

独自技術による  
「Chemicals Informatics」の  
「革新性」に迫る

Chemicals  
Informatics™



◎ 株式会社 日立ハイテクソリューションズ

# 1 マテリアルズ・インフォマティクスとは？

マテリアルズ・インフォマティクス (MI : Materials Informatics) とは、統計分析などを活用したインフォマティクス（情報科学）の手法を用いて、材料開発を高効率化する取り組みです。

人工知能 (AI : Artificial Intelligence) 関連技術の機械学習に加え、物性理論、実験、シミュレーション、データベース、クラウド、セキュリティー分野などのさまざまな技術により成り立っていることが特徴です。

グローバル化の加速に伴い、新素材の開発競争は近年激化しています。MIが注目される以前の各企業は、実験を繰り返しながらさまざまなトライアル＆エラーを経て開発をおこなっていました。こうした新素材の開発には十年単位の期間に加え、数十億から数百億円の費用がかかることも珍しくないと言われています。

2010年代に入り、米国マサチューセッツ工科大学 (MIT) などの研究グループがMIを活用した共同研究を実施。実験をおこなわず、わずか1年で新材料を見つけるという快挙を成し遂げ、世界に衝撃を与えました。

現在では米国、欧州、韓国、中国などの各国で研究・開発が進んでいるほか、我が国、日本においても政府によりマテリアル革新力強化の国家戦略として位置付けられている、最先端の開発手法です。

## 2 Chemicals Informatics(ケミカルズ・インフォマティクス)とは？

日立ハイテクソリューションズが開発したChemicals Informaticsは、MIの分野における独自技術です。特許や論文などの公開文献から、独自の自然言語処理（NLP：Natural Language Processing）技術を用いて蓄積した1億を超える既知化合物データやAIが生成した新規化合物データの中から、目的の特性をもつ化合物や複合材の組み合わせを高速で探し出します。研究開発の初期段階において有望な化合物候補を選定する化合物スクリーニングや、研究開発の方向性を定める特許戦略の立案などを支援します。

既存のMIによる探索では、既存化合物の周辺の化合物しか見つけられない、複合材の探索ができないなどの課題もあり、狭い範囲の活用にとどまっていました。また、材料の配合比や製造条件のチューニングにフォーカスしたMIでは、そもそも良特性を得にくい化合物の組み合わせだと、特性値の改善効果が限定的である場合もありました。

Chemicals Informaticsは、有機・無機、モノマー・ポリマーを含む幅広い化合物データの中から既存化合物のアイデアの延長線上にない化合物や、良特性を得やすい化合物を組み合わせた複合材の探索を網羅的かつ高速におこなうことができます。次項が主な特長です。

### 3 良特性の組み合わせを発見する「掛け合わせ探索」をサポート

良特性を実現する構造や化合物の組み合わせを発見することを目的とし、その実現手段として「掛け合わせ探索」をサポートしています。

複数の部分構造を併せもった単一の化合物を探索する「単一化合物探索」と、複数の化合物を組み合わせた複合材を探索する「複合材探索」の2種類をサポートした「掛け合わせ探索」をおこなえるのが、既存のMIと異なるChemicals Informaticsの大きな特長です。

構造を掛け合わせた「単一化合物探索」は配偶子の減数分裂による交叉、化合物を掛け合わせた「複合材探索」は配偶子の接合による2倍体化と、生物の進化を模した探索でより良い特性をもつた新たな化合物、複合材の発見を支援します。

