



# データ科学による新物質の予測と発見

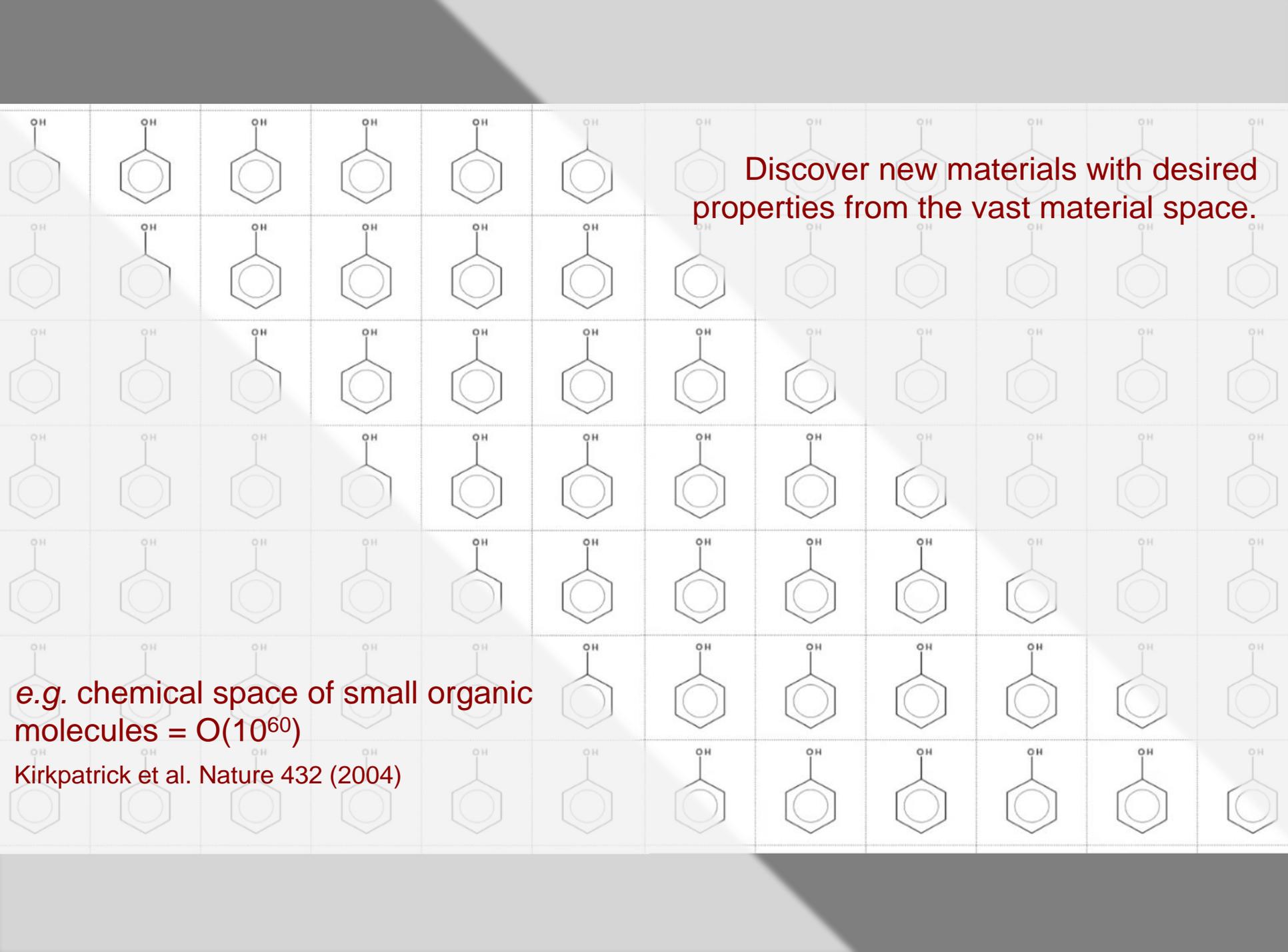
吉田 亮<sup>1,2,3,4</sup> [yoshidar@ism.ac.jp](mailto:yoshidar@ism.ac.jp)

<sup>1</sup> 情報・システム研究機構 統計数理研究所 データ科学研究系

<sup>2</sup> 情報・システム研究機構 統計数理研究所 ものづくりデータ科学研究センター

<sup>3</sup> 総合研究大学院大学 複合科学研究科 統計科学専攻

<sup>4</sup> 物質・材料研究機構



Discover new materials with desired properties from the vast material space.

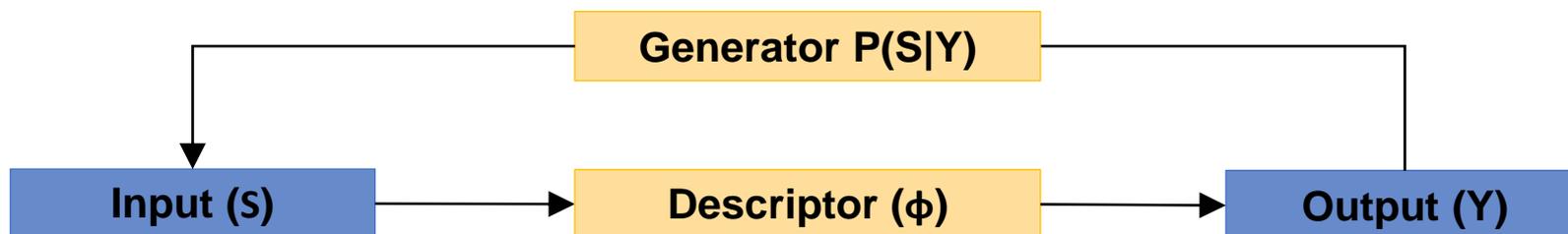
e.g. chemical space of small organic molecules =  $O(10^{60})$

Kirkpatrick et al. Nature 432 (2004)

# マテリアルズインフォマティクスの順問題と逆問題

様々な物質・材料の表現・学習・生成

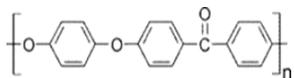
Backward prediction:  $S = f^{-1}(Y)$



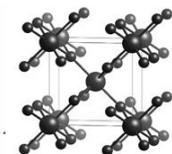
Forward prediction:  $Y = f(S) = f \circ \phi(S)$



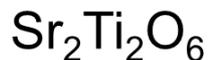
Chemical structure



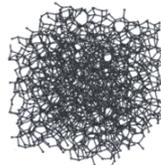
Crystal



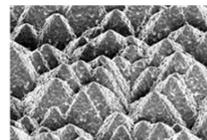
Composition



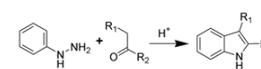
Amorphous



Microstructure

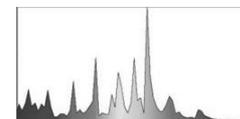


Chemical reaction



$\Theta, T, P$

Spectrum



# 予測 ~ データ科学的視点から“順問題・逆問題”を発掘する

## Process & Composition Microstructure

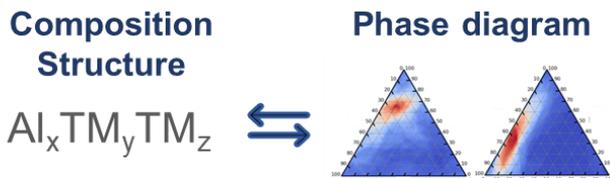


Iwayama et al. J. Chem. Inf. Model. (2022)

## Chemical structure Spectrum



Iwayama et al. J. Chem. Inf. Model. (2022)



Liu et al. Adv. Mater. 33(36):e2102507 (2021)

Liu et al. Quasicrystals predicted and discovered by machine learning (2022)

Aoki et al. Predicting polymer-solvent miscibility using machine-learned Flory-Huggins interaction parameters (2022)

## Composition Crystal structure



Kusaba et al. Comput Mater Sci. 211:111496 (2022)

Liu et al. Shotgun crystal structure prediction using machine-learned formation energies (2022)

## Reactants Synthetic product



Guo et al. J. Chem. Inf. Model. 60(10):4474-4486 (2020)

Zhang et al. ChemRxiv. 10.26434/chemrxiv-2022-5qp9v (2022)

## Crystals or molecules Property



Yamada et al. ACS Cent. Sci. 5(10):1717-1730 (2019)<sup>26</sup>

Ju et al. Phys. Rev. Mater. 5:053801 (2021)

Minami et al. AAAI. 35(10):8992-8999 (2021)

Ikebata et al. J. Comput. Aided Mol. Des. 31:379-391 (2017)

Wu et al. npj Comput. Mater. 5:66 (2019)

Wu et al. Mol Inform. 39:1-2 (2020)

Torres et al. J Phys Condens Matter. 34(13):135702 (2022)

Hayashi et al. npj Compt Mater (2022)



データ駆動科学とは森を見て木を知るすべてである



新学術領域「ハイパーマテリアル:捕空間が創る新物質科学」(田村隆治 東京理科大)

2019年8月~計画研究A03“ハイパーマテリアルズインフォマティクス”(代表:吉田)

# 統計的機械学習による準結晶の予測・理解・発見



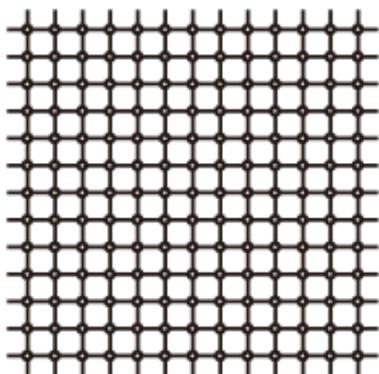
Liu et al. *Advanced Materials*.  
33(36):e2102507 (2021)



# ハイパーマテリアルとは

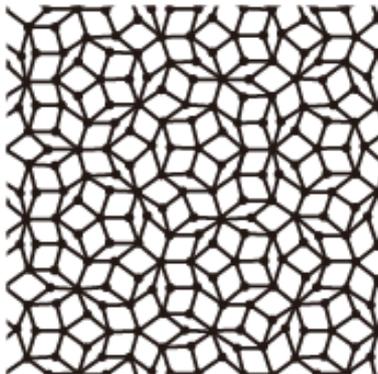
- 準結晶：通常の周期結晶には許されない対称性を持つ長距離秩序を有する第3の固体
- 近似結晶：準結晶と類似した組成と局所構造を持つ。準結晶の前駆体物質

結晶



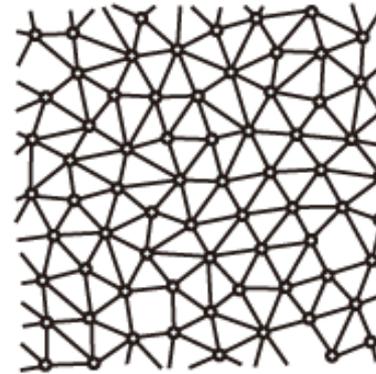
規則正しい

準結晶



規則正しい？  
正しくない？

非結晶



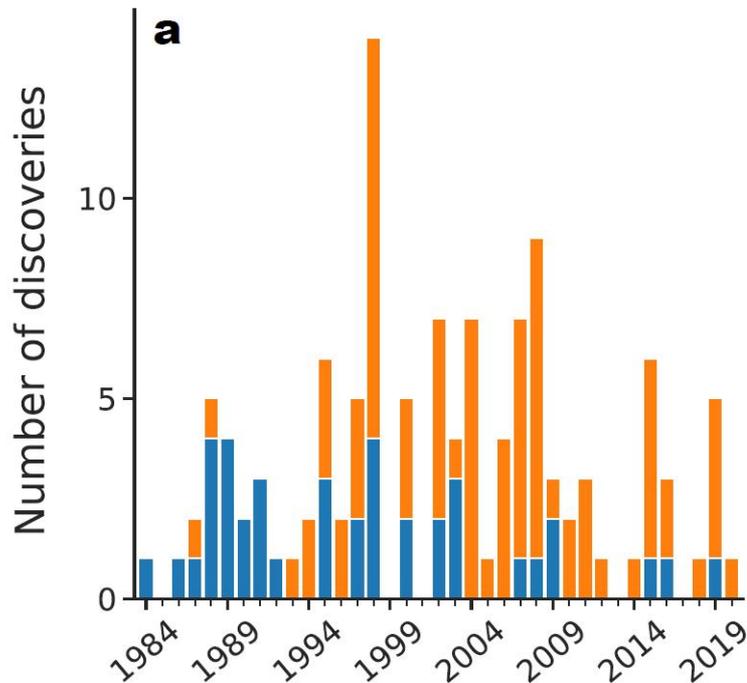
規則正しくない

斎藤勝司「【2011年ノーベル化学賞】結晶とも非結晶とも違う「準結晶」を発見」  
<https://www.kodomonokagaku.com/read/30301/>

