

小規模データの マテリアルズインフォマティクス

おあき ゆうや
緒明 佑哉

慶應義塾大学 理工学部 応用化学科 准教授

本説明会の内容

0. 自己紹介
1. マテリアルズインフォマティクス(MI)に関する動向
2. MIを活用したプロセス最適化事例:
 ナノシート材料の収率の向上・サイズ制御
3. MIを活用した物質探索事例:
 高性能リチウムイオン二次電池有機正・負極の探索
4. おわりに

自己紹介



慶應義塾大学
理工学部
応用化学科



東京大学
大学院工学系研究科
化学生命工学専攻

2006 博士学位取得 \longrightarrow 2007 博士研究員 (学振PD)

2009~2012 助教

2012~2016 専任講師

2016~ 准教授

2016~2020 JSTさきがけ研究者 **マテリアルズインフォマティクス (MI)**

2018~2020 文科省研究振興局学術調査官

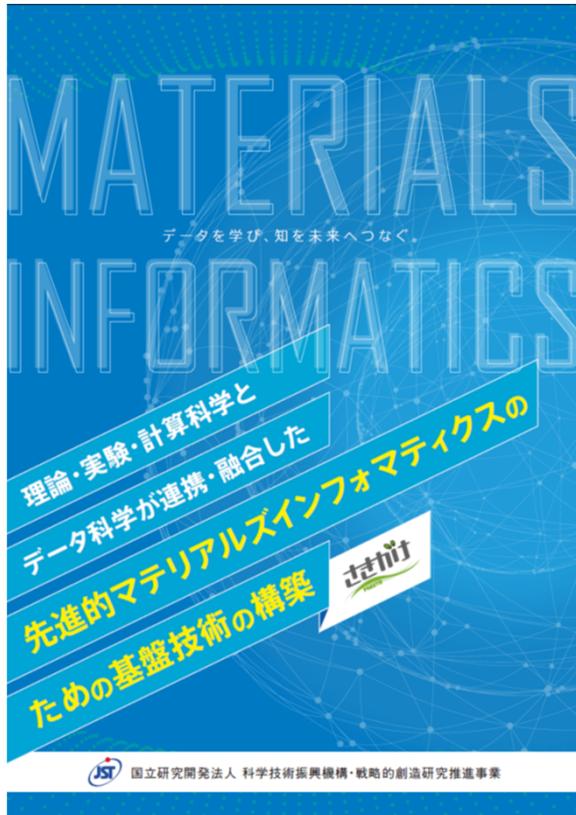
材料 共役高分子材料 / 層状物質・ナノシート材料 (2D材料) **実験研究者**

手段 ナノ構造の設計と制御 / 形態制御 / **MI**

応用 電池 / キャパシタ / センサ / 触媒 (資源・環境・エネルギー関係)

MIをはじめたきっかけ

JSTさきがけ



実験・理論・計算・データ科学者
物理・情報・化学・材料分野
の若手研究者が融合

理論・**実験**・計算科学とデータ科学が
連携・融合した
先進的マテリアルズインフォマティクス
のための基盤技術の構築

領域設定期間：2015~2020年度
研究総括：常行真司先生
(東京大学 大学院理学系研究科 教授)

領域アドバイザー (敬称略)		田中 功	京都大学大学院 工学研究科 教授
青柳 岳司	産業技術総合研究所 機能材料コンピューショナルデザイン 研究センター総括研究主幹	知京 豊裕	物質・材料研究機構 統合型材料開発・情報基盤部門 情報統合型物質・材料研究拠点 副拠点長 MANA半導体デバイス材料グループグループリーダー
射場 英紀	トヨタ自動車株式会社 先進技術開発カンパニー 電池材料技術・研究部長	津田 宏治	東京大学大学院 新領域創成科学研究科 教授
伊藤 聡	物質・材料研究機構 統合型材料開発・情報基盤部門 情報統合型物質・材料研究拠点 拠点長 材料データプラットフォームセンター センター長	寺倉 清之	物質・材料研究機構 フェロー
小谷 元子	東北大学 大学院理学研究科 教授 東北大学 材料高等研究所 研究所所長	中井 浩巳	早稲田大学 理工学術院 教授
佐藤 寛子	情報・システム研究機構 国立情報学研究所 准教授	樋口 知之	情報・システム研究機構 統計数理研究所 所長

本説明会の内容

0. 自己紹介

1. マテリアルズインフォマティクス(MI)に関する動向

2. MIを活用したプロセス最適化事例:

ナノシート材料の収率の向上・サイズ制御

3. MIを活用した物質探索事例:

高性能リチウムイオン二次電池有機正・負極の探索

4. おわりに

MIに関連した歴史

第4の科学の活用

- ・経験科学
- ・理論科学
- ・計算科学
- ・データ科学

～インフォマティクス

- ・ケモインフォマティクス(1980年頃～)
計算化学と理論化学主導
- ・バイオインフォマティクス(2000年頃～)
生物分野: 遺伝情報=ビッグデータ
- ・マテリアルズインフォマティクス(2010年頃～)
データの洪水に気付く→AIで材料開発(?)

MIの国内外動向

- ・マテリアルズゲノムプロジェクト(米国)(2011年頃～)
- ・JST戦略的創造研究推進事業 CRESTやさきがけ(2015年頃～)
- ・NIMS 情報統合型物質 材料開発イニシアティブ(MI²)設置
- ・学会や雑誌でも多数の特集企画

最近の課題



日本は世界をリードできるか？

MIで本当に世の中を変える材料ができるのか？

誰でも使えるのか？どう使うのか？

MIへの期待

目的 ビッグデータ・データ科学的手法を用い材料の研究開発を加速
理論・計算科学や既存データ + データ科学手法



JSTさきがけ (2016-2020)

理論・実験・計算科学とデータ科学が連携・融合した
先進的マテリアルズインフォマティクスのための基盤技術の構築

私の関心 ➡ 実験者として材料・実験研究にいかにMIが活用できるか？

従来技術とその問題点: MIの死角を考える

古典的MIのあるある事例

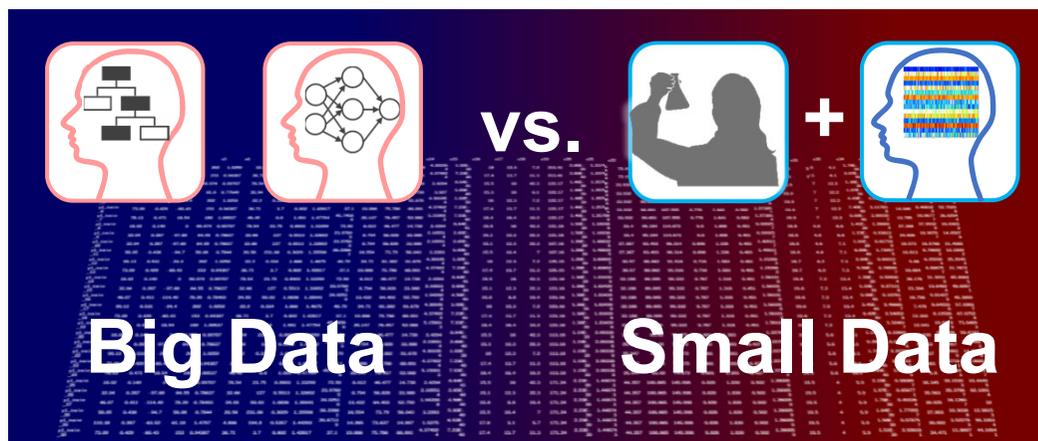
- ・ビッグデータが必要
- ・因果関係はわからない
- ・ブラックボックスなプロセス
- ・AIによる予測の信頼性
- ・AIの活用が目的
- ・熟練研究者より効率的