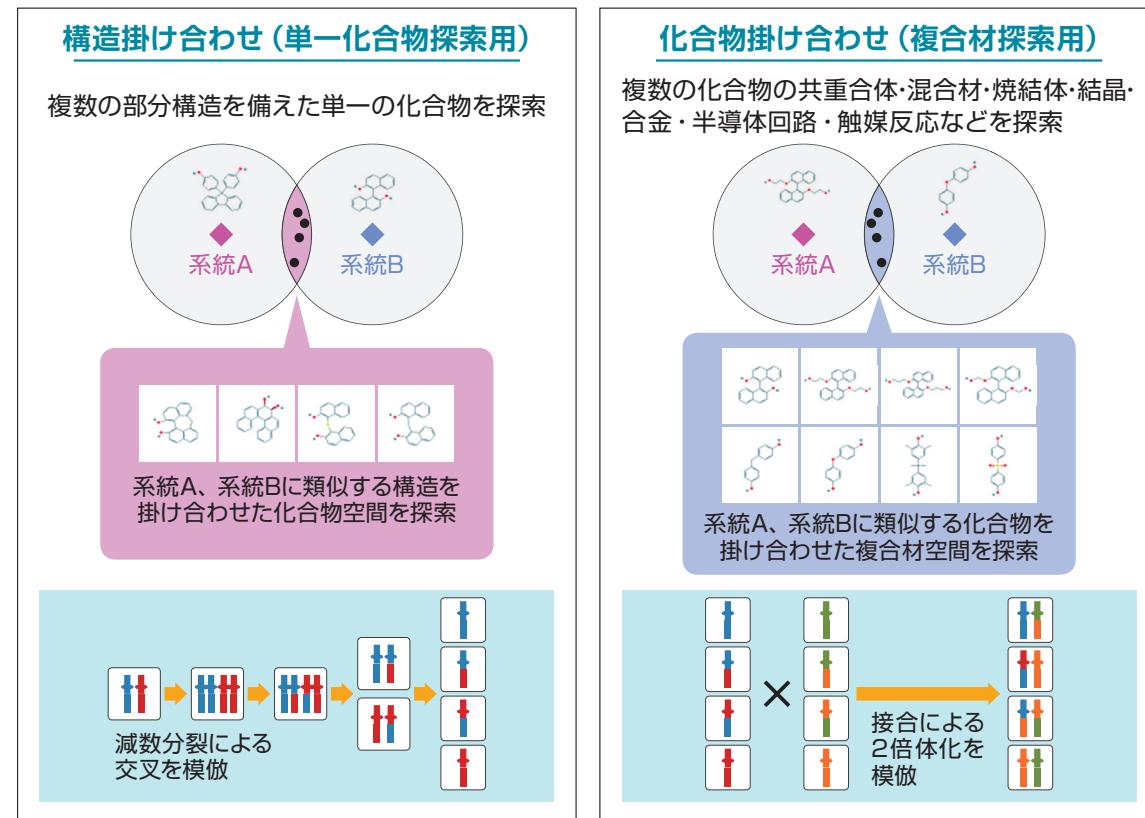


図1 構造を掛け合わせた「単一化合物探索」と、化合物を掛け合わせた「複合材探索」

## 3-1 構造を掛け合わせた「単一化合物探索」

これまでのMIでは、化合物を探索する際、目的に近い 化合物を指定し、その周辺にある化合物を探索する近傍探索の手法が用いられてきました。近傍探索では探索の範囲を広げようとすると候補化合物が指数的に増えてしまい、絞り込みが困難となるのが課題でした。

Chemicals Informaticsでは、最大64×64通りの化合物の「掛け合わせ探索」を用いることで、幅広い範囲からの有望空間の絞り込みを実現しました。

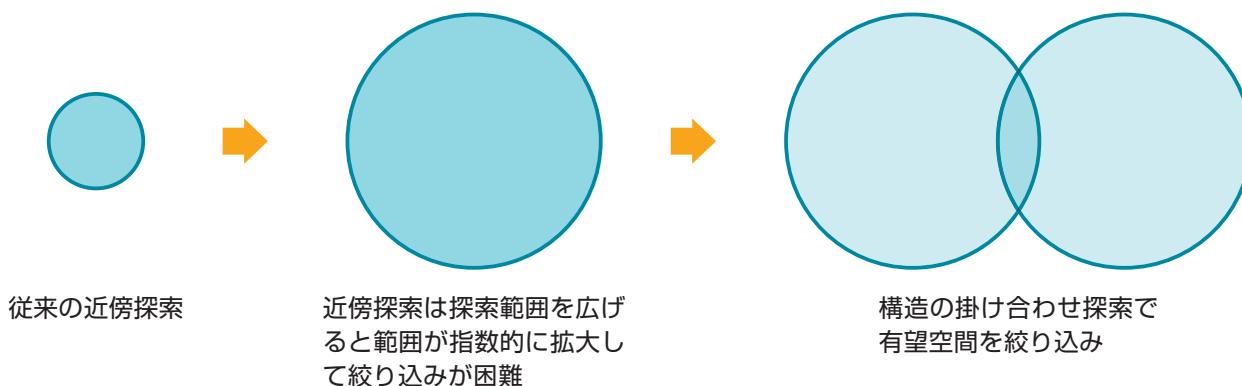


図2 構造の「掛け合わせ探索」のメリット

## 3-2 化合物を掛け合わせた「複合材探索」

化合物を指定し、その周辺にある化合物を探索する近傍探索の手法を用いるMIでは、2系統以上の化合物を組み合わせた複合材を探索することは原理的に不可能でした。また、材料の配合比や製造条件のチューニングにフォーカスしたMIでは、そもそも良特性を得にくい化合物の組み合わせだと、特性値の改善効果が限定的である場合もありました。