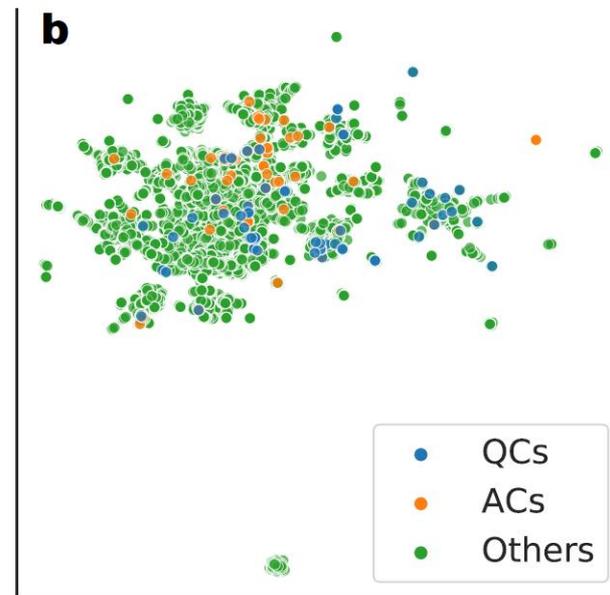
# 歴史：準結晶の発見

- 1984: Dan Shechtman博士（2011:ノーベル化学賞）による最初の準結晶の発見
- これまでにおよそ100個の安定準結晶が見つかった。
- 発見のペースが鈍化：約40年の歴史の中で普遍的な設計原理を確立できなかった。

AIを含む準結晶と近似結晶の発見数

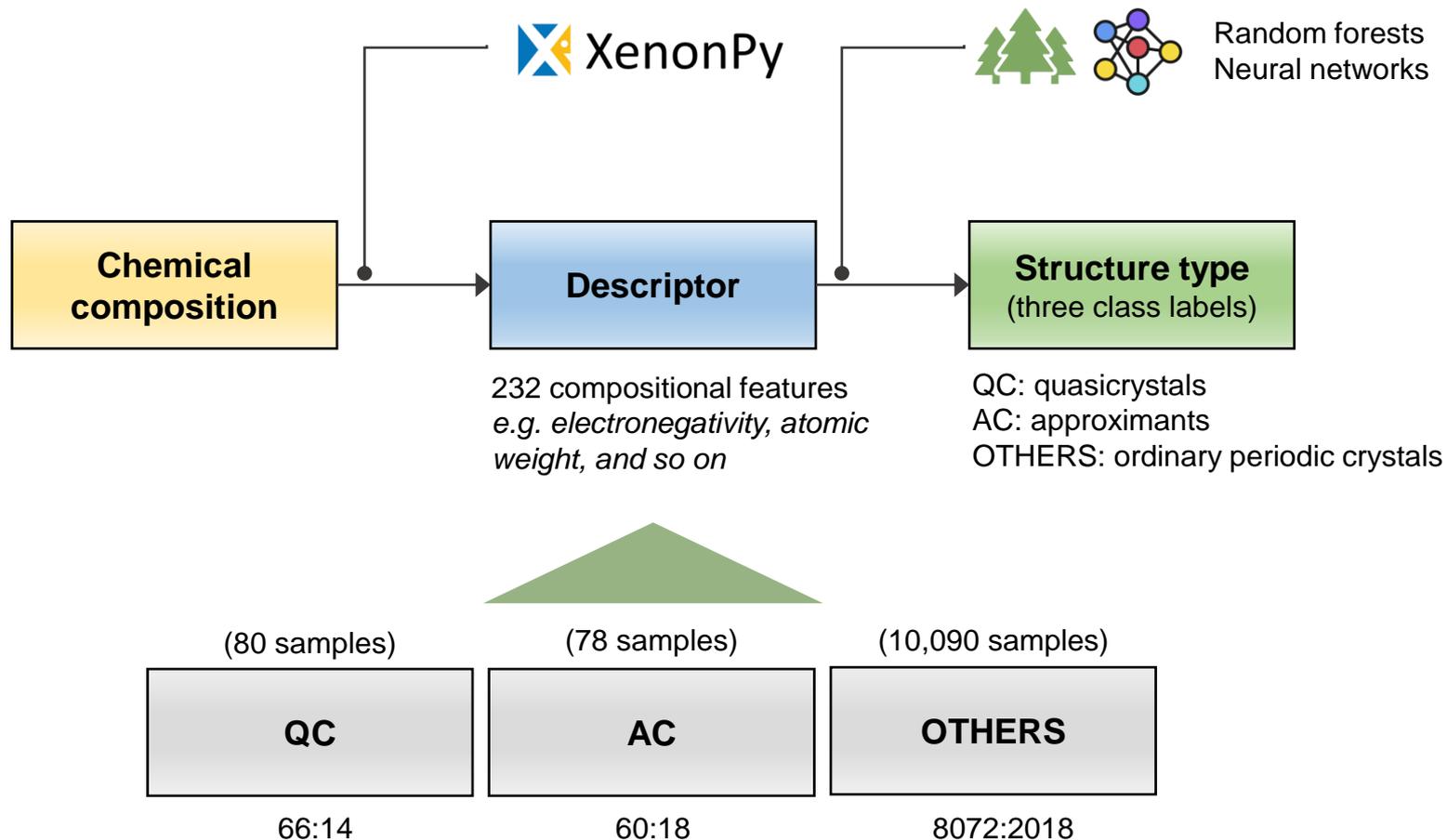


準結晶・近似結晶・結晶の組成分布



# 統計的機械学習で準結晶の発見を加速する

入力組成から準結晶・近似結晶・通常の結晶という3クラスのカテゴリを構築する



# Compositional Descriptors

Yamada et al. ACS Cent Sci. 5(10) 1717-1730 (2019), Liu et al. Adv Mater. 33(36):e2102507 (2021)

**S: composition**

The  $i$ -th element of a descriptor vector



$$\phi_i(S) = f\left(\underbrace{a, b, c, d, e}_{\text{Fraction}}, \underbrace{\eta_{\text{Ni}}^i, \eta_{\text{Ti}}^i, \eta_{\text{Cu}}^i, \eta_{\text{Fe}}^i, \eta_{\text{Pd}}^i}_{\text{Element features}}\right)$$

**Fraction**

**Element features**

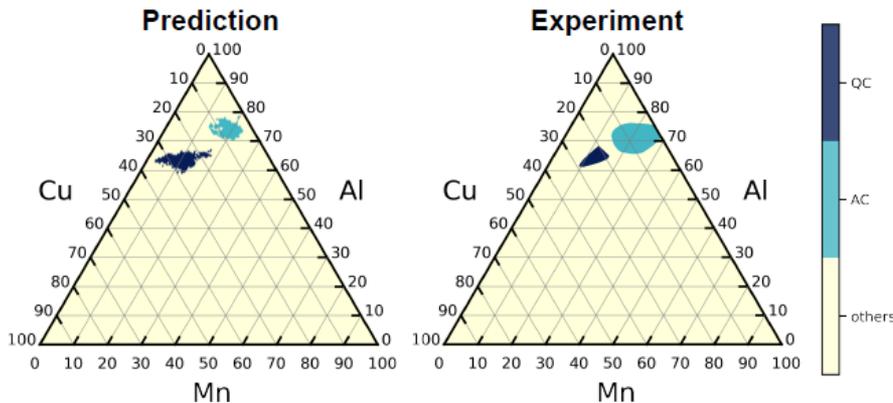
atomic weight, electronegativity, etc

$f(\cdot)$ : weighted mean, variance, max, min, covariance between features  $(i, j)$ , and so on

# Predictive Performance in Phase Diagram Drawing

- Hypermaterials (QC/AC) vs others: almost **PERFECT** prediction!
- Quasicrystalline phases can be predicted with the accuracy of about 70%
- In ~73% of the systems, the predicted QC and AC phases coexist in the same systems

Despite no training data were given, QC/AC phases could successfully be predicted in many systems.



		Prediction		
		QC	AC	OTHERS
True	QC	9.63 (1.64)	3.24 (1.34)	3.13 (1.18)
	AC	3.11 (1.55)	9.73 (1.80)	3.16 (1.57)
	OTHERS	0.76 (0.89)	0.42 (0.61)	2016.82 (1.02)

	QC	AC	OTHERS
Precision	0.722 (0.090)	0.731 (0.089)	0.997 (0.001)
Recall	0.602 (0.103)	0.608 (0.113)	0.999 (0.001)
F <sub>1</sub>	0.650 (0.076)	0.658 (0.088)	0.998 (0.001)



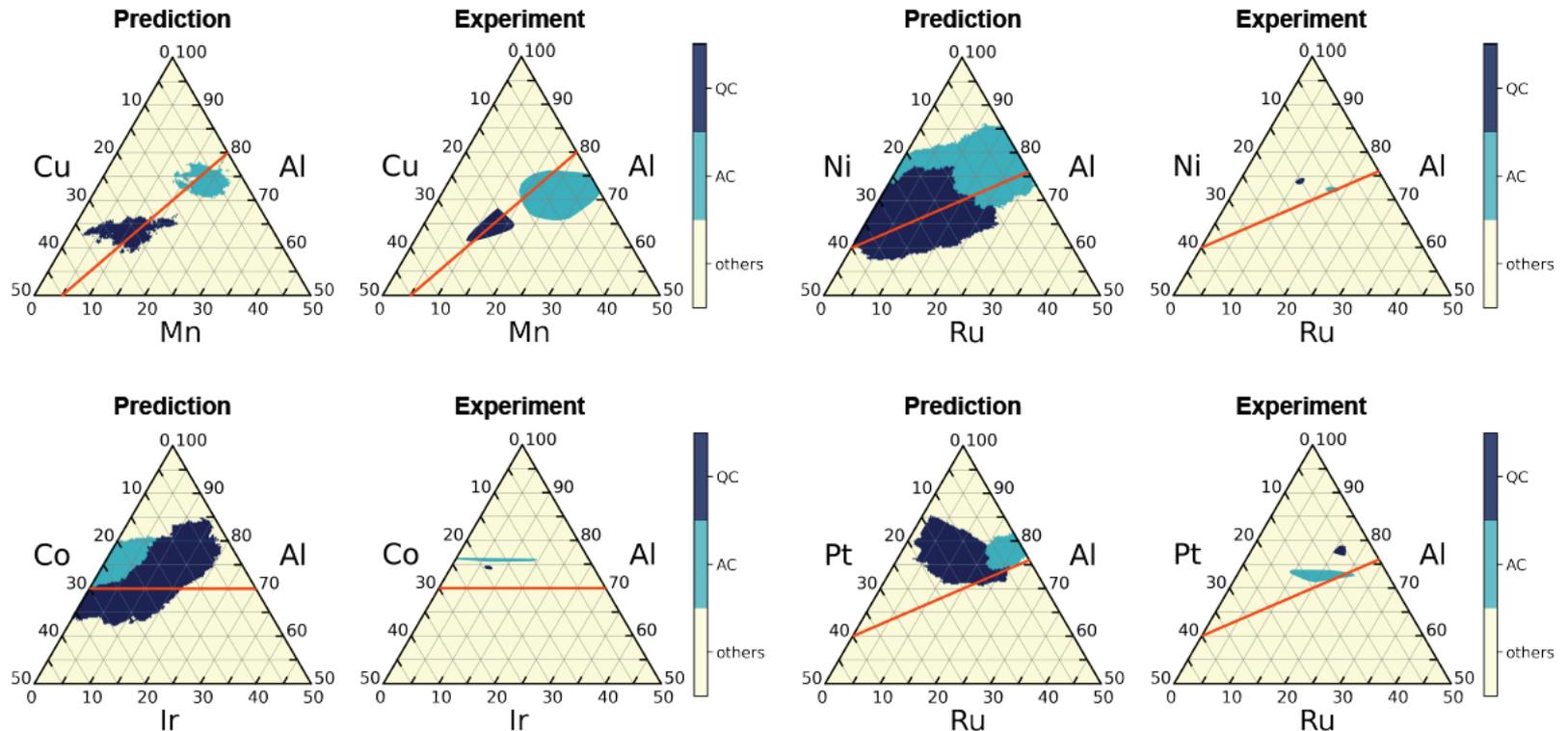
ハイパーマテリアル  
補空間が創る新物質科学

# 機械学習による“法則”の発見～知識獲得

# モデルはヒューム=ロザーリの電子濃度則を学習

Tsai, Sci Technol Adv Mater. 9, 013008 (2008)

