



IBM Research

# AI分子デザインが拓く これからのマテリアル・デザイン

日本IBM 東京基礎研究所 武田征士

実践データ駆動科学オンラインセミナー  
23 August, 2021

# Self Introduction

武田 征士  
(Takeda, Seiji)

2010年           応用量子光学（レーザ物理）で博士号取得 @ 慶應義塾大学大学院  
2011-2012 年   特任助教（慶應）・客員研究員（Ecole Centrale Lyon）  
2012年           日本 IBM 中途入社（東京基礎研究所）  
                  光インターコネクト、シリコンフォトニクスの研究開発に従事  
2014年           Material Discovery プロジェクト立ち上げ  
2015-2021 年   IBM Research グローバルチームにて、  
                  Material Discoveryの戦略立案・プロジェクト推進



Seiji Takeda  
@SeijiTkd

# 3000

Researchers

# 19

Locations

# 6

Continents

6 Nobel Laureates  
10 Medals of Technology  
5 National Medals of Science  
6 Turing Awards



- (1) Future of Computing, Accelerated Discovery
- (2) AIによる分子デザイン手法
- (3) IBM 東京基礎研究所の分子デザイン技術
  - 手法概要
  - 事例紹介
- (4) これからのマテリアル・デザイン – AIと人の共創

(1) Future of Computing, Accelerated Discovery

(2) AIによる分子デザイン手法

(3) IBM 東京基礎研究所の分子デザイン技術

- 手法概要
- 事例紹介

(4) これからのマテリアル・デザイン－AIと人の共創

## 5 in 5

IBM Research

持続可能な  
未来を実現  
するために  
新たな材料  
の発見プロ  
セスを加速

<https://www.ibm.com/blogs/think/jp-ja/ibm-5-in-5-accelerating-process-of-discovery/>

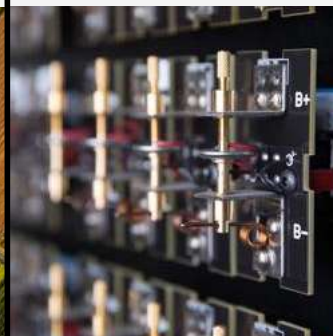
気候変動を緩和するために、二酸化炭素を回収し、有用なものに変える

炭素排出を削減しながら、増加する人口を養うため、母なる大地をモデリングする

世界を見直さなければならなくなる前に、バッテリーを再考する

持続可能な材料や製品を開発し、持続可能な地球を実現する

より健康な未来のために、過去から学ぶ





# Accelerated Material Discovery は 材料開発による持続的な世界の実現を目指します



# 科学的発見には多くの時間がかかります

1つの新しい材料を発見するのに、  
10-100億円以上の投資と  
約10年間の歳月がかかります





# 科学は、「発見の加速」の時代へと 移りつつあります

スピード、自動化、規模の拡大



Accelerated  
Discovery

## 経験的科学 第1のパラダイム



天体観測



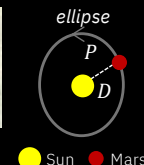
- ・ 観察
- ・ 実験

## 理論的科学 第2のパラダイム

ケプラーの法則  
(Kepler, 1600 – 1619)



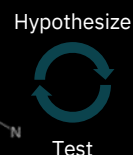
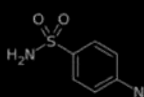
$$\frac{D^3}{P^2} = k$$