導入への不安と障壁

- ・スモールデータは使えない？
- ・全面的に信じられない？
- ・結果が解釈できない？
- ・経験や勘とのミスマッチ？
- ・AIが手段にならない？
- ・研究者はいらない？

限定的な適用事例や材料系の成功例に注目が集まりがち



我々のMI

- ・小規模データ
- ・研究者の考察重視
- ・高解釈性

新技術の特徴：研究者の経験と勘を重視



小規模データから研究者が主導する単純明快なMIを開拓

本説明会の内容

0. 自己紹介
1. マテリアルズインフォマティクス(MI)に関する動向
2. MIを活用したプロセス最適化事例:
ナノシート材料の収率の向上・サイズ制御
3. MIを活用した物質探索事例:
高性能リチウムイオン二次電池有機正・負極の探索
4. おわりに

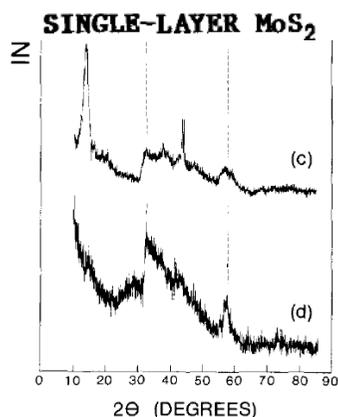
ナノシート材料の研究背景

層状構造のはく離

単～数原子層(数 nm)の2次元ナノ構造

多様な物質群での作製と機能開拓

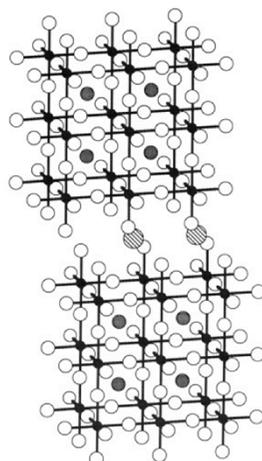
金属カルコゲナイト



R. Schöllhorn et al.
Inorg. Chem. **1977**

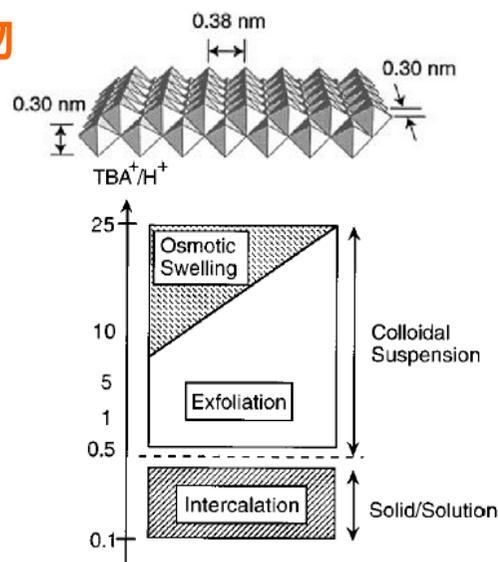
S. R. Morrison et al.
Mater. Res. Bull. **1986**

粘土化合物
層状複水酸化物



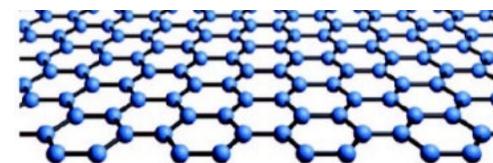
A. J. Jacobson et al.
Chem. Mater. **1990**

遷移金属酸化物



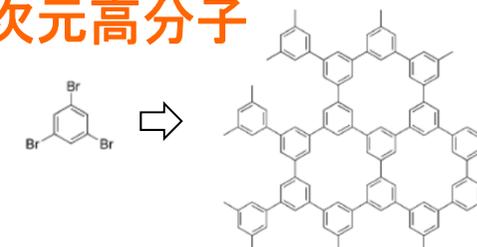
T. Sasaki et al.
J. Am. Chem. Soc. **1997**

グラファイト



K. S. Novoselov et al.
Science **2004**

2次元高分子



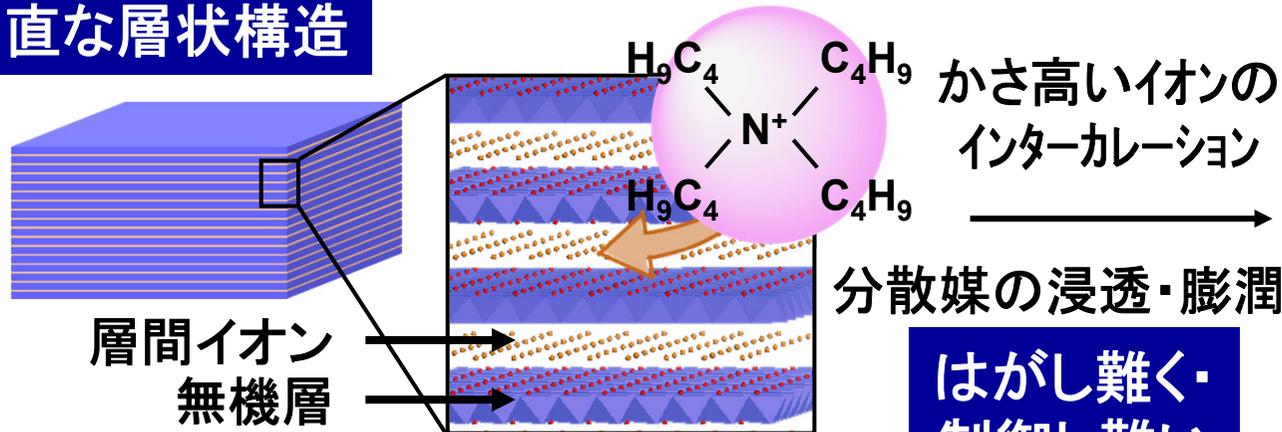
A. D. Schlüter et al.
Angew. Chem. Int. Ed. **2009**

特異な性質: 2次元異方性・高比表面積・柔軟性・量子サイズ効果など

課題 ➡ プロセス制御(収率向上・厚さや横幅制御)

