

# 第一原理計算に基づいたマテリアルズ・インフォマティクス Materials Informatics based on First Principles Calculations

田中 功, 世古 敦人, 大場 史康, 東後 篤史 京都大学

Isao TANAKA, Atsuto SEKO, Fumiyasu OBA and Atsushi TOGO Kyoto University

## 概要 / Abstract

多数の第一原理計算により系統的に獲得したデータをもとに、データマイニング手法を用いて、効率的な新材料開発に繋げるマテリアルズ・インフォマティクスの現状と将来展望を紹介する。この手法は、多種多様な元素、組成、構造の中から望みの特性を最適化する組合せを探索するという材料の研究開発において有用であり、今後重要性を大きく増すと予想されるが、米国をはじめとする諸外国に比べ、わが国でのこの分野の研究者は極めて少ない。

Current status and future prospects of materials informatics based on massive sets of first principles calculations are given. When data mining and high-throughput screening techniques are applied to the first principles database, selection or discovery of materials from a diversity of chemistry, composition, and structures is expected to be made efficiently. Despite the importance of the technique, however, the number of researchers in this field is currently very limited in Japan as compared to the US and other nations.

## 第一原理計算に基づいたマテリアルズ・インフォマティクスとは

材料に関する様々なデータを計算機によって統合・整理し、データマイニング技法によって必要な知識を取り出す手法のことを一般的にマテリアルズ・インフォマティクスと呼んでいる。近年、高精度な第一原理計算が現実的な時間内に多数実行できるようになり、それを活用したマテリアルズ・インフォマティクスに基づく材料開発研究が視野に入ってきた。図1に、筆者らの考える材料研究・開発の流れを示す。多数の第一原理計算結果をもとに、フォノン計算やクラスター展開法を適用することで、固溶体で有限温度といったような現実系でのデータを温度や圧力、化学ポテンシャルの関数として獲得する。そして、そのようなデータをもとに、適切な記述子(descriptor)の導入あるいは、非線形回帰分析のようなデータマイニング技法を用い、そのうえで、ハイスループット・スクリーニングを行うことで適切な実験計画(Design of Experiments)を行うのである。

わが国では、未だマテリアルズ・インフォマティクスを単なるデータベース作りというように時代錯誤に認識されることが多いが、図1に示す1次データベースを作るだけでは、材料探索には全く不十分であることを強調したい。多くの材料機能は、様々なスケールでの階層構造の性質が創発的に現れたものであり、物質の一次情報と材料機能を結ぶ物理モデルは、大抵の場合に自明ではないからである。

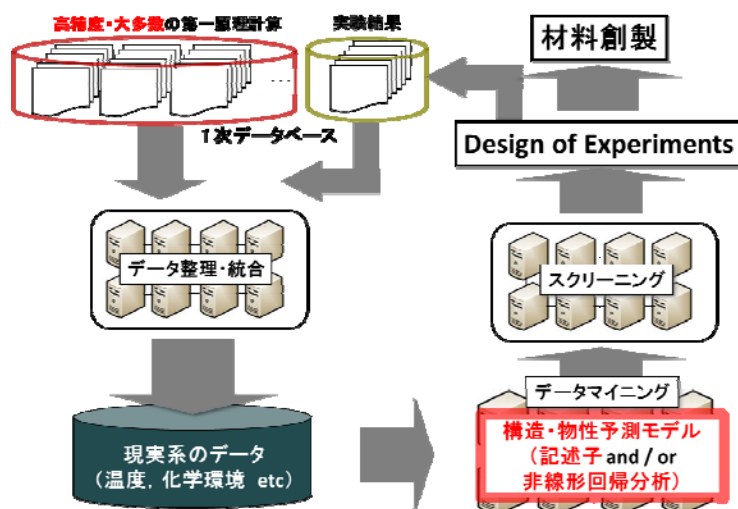


図1 第一原理計算に基づいた材料創製の流れ

## 諸外国における現況

最近、米国を中心に Materials Genome, Materials Informatics, Inverse Materials Design などのタイトルのもとで、計算機を利用して材料情報を整理し、材料開発に役立てようという大型プロジェクトが開始している。MATERIALS PROJECT[1]と AFLOWLIB[2]という WEB サイトには、それぞれ 30000 件余、16500 件のデータが収録・公開されており、それを使った精力的な研究が進められている。欧州では、CECAM の 2011 年の会議に Materials Informatics というテーマが取り上げられ、韓国では、最近 KIST に MIDAS[3]という WEB サイトが構築された。学会では、2012 年秋の MRS において、Materials Informatics ならびに Advanced Multiscale Materials Simulation—Toward Inverse Materials Computation という 2 つのシンポジウムが開催され、2013 年 6 月の PACRIM で Ceramics by Genome, 同年秋の MRS で Strategies and Techniques to Accelerate Inorganic Materials Innovation と次々に計画されており、材料科学者と情報科学者の連携が進もうとしている。