- ・ 科学法則
- ・ 天文学、物理、化学  
生物学

## 計算的科学 第3のパラダイム

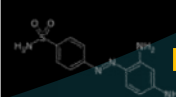
Drug design and  
discovery



- ・ シミュレーション
- ・ 分子動力学
- ・ 機械的モデル

## ビッグデータによる科学 第4のパラダイム

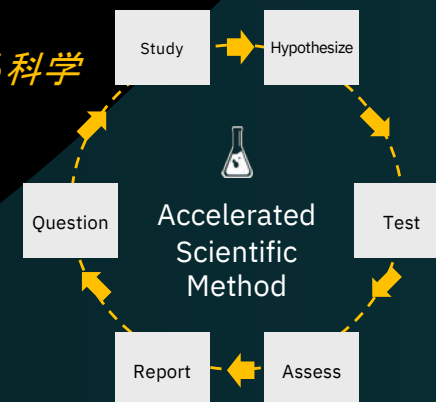
Today: AI-transformation in science



Big data

- ・ ビッグデータ、機械学習
- ・ パターン、異常検知
- ・ 可視化

AI



- ・ 大規模な科学的知識
- ・ AIによる仮説生成
- ・ 自動検証

1つの新しい材料を発見するのに、  
10-100億円以上の投資と  
約10年間の歳月がかかります



時間とコストの両面において  
90%の削減を目指します







スーパーコンピュータ



量子コンピュータ



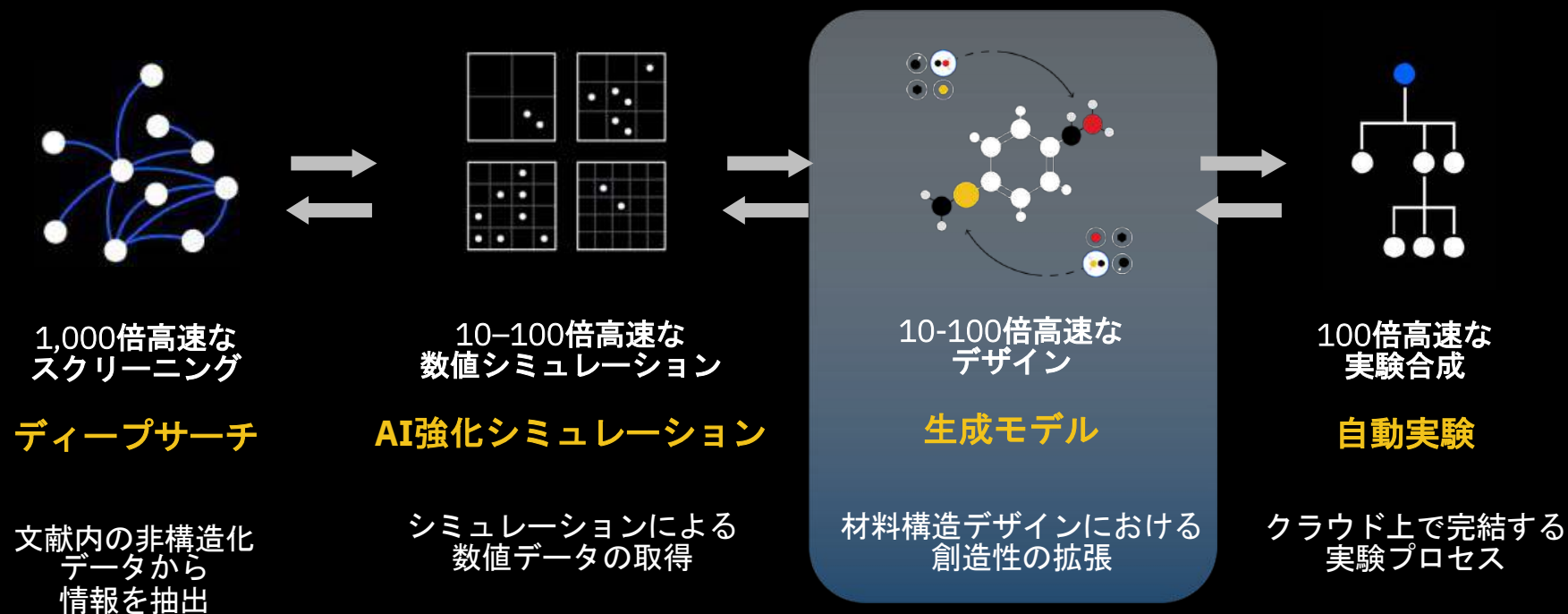
人工知能



ハイブリッドクラウド

# Future of Computing Accelerated Discovery for Materials

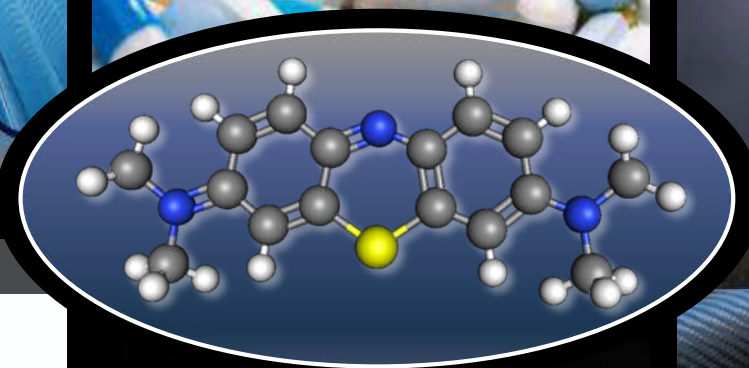
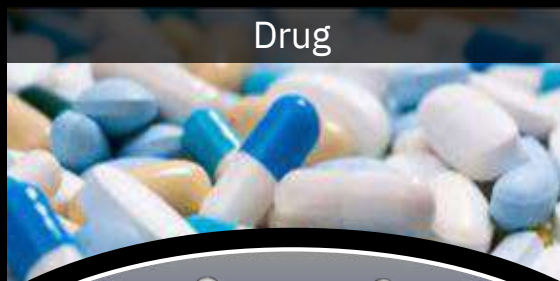
IBM Research