ナノシート合成へのMIの適用

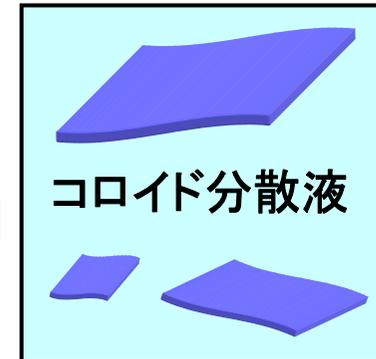
剛直な層状構造



T. Sasaki et al., *J. Am. Chem. Soc.* 1996, 113, 8129.

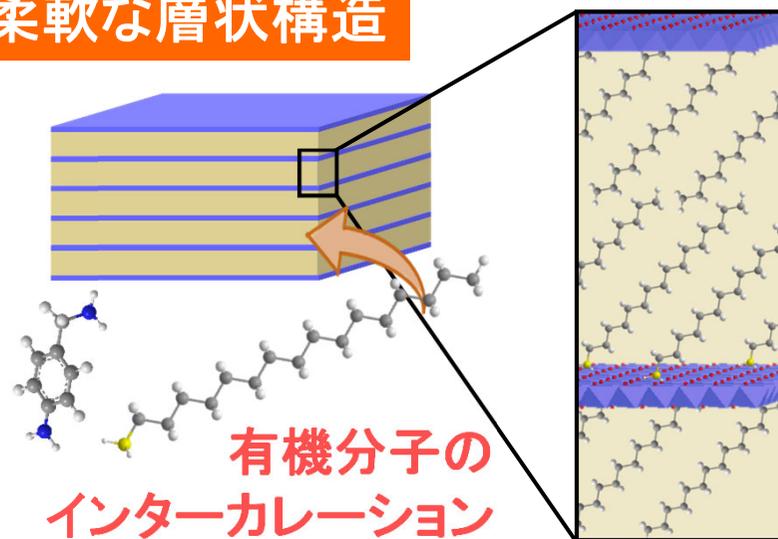
はがし難く・
制御し難い

ナノシート



水・極性分散媒

柔軟な層状構造

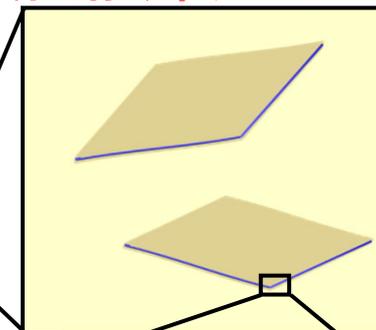


分散媒の
浸透・膨潤

非極性分散媒

はく離挙動の制御

表面修飾ナノシート



Y. Oaki et al. *Chem. Mater.* 2014, 26, 3579. *Chem. Commun.* 2015, 51, 10046.

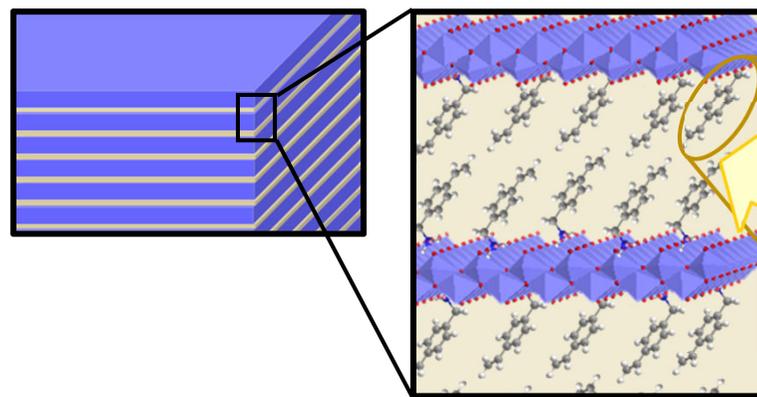
(総説) Y. Oaki *Chem. Lett.* 2021, 50, 305 (50周年記念Highligh Review).

ナノシート合成へのMIの適用

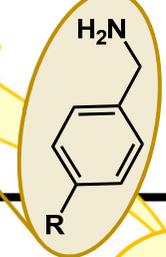
①前駆体合成

ホスト層

層状複合体



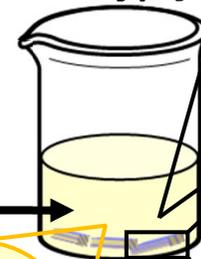
ゲスト分子



分散媒

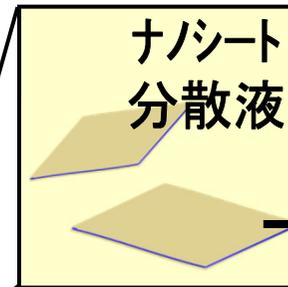
②はく離

温度
時間

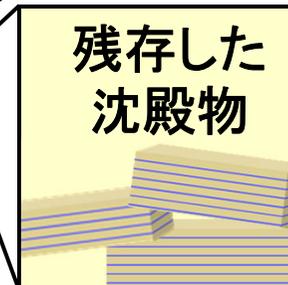


③精製

ナノシート
分散液

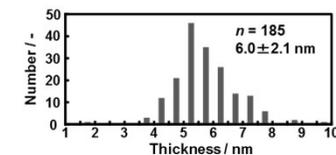
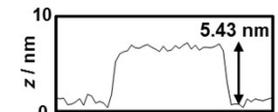
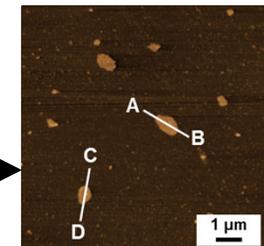