## 筆者らの取り組み

上述の 2 つのサイトに公開されている第一原理計算の結果は数万種類である。この程度の一次データ獲得は比較的容易と筆者らは考えている。図 2 には、東後が開発した自動計算オープンソース cogue[4]を利用し、立方晶ペロブスカイト結晶 530 種類について GGA レベルでの第一原理計算を VASP コードによって計算し、その格子体積と体積弾性率の関係をプロットしたものである。この計算は、8 ノードの PC により、全体が 6 時間程度で終了した。

このような一次データを利用して、特性値を予測した結果を紹介する。化合物の融点は、実験結果が豊富であるが、第一原理計算からは直接には求めにくいものである。融点の物理モデルとしては、Lindemann rule が古くから知られているが、定量的な議論に耐えるものではない。ここでは、226 種類の単体および 2 元化合物について、構成元素の原子番号、原子量、価電子数、元素の周期、化合物組成など根源的な変数のほかに、網羅的な第一原理計算で得た凝集エネルギー、体積弾性率、格子体積、最近接原子間距離を説明変数の候補とし、Sequential Forward Floating Selection 法で説明変数を取捨選択し、サポートベクトル回帰法を用いて融点を予測した。その結果を図 3 に示す。Training data が学習に用いたデータであり、Test data は検証に用いたデータである。このような粗い手法でも、平均二乗誤差 270K で予測することが可能であった[5]。

現段階では、第一原理計算に基づいたマテリアルズ・インフォマティクスは端緒についたばかりであるが、世界の趨勢を見るにつけ、適用例を増やし、経験を積んでいくことが急務と考えている。

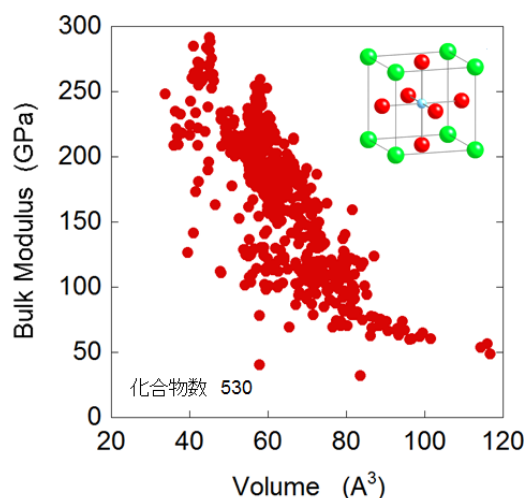


図 2 立方晶ペロブスカイト型化合物 530 種類の第一原理計算結果

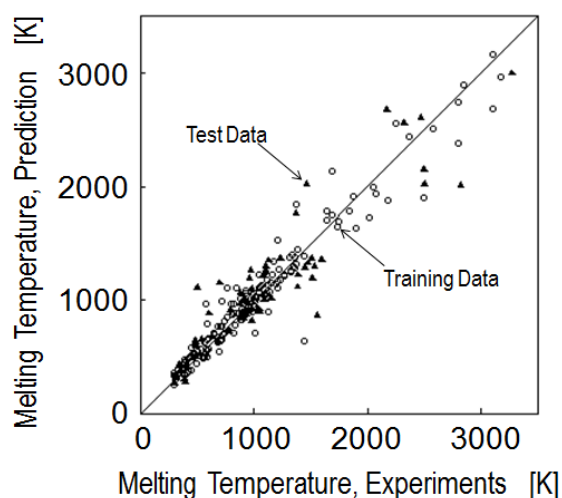


図 3 226 種類の単体および 2 元化合物の融点について機械学習による予測結果

[1] <http://www.materialsproject.org/>

[2] <http://www.aflowlib.org/>

[3] <http://cmd.kist.re.kr/>

[4] <https://github.com/atztogo/cogue>

[5] T. Maekawa, A. Seko *et al.* unpublished 2012