すべてのプロセスは、オンプレミス、クラウド、HPC、AI、量子コンピュータをふくむ  
ハイブリッドクラウド上で動作します

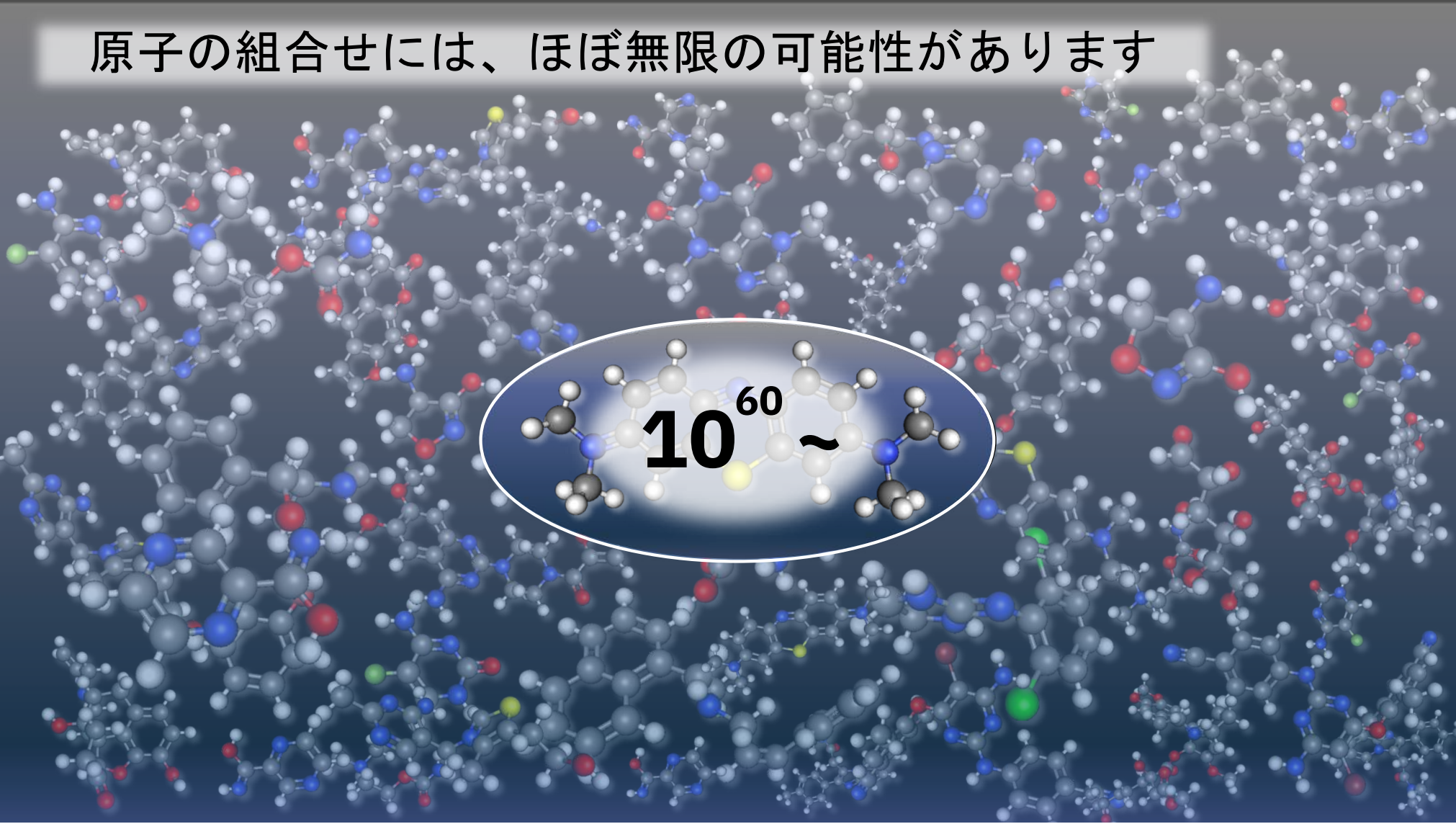


# Molecular design is a key technology across broad material domains





原子の組合せには、ほぼ無限の可能性がありますが



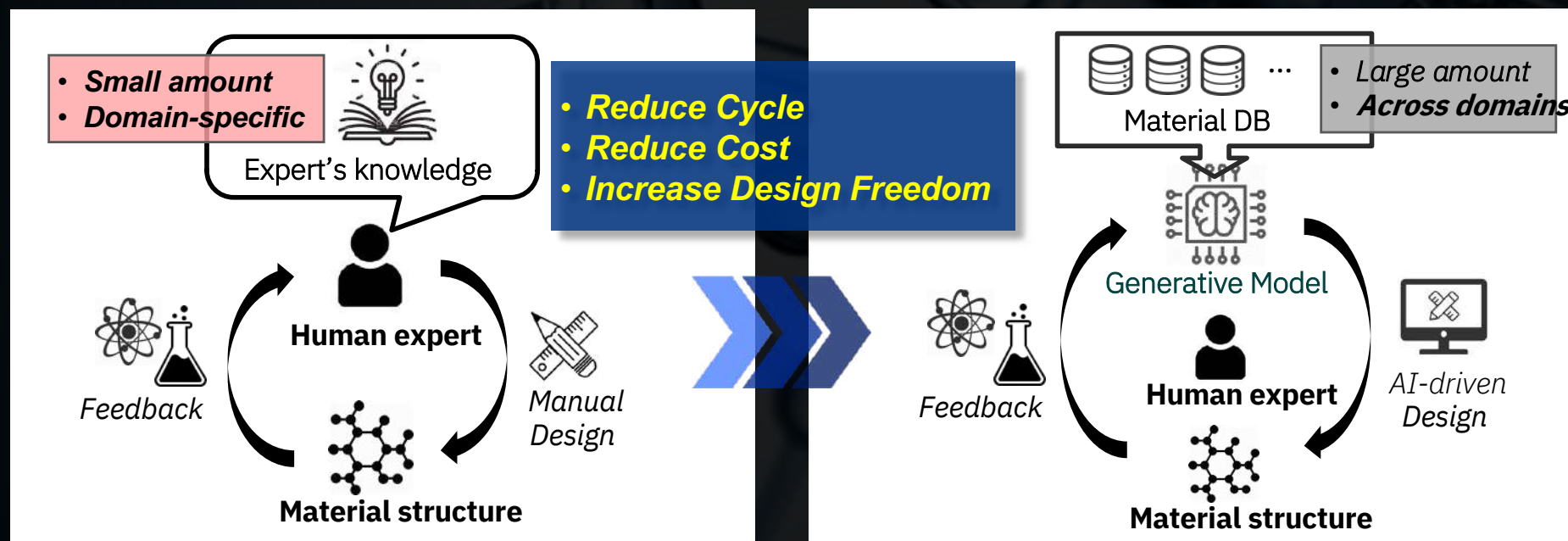
**10<sup>60</sup> ~**

# AI-driven Molecular Design



Today's total lead time to develop new material is **10-20 years**.

- Trial-and-error process by human expert's knowledge, experience, and intuition.
- The variety of design is limited by expert's unconscious bias.