残存した
沈殿物



④評価

AFM観察



具体的な目標

【ナノシートの収率
ナノシートのサイズ】を物質とプロセスに関するパラメータで予測・制御

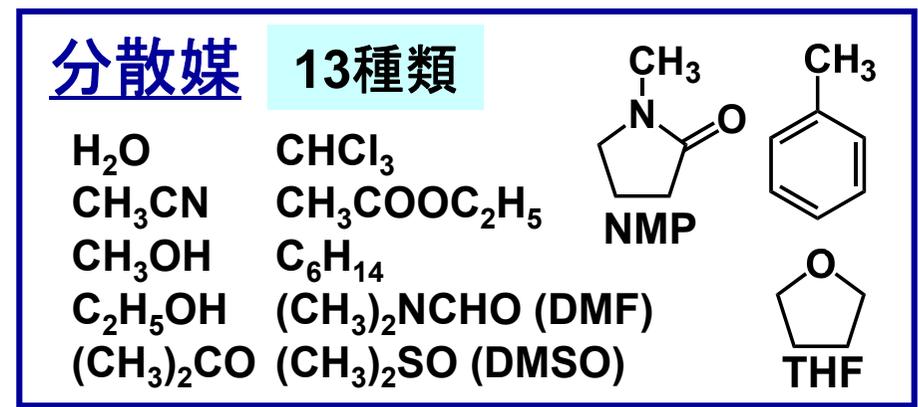
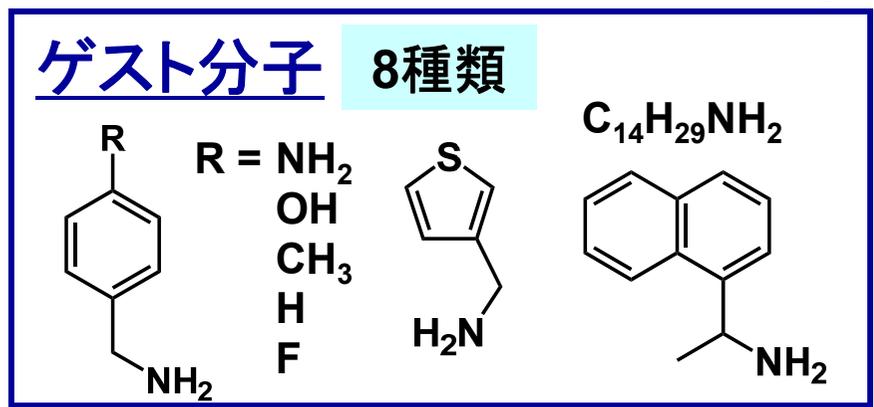
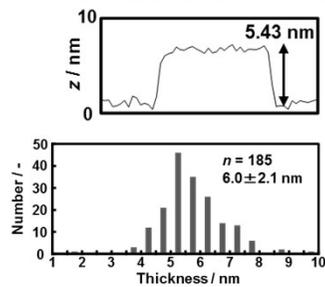
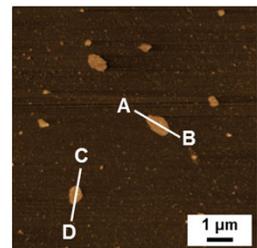
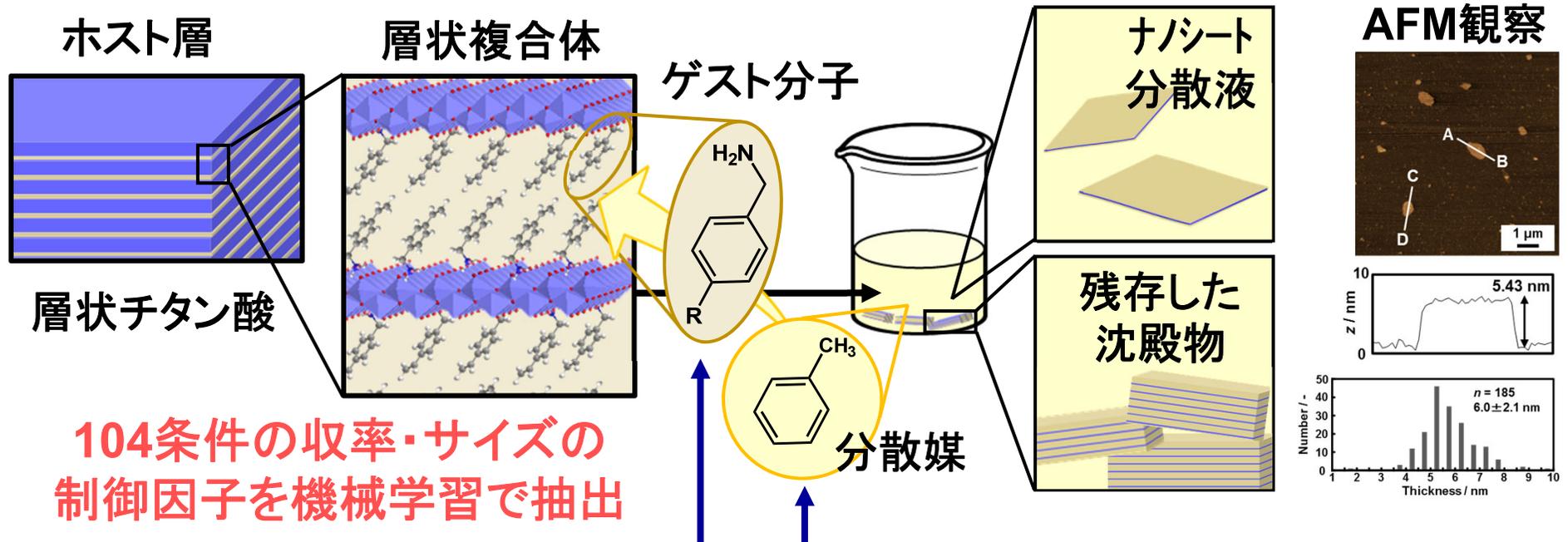


未知の層状前駆体から、高収率でサイズ制御されたナノシートを最少実験数で得る。

自前スモールデータセットの準備

実験による収率測定

収率測定 DLS粒度分布



良質なデータセットの効率的な作成

動的光散乱(DLS) → 粒度分布 0.1~1 μmに分布
 ナノシートの収量 → 収率 2%以上
 極端な低収率条件の排除

ゲスト分子	C ₁₄ -NH ₂		NH ₂ -BA		CH ₃ -BA		H-BA		F-BA		OH-BA		Naph-EtNH ₂		Tp-CH ₂ NH ₂	
分散媒	^a Size / nm	^b Yield / %	^a Size / nm	^b Yield / %	^a Size / nm	^b Yield / %	^a Size / nm	^b Yield / %	^a Size / nm	^b Yield / %	^a Size / nm	^b Yield / %	^a Size / nm	^b Yield / %	^a Size / nm	^b Yield / %
Water	262	≈ 0	890	62.1	^d 477	10.0	^e 441	4.2	689	5.3	578	65.8	^c 151	10.5	^d 385	59.7
DMSO	^c 237	4.5	539	36.6	890	4.5	^d 390	9.5	1325	2.0	716	73.1	^c 134	7.4	901	15.4
DMF	^d 433	22.6	464	52.7	564	2.2	699	1.4	^c 133	21.8	^d 261	67.1	^d 342	44.1	^c 248	22.1
Acetonitrile	^d 395	0.8	651	10.8	2074	7.8	540	2.8	^c 160	3.3	^d 781	1.4	^c 206	5.1	^d 304	3.6
Methanol	^f N. A.	1.0	462	2.2	^c 242	14.1	^g 670	21.5	840	8.4	502	48.8	522	62.1	^d 507	21.2
NMP	^c 188	13.6	^g 605	12.0	631	7.3	553	4.1	581	25.3	^c 113	57.1	1086	44.1	^c 136	14.3
Ethanol	3500	0.2	2061	1.6	540	2.0	^d 369	6.4	^c 142	6.2	473	45.4	^d 328	32.0	870	7.2
Acetone	606	≈ 0	^c 204	≈ 0	1655	2.6	^c 172	12.2	^e 133	3.7	^c 224	7.2	^c 204	13.0	^d 510	5.4
THF	^d 411,	0.4	599	2.6	^d 359	≈ 0	490	1.9	^c 135	1.2	^g 766	4.6	625	1.4	^d 574	0.7
Ethyl acetate	^g 562	2.7	576	1.6	^f N. A.	≈ 0	548	0.7	523	1.9	^d 503	1.1	^c 604	1.2	^d 497	0.9
Chloroform	^d 330	2.8	^g 4303	2.5	1582	0.7	^d 142	2.3	260	2.7	326	≈ 0	^c 134	55.3	^c 169	0.2
Toluene	348	1.1	12.1	1.3	^d 24.5	0.9	^d 227	≈ 0	^g 551	0.4	^f N. A.	0.6	^d 236	≈ 0	^f N. A.	≈ 0
Hexane	^d 322	4.4	>10 ⁴	0.9	^d 318	0.0	448	3.7	^f N. A.	≈ 0	^d 563	0.9	373	≈ 0	^d 481	≈ 0

G. Nakada, Y. Igarashi, H. Imai, Y. Oaki, *Adv. Theory Simul.* 2019, 2, 1800180.

60 / 104条件



訓練データの作成と機械学習

自前スモールデータセットの準備

機械学習の準備 データセットの作成とモデルの選択

目的変数: ナノシートの収率 (スクリーニング後 y , 60個)