Chemicals Informaticsでは、最大4系統の複数化合物同士の組み合わせから、良特性に寄与する化合物の組み合わせを抽出した上で、それらの化合物に類似した化合物同士の組み合わせから、他社が押さえていない特許空白地帯<sup>1</sup>の新しい組み合わせを抽出することが可能です。

この複数の化合物を掛け合わせた「複合材探索」は、Chemicals Informaticsならではの強みです。共重合体・混合材・焼結体・結晶・合金・半導体回路・触媒反応などの幅広い分野で、良特性に寄与する化合物の組み合わせを探索することが可能です。

1 Chemicals Informatics の探索結果で特許情報欄が空白となっている(=特許未出願)地帯を指す

### 3-3 膨大な公開データとAIが生成した新規化合物データを活用

Chemicals Informaticsでは、米国の公的データベースが登録する膨大な化合物データと特許・論文などの文献データの中から、独自のNLP技術で化合物の特性値などを抽出し、Chemicals Informaticsの化合物データベースに収録しています。また、独自のAIが生成した新しい構造の化合物も活用することができます。近年、さまざまな新構造の生成手法が生まれていますが、Chemicals Informaticsでは、すでに何らかの特性が判明している既知の化合物の構造を一部置換したり、付加したりすることによって新たな構造を生成する手法を採用しています。これにより、実際に合成した際に特性をもつ可能性の高い化合物のみを生成します。また、AIが生み出した化合物の構造はすべて既存化合物と比較し、新規性を確認しています。これには高度な技術と計算リソースが必要なため省略しているケースもありますので、注意が必要です。

1億1680万個の既存化合物データと、あらかじめAIで生成した新構造1140万と、105か国2500万特許から抽出した61種類4億個の特性データの中から、最大64×64通りの構造を掛け合せた良特性の化合物や、最大4系統の化合物を掛け合せた良特性の複合材を、1回あたり2~40分で高速に探索します。これにより、化合物1.28億個の2~64乗通りに相当する幅広い範囲から、良特性を実現する構造・化合物の新たな組み合わせを抽出することができます。

## 4 「Chemicals Informatics」の広い探索領域

Chemicals Informaticsは、構造・化合物の掛け合わせ探索によって、自社だけでなく他社も見落としている広大な化合物・複合材空間を高速に探索します。

自社プライベートデータが中心のMIの探索領域は、プライベートデータが存在する領域の周辺を深堀することに長けますが、自社が未だカバーしていない広範な領域の探索は難しいものでした。Chemicals Informaticsは、膨大な公開データを活用し、進化の樹形図に沿って過去からフロンティア領域に向けて掛け合わせ探索を繰り返すことで、良特性が期待できる特許空白地帯の構造・化合物の組み合わせを、幅広い範囲から網羅的に見つけ出すことができます。

Chemicals Informaticsでは、指定した構造・化合物の組み合わせごとに61種類の特性値を予測し、良特性に寄与する組み合わせを抽出します。さらに、その近傍の特許空白地帯の構造・化合物の組み合わせを、高速に探索することができます。

また、予測の根拠として特性値ごとの特許分布数と特許リンクを表示し、特性値ごとに分類された特許リンクから製法を確認することができます。注目する化合物・複合材の特許が存在しない場合でも、類似の化合物・複合材から製法を取得可能です。加えて、注目する化合物・複合材の別の用途を61特性の特許リンクから確認することもできます。