# AIによる分子デザインに求められるもの

基本性能：所望の性質を持つ分子構造（および配合比など）を、高速かつ多岐にわたり、自動でデザインする機能

## デザイン・スペック

- デザイン速度
- 構造の多様性

## 分子の妥当性

- 安定性・合成コスト
- 分野固有の妥当性

## モデルの透明性

- 原子単位での微調整
- モデルの解釈可能性

## データのマルチモダリティ

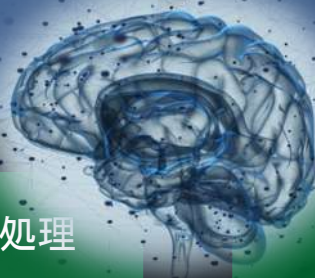
- 分子グラフ
- 画像(X線イメージ, etc.)
- スペクトル

## 構造デザイン以外の処理

- 分野固有の構造処理
- ワークフロー設計

## 利用の簡便性

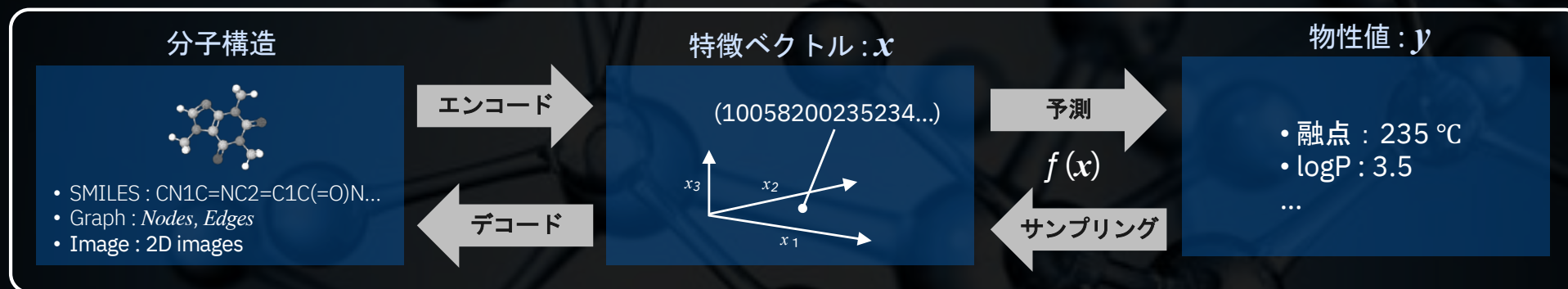
- 低いAI学習コスト
- ユーザーインターフェース





# AIによる分子デザインの基本コンセプト

- 分子から物性値を予測するモデルを作成
- そのモデルを元に、様々な生成アルゴリズムを用いてデザイン

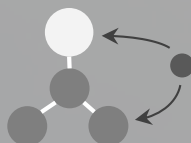


## DNN以外のAIおよび古典的手法

分子グラフを直接操作するため、デザイン過程の解釈性が高く、微調整が可能。  
スピードがやや遅い。部分構造情報なので予測精度高くない。

- バーチャル・スクリーニング
- 進化的アルゴリズム
- 分子グラフの逆問題

...



## Deep Neural Network (DNN)による手法

SMILESや分子グラフをデザインする生成モデル。  
事前学習が必要かつ、構造の微調整が難しい。  
スピード速い、生情報扱うので予測精度高い可能性。

- 変分オートエンコーダ (VAE)
- 敵対的生成ネットワーク (GAN)
- トランスフォーマー

...



(1) Future of Computing, Accelerated Discovery

(2) AIによる分子デザイン手法

(3) IBM 東京基礎研究所の分子デザイン技術

- 手法概要
- 事例紹介

(4) これからのマテリアル・デザイン – AIと人の共創

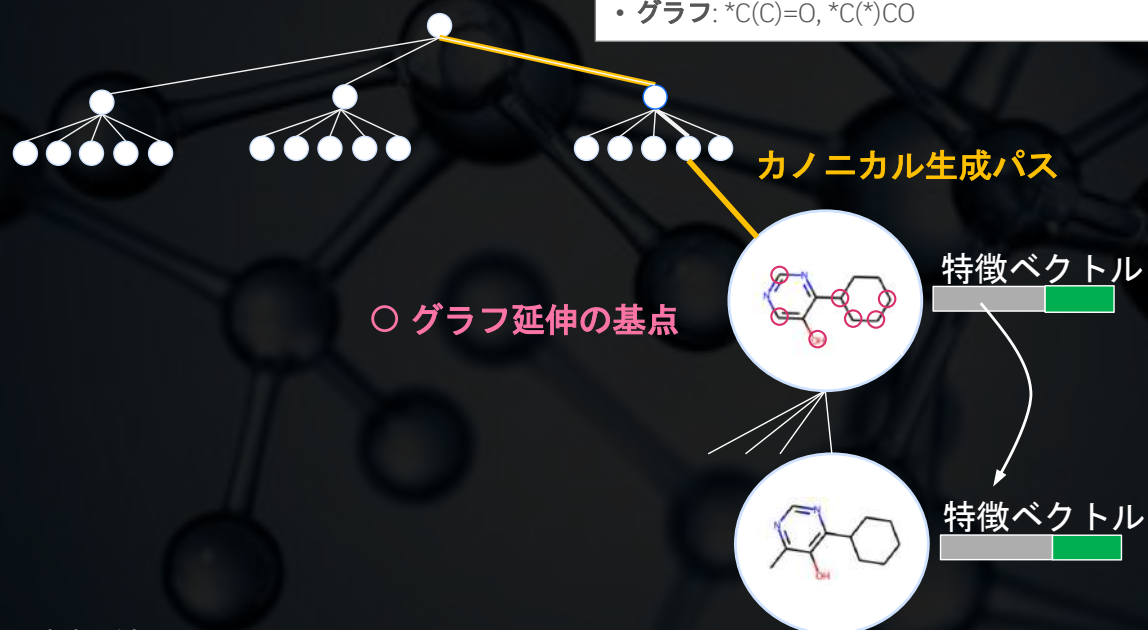


# グラフ理論を元にした分子生成アルゴリズム

- 与えられたグラフリソースを、**漏れなく・重複なく**連結
- 部分構造や原子数の制約を導入可能

グラフリソース (building block)

- 原子: C, O, N, F ...
- リング: c1ccccc1, C1CCCC1, ...
- グラフ: \*C(C)=O, \*C(\*)CO



固定特徴ベクトル

- ユーザー定義特徴量
- ...