説明変数: 実験値・文献値・計算値 (x_n , 35個)

データ
セット

	1	2	3	4	5	6	36	37
1	サンプル名	分子量(分散媒)	分子長(分散媒)	融点(分散媒)	沸点(分散媒)	密度(分散媒)	HSP Distance	収率
2	C14NH2-7	88.11	0.569	-83.55	77.11	0.89455	4.355	2.7
3	C14NH2-8	119.38	0.287	-63.52	61.18	1.4797	3.842	2.81
4	C14NH2-10	86.18	0.82	-95.32	68.74	0.65484	5.278	4.36
5	C14NH2-11	99.13	0.531	-24.4	202	1.0259	11.138	13.64
6	C14NH2-12	73.09	0.429	-60.43	153	0.94387	13.732	22.55

- 目的
- ・現象全体(y , 収率)を支配している因子 x を知りたい(考察)
 - ・次の実験や材料設計の指針を立てたい(予測)
 - ・ある程度予測できればよい(実験数削減)

➡ 少数の因子(記述子)でわかりやすく(線形)記述するモデル

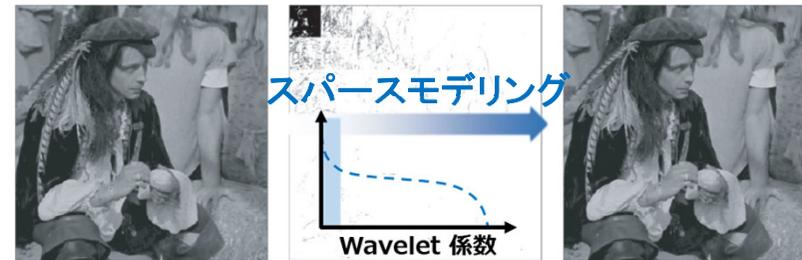
スパースモデリングと線形回帰

スパースモデリングとは？

高次元なデータから全体を説明可能な少数の本質的なパラメータを抽出するための機械学習

応用：データ圧縮, 画像のノイズ除去

元画像の2.5%の抽出データから元画像を再現



E. Candès, J. Romberg, *Inverse Problems* 2007, 23, 969.

線形の回帰モデル

構築したモデルの解釈が容易

目的変数 説明変数

$$y = \underbrace{a}_{a=0}x_1 + \underbrace{b}_{b \neq 0}x_2 + \underbrace{c}_{c \neq 0}x_3 + \underbrace{d}_{d=0}x_4 + \underbrace{e}_{e=0}x_5 + \dots + \underbrace{x}_{x \neq 0}x_n$$

係数：相関関係の重み

係数が0の項: 影響なし

係数が非0の項: 影響あり

説明変数の刈り込み
(スパースモデリング)

機械学習

- ・変数選択可能な線形回帰 (LASSO, MC+等)
- ・全状態探索

$$y = \underbrace{b}_{b \neq 0}x_2 + \underbrace{c}_{c \neq 0}x_3 + \dots + \underbrace{x}_{x \neq 0}x_n$$

説明変数の刈り込み

考察・経験

- ・化学的な妥当性
- ・経験に基づく重要性

$$y = \underbrace{c}_{c \neq 0}x_3 + \dots + \underbrace{x}_{x \neq 0}x_n$$



少数の記述子(2~5)による解釈性の高い予測モデル

全状態探索によるスパースモデリング

全状態探索

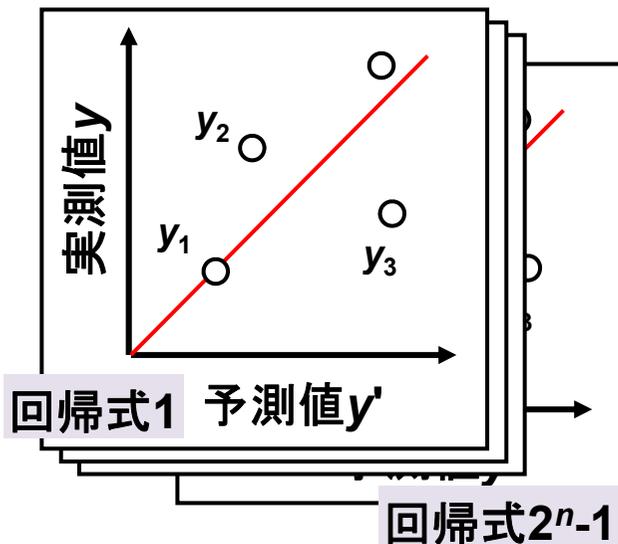
データ内で考え得る全ての場合の線形重回帰を行う

目的変数 説明変数 n 個 $\rightarrow 2^n - 1$ 通り

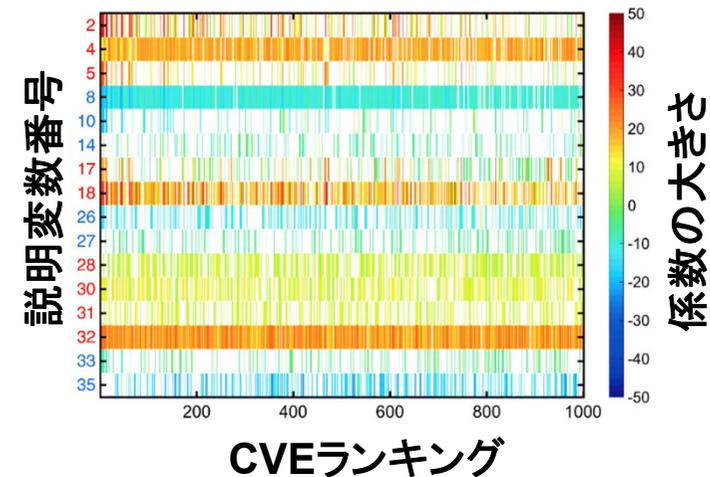
$$y = ax_1 + bx_2 + cx_3 + dx_4 + ex_5 + \dots + x_n$$

$$\left(\begin{array}{ccccc} a = 0 & b = 0 & c = 0 & d = 0 & e = 0 & x \neq 0 \\ \text{or} & \text{or} & \text{or} & \text{or} & \text{or} & \text{or} \\ a \neq 0 & b \neq 0 & c \neq 0 & d \neq 0 & e \neq 0 & x \neq 0 \end{array} \right)$$

No.	x1	x2	...	x _n	y
T01					y ₁
T02					y ₂
...					