4 「Chemicals Informatics」とプライベートDBを用いたMIの探索領域の差異

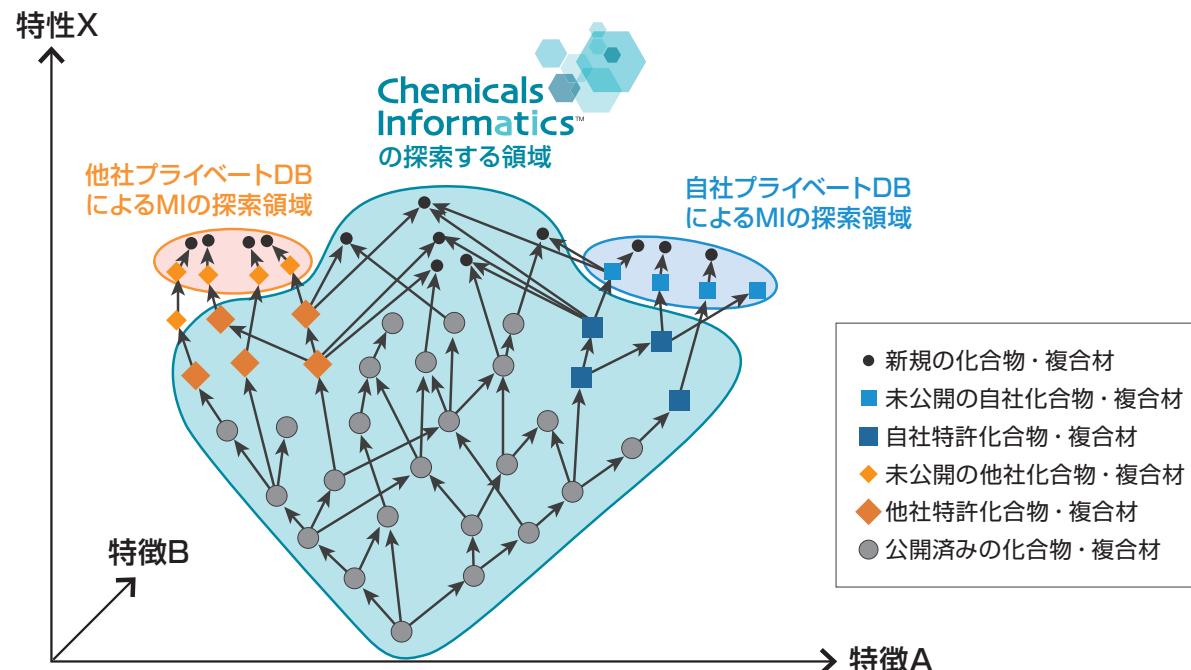


図3 「Chemicals Informatics」とプライベートDBを用いたMIの探索領域の差異

# 5 「Chemicals Informatics」の組み合わせ最適化 V.S. 配合比チューニング

下図は、Chemical Informaticsの大局的な組み合わせ最適化と、配合比チューニングの探索結果にどのような差があるかを示します。

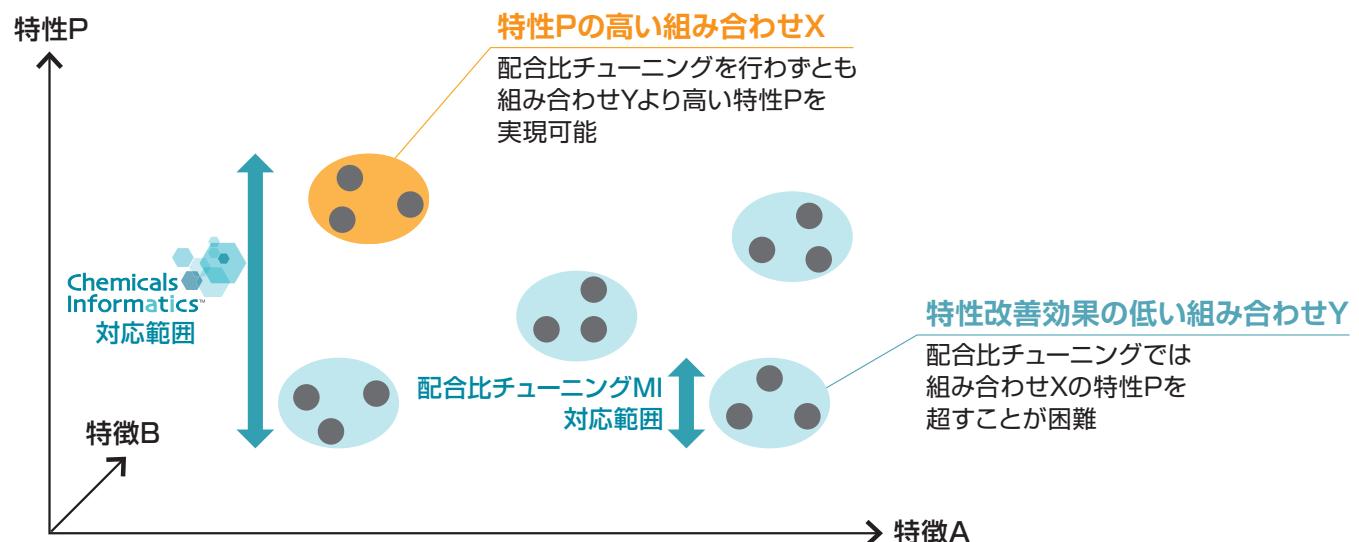


図4 CIの組み合わせ最適化 V.S. 配合比チューニング(その1)