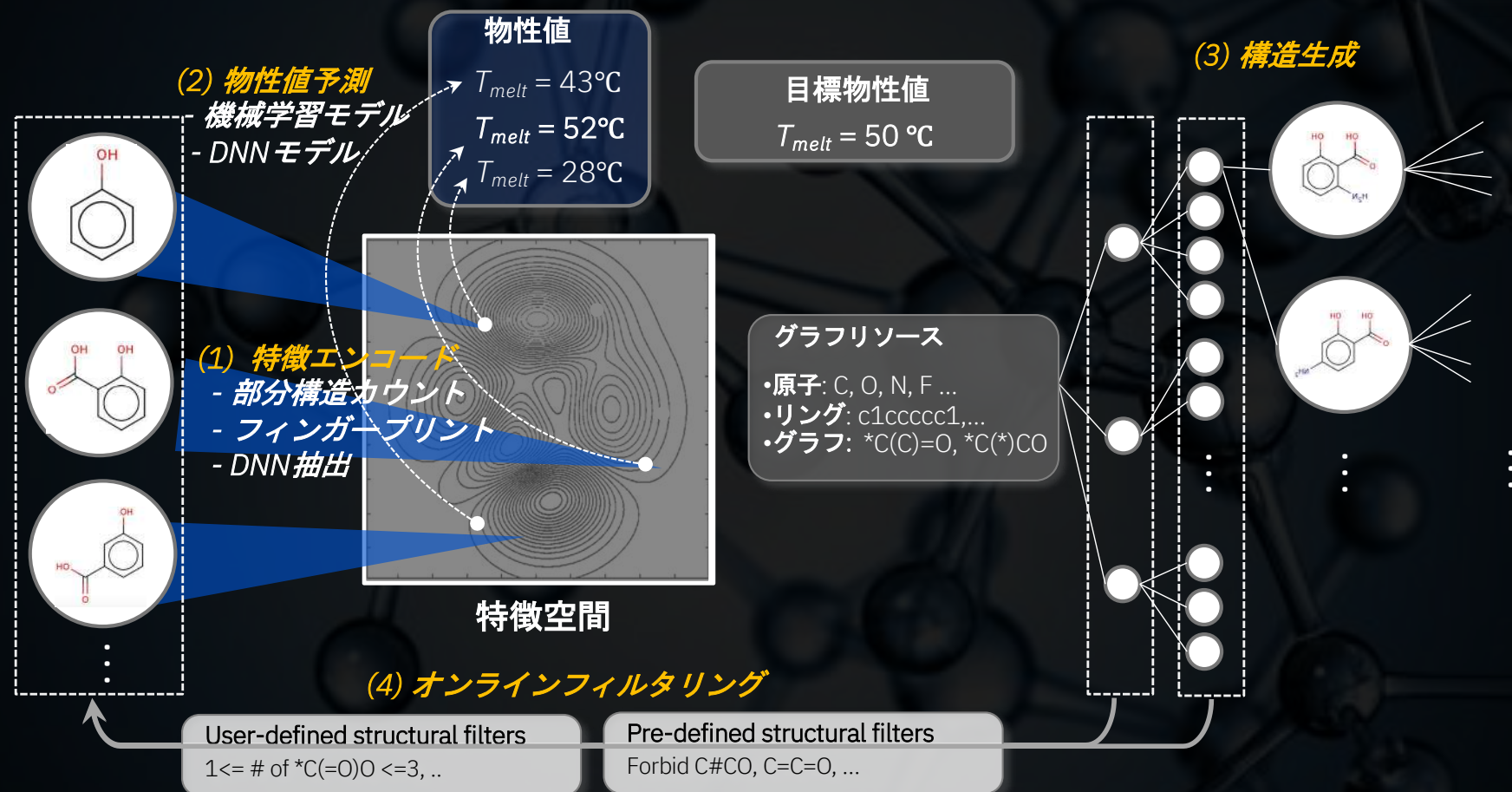
動的特徴ベクトル

- 原子数
- 部分構造数
- リング数
- ...

Reference for the fundamental algorithm

- B. D. McKay, "Isomorph-free exhaustive generation", *Journal of Algorithms*, 26 (1998) 306-324.
- S. G. Hartke and A. Radcliffe, "McKay's canonical graph labeling algorithm", In *Communicating Mathematics*, 479, 99-111. American Mathematical Society, 2009.

# グラフ理論を元にした分子生成アルゴリズム



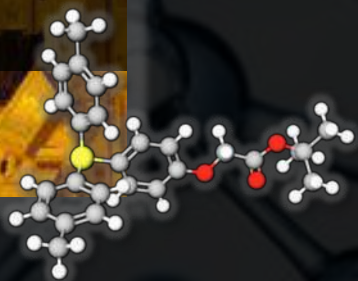
# 事例紹介

IBM



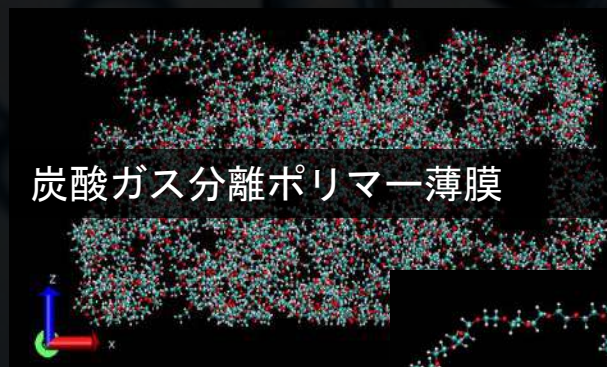
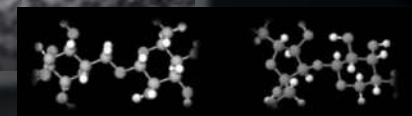
フォトレジスト材料  
IBM Project Photoresist

<https://www.research.ibm.com/science/photoresist/>



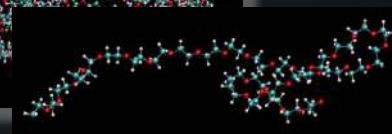
食品（糖質）

<https://tabrasa.jp/analytics>



炭酸ガス分離ポリマー薄膜

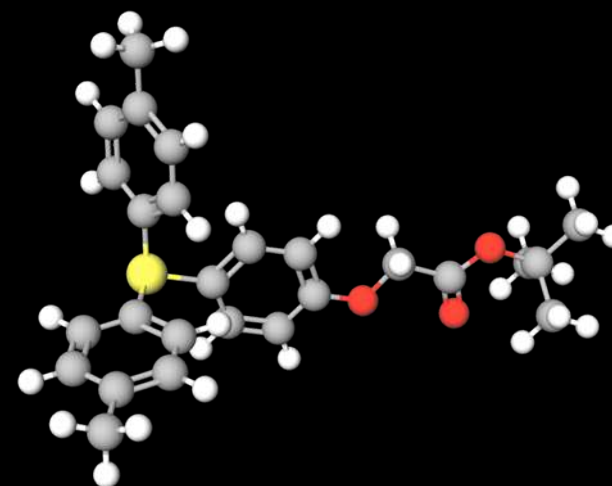
<https://research.ibm.com/blog/carbon-capture-separation-storage>





# 半導体プロセス用の材料（フォトレジスト）デザイン

- 微細配線パターン形成に必要なフォトレジスト内に含まれるPAG（光酸発生剤）は、その毒性ゆえに近年、環境規制の対象とされています。

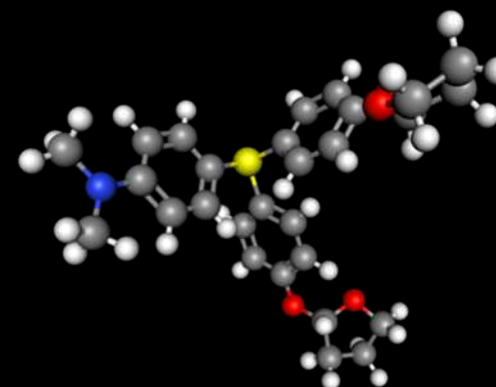


- Research website (Project Photoresist) <https://www.research.ibm.com/science/photoresist/>
- THINK Blog Japan 「ガジェットをより持続可能にする材料を発見」  
<https://www.ibm.com/blogs/think/jp-ja/accelerated-discovery-sustainable-pags/>

