

## 科学研究費助成事業 研究成果報告書

令和 5 年 6 月 4 日現在

機関番号：12601

研究種目：研究活動スタート支援

研究期間：2021～2022

課題番号：21K20551

研究課題名（和文）原子構造のユニーク識別子の開発によるマテリアルデータベース横断研究

研究課題名（英文）Development of a unique identifier for atomic structures for meta analysis of material databases

研究代表者

村岡 恒輝（Muraoka, Koki）

東京大学・大学院工学系研究科（工学部）・助教

研究者番号：10912074

交付決定額（研究期間全体）：（直接経費） 2,400,000円

研究成果の概要（和文）：計算速度、ストレージ、およびアルゴリズム開発の進歩によって、材料の計算データベースが急速に拡張している。しかし、無機結晶材料の標準化された識別子はなく、データベースの効率的な利用が妨げられている。この問題に対処するために、マテリアルIDを開発した。この識別子は、優れたスケーラビリティを示し、構造的およびエネルギー的に異なる材料を正確に識別できた。これは、バルク結晶、表面構造、および分子を含むさまざまな化学構造に適用できる。さらに、この識別子を使用して材料データベースと他の結晶構造を結合するWebアプリケーションを開発した。

研究成果の学術的意義や社会的意義

結晶性無機材料の識別子の開発は、材料科学における標準化された識別子の欠如を解決した。既存の方法を上回る性能を持ち、スケーラブルで、さまざまな化学構造に適用できるため、学術的にも社会的にも意義深い。この研究成果により、材料のシミュレーションとデータ分析が加速され、計算データベースの効率的な利用が可能となる。このため、学术界内外でデータベースを活用した材料開発がより円滑になる。これにより、様々な産業において潜在的な応用がある新しい材料の発見が早まる。また、材料を検索するオンラインサービスにより、この研究成果はアカデミア内外、国内外で広く利用できるようになっており、成果の社会への還元を行っている。

研究成果の概要（英文）：The rapid expansion of computational materials databases, fueled by advancements in computational speeds, storage, and algorithm development, has enabled the unified treatment of materials with diverse compositions and structures. However, the efficient utilization of these databases is hindered by the lack of standardized identifiers for crystalline inorganic materials. To address this issue, we propose a generative identifier that exhibits excellent scalability and accurately differentiates between structurally and energetically distinct materials, outperforming symmetry-based methods and pairwise methods. This identifier can be applied to various chemical structures, including bulk crystals, surface structures, and molecules. Additionally, we provide an online service that uses graph ID to link material databases with user-supplied crystal structures, resulting in accelerated material simulation and data analysis.

研究分野：マテリアルズ・インフォマティクス

キーワード：識別子 化学構造 結晶構造 マテリアル データベース

## 様式 C-19、F-19-1、Z-19 (共通)

### 1. 研究開始当初の背景

人類活動の爆発的な拡大、経済競争の加速、情報通信による社会システムのかつてない高速な相互連動によって、我々は気候変動やエネルギー供給の問題といった地球規模の社会問題に直面している。このような予測不能な世界における社会問題を解決することが材料科学に期待されている。

新材料の探索は、主に出発物質や調製方法の組み合わせを試すことによって行われている。しかしながら、新材料の発見や最適化は多くの試行錯誤を必要とし、商用化には大きな経済的、時間的、人的資源の投入が必要になる。変動の激しい現代の社会問題に対応するには、より高速なオンデマンドの技術開発が求められている。

このような背景のもと、第一原理計算をはじめとする計算化学的手法と機械学習や統計的手法によって材料探索を行う、マテリアルズ・インフォマティクスに注目が集まっている。その一環として、無数の候補材料の中から有用なものを見つけ出すために、理論計算と機械学習を大規模に適用し、データベースとして公開することが近年盛んに行われている<sup>1-8</sup>。しかしながら、既存の材料データベースは共通の ID を有していないため、

1. 手元の化学構造と合致する情報を検索する
2. データの重複を見つける
3. 複数のデータベースを組み合わせる

というデータベースの基本機能を欠いてしまっていた。

### 2. 研究の目的

本研究は、あらゆる化学構造に対して一意の文字列を付与する「マテリアル ID」を開発し、その成果を用いて無機結晶材料の検索エンジンを構築することを目的とした。

### 3. 研究の方法

Materials Project<sup>1,9</sup>、OQMD<sup>10</sup>、AFLOW<sup>2</sup> といったデータベースに蓄積された無機結晶、PubChem に蓄積された有機分子を用いて、ID の一意性を検証した。また、クラウドアプリケーションを開発することにより、様々な無機結晶を検索可能な検索エンジンを開発した。

