしに統合する手法 Total Energy Alignment (TEA)を開発し、これを用いて有機系・無機系を単一のニューラルネットワークで扱える世界初の汎用MLIP「MACE-Osaka24」を構築、MITライセンスで公開しています[2]。MLIPコミュニティ全体の進展も伴って上記の二点については、大きな改善が見られています。

残された大きな課題の一つは、元素カバレッジの拡張です。既存の汎用MLIPの学習データは、アクチノイドをはじめとする重元素をほとんど含んでいません。とりわけ超ウラン元素を含む系は、放射性ゆえに実験が極めて困難でシミュレーションへの期待が大きい一方、出発点となる実験データが少ないこともあり第一原理計算のデータ整備そのものが進んでいませんでした。そこで本年度は、アクチノイドを含む重元素系の第一原

理計算データセット HE26 を新たに整備し、これを TEA により既存の有機系・無機系データセットと統合したうえで、97 元素を網羅する汎用 MLIP 「MACE-Osaka26」を構築しました[3]。

2. データセット統合法 Total Energy

Alignment (TEA)

このセクションでは、本年度の成果の前提となる TEA について説明します。第一原理計算/量子化学計算は有機系と無機系とで異なるソフトウェアエコシステムを形成しながら発展してきました。そのため、有機分子系の公開データセットと無機系のそれとの間には全エネルギーの基準や内殻電子の扱いに起因する深刻な非互換性があります。例えば代表的な無機系データセット MPTrj [4]は VASP という平面波基底と内殻を PAW 法という方法で取り扱い、密度汎関数法(DFT)の交換相関汎関数法の交換相関には PBE を用いています。一方で、有機系データセット OFF23 [5]は、Psi4 という Gauss 基底を用い内殻電子も露わに取り使った全電子計算となっています。汎関数も ω B97M (と D3(BJ)補正) というものが使われており、これは PBE とは出自や性質が大きく異なります。この様々な違いの結果、分子データセット QM9 [6]を両者の設定で計算して比較すると決定係数 R2 は-103.6 にも達します。これが意味するのは、これらのデータを単純に混合してモデルを学習することはできないということです。この問題に対する、素朴な対処法は、有機系のデータセットを無機系のデータセットで使われているソフトウェアと計算条件で全て再計算して作り直すことですが、現代の大規模データセットは 1 億以上の構造を含むものもあり、そうした大規模データセットを作り直す計算資源を確保することは Meta のような大手 IT 企業を除けば極めて困難です。

TEA は、この問題に対応するために開発されました。詳細は省きますが、TEA では、内殻電子の取り扱いの違いを整合させる Inner Core Energy

Alignment (ICEA) と、孤立原子エネルギーおよび原子化エネルギーに基づいて電子状態計算手法の違いを線形変換で補正する Atomization Energy Correction (AEC) の二段階により、この非互換性を解消します。前述の QM9 を用いた検証では TEA の適用により R2 は 0.9965 (RMSE 0.84 eV) まで改善し、補正パラメータの決定に必要な参照計算も実用上 1000 点程度で済みます。TEA の補正は原子化エネルギーなどの相対量のみに基づくため任意のデータセットの組に適用可能であり、「データ整備の民主化」を実現する基盤技術となっています[2]。

3. 重元素データセット HE26 と汎用 MLIP 「MACE-Osaka26」の構築

HE26 は、本研究で開発した重元素・アクチノイド領域向けの第一原理計算データセットである。既存の大規模データセットです。MPTrj や OFF23 は、一般的な結晶材料や有機分子を広く含む一方で、Am、Cm のような minor actinide や Cf のような超ウラン元素、さらに Fr、At、Ra、Po、Rn といった重元素をほとんど含んでいませんでした。そのため、既存の汎用 MLIP は周期表の広い範囲を扱えるように見えても、核燃料、核廃棄物、アクチノイド含有セラミックスなどに重要な重元素化合物には適用が難しい状況にありました。HE26 はこの欠落を埋めるために作られたデータセットです。VASP を用いた 59,438 構造の DFT 計算の結果からなります。このデータセットでは、既存データでは希少な 8 つの重元素、Am、Cm、Cf、Fr、At、Ra、Po、Rn が明示的に含まれているのが特徴です。

HE26 は大きく三つのサブセットから構成される。第一の BHE (Basic Heavy Element) は、重元素の元素固体や二元酸化物など、最も基本的な構造を集めたものです。第二の CHE (Complex Heavy Element) は、COD、文献、無機物質データベースなどから収集した、重元素を含む多元素・複雑組成の化合物構造を含みます。第三の CCFO は

Compositionally Complex Fluorite Oxides で、螢石型 ThO₂ を出発点として、アクチノイド、ランタノイド、希土類、Zr、Nbなどを置換した混合酸化物群となっています。特に CCFO は、minor actinide を含む混合酸化物燃料や、アクチノイド含有高エントロピーセラミックスを意識して設計されています。したがって HE26 は、単なる元素拡張用の補助データではなく、核材料分野で重要な組成複雑性と重元素化学を MLIP に学習させるための基盤データと位置づけることができます。

この HE26 を用いて作った MLIP が MACE-Osaka26 です。ここでは、HE26 は単独で使われるのではなく、結晶材料データである MPtrj、有機分子データである OFF23 と統合されています。MPtrj と OFF23 の間のエネルギースケールの不整合は先述の TEA プロトコルによって解消されており、HE26 は MPtrj と互換性のある計算条件で生成されているため、この統合データセットへ補正なしに直接組み込むことができるようになっています。こうして分子・結晶・重元素化合物という異なるドメインを一つの MACE モデルに学習させた結果、MACE-Osaka26 は従来モデルでは扱えなかった重元素を含めて 97 元素をカバーするモデルとなりました。