# ターゲット設定・データ取得

## (1) 5つの目標物性値にフォーカス

- LogWS (水溶性)
- LogP (脂溶性)
- BioDeg\_LogHalfLife (生分解性)
- CATMOS\_LD50 (毒性)
- Energy\_gap (HOMO-LUMO エネルギー差)



## (2) Deep SearchによりUS特許文献からデータ抽出

A typical structure of a cation of ionic PAG

## (3) AI強化シミュレーションにより、Energy gapなど不足情報を補完

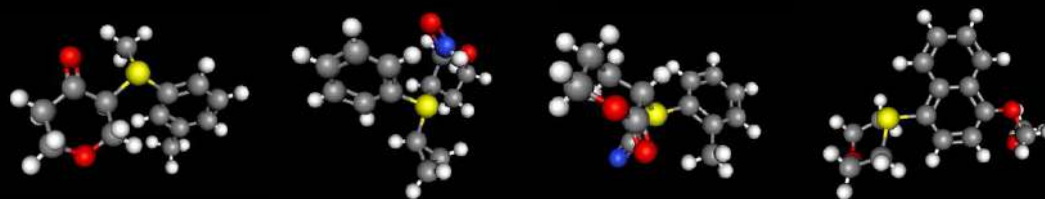
|   | A          | B                                                                       | C          | D          | E          | F          | G          | H          | I         | J         | K         | L         | M         | N          |
|---|------------|-------------------------------------------------------------------------|------------|------------|------------|------------|------------|------------|-----------|-----------|-----------|-----------|-----------|------------|
|   | MoleculeID | SMILES                                                                  | MP         | BP         | LogWS      | LogP       | LogKoc     | BioDeg_Log | CATMoS_VT | CATMoS_NT | CATMoS_EP | CATMoS_GH | CATMoS_LD | LogBCF     |
| 1 | AUTOGEN_o  | CCCCCc1ccc([I+](c2ccccc2)C1=CC=CC=C2)C1=CC=CC=C2                        | 63.1252109 | 347.8132   | -6.8673989 | 5.25234019 | 4.04945158 | 1.99799145 | 0         | 0         | 2         | 3         | 271       | 2.93534508 |
| 2 | AUTOGEN_o  | C1=CC=CC=C1C([SH+])C1=CC=CC=C1                                          | 85.2743317 | 335.686548 | -3.2566832 | 3.5842302  | 3.31045825 | 1.0856501  | 0         | 0         | 3         | 4         | 1452      | 2.30183493 |
| 3 | AUTOGEN_o  | C1=CC=CC=C1C([I+](c2ccccc2)F)C1=CC=CC=C1                                | 70.2147453 | 333.795872 | -5.4762933 | 5.48653884 | 3.14939592 | 1.99776864 | 1         | 0         | 1         | 2         | 38        | 2.98998536 |
| 4 | AUTOGEN_o  | c1ccc(COc2ccc([S+](c3ccccc3)C1=CC=CC=C2)C1=CC=CC=C2)cc1                 | 154.052828 | 471.176687 | -5.1040206 | 6.43518079 | 5.12801727 | 2.19046646 | 0         | 1         | 3         | 5         | 2737      | 2.41810413 |
| 5 | AUTOGEN_o  | C1=CC=CC=C1C([S+](c2ccc(C1=CC=CC=C2)C1=CC=CC=C2)C1=CC=CC=C2)C1=CC=CC=C1 | 255.849236 | 449.38699  | -7.2613809 | 7.49537151 | 5.40563659 | 2.49282343 | 0         | 1         | 3         | 5         | 2772      | 5.56330513 |
| 6 | AUTOGEN_o  | O=C(c1ccc(OCC)cc1)c1ccc1                                                | 253.838921 | 472.46948  | -5.68993   | 7.9259928  | 5.11514936 | 2.40790114 | 0         | 1         | 4         | 5         | 6048      | 2.77368496 |
| 7 | AUTOGEN_o  | CCCCCCCCC1(c2ccccc2)C1=CC=CC=C2                                         | 126.183792 | 338.507254 | -6.5546511 | 4.00108218 | 2.98685063 | 1.12818432 | 0         | 0         | 2         | 3         | 230       | 2.78084846 |
| 8 | AUTOGEN_o  | CS([S+])(C)C                                                            | -63.844711 | 146.946482 | -1.8659851 | 1.44607032 | 2.15037486 | 1.84669661 | 0         | 0         | 3         | 4         | 994       | 0.76820709 |



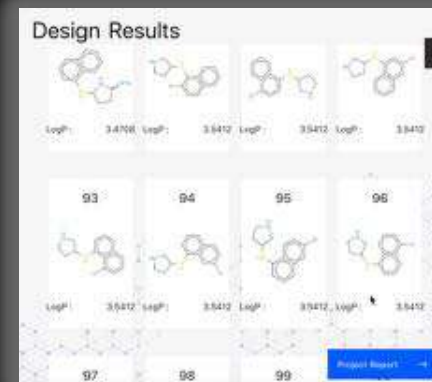
# 分子構造のデザイン

データ取得から合成まで数年かかるフォトレジスト材料の発見が、半年間に短縮されました

- 約 **3,000 種類** の多様性に富んだPAG構造が、高速にデザインされました。
- **1分子/7秒** の速度でデザイン – 数日間かかる初期構造デザインプロセスを**10-100倍程度短縮**。
- デザインされたPAG分子は、RoboRXNにて合成されました。