### 4. 研究成果

開発したマテリアル ID は、様々なデータベースに対して一意性を有しているとともに、並進対称性のある構造を同一視することができた。これを用いて、Materials Project<sup>1,9</sup>、OQMD<sup>10</sup>、

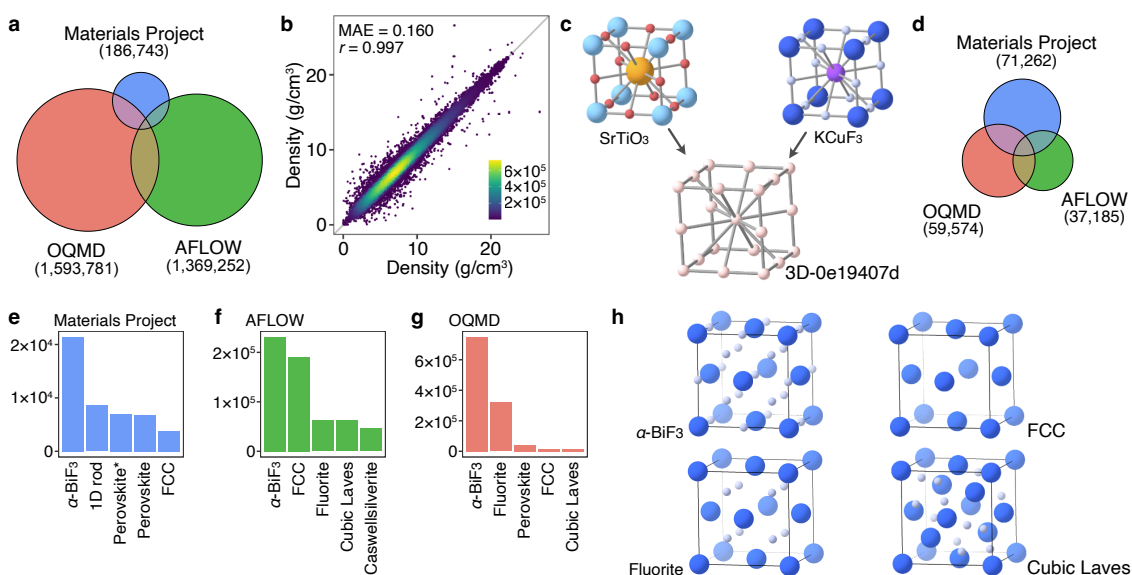


図 1 Linking structures in Materials Project, AFLOW, and OQMD using Material ID. **a** Number of unique Material IDs among three databases. **b** Comparison of densities of duplicate structures identified by Material ID. **c** Two perovskite structures with different compositions can be represented by a same Material ID by ignoring their chemistry. **d** Number of unique topologies among three databases. **e-g** Top five topologies and their number of structures in Materials Project **e**, AFLOW **f**, and OQMD **g**. **h** Some of popular topologies in the three databases.

AFLOW<sup>2</sup> の 3 つのデータベースに含まれる構造を分析した。それぞれのデータベース内のエントリーの大部分は、他のデータベースには見られない独自の構造であった (図 1a)。これらのデータベースを組み合わせることで、より広範な化学空間の探索が可能になる。重複構造のペアの密度を図 1b に示す。ほとんどのデータ点が対角線に沿って分布しており、MAE が  $0.160 \text{ g/cm}^3$ 、相関係数が 0.997 であることから、マテリアル ID がデータベース間の重複構造を正しく判定できたことがわかる。