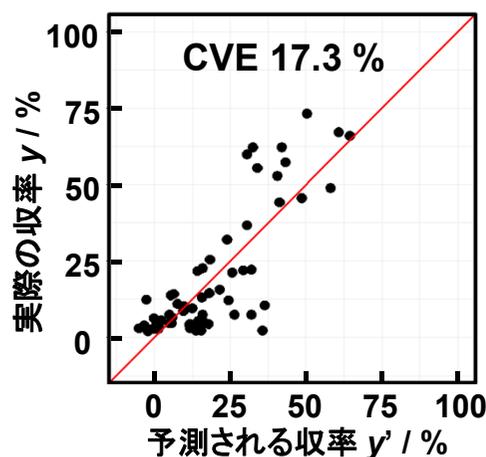
ウェイトダイヤグラム



CVE値が小さい順に各回帰式の係数の大小を表示

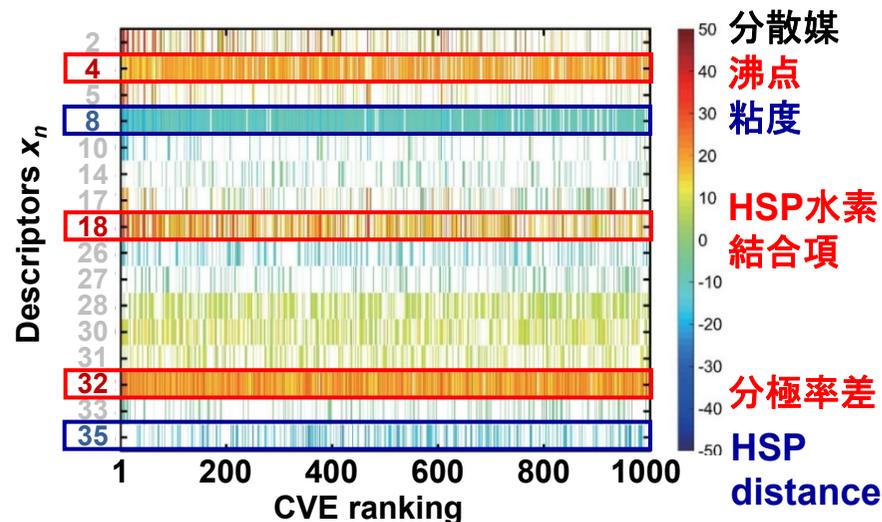
記述子の抽出と予測モデル構築

罰則項付き 重回帰分析(MC+)



16 / 35個の
記述子を抽出

全状態探索



考察 記述子の抽出

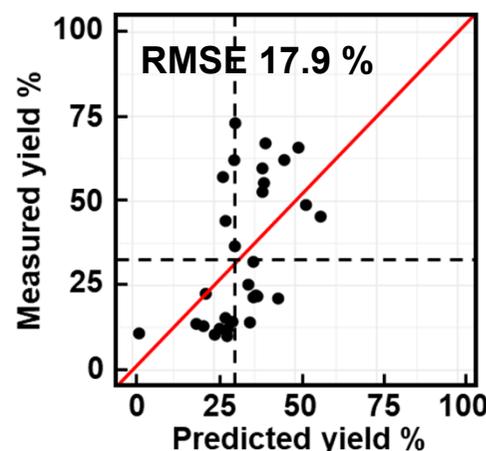
訓練データの閾値再設定

2%以上 (データ数: 60)

➡ 10%以上 (データ数: 30)

$$y = ax_{18} + bx_{35} + c$$

化学的に妥当な記述子の抽出



2 / 35個の
記述子を抽出

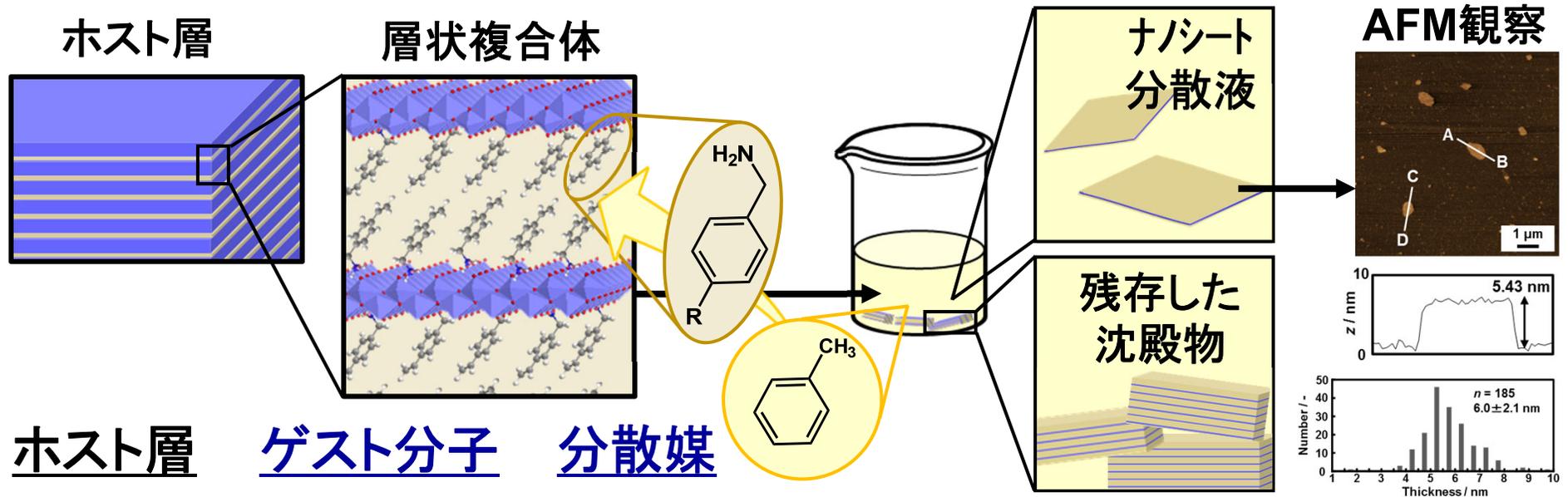
収率30%以上の
予測精度

12 / 19 (63.2%)

予測モデルの未知な層状物質への適用

テストデータへの適用

予測により最少実験数で高・低収率を実現



ホスト層

ゲスト分子

分散媒

TiO₂

8種類

13種類

= 104 → 60, 30 (訓練データ)

➡ 収率予測モデル

$$y = ax_{18} + bx_{35} + c$$

TiO₂

WO₃

10種類

26種類

260 - 49 (訓練データ分) = 211 × 2 (テストデータ)

MnO₂

10種類

40種類

156 - 24 (訓練データ分) = 132 (テストデータ)

収率予測モデル適用

➡ 予測収率高・低の上位条件のみ実験

予測に基づく最小実験数でのテスト

高収率上位10条件(Top 10)

Entry	ゲスト分子	分散媒	予測収率 / %	実測収率 / %
1	OMe-BA	Benzyl alcohol	55.02	28.77
2	Im-PrNH ₂	Benzyl alcohol	55.00	5.51
3	Im-PrNH ₂	2-Methoxyethanol	52.31	22.86
4		ol	52.09	4.58
5		ol	51.70	38.28
6		l	50.20	46.75
7			49.70	25.26
8	Im-PrNH ₂	1-Propanol	49.60	20.06
9	OMe-BA	2-Aminoethanol	49.22	3.74
10	VBA	Benzyl alcohol	48.81	43.78

平均値をとると
高収率と低収率
で差がある。

低収率下位10条件(Worst 10)

Entry	ゲスト分子	分散媒	予測収率 / %	実測収率 / %
1	Im-PrNH ₂	Heptane	-7.53	1.53
2	OMe-BA	Heptane	-5.49	2.41
3	Im-PrNH ₂	Cyclohexane	-3.68	0.26
4			-0.48	4.70
5			-0.02	0.22
6			2.95	0.08
7			6.60	0.14
8	Im-PrNH ₂	Benzene	6.96	3.45
9	VBA	Cyclohexane	8.45	8.98
10	C ₁₄ -NH ₂	Benzonitrile	8.78	49.08