デザインされたPAG構造の例



MolGX の画面



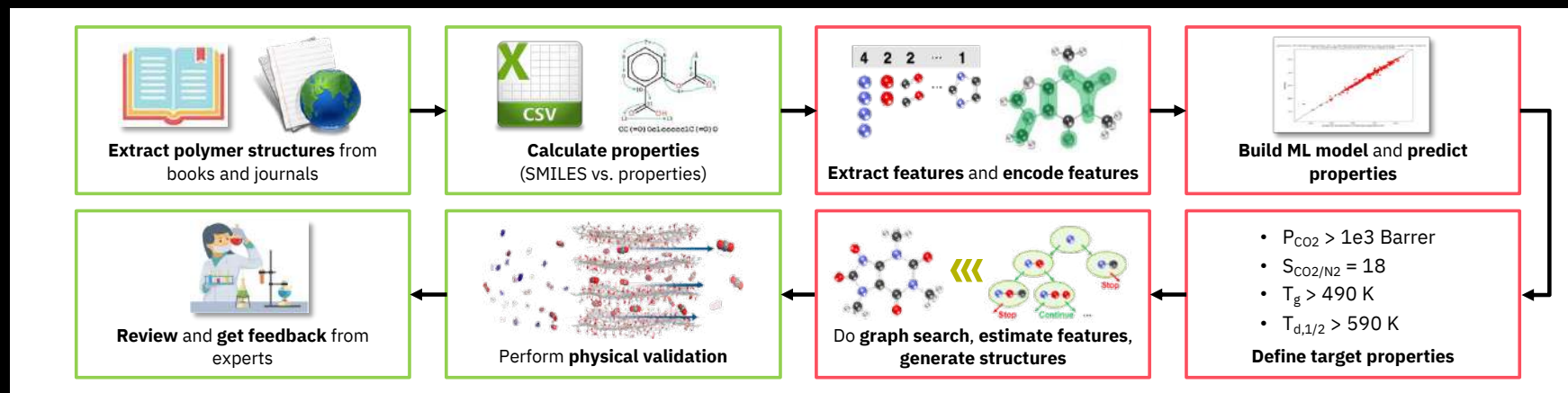
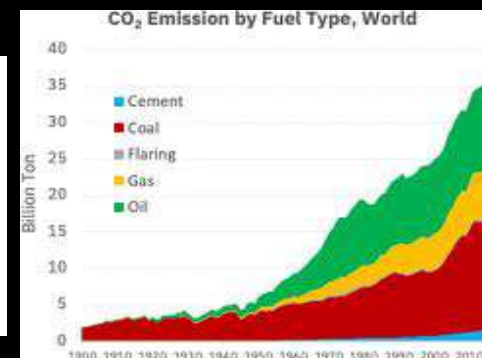
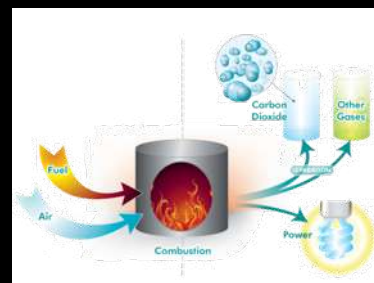
RoboRXN

# CO<sub>2</sub>分離用ポリマー薄膜のデザイン #1

高い**選択性 (selectivity)** と **透過性(permeability)** をもつCO<sub>2</sub>の分離において、ポリマー薄膜をデザインしました。

なぜポリマー薄膜なのか？

- 保管や取り扱いが簡単で、危険性が低い
- 自身は化学反応をしない、シンプルな passive operation
- SO<sub>x</sub> や NO<sub>x</sub> への耐性が高い
- 燃焼後回収 (Post Combustion Process)に適している

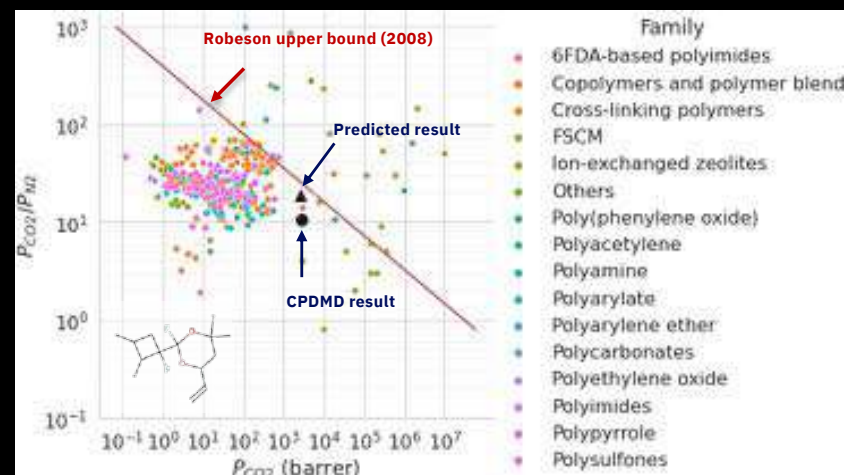
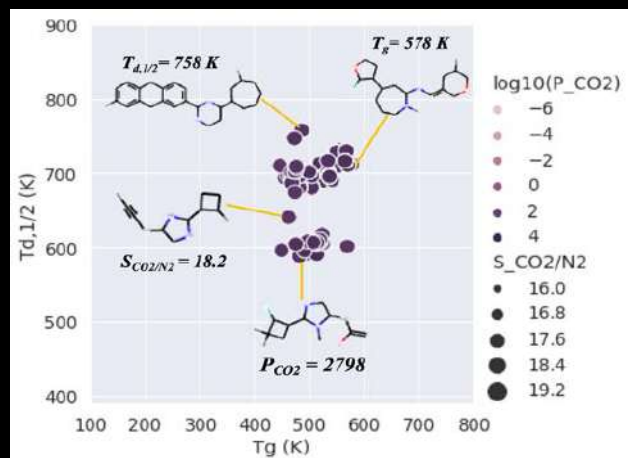


# CO<sub>2</sub>分離用ポリマー薄膜のデザイン #2

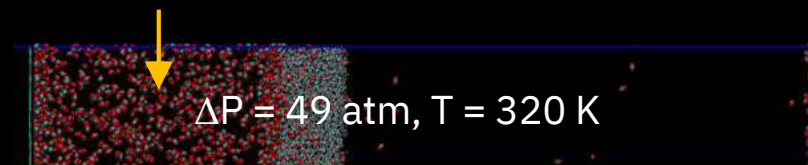
## 目標物性値

- CO<sub>2</sub>透過率:  $P_{\text{CO}_2} > 10^3$  Barrer
- CO<sub>2</sub>/N<sub>2</sub> 選択率:  $S_{\text{CO}_2/\text{N}_2} > 18$
- ガラス転移点:  $T_g > 490$  K
- 熱分解温度:  $T_{d,1/2} > 590$  K