データベース間の違いをさらに理解するために、構造からトポロジー情報のみを抽出してグラフ ID を計算した (図 1c)。これにより、異なる組成で同じトポロジーの構造を同じグラフ ID に変換できるようになる。この「匿名化された」グラフ ID を使用して、データベースから重複を削除した。その結果、Materials Project、OQMD、および AFLOW にはそれぞれ 71,262、59,574、37,185 のユニークなトポロジーが含まれていることがわかった。Materials Project は、組成と原子位置を考慮したユニークなグラフ ID の数が最も少ないものの (図 1a)、他の 2 つのデータベースと比較してトポロジーの数の面でより多様性があることが、図 1d で示された。共通部分の割合は図 1a と比較して増加したが、それでも各データベースに独自のトポロジーが存在した。それぞれのデータベースで多くの対応する構造を持つトップ 5 のトポロジーが図 1e-g に示されている。 $\alpha\text{-BiF}_3$  トポロジー (図 1h) は、すべてのデータベースで最も一般的なトポロジーであり、Materials Project のユニークなトポロジーの 10 パーセント、AFLOW の 12 パーセント、OQMD の 38 パーセントを占めていた。AFLOW と OQMD の  $\alpha\text{-BiF}_3$  トポロジーだけのエントリー数は、Materials Project の総エントリー数を上回っている。これらのデータベースには、このトポロジーをプロトタイプとして計算的に生成された多数の置換構造が含まれている可能性がある。 $\alpha\text{-BiF}_3$  の結晶構造は Bi の FCC 構造に基づいており、隣接する Bi 原子間の中点および 4 つの Bi 原子の中心に F 原子が導入された構造と見なすことができる。FCC は、Materials Project で 5 位、AFLOW で 2 位、OQMD で 4 位のトポロジーで、多くのエントリーがあるプロトタイプである。これらのエントリー数が多いのは、自然界でこのような FCC 様のプロトタイプ結晶構造がよく見られ、また応用研究の重要性ゆえであろう。同様に、Fluorite と Cubic Laves は、

**Matfinder**

**Analysis:**

Composition	Si36 O72
Space Group	71
Sites	108
Inequivalent Sites	12

**Search results:**

Name	Link
aflow:cc27b2d9d0ac0b71	<a href="#">AFLOW</a>
FER	<a href="#">International Zeolite Association</a>
mp-558025	<a href="#">The Materials Project</a>
mp-15078	<a href="#">The Materials Project</a>
aflow:a27f22612a8a931f	<a href="#">AFLOW</a>
59185	<a href="#">OQMD</a>
4113638	<a href="#">COD</a>
4113639	<a href="#">COD</a>
4113637	<a href="#">COD</a>

図 2 Web UI of the application.

FCC 構造に原子を追加して構築できる結晶であり、AFLOW および OQMD データベースに多数含まれる。このようなデータベース間の分析は、過剰な計算時間や精度不足などの問題があるため、これまで困難であった。非常に正確でスケーラブルな識別子によって、それを実行することができた。

この正確性とスケーラビリティを用いて、大規模な材料データベースから同一および類似構造を検索するウェブアプリケーションを開発した。図 2 にアプリケーションのユーザーインターフェースを示す。ユーザーはまず、構造ファイルをアップロードする。次に、アプリケーションは、指定されたファイルのグラフ ID を生成し、データベース内の同一構造を検索する。その後、ユーザーは検索結果をクリックしてデータソースの詳細ページにアクセスできる。現在、データベースには 5,000,000 以上の結晶構造が含まれている。データは、Materials Project、AFLOW、OQMD、International Zeolite Association、COD<sup>11</sup>、および PCOD<sup>12</sup> を含む複数のデータソースから取得されている<sup>9</sup>。ウェブアプリケーションは、<https://matfinder.net> で公開されている。

## 参考文献

- [1] Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002.
- [2] Calderon, C. E.; Plata, J. J.; Toher, C.; Oses, C.; Levy, O.; Fornari, M.; Natan, A.; Mehl, M. J.; Hart, G.; Nardelli, M. B.; Curtarolo, S. The AFLOW standard for high-throughput materials science calculations. *Comput. Mater. Sci.* **2015**, *108*, 233–238.
- [3] Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **2015**, *1*, 15010.
- [4] Draxl, C.; Scheffler, M. NOMAD: The FAIR concept for big data-driven materials science. *MRS Bull.* **2018**, *43*, 676–682.
- [5] Hellenbrandt, M. The Inorganic Crystal Structure Database (ICSD)—Present and Future. *Crystallogr. Rev.* **2004**, *10*, 17–22, Publisher: Taylor & Francis.
- [6] Zhang, T.; Jiang, Y.; Song, Z.; Huang, H.; He, Y.; Fang, Z.; Weng, H.; Fang, C. Catalogue of topological electronic materials. *Nature* **2019**, *566*, 475–479.
- [7] Talirz, L. et al. Materials Cloud, a platform for open computational science. *Scientific Data* **2020**, *7*, 299.
- [8] Huber, S. P. et al. Common workflows for computing material properties using different quantum engines. *npj Computational Materials* **2021**, *7*, 136.
- [9] Jain, A.; Persson, K. A.; Ceder, G. Research Update: The materials genome initiative: Data sharing and the impact of collaborative ab initio databases. *APL Materials* **2016**, *4*, 053102.
- [10] Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501–1509.
- [11] Downs, R. T.; Hall-Wallace, M. The American Mineralogist Crystal Structure Database. *American Mineralogist* **2003**, *88*, 247–250.
- [12] Pophale, R.; Cheeseman, P. A.; Deem, M. W. A database of new zeolite-like materials. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12407–12412.