ただし、MACE-Osaka26 はまだまだ課題があります。HE26 の参照計算は PBE に基づき、一部の遷移金属を含む酸化物・フッ化物には DFT+U 補正が適用されているものの、アクチノイドで重要になるスピン軌道相互作用を扱っておらず、そもそも f 電子系における PBE 汎関数の信頼性も十分とは言えない。加えて、HE26 はデータ数が MPtrj や OFF23 に比べて小さく、広範な化学環境を提供する訓練データとはなっていない。HE26 は、今後の重元素データセット開発のあくまで出発点であり、ここで挙げた課題の解決は今後おこなっていく計画です。

4. MACE-Osaka26 モデルの検証と応用

構築した MACE-Osaka26 モデルの検証としては、まず、HE26 に含まれる重元素系を対象とした格子定数計算を実施した。まず BHE では格子定数の RMSE は 0.4488 Å となり、CHE では 133 構造を評価対象として格子定数の RMSE が 0.0477 Å となった。CCFO については、fluorite 型酸化物を対象に、Th から Cf までのアクチノイド系列で格子定数を評価したところ、MACE-Osaka26 の DFT に対する RMSE は 0.0067 Å に抑えられていることが確認された。ランタノイド、Y、Zr、Nb を含む fluorite 型酸化物でも、RMSE は 0.0124 Å と同程度の精度が得られていることも確認している。

次に、AnO₂ (An = Th, Pa, U, Np, Pu, Am, Cm, Cf) に対して二次および三次の原子間力定数計算もおこなった。計算には phono3py[7]を用い、96 原子の 2×2×2 fluorite supercell、変位幅 0.03 Å、10×10×10 q 点メッシュを採用した。Am_{0.25}U_{0.75}O₂ については、12 原子の U₄O₈ conventional cell 中の U を 1 つ Am に置換して U₃Am₁O₈ とし、これを 2×2×2 に拡張したモデルで扱った。得られた力の定数から、格子熱伝導率を Wigner transport equation を解くことで算出した。その結果、8 種の AnO₂ すべてにおいて虚数フォノン振動数が一切現れず、単一の普遍的モデルでアクチノイド系列全体の動的安定性を再現できることが確認されました。得られた格子熱伝導率の温度依存性は、実験データが存在する UO₂、ThO₂、NpO₂、PuO₂、AmO₂、CmO₂ について、300–1500 K の広い温度範囲で実験報告と整合しています。また、Am_{0.25}U_{0.75}O₂ の格子熱伝導率を純粋な UO₂ および AmO₂ と比較したところ、Am ドープによる熱伝導率の大幅な低下が定量的に再現され、Am 濃度がわずかに異なる実験値 (Am_{0.20}U_{0.80}O₂) と良好な一致を示しました。

5. おわりに

本稿では、有機・無機化合物を包括する「化学分野の基盤モデル」の構築に向けた本年度の取り

組みとして、アクチノイドを含む重元素の第一原理計算データセット HE26 の整備と、これを TEA により既存の有機系・無機系データセットと統合して学習した、97 元素対応の汎用 MLIP「MACE-Osaka26」の構築について報告しました。MACE-Osaka26 は、既存の汎用 MLIP がほとんど扱えなかった重元素領域においても格子定数を DFT と整合する精度で再現し、アクチノイド酸化物系列全体の動的安定性や格子熱伝導率の温度依存性を単一のモデルで記述できることを確認しました。特に、マイナーアクチノイドを含む混合酸化物燃料を模した Am-U-O 系の熱伝導率評価は、実験が極めて困難な系をシミュレーションで定量的に評価できることを示すと同時に、汎用 MLIP が有機・無機化学の枠を超えて原子力材料のような応用分野へ直接貢献しうることを示すものであり、化学の基盤モデルの射程の広さを物語る成果と考えています。

一方で、第 3 節で述べたとおり、スピン軌道相互作用の取り扱いや f 電子系に対する交換相関汎関数の信頼性、重元素領域の訓練データの規模など、残された課題も少なくありません。今後は、より信頼性の高い参照計算に基づくデータの拡充を進めるとともに、応用先のさらなる開拓を通じてモデルの検証と改良を重ねていく計画です。MACE-Osaka26 は前年度の MACE-Osaka24 と同様にオープンなモデルとして公開しており、周期表のほぼ全域を誰もが利用できる形でカバーする、真の意味での「化学分野の基盤モデル」の確立に向けて、今後も研究を進めていきます。

参考文献

- (1) I. Batatia, et al., J. Chem. Phys., **163**, 184110, (2025).
- (2) T. Shiota, K. Ishihara, T. M. Do, T. Mori and W. Mizukami, arXiv:2412.13088, (2024).
- (3) N. Kuroda, K. Ishihara, T. Shiota and W. Mizukami, arXiv:2603.03223, (2026).

- (4) B. Deng, et al., Nat. Mach. Intell., **5**, 1031, (2023).
- (5) D. P. Kovács, et al., J. Am. Chem. Soc., **147**, 17598, (2025).
- (6) R. Ramakrishnan, et al., Sci. Data, **1**, 140022, (2014).
- (7) A. Togo, et al., J. Phys.: Condens. Matter, **35**, 353001, (2023).