## 5 「Chemicals Informatics」の組み合わせ最適化 V.S. 配合比チューニング

配合比チューニングMIを用いてあらかじめ定めた化合物の組み合わせの最適な配合比をチューニングする際には、その組み合わせでは他社特許がすでに数多く存在していたり、特性改善効果が限定的であったりするケースも見受けられます。

そのため、Chemicals Informaticsは、組み合わせる化合物自体を変更する組み合わせの最適化を、チューニングより前の段階でおこなうとより効率が良くなるものと考え、良特性で特許空白地帯が多くより良い特性が期待できる組み合わせを高速でスクリーニングします。これにより、後続の配合比チューニングの効率を高め、早期に良い結果を導き出すことが可能となります。

また、組み合わせを最適化するChemicals Informaticsは、意図した特性を平均的な製法で再現できる最適な組み合わせを幅広い範囲から大局的に探索します。製法・配合比は特性値ごとに分類した特許リンクから取得可能です。

一方、配合比をチューニングするMIは、局所的なチューニングにフォーカスしています。つまり、製法や配合比の工夫で特性を高めても、改善効果が限定的な可能性もあることを右図は端的に表しています。

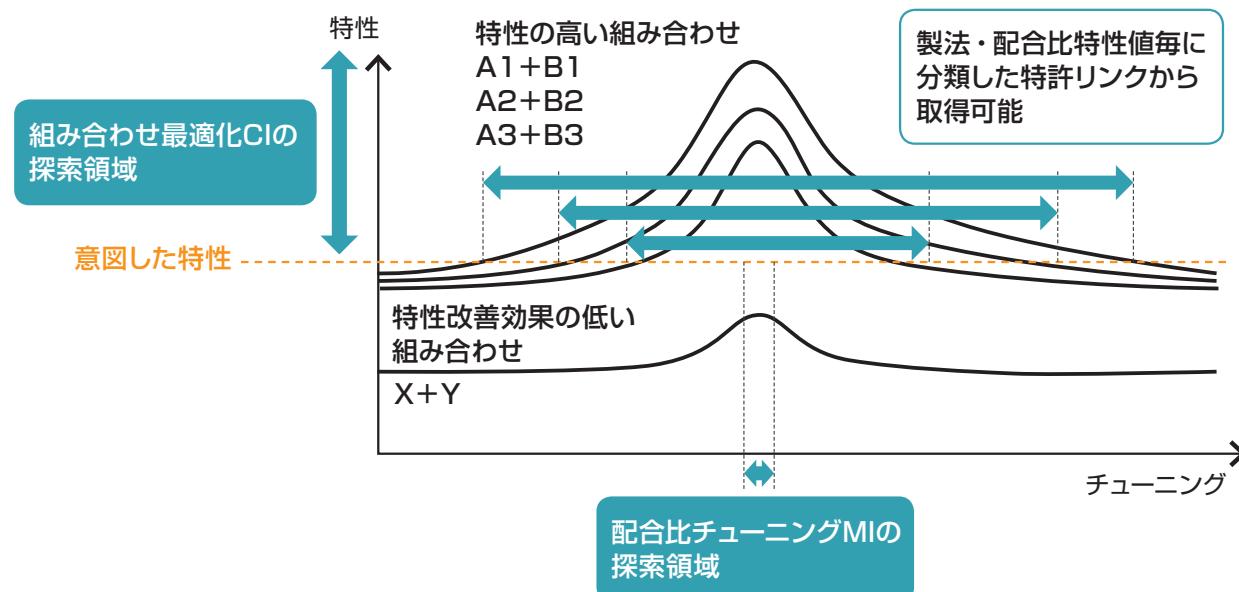


図5 CIの組み合わせ最適化 V.S. 配合比チューニング(その2)

## 6 まとめ

「Chemicals Informatics」の大きな特長は、膨大な既存化合物データと、AIで生成した新構造データと、105か国の特許から抽出した特性データを用いて、構造・化合物の掛け合わせ探索により、良特性に寄与する新たな構造・化合物の組み合わせを網羅的かつ高速に探索できる点です。

すでにさまざまな企業の特許化合物を起点とする探索をおこない、未だ特許化されていない有望な新構造の化合物や、新しい組み合わせの複合材を多数発見しており、Chemicals Informaticsが意図通りに広範囲を探索できていることを確認しています。

分野数×特性数の掛け算により、Chemicals Informaticsの適用範囲は今後もあらゆる領域へと広がっていきます。幅広い用途でのご利用を目指し、さらなる拡張を進めていく予定です。