新規に生成された構造  
(~ 500 個)

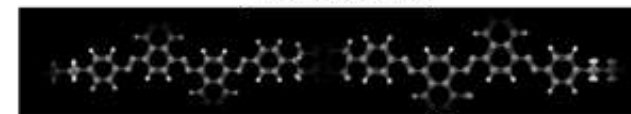
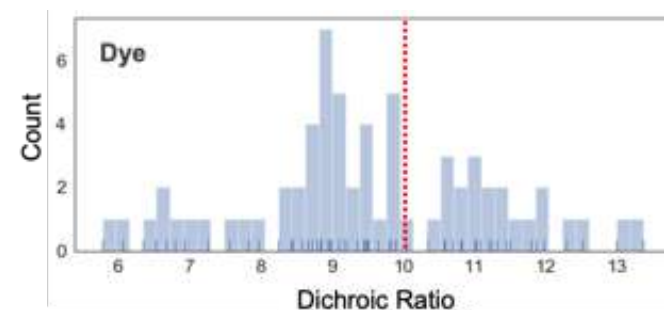
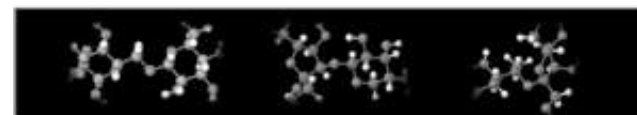
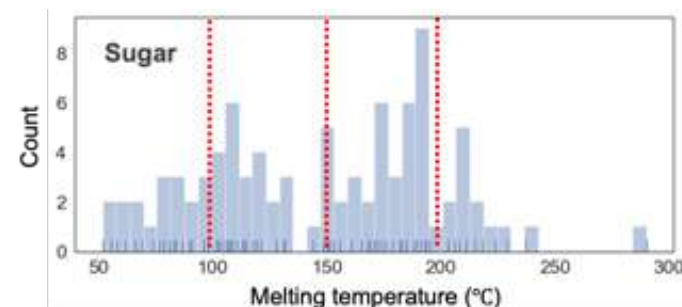


1000 CO<sub>2</sub> molecules



## 糖構造、色素構造

- 狙った融点をもつ**糖構造**、2色比をもつ**色素構造**の設計について動作が確認されました。
- 人間の専門家と比較して、**数10倍の構造デザイン速度**、より豊かな構造バリエーションが実証されました。
- 長瀬産業様の材料発見プラットフォームTABRASAのコアエンジンとして採用頂きました。



S. Takeda et al. "Molecular Inverse-Design Platform for Material Industries", KDD 2020.



(1) Future of Computing, Accelerated Discovery

(2) AIによる分子デザイン手法

(3) IBM 東京基礎研究所の分子デザイン技術

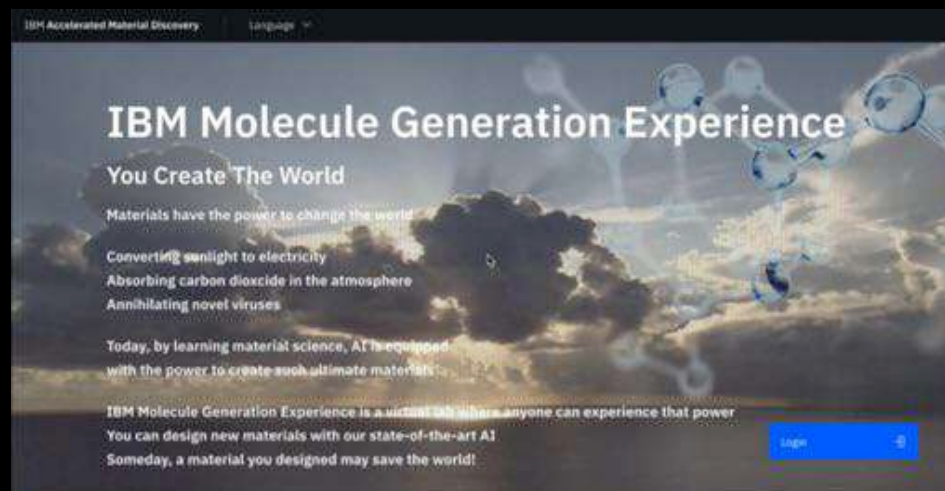
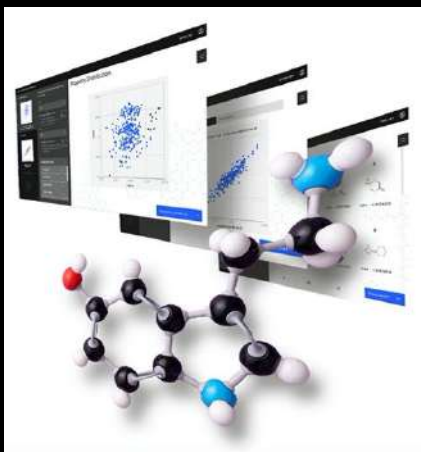
- 手法概要
- 事例紹介

(4) これからのマテリアル・デザイン – AIと人の共創

# IBM Molecule Generation Experience (MolGX) Webアプリ体験版

<https://molgx.draco.res.ibm.com>

IBMのAI分子生成モデルを、誰もがシンプルな  
ユーザーインターフェースにより体験できるWebアプリケーションです。



# 次世代のマテリアルデザインは、 人同士、人とAI・コンピュータの協力によって成立します

## 人とAIの協力

- 学習すべきデータの選定
- ターゲット（構造・物性値）の設定
  - 分子構造のどの部分をデザインするか？
  - 実験条件・配合比は入れるか？
  - ...
- 結果のレビュー、モデルへのフィードバック
  - 物性値予測の精度
  - 構造の制約を導入
  - ...



## 材料分野のオープンイノベーション環境

- データベース
  - 共有・検索可能なフォーマット統一
  - ...
- MIツールの共有
  - オープンソース・APIの共有
  - データ可視化・GUIの整備
  - ...
- 成果・手法の透明化
  - 異分野のデータ・手法の取り込み
  - ...





think