- Tsaiの経験則：アルミニウム合金（正二十面体・Mackayクラスタ）では平均遍歴価電子数  $e/a$  が 1.8 の付近に準結晶・近似結晶相が出現
- 後に理論的説明 Kitahara et al. Cryst Mater 232 507-513 (2017)
- 準結晶・近似結晶相の予測領域の多くも  $e/a = 1.8$  付近に出現



# 準結晶の形成・安定化に関する新法則を発見する

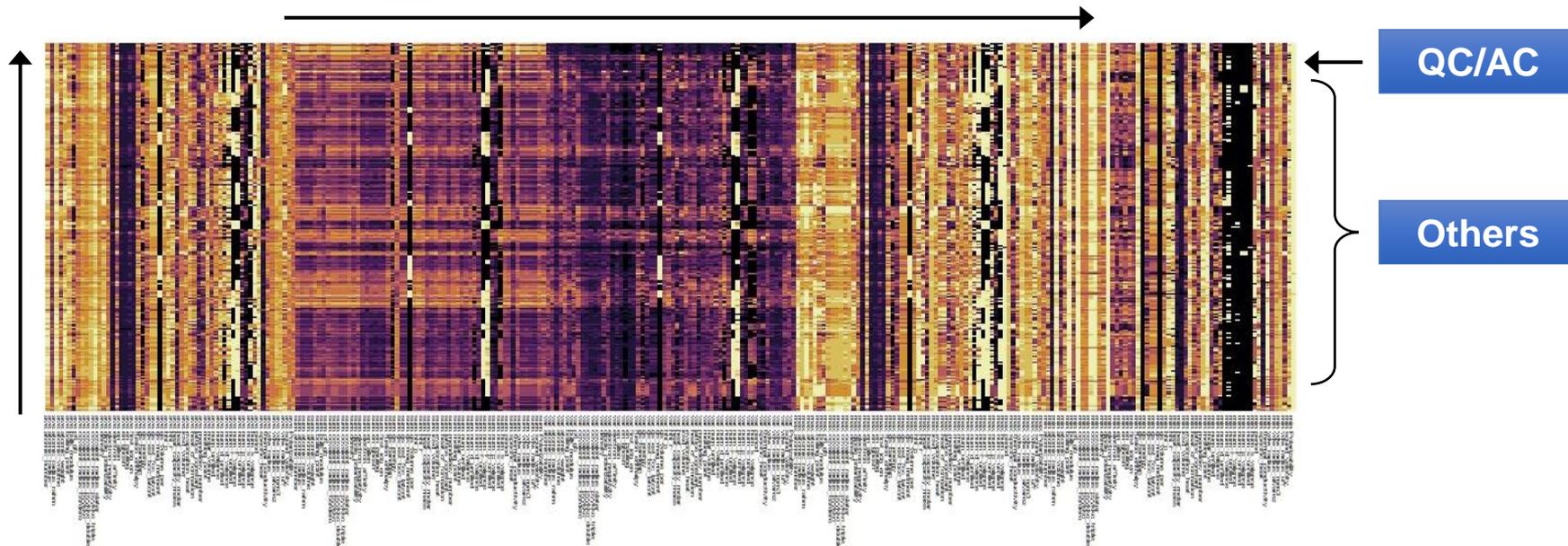
問. モデルは何を根拠に準結晶相の有無を判定しているのか？（言語化された知識）

まずは、重要特徴量（クラス分布の重なりが小さい）を同定 (MIC: Reshef et al. Science. 2011)

- Hypermaterials vs OTHERS
- Quasicrystals vs Approximants

XenonPy 232次元組成記述子

スクリーニング対象のAI-TM-TM



# Feature Relevance

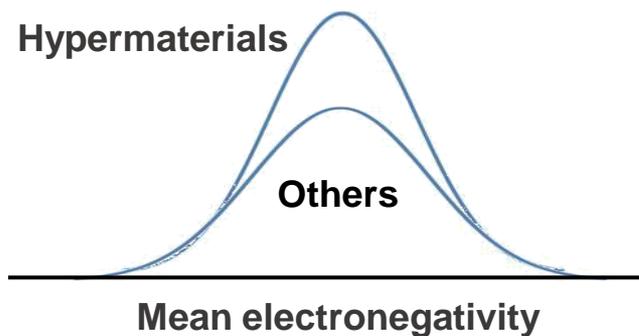
MIC (Maximum Information Coefficient: Reshef et al. Science. 2011)

**Mutual Information:**

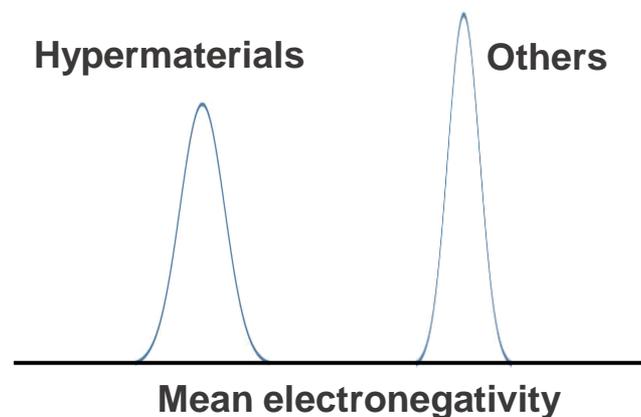
$$I(Y, X) = \int p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad p(x, y) = p(x)p(y) \Leftrightarrow I(X, Y) = 0$$

Independence between X (candidate compositions) and Y (predicted class labels)

Relevance is low



Relevance is high



**Others:** TIC<sub>e</sub> (Yakir et al. JMLR. 2016); S<sup>DDP</sup> (Heller et al. JMLR. 2016); MMD (Gretton et al. NIPS. 2019)

# 重要特徴量の同定とルールの言語化

例1：元素の電気陰性度の加重平均が $0.15 \pm 0.01$ であることが、QC/ACの必要条件

例2：p軌道の平均価電子数が $0.71 \pm 0.06$ であることが、QC/ACの必要条件

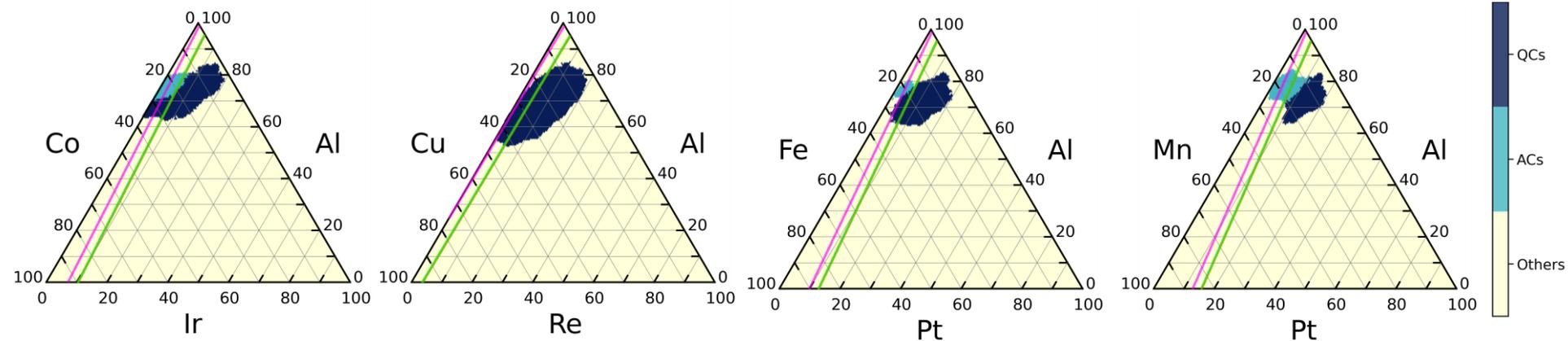
Descriptor ID	Description (unit)	MIC	QC	AC	Others
ave:vdw_radius_uff	Van der Waals radius from the UFF (pm)	0.43	409.05 (3.37)	406.49 (6.81)	382.297 (40.660)
ave:en_ghosh	Ghosh's scale of electronegativity	0.42	0.15 (0.00)	0.15 (0.01)	0.158 (0.021)
ave:first_ion_en	First ionisation energy (eV)	0.41	6.49 (0.09)	6.53 (0.17)	6.843 (0.580)
ave:mendeleev_number	Mendeleev's number	0.41	75.94 (0.36)	75.86 (1.47)	73.323 (4.326)
ave:specific_heat	Specific heat at 20 °C (J/(g mol))	0.40	0.74 (0.02)	0.73 (0.04)	0.658 (0.151)
ave:num_p_valence	Number of filled p valence orbitals	0.40	0.71 (0.06)	0.73 (0.05)	0.569 (0.247)
ave:num_p_unfilled	Number of unfilled p valence orbitals	0.40	3.53 (0.30)	3.63 (0.24)	2.846 (1.237)
ave:heat_capacity_mass	Specific heat capacity at STP <sup>1</sup> (J/mol-K)	0.40	0.74 (0.02)	0.73 (0.04)	0.655 (0.150)
ave:covalent_radius_cordero	Covalent radius by Cordero et al. (pm)	0.39	126.06 (1.19)	126.38 (2.13)	129.510 (6.311)
ave:vdw_radius	Van der Waals radius (pm)	0.37	189.51 (0.55)	190.67 (1.81)	193.799 (6.328)
ave:gs_energy	DFT energy per atom (raw VASP value) of T=0K ground state (eV/atom)	0.37	-4.57 (0.18)	-4.69 (0.28)	-5.192 (1.118)
ave:thermal_conductivity	Thermal conductivity at 25 °C (W/(m K))	0.36	221.35 (21.74)	201.23 (13.96)	170.722 (60.677)
ave:covalent_radius_slater	Covalent radius by Slater (pm)	0.35	127.92 (1.19)	128.08 (0.93)	130.396 (3.376)
ave:period	Period in periodic table	0.35	3.40 (0.06)	3.52 (0.19)	3.725 (0.553)
var:num_p_valence	Number of filled p valence orbitals (pm)	0.34	0.20 (0.02)	0.20 (0.02)	0.184 (0.071)
ave:num_d_valence	Number of filled d valence orbitals (pm)	0.34	2.30 (0.60)	2.08 (0.52)	3.149 (1.901)
ave:heat_capacity_molar	Molar heat capacity at STP (J/mol-K)	0.34	24.44 (0.10)	24.47 (0.14)	24.807 (0.580)
ave:density	Density at 295K (g/cm <sup>3</sup> )	0.34	4.94 (0.32)	5.55 (1.24)	6.700 (3.347)
var:num_p_unfilled	Number of unfilled p valence orbitals	0.34	5.09 (0.60)	4.91 (0.48)	4.602 (1.767)
ave:hhi_p	HerfindahlHirschman Index (HHI) production values	0.33	1810.99 (242.60)	2106.51 (274.67)	2196.878 (706.345)

# 例：平均電気陰性度のルール

経験式 = 組成記述子の目標値を満たす組成比の部分空間

$$\mathcal{S} = \{(c_1, c_2, c_3) \mid c_1\eta_1^{\text{en}} + c_2\eta_2^{\text{en}} + c_3\eta_3^{\text{en}} \simeq 0.15\}$$

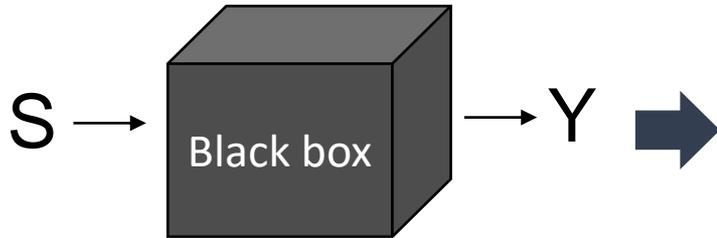
ほぼ全てのAI-TM-TMにおいて、経験式の直線は予測されたQC/AC相にオーバーラップ



Pink: target = Mean QC; Green: Mean QC/AC

# 準結晶の形成・安定化に関する五つのルール

アルミニウム合金の組成空間では、次のような近似（局所線形近似）が成り立つ。



$$\begin{aligned}
 c^1\eta_{en}(S^1) + c^2\eta_{en}(S^2) + c^3\eta_{en}(S^3) &= 0.15 \pm 0.00 \\
 c^1\eta_p(S^1) + c^2\eta_p(S^2) + c^3\eta_p(S^3) &= 0.71 \pm 0.06 \\
 c^1\eta_{vdw}(S^1) + c^2\eta_{vdw}(S^2) + c^3\eta_{vdw}(S^3) &= 409.05 \pm 3.37 \text{ [pm]} \\
 c^1\eta_{ie}(S^1) + c^2\eta_{ie}(S^2) + c^3\eta_{ie}(S^3) &= 6.49 \pm 0.09 \text{ [eV]} \\
 c^1\eta_{gs}(S^1) + c^2\eta_{gs}(S^2) + c^3\eta_{gs}(S^3) &= -4.57 \pm 0.18 \text{ [eV]}
 \end{aligned}$$

■ van der Waals radii - UFF   
 ■ Electronegativity - Ghosh   
 ■ First ionisation energy   
 ■ Filled p valence orbital   
 ■ Ground state energy

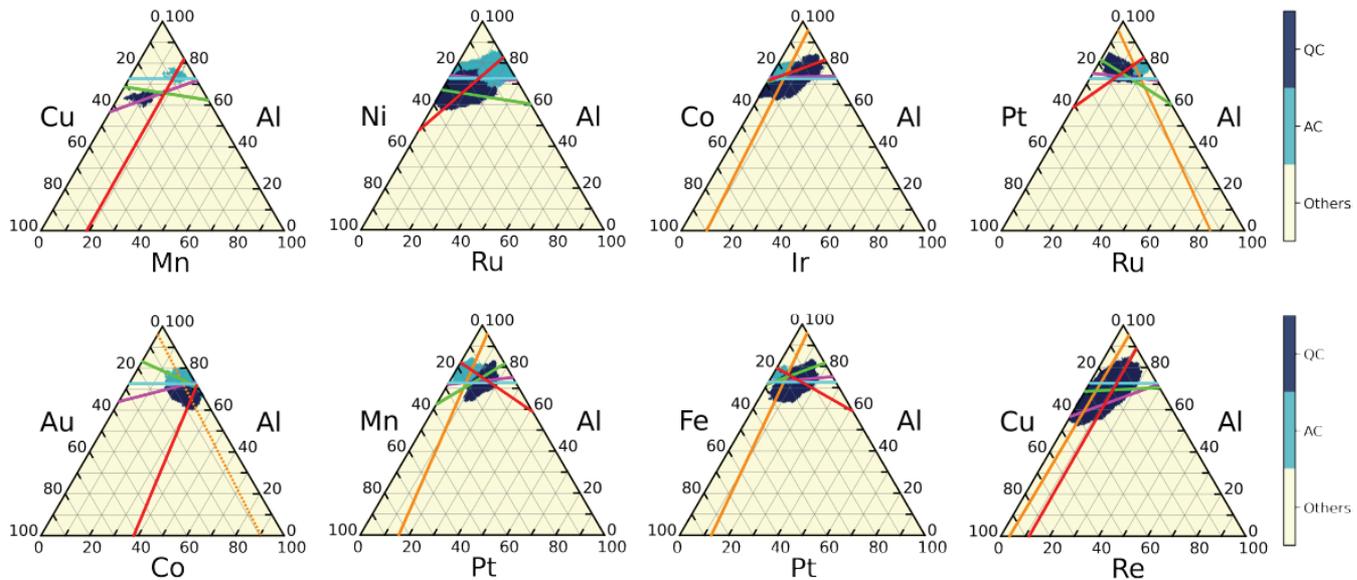


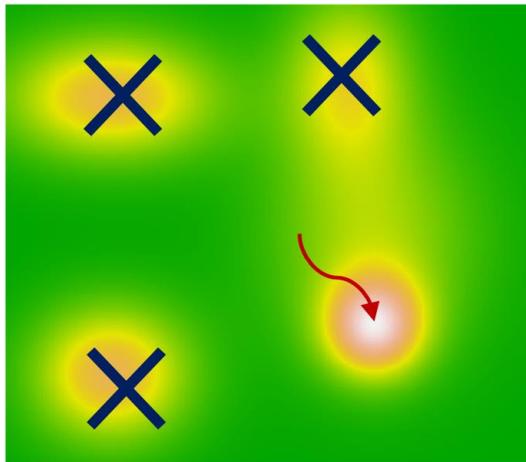
Figure 6 in Liu et al. Advanced Materials. 33(36):e2102507 (2021)

# 逆問題を解く二つの方法

	順問題	逆問題
決定論的モデル	$Y = f(S)$	$S = f^{-1}(Y^*) \quad S \approx f^{-1}(Y^*)$
確率モデル	$p(Y S, f)$	$p(S Y = Y^*, f)$

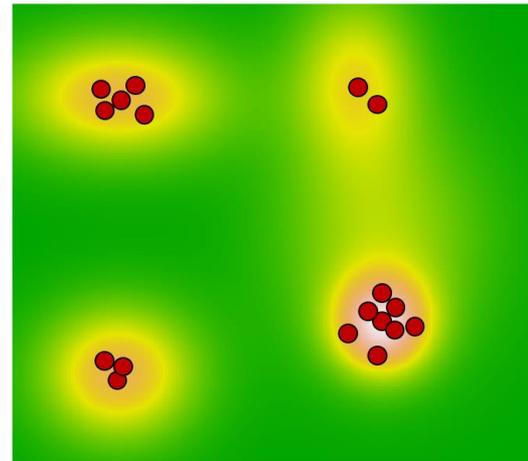
## 点推定 (最適化)

- Evolutionary (genetic) algorithm
- Gradient descent
- Combinatorial optimization



## 分布の推論・解の数え上げ

- MCMC
- Sequential Monte Carlo
- Enumeration



# 不完全モデルからの知識発見

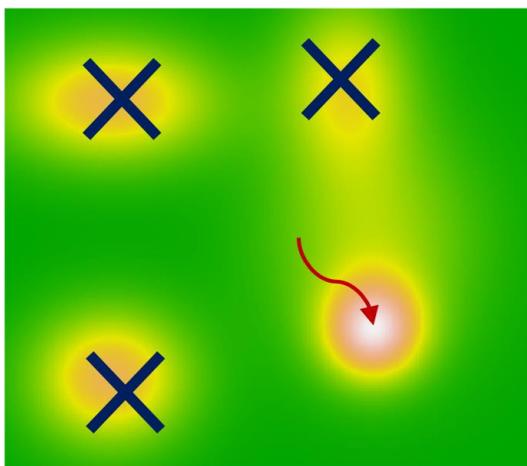
分布（地図）に基づく確率的推論が必要不可欠

モデルの世界の最適解は現実系において最適解ではない

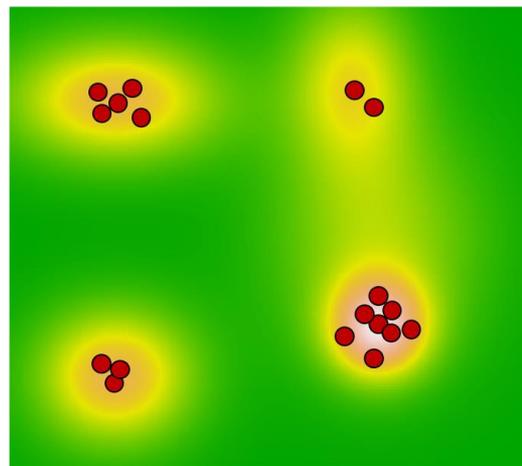
$$\arg \max_X f_{\text{model}}(X) \neq \arg \max_X f_{\text{true}}(X)$$



点推定 ～ 木を見て森を見ず



分布推論 ～ 個々の予測値ではなく、  
分布をみることで知識発見を促進

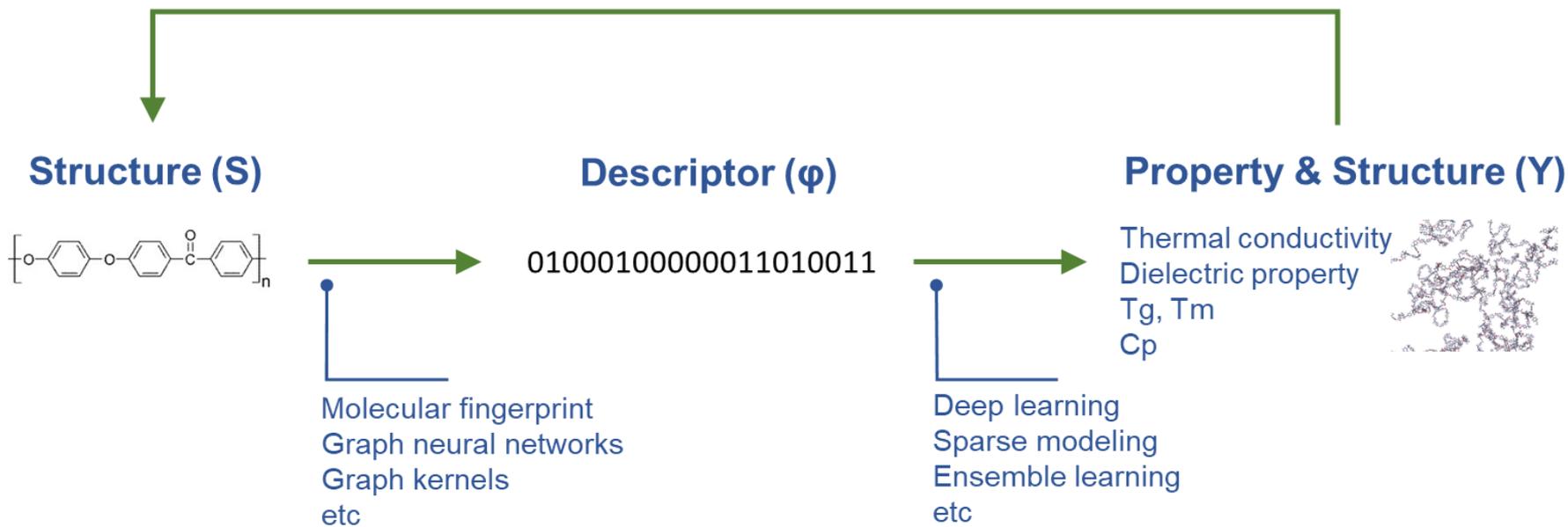


データ科学・シミュレーション・スパコン

不完全なモデルと不確かな現実系の壁を乗り越える

# Prediction and Discovery of New Polymers

Inverse problem: find  $S$  with desired  $Y^*$  with  $S = f^{-1}(Y^*)$



Forward problem: with a given dataset on (Y, S), build  $Y = f(S)$

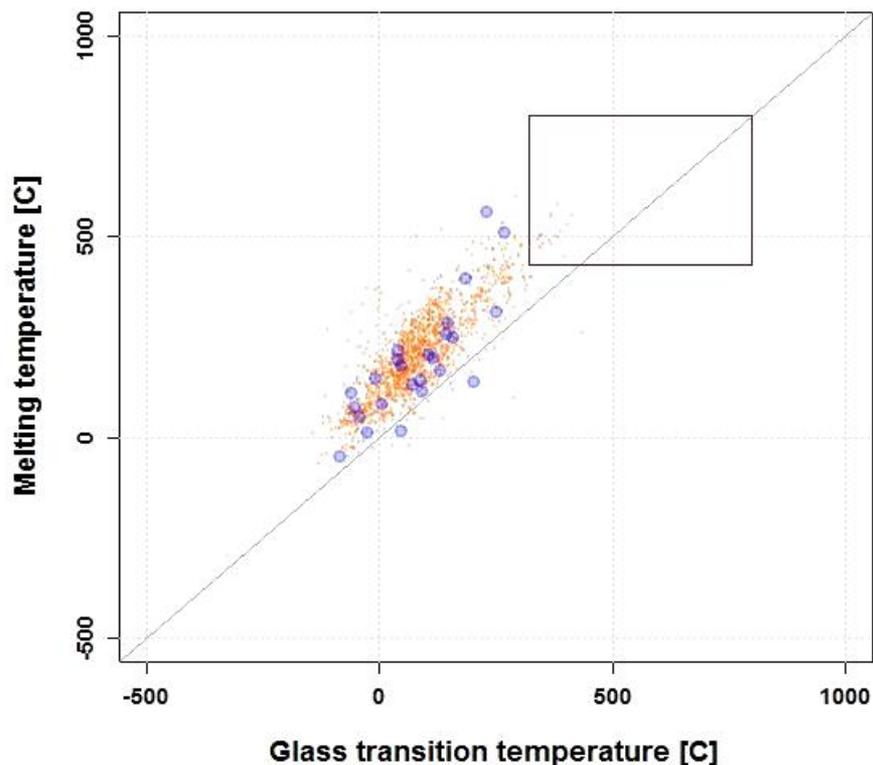
- Guo et al. Bayesian algorithm for retrosynthesis. *J Chem In. Model.* 60:4474–4486 (2020)
- Wu et al. iQSPR in XenonPy: a Bayesian inverse molecular design algorithm. *Mol Inform.* 39:1900107 (2019)
- Yamada et al. Predicting materials properties with little data using shotgun transfer learning. *ACS Cent Sci.* 5:1717-1730 (2019)
- Wu et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput Mater.* 5:66 (2019)
- Ikebata et al. Bayesian molecular design with a chemical language model. *J Comput Aided Mol Des.* 31:379-391 (2017)

# Molecular Design using Generative Machine Learning

Ikebata et al. Bayesian molecular design with a chemical language model. *J Comput Aided Mol Des.* 31:379-391 (2017)

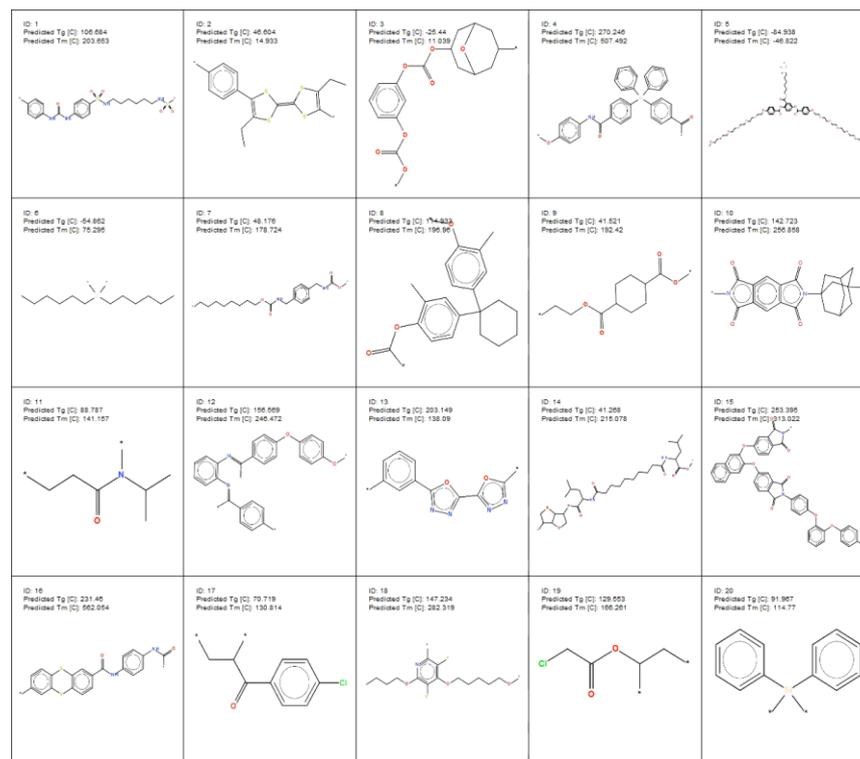
## Refinement of polymer properties

$T_g \nearrow$  and  $T_m \nearrow$



## Designed molecules (monomers)

Generative models of chemical structures



## Discovery of thermally conductive amorphous polymers with $T_c > 0.4 \text{ WmK}$

Wu et al. Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput Mater*, 5:66 (2019)

# 高分子材料インフォマティクスにおけるデータ資源の不足

無機材料系では第一原理計算データベースがMIの発展を牽引

## Polymer Property Database

### ■ PoLyInfo (literature survey)

~100 properties for 18,015 polymers

NO API

<https://polymer.nims.go.jp/>

### ■ Polymer Genome (DFT)

~10 properties for ~800 polymers

NO API

<https://www.polymergenome.org/?m=home>

### ■ CROW (literature survey)

NO API

<http://www.polymerdatabase.com/>

### ■ CRIPT (not yet open)

MIT + Citrine Informatics

<https://cript.mit.edu/>

## Inorganic Compounds

### ■ Materials Project (DFT)

~130,000 virtual and real compounds

Accessible via API

<https://materialsproject.org/>

### ■ AFLOW (DFT)

~3,000,000 virtual and real compounds

Accessible via API

<http://www.aflowlib.org/>

### ■ OQMD (DFT)

~600,000 virtual and real compounds

Accessible via API

<http://oqmd.org/>

### ■ ICSD (experiment)

~240,000 real compounds

Accessible via API

<https://icsd.products.fiz-karlsruhe.de/>

# 短中期的には大学のラボや1企業で生産可能なデータが標準に

このままでは高分子材料のマテリアルズインフォマティクスは普及・発展しない

## コスト

実験・シミュレーション・試料作製・物性評価に要するコストが高い。

## ニーズ 多様性

研究者の興味や設計変数（材料種，試料の作製方法など）が多様  
コモンデータを創出しようという動きが起きにくい。

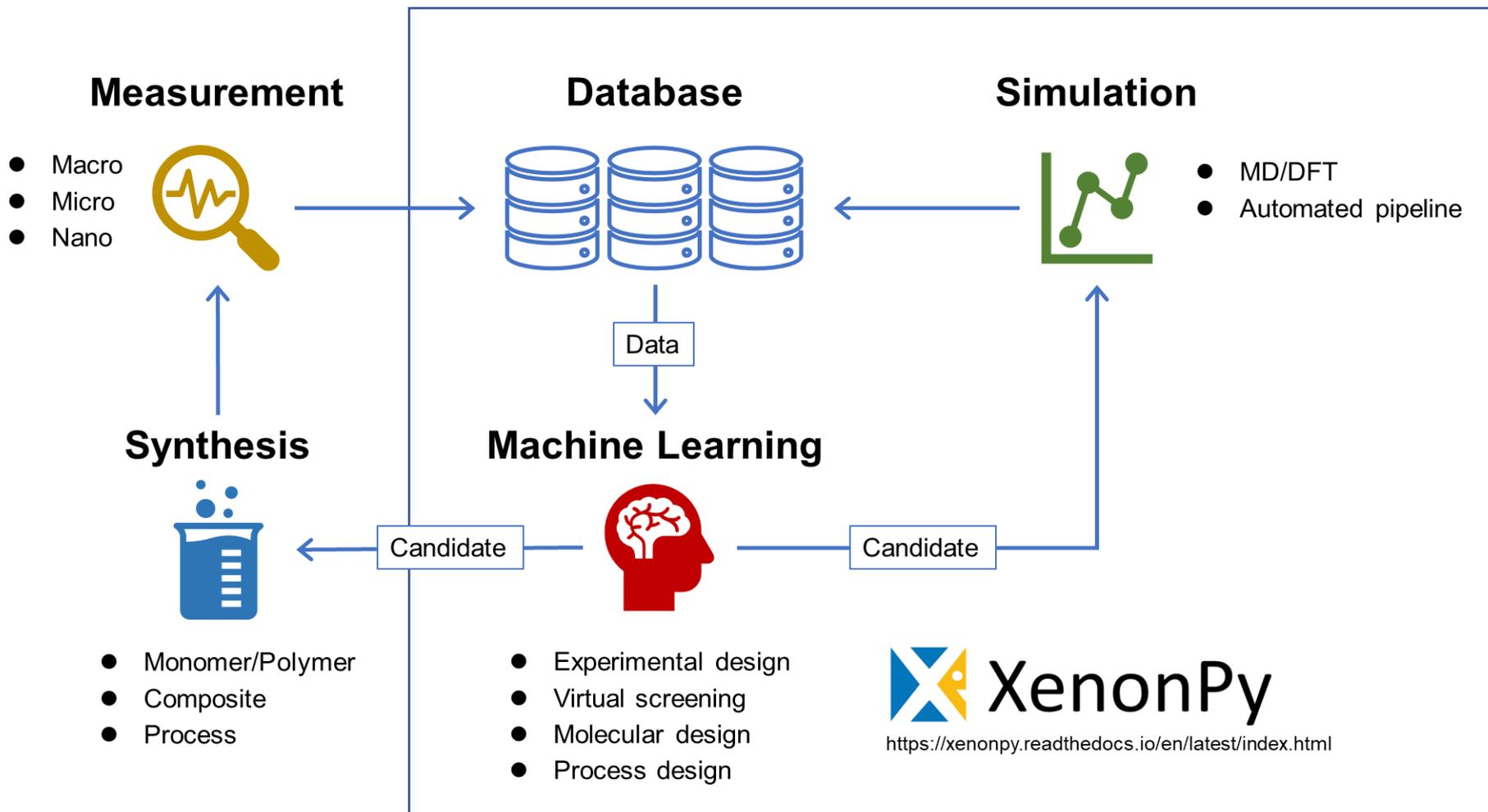
## インセン ティブ欠如

基礎と応用の垣根が低い。  
競合相手に対する情報秘匿の意識が高い。  
データを公開するインセンティブが研究者に働きにくい。

# 高分子材料データ駆動研究プラットフォーム

JST-CREST 熱制御「高分子の熱物性マテリアルズインフォマティクス」(2019.10-) 代表：森川淳子（東工大）

機械学習・計算科学グループ：吉田亮・林慶浩・Wu Stephen・野口瑤・Chang Liu・塩見淳一郎・川内進





# 全原子分子動力学法による高分子物性計算の自動化

RadonPy

Hayashi et al. RadonPy: Automated Physical Property Calculation using All-atom Classical Molecular Dynamics Simulations for Polymer Informatics. npj Comput Mater (2022) in press.

GitHub: <https://github.com/RadonPy/RadonPy>

## Latest: 15 properties implemented

(2022/03/31)

- ◆ Thermal conductivity
- ◆ Thermal diffusivity
- ◆ Density
- ◆ Radius of gyration
- ◆ Specific heat capacity  $C_p$
- ◆ Specific heat capacity  $C_v$
- ◆ Compressibility (isothermal)
- ◆ Isentropic compressibility
- ◆ Bulk modulus (isothermal)
- ◆ Isentropic bulk modulus
- ◆ Self-diffusion coefficient
- ◆ Thermal expansion coefficient
- ◆ Linear expansion coefficient
- ◆ Dielectric constant (static)
- ◆ Refractive index

## Polymer systems

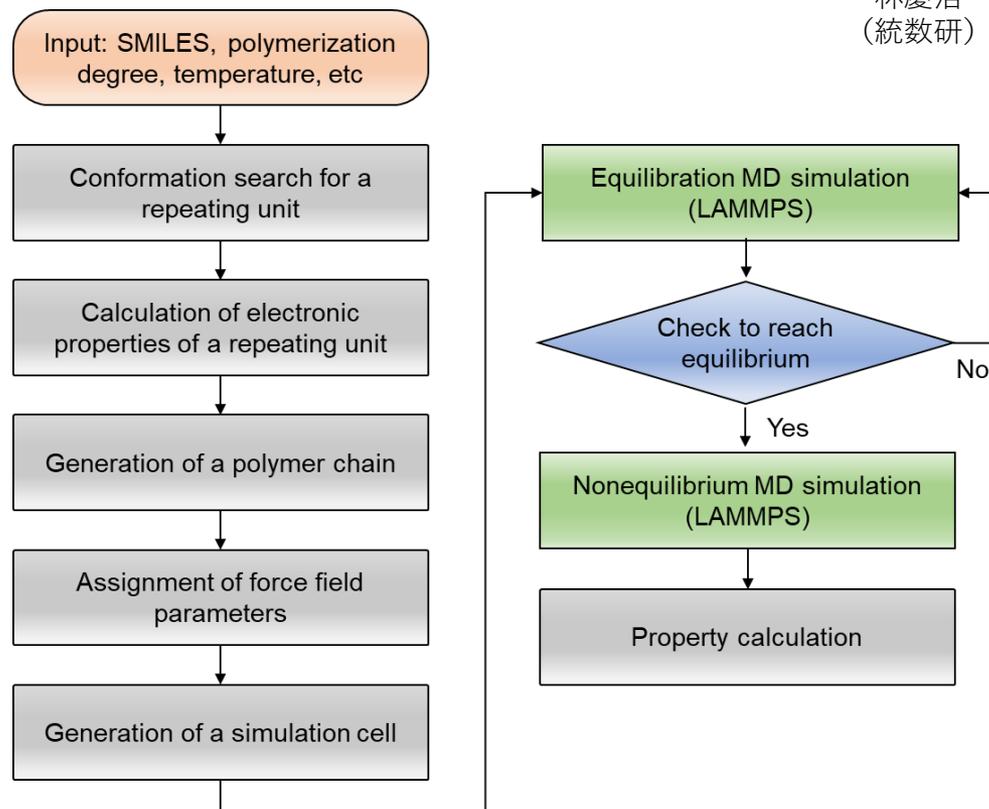
- ◆ Amorphous/crystalline states
- ◆ Oriented structures
- ◆ Mixtures

## RadonPy (OSS)

Python library for the automation of all-atom MD simulations using LAMMPS



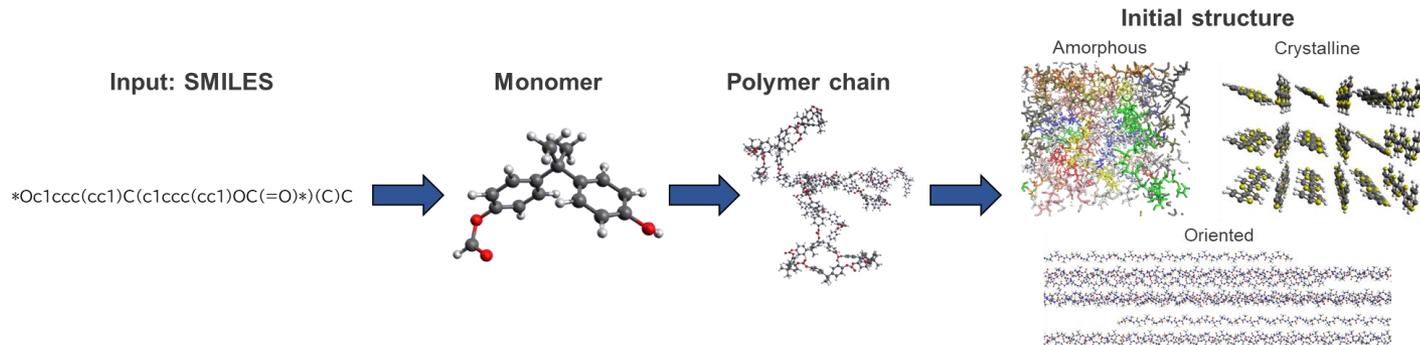
林慶浩  
(統数研)



# 高分子物性自動計算の難しさ：ノウハウとコスト

## 計算条件・パラメータ設定の「ノウハウ」やパイプライン構築のツールが必要

- シミュレーション条件（アニーリング条件・重合度・時間刻み幅・長距離相互作用など、物性値毎に検討が必要）
- 系の平衡化の終了判定（各種物理量の収束状況から判定）
- MD計算の妥当性の評価（温度勾配・配向度など、物性値毎に検討が必要）



## 膨大な計算コスト

- 1ポリマーの計算に100時間以上（延伸配向したポリマーの熱伝導率@分子研スパコン）
- 延伸配向した2,000ポリマーの計算 = ABCIで換算すると2,500万円分
- 遅い緩和時間の問題であり、スパコンの並列数が増えても解決しない。

# 産学連合体による高分子物性オープンデータベースの共創

2021- RadonPy DB 共同開発コンソーシアム@統計数理研究所ものづくりデータ科学研究センター

10<sup>5</sup>-10<sup>7</sup>ポリマーの包含する世界初の体系的な高分子物性データベースの創出

- 統数研・3大学・19企業が組織の垣根を越えてデータを共同生産・共有
- 「富岳」成果創出加速プログラム「データ駆動型高分子材料研究を変革するデータ基盤創出」

計算候補ポリマーライブラリ

- ◆ 市販ポリマー・公共データベースのポリマー
- ◆ 機械学習で生成した仮想ポリマー



機械学習（実験計画法）



1国研・3大学・19企業（R3年度）によるデータの共同生産

コントリ  
ビュータ1

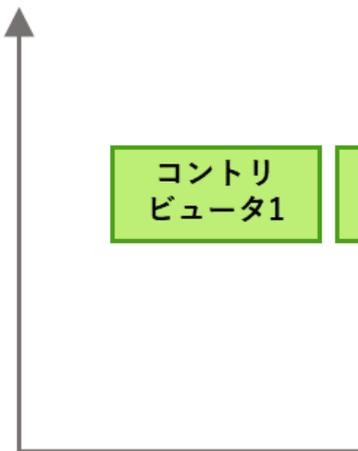
コントリ  
ビュータ2

コントリ  
ビュータ3



包括的な高分子物性データベース

- ◆ 10万種類以上の高分子骨格
- ◆ 高分子集合系，高分子液晶，架橋高分子
- ◆ 熱物性・力学的特性など

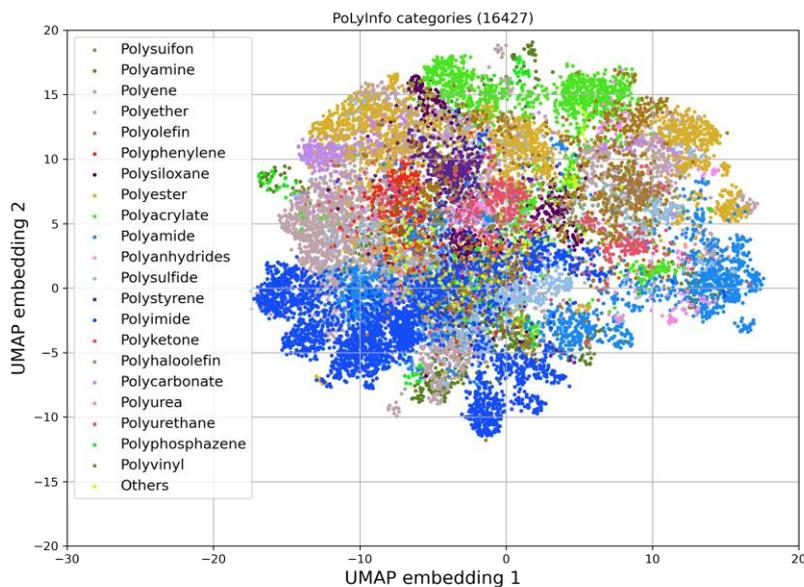


# 高分子物性大地図の作成に向けて

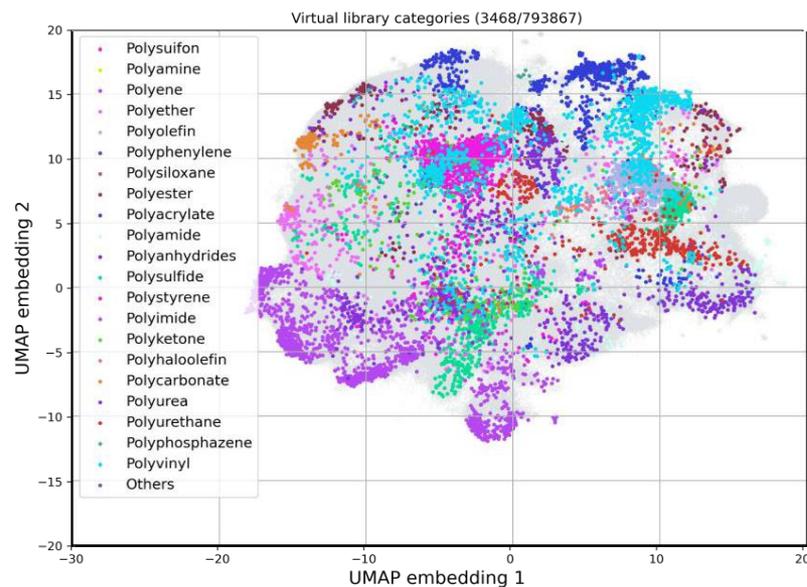
10<sup>5</sup>-10<sup>7</sup>個の分子骨格が張るケミカルスペースの全体像を明らかにする

- 複数物性の同時分布を観測し，パレートフロンティアの位置や構造的特徴を明らかに
- 特異な特性を持つポリマーを予測・合成
- 機械学習のデータ資源としての利用

これまでに合成された高分子の分布  
(PoLyInfo: 16,427)



2022/09/13  
計算完了ポリマー：30,982  
(平衡MD: 18586, 非平衡MD: 12,396)



# 計算物性のバイアスとバリエーション



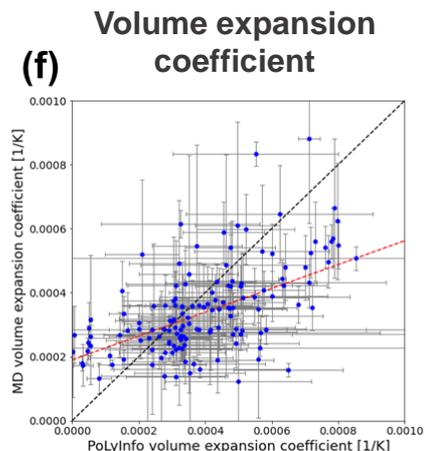
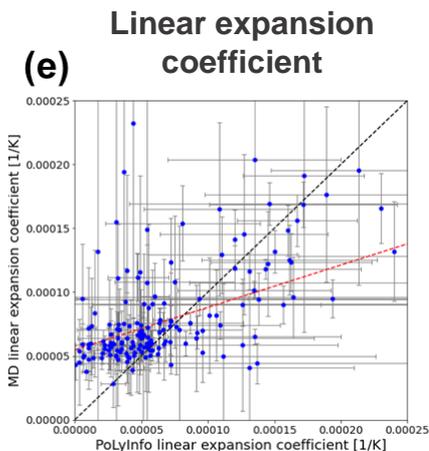
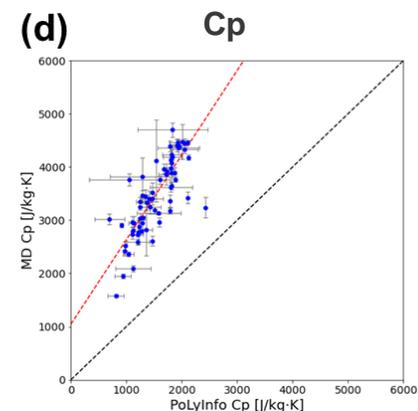
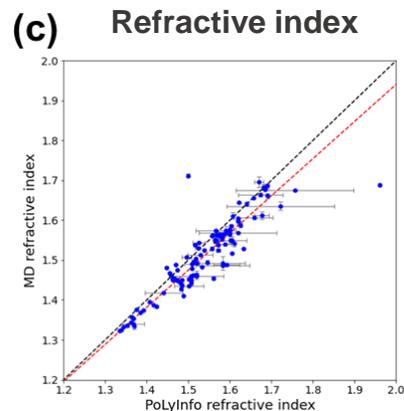
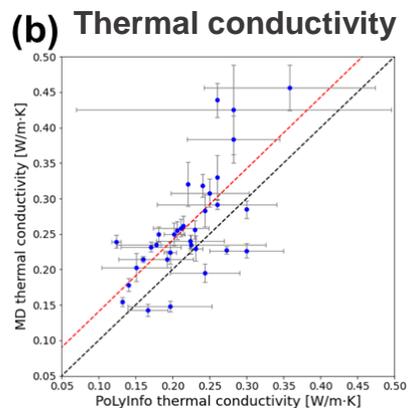
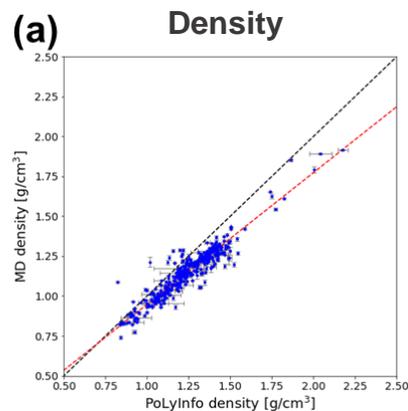
vs



RadonPy

Hayashi et al. npj Comput Mater (2022) in press.

- PoLyInfo 1,070 ポリマー, 15物性の計算, 5回の独立な試行
- 実験値との比較により計算条件 (プリセット) の妥当性を検証

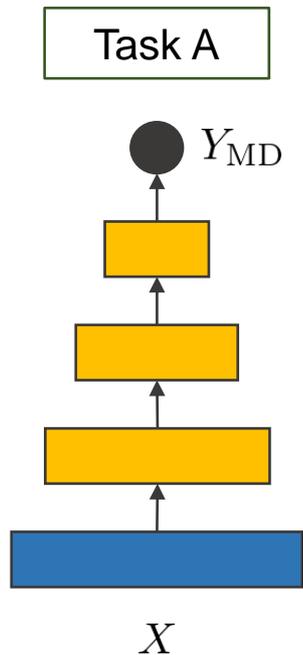


Vertical axis: predicted properties (RadonPy)  
Horizontal axis: experiments (PoLyInfo)

# キャリブレーション：転移学習

- 計算物性のデータ集合から特徴抽出器を学習  $f_s(X)$
- 実験データへの変換関数  $g$  を推定  $Y_{\text{EXP}} = g(f_s(X))$

シミュレーションデータ

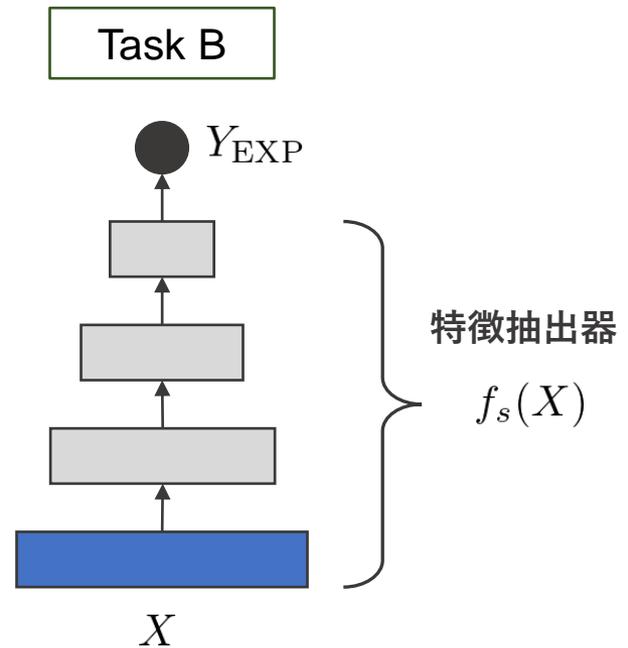


ポリマー化学構造

転移学習



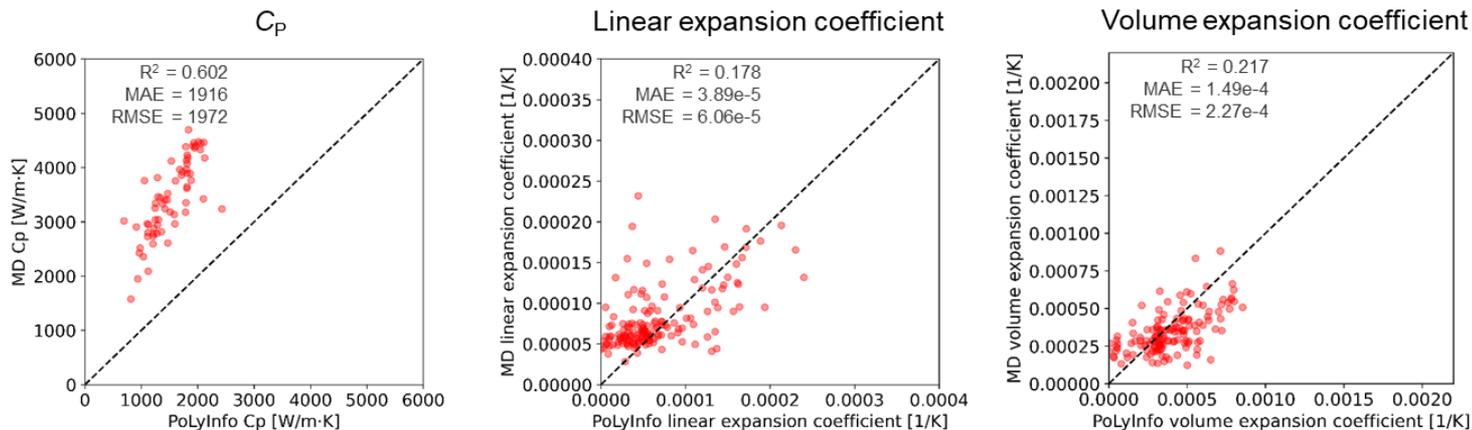
実験データ



# 計算物性・実験値のバイアスとバリエーションが大幅に減少

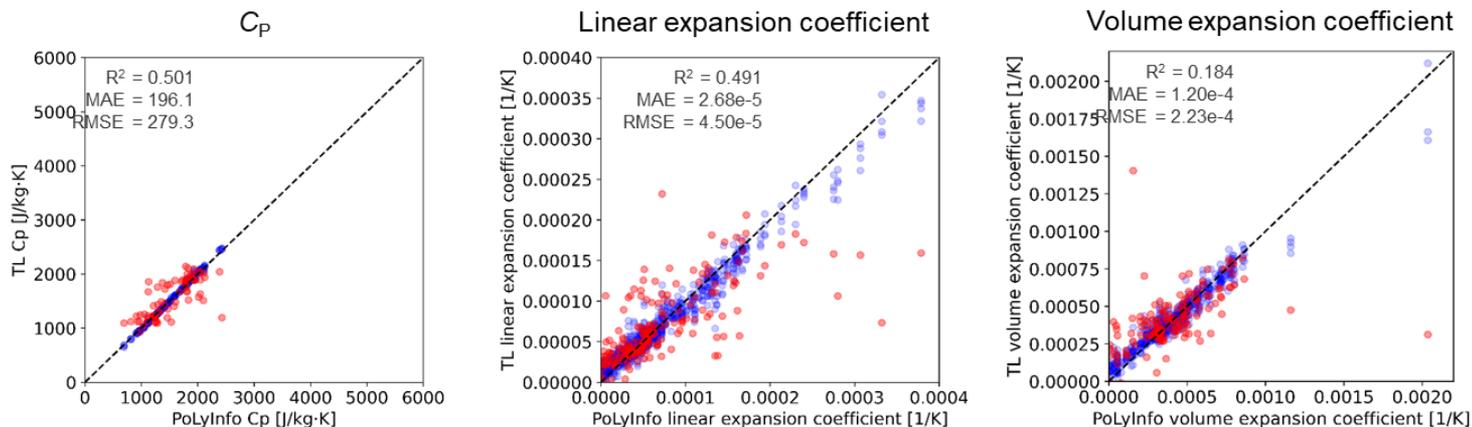
Hayashi et al. npj Comput Mater (2022) in press.

## (a) MD simulation



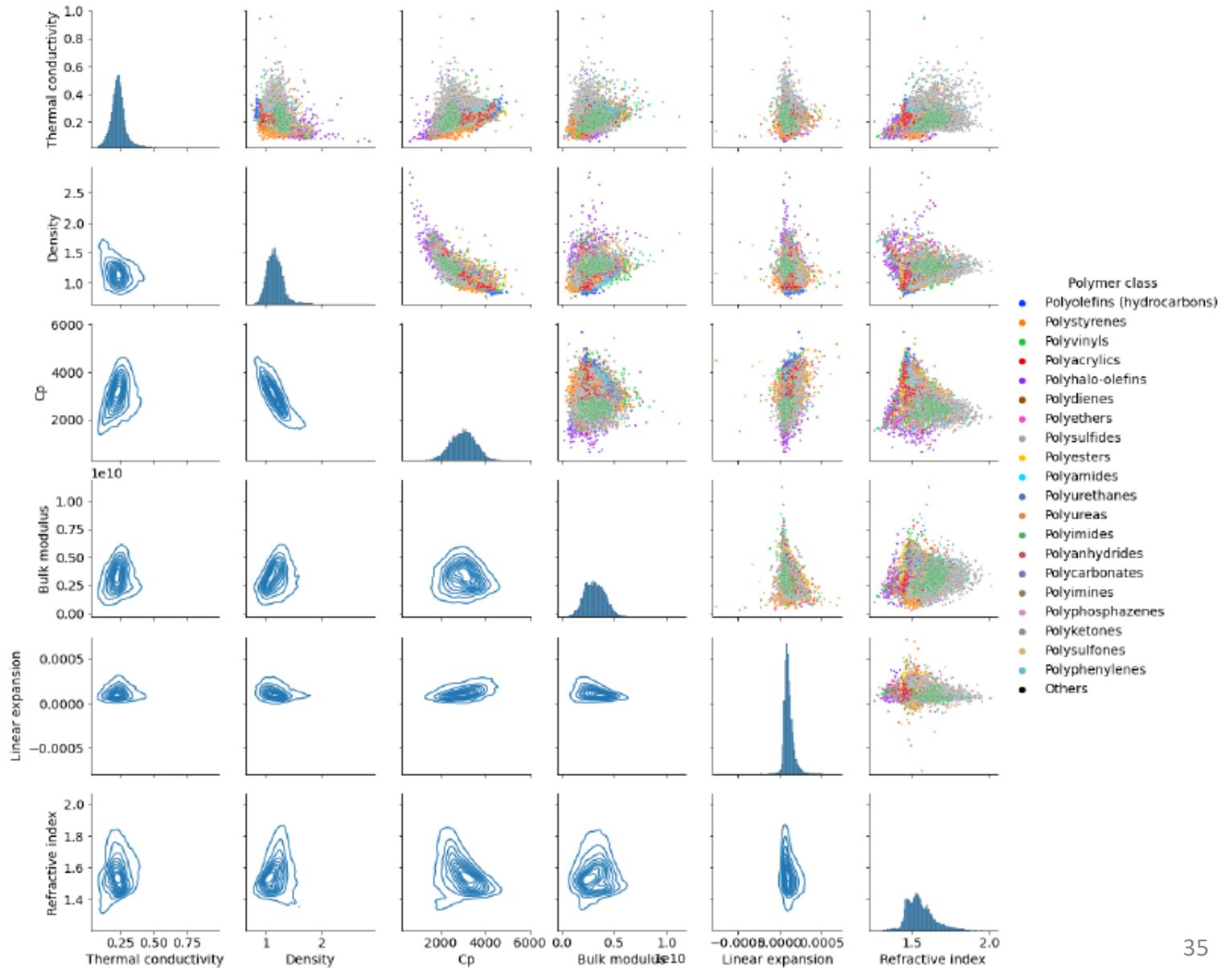
**Reduction of bias and variance**

## (b) Transfer learning



# 複数物性のパレートフロントがみえてくる

Hayashi et al. npj Comput Mater (2022) in press.



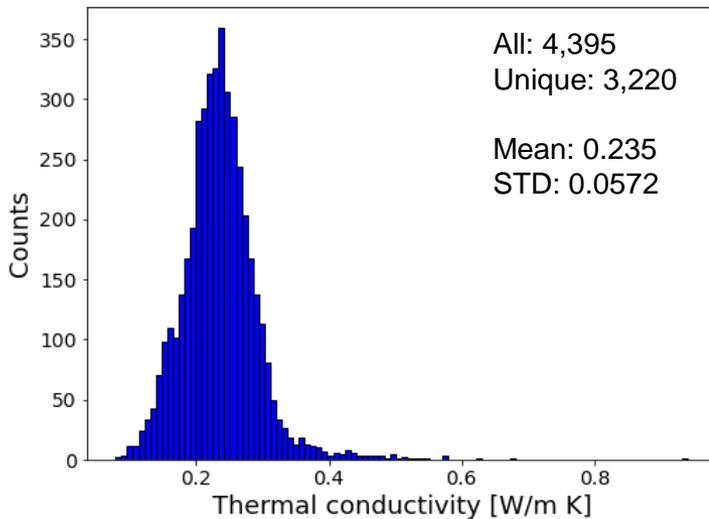
# 特異な物性を持つポリマーの同定

Hayashi et al. npj Comput Mater (2022) in press.

熱伝導率が0.5 W / (m · K)を超えるアモルファスポリマーが存在するという予想

- ✓ 通常のアモルファスポリマーの熱伝導率は0.2-0.3 W/mKほど
- ✓ 水素結合可能なユニットが高密度で存在：水素結合やdipole-dipole相互作用を介した熱伝導
- ✓ 分子骨格の剛直性により共有結合を介した熱伝導

3,220ポリマーの熱伝導率の分布



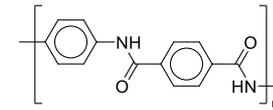
PI1 (0.456 W/m·K)



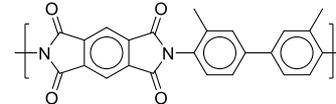
PI241 (0.439 W/m·K)



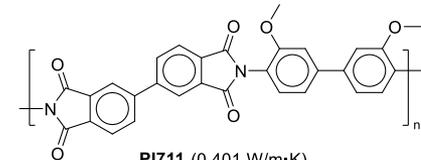
PI305 (0.503 W/m·K)



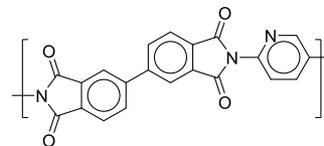
PI626 (0.473 W/m·K)



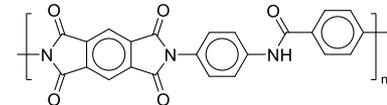
PI687 (0.581 W/m·K)



PI711 (0.401 W/m·K)



PI715 (0.483 W/m·K)



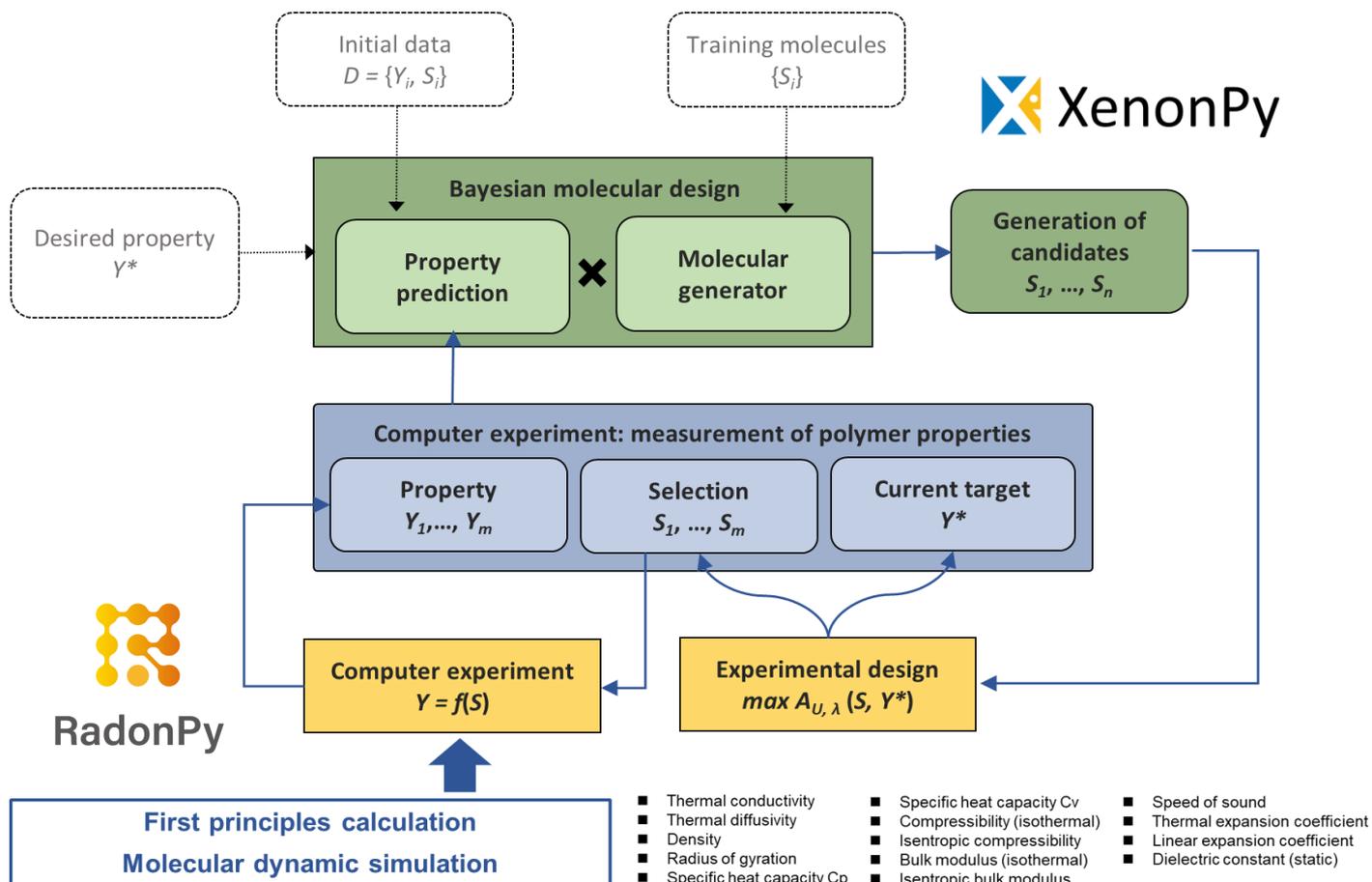
PI1093 (0.501 W/m·K)



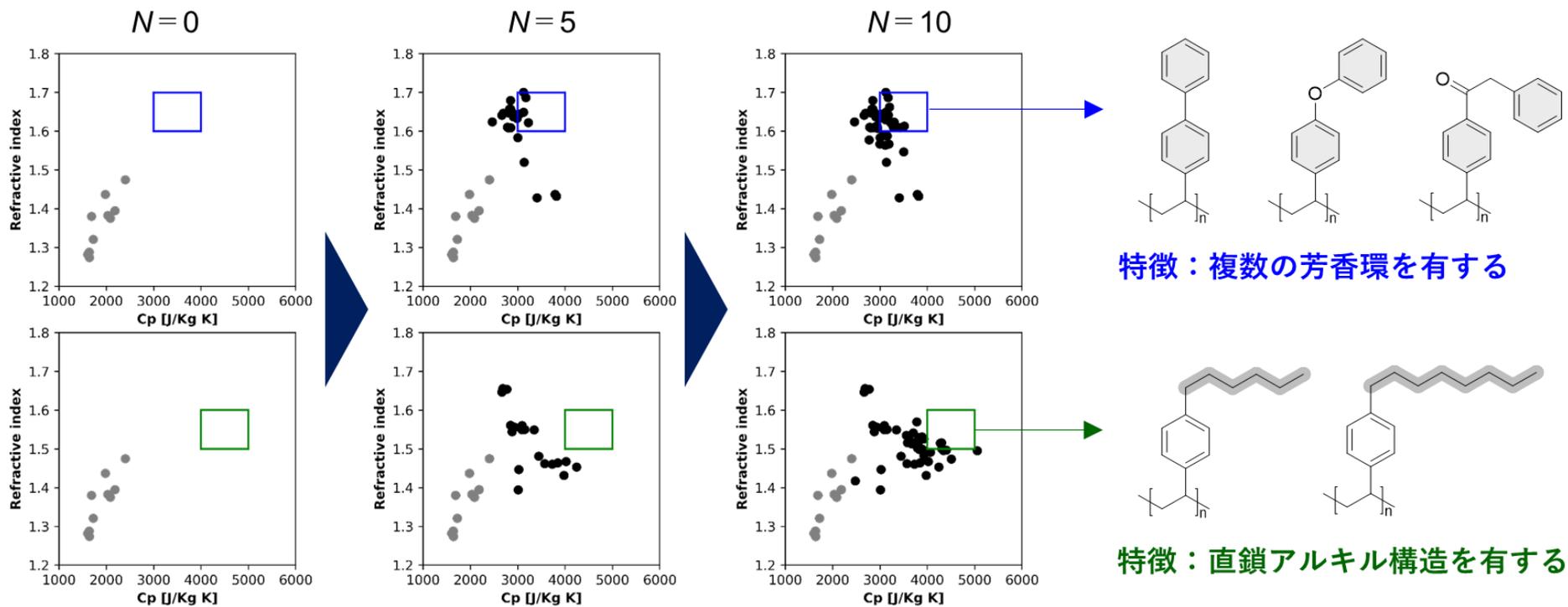
# 実験・シミュレーション・機械学習の融合

データ科学の内挿的予測の限界を乗り越える

- ✓ 適応的実験計画法に基づき、データが存在しない未踏領域を埋めていく
- ✓ 高分子物性自動計算 + 機械学習による自動分子設計 は未踏課題

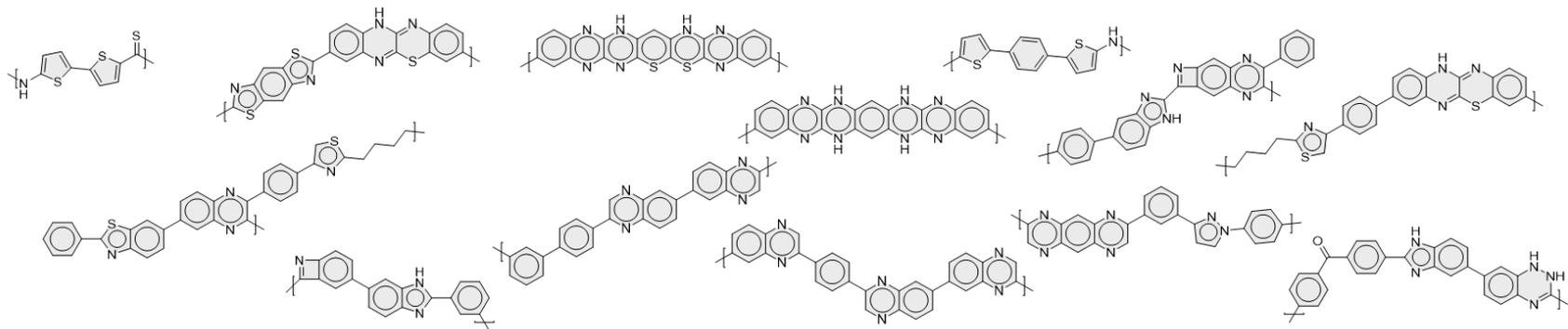


# MD計算+機械学習でケミカルスペースを自由自在に走査

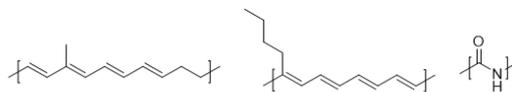
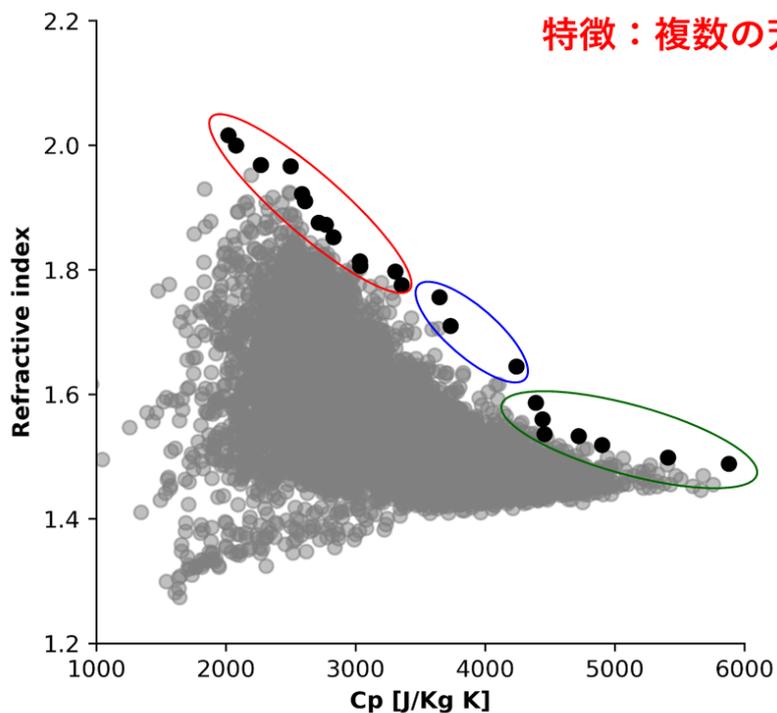


初期データ（灰色）からスタートし、要求物性（四角）を満たす高分子を自動設計

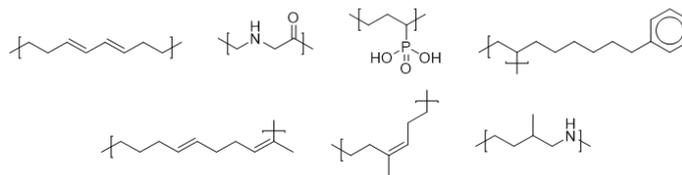
# パレートフロンティアを構成する高分子を探索する



特徴：複数の芳香環から構成される



特徴：sp<sup>2</sup>炭素の割合が高い

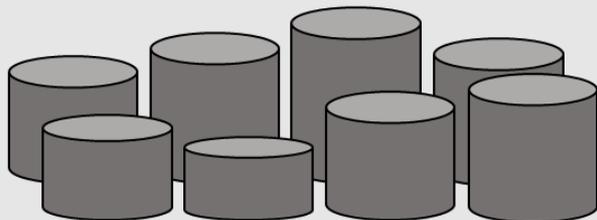


特徴：sp<sup>3</sup>炭素の割合が高い

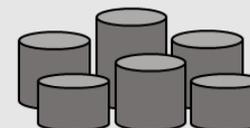
- ✓ 不完全なモデルと不確かな現実系：誤った「木」を観察しているだけでは何も得られない。データ科学は「森を見て木を知る」ための有用な道具
- ✓ 社会が協調してデータ駆動型研究に資する体系的・包括的なデータ群を創出していくことがいまやるべきこと

オープン領域

シミュレーションによる基礎物性・構造データベース  
第一原理電子状態計算・分子動力学法



中規模な実験データベース  
オープン化は中々進まない  
文献・特許・自動実験



統合解析のためのデータ科学

クローズ領域

素材企業



大学の研究室



中規模国プロ



コンソーシアム



AIスタートアップ