## 5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 1件/うち国際共著 1件/うちオープンアクセス 0件）

1. 著者名 Xia Wei, Hou Zhufeng, Tang Jing, Li Jingjing, Chaikittisilp Watcharop, Kim Yena, Muraoka Koki, Zhang Hongjuan, He Jianping, Han Buxing, Yamauchi Yusuke	4. 巻 94
2. 論文標題 Materials informatics-guided superior electrocatalyst: A case of pyrolysis-free single-atom coordinated with N-graphene nanomesh	5. 発行年 2022年
3. 雑誌名 Nano Energy	6. 最初と最後の頁 106868 ~ 106868
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.nanoen.2021.106868	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Asada Daiki, Ikeda Tatsushi, Muraoka Koki, Nakagawa Yoshinao, Tomishige Keiichi, Nakayama Akira	4. 巻 126
2. 論文標題 Density Functional Theory Study of Deoxydehydration Reaction by TiO <sub>2</sub> -Supported Monomeric and Dimeric Molybdenum Oxide Catalysts	5. 発行年 2022年
3. 雑誌名 The Journal of Physical Chemistry C	6. 最初と最後の頁 20375 ~ 20387
掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acs.jpcc.2c06018	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計9件（うち招待講演 0件/うち国際学会 2件）

1. 発表者名 村岡 恒輝, W. Chaikittisilp, 大久保 達也
2. 発表標題 データ駆動による設計的ゼオライト合成
3. 学会等名 第37回ゼオライト研究発表会
4. 発表年 2021年

1. 発表者名 朝田 大生, 池田 龍志, 村岡 恒輝, 中川 善直, 冨重 圭一, 中山 哲
2. 発表標題 脱酸素脱水反応における MoOxHy@TiO <sub>2</sub> 触媒の構造および反応機構に関する理論的研究
3. 学会等名 第129回触媒討論会
4. 発表年 2022年

1. 発表者名 横尾 陸、池田 龍志、村岡 恒輝、中川 善直、富重 圭一、中山 哲
2. 発表標題 MoOx@Rh, ReOx@Rh, ReOx@Ir 触媒による水素化分解反応の理論的解明
3. 学会等名 第129回触媒討論会
4. 発表年 2022年

1. 発表者名 村岡 恒輝、宗形 翼、中山 哲
2. 発表標題 材料データベースを結合する識別子の開発
3. 学会等名 化学工学会 第53回秋季大会
4. 発表年 2022年

1. 発表者名 大石 宏太、村岡 恒輝、中山 哲
2. 発表標題 共晶ゼオライトのin silicoハイスループットスクリーニング
3. 学会等名 化学工学会 第53回秋季大会
4. 発表年 2022年

1. 発表者名 八日市屋 朋子、池田龍志、村岡恒輝、中山哲
2. 発表標題 吸着種間相互作用を取り込んだkinetic Monte Carloシミュレーション手法の開発：ニューラルネットワークポテンシャルの適用
3. 学会等名 第25回理論化学討論会
4. 発表年 2023年

1. 発表者名 Daiki Asada, Tatsushi Ikeda, Koki Muraoka, Yoshinao Nakagawa, Keiichi Tomishige, Akira Nakayama
2. 発表標題 Deoxydehydration reaction by TiO <sub>2</sub> -supported monomeric and dimeric molybdenum oxide: A DFT study
3. 学会等名 The 19th Korea-Japan Symposium on Catalysis (国際学会)
4. 発表年 2023年

1. 発表者名 T. Yokaichiya, T. Ikeda, K. Muraoka, and A. Nakayama
2. 発表標題 Effects of Lateral Interactions on Surface Chemical Reactions Revealed by Kinetic Monte Carlo Simulations with Neural Network Potentials
3. 学会等名 10th Asia-Pacific Association of Theoretical and Computational Chemistry Conference (APATCC-10) (国際学会)
4. 発表年 2023年

1. 発表者名 大石宏太, 村岡恒輝, 中山哲
2. 発表標題 ゼオライトヘテロ界面の構造予測とその評価
3. 学会等名 第38回ゼオライト研究発表会
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<a href="https://matfinder.net">https://matfinder.net</a>
---

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------