個別の予測値と実測
値の整合性はそこま
で高くない。

MnO₂ 10 / 132 条件 平均: 24.0 %

TiO₂ 10 / 211 条件 平均: 41.7 %

WO₃ 10 / 211 条件 平均: 40.6 %

MnO₂ 10 / 132 条件 平均: 7.09 %

TiO₂ 10 / 211 条件 平均: 0.97 %

WO₃ 10 / 211 条件 平均: 1.25 %

(参考) 訓練データ(60条件)の全平均収率: 19.3 % 実験数削減はできている。

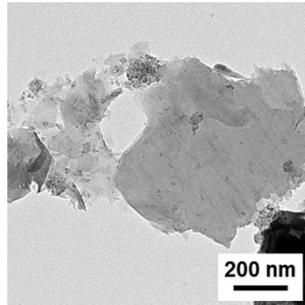
60 / 554条件のみで実験 → 高収率・低収率条件で有意差

実験的な条件の最適化 60 °C, 5日: 74.2 % → 180 °C, 30分: 57.4 %

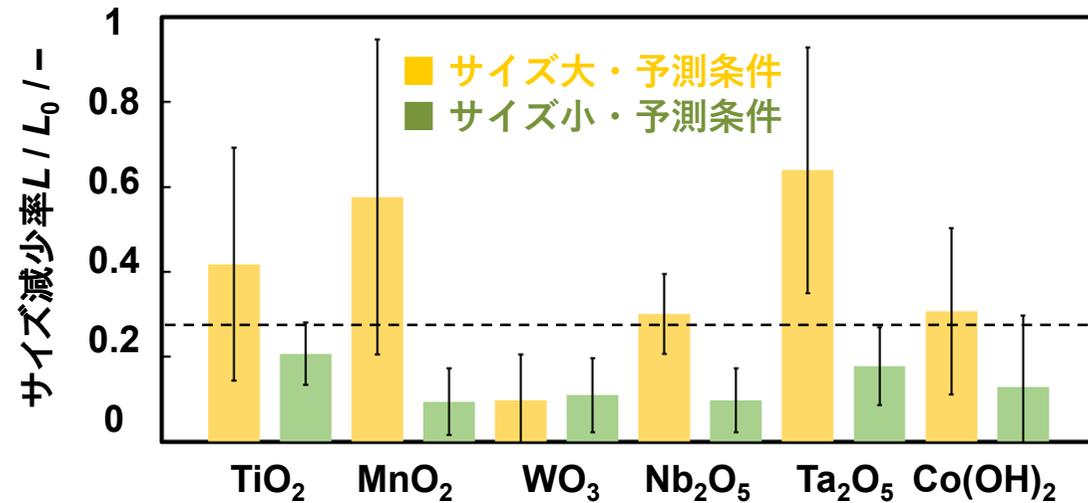
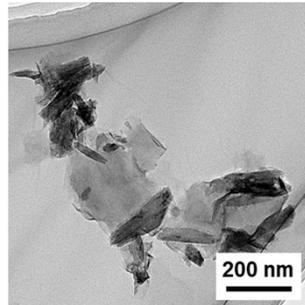
予測モデル構築とナノシート材料の合成

横幅サイズ

サイズ大



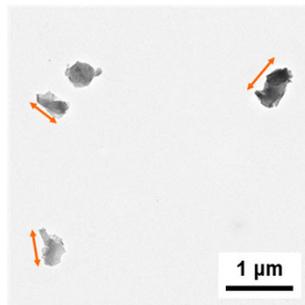
サイズ小



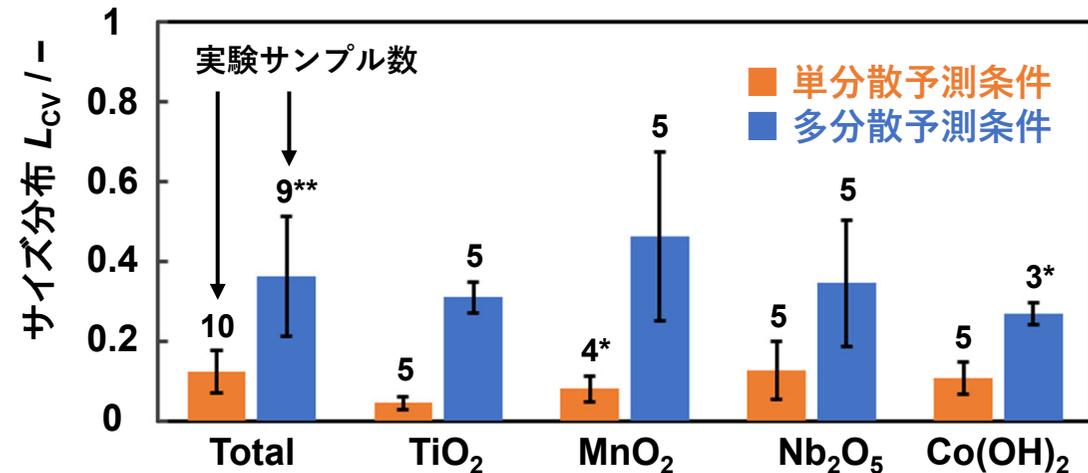
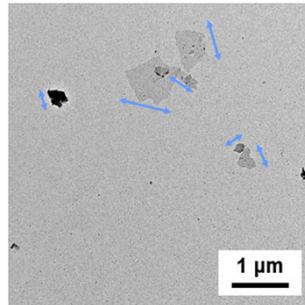
R. Mizuguchi, Y. Igarashi, H. Imai, Y. Oaki, *Nanoscale* 2021, 13, 3853.

サイズ分布

単分散



多分散



Y. Haraguchi, Y. Igarashi, H. Imai, Y. Oaki, *Adv. Theory Simul.* 2021, 4, 2100158.

➡ 候補条件の2~3%程度の実験量で選択的合成を実現

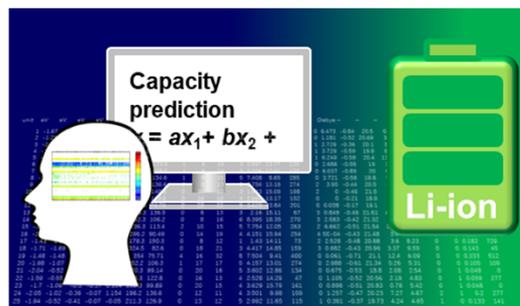
本説明会の内容

0. 自己紹介
1. マテリアルズインフォマティクス(MI)に関する動向
2. MIを活用したプロセス最適化事例:
ナノシート材料の収率の向上・サイズ制御
3. MIを活用した物質探索事例:
高性能リチウムイオン二次電池有機正・負極の探索
4. おわりに

物質探索への応用：有機リチウムイオン電池

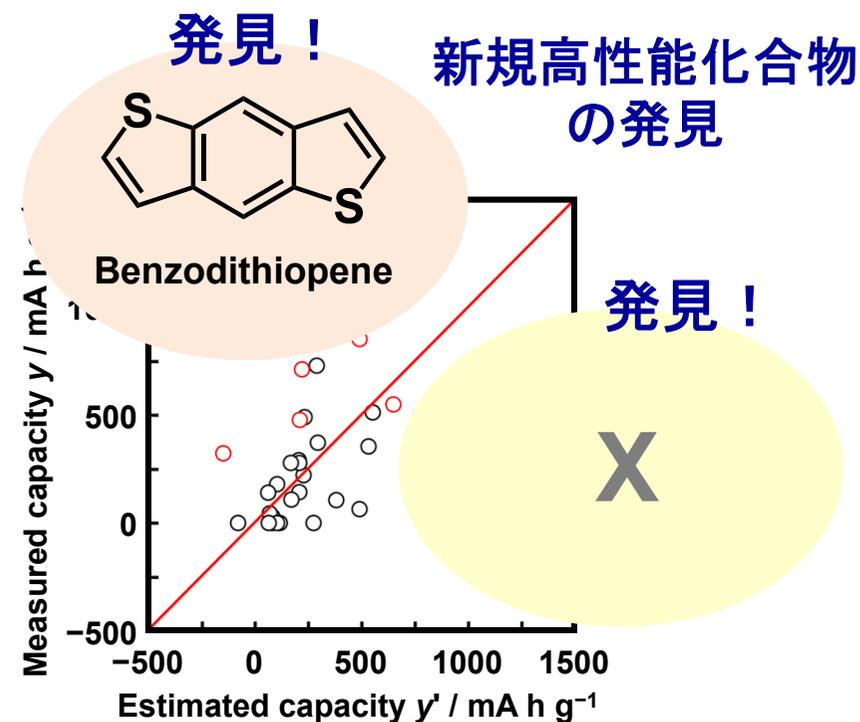
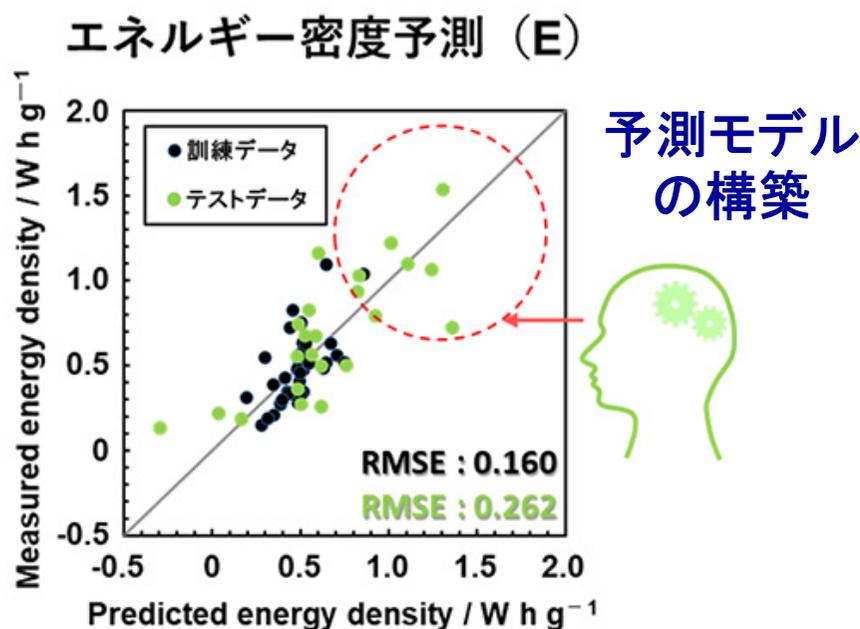
正極

既存の無機正極を
超える容量が必要



負極

分子設計指針が不明確



2021年11月2日
ソフトバンク株式会社
プレスリリース参照

H. Numazawa, Y. Igarashi, K. Sato, H. Imai, Y. Oaki,
Adv. Theory Simul. 2019, 2, 1900130.

本説明会の内容

0. 自己紹介

1. マテリアルズインフォマティクス(MI)に関する動向

2. MIを活用したプロセス最適化事例:

ナノシート材料の収率の向上・サイズ制御

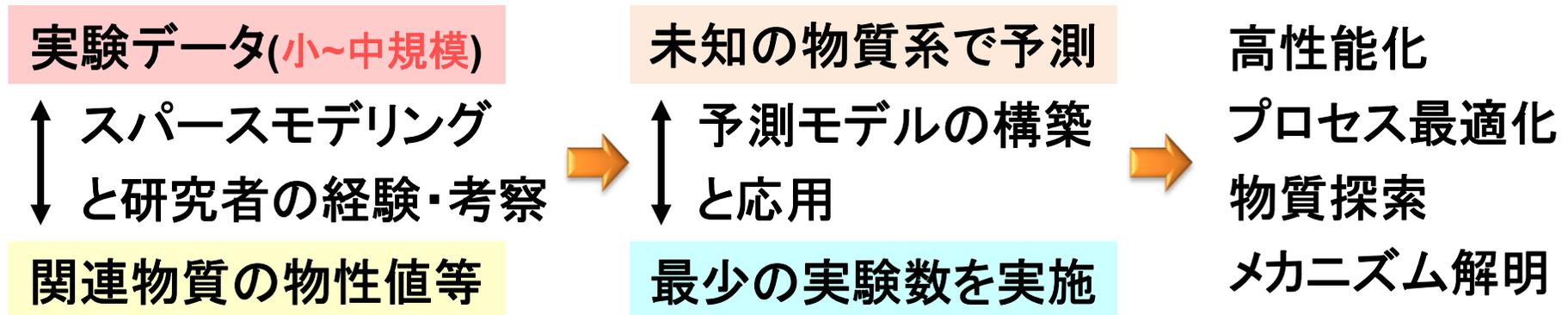
3. MIを活用した物質探索事例:

高性能リチウムイオン二次電池有機正・負極の探索

4. おわりに

実験主導MIのススメ

そのデータ小規模でも使えます！



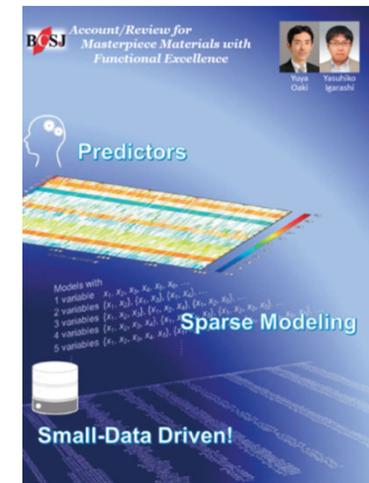
- ・スパースモデリング → 少数の本質的な記述子
- ・線形回帰 → 単純化と外挿性への期待
- ・予測精度 → そこそこの精度・実験数の削減

(総説) Y. Oaki, Y. Igarashi, *Bull. Chem. Soc. Jpn.* 2021,94, 2410.

悩み1: ラボ内のデータをもとに目標性能に早く到達したい...

悩み2: 性能向上の要因がわかりそうでわからず苦戦している...

悩み3: 社内でMIで貢献するよう命じられてたがデータが少なく困っている...



手段としてのMIでR&Dにおける問題を共同で解決できる可能性

お問い合わせ先

慶應義塾大学

研究連携推進本部 知的資産担当

T E L 03-5427-1439

F A X 03-5440-0558

e-mail toiawasesaki-ipc@adst.keio.ac.jp