機械学習ポテンシャルを活用した結晶構造予測

劉　暢　　ものづくりデータ科学研究センター　特任助教

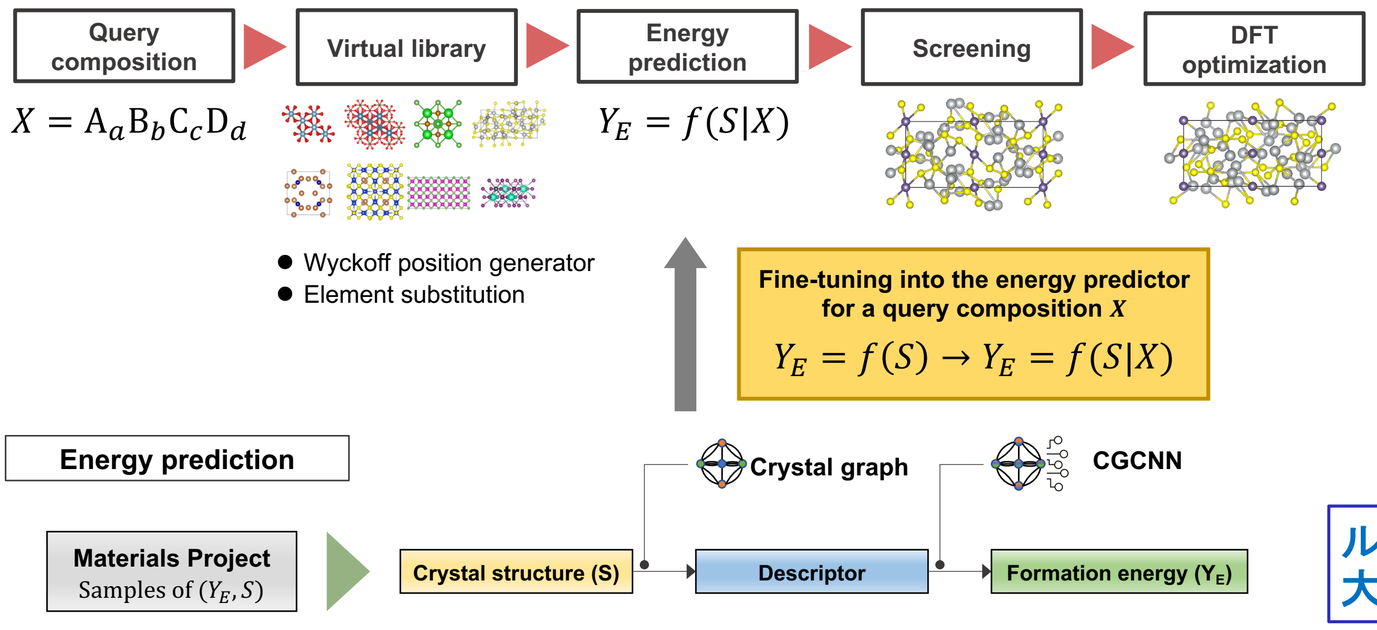
Introduction

結晶構造予測の目標は，標的分子の構造式の情報のみから，その分子の安定に存在する結晶構造を提案することである．近年では，機械学習による探索手法が数多く提案されたが，予測精度や計算コストなどの問題は依然として存在する。結晶構造予測の効率・高速化は、産業界から強く期待されている．

現段階では以下の成果を得ている．

1. C. Liu *et al.*, “Shotgun crystal structure prediction using machine-learned formation energies,” (submitted), DOI: 10.48550/arXiv.2305.02158
2. M. Kusaba *et al.*, “Crystal structure prediction with machine learning-based element substitution,” Comput. Mater. Sci., DOI: 10.1016/j.commatsci.2022.111496
3. Code: https://github.com/yoshida-lab/XenonPy/blob/master/samples/CSP_with_element_substitution.ipynb

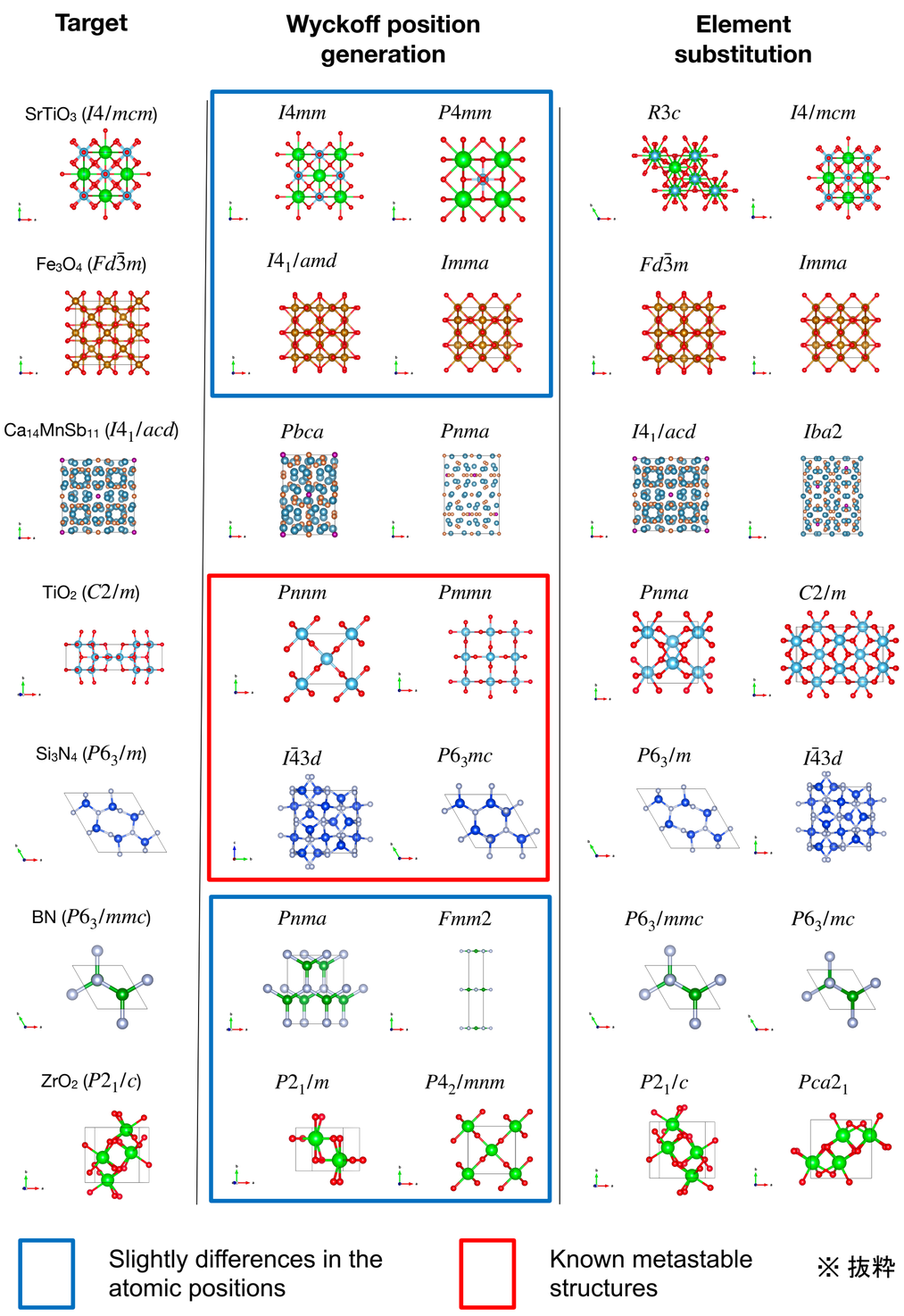
Workflow



1. 化学組成からバーチャルライブラリーを生成する
   - Wyckoff position generator
   - Elements substitution
2. バーチャルライブラリーの構造エネルギーを評価し構造候補を選択する
   - CGCNN (転移学習)
3. 候補構造を最適化して，安定構造を提案する

ループ計算が要らないため，構造最適化にかかる時間が大幅に削減！

Results



Dataset I. 40 benchmark crystals selected based on the diversity of space groups, constituent elements, number of atoms, and element species; and the diversity of applications such as battery and thermoelectric materials

Composition	Number of atoms	Space group	Wyckoff position generation	Element substitution	USPEX R1	USPEX R2
C	4	$R\bar{3}m$	✓ (✓)	✓ (✓)	×	×
Si	2	$Fd\bar{3}m$	✓ (✓)	✓ (✓)	✓	✓
GaAs	2	$F\bar{4}3m$	✓ (✓)	✓ (✓)	✓	✓
ZnO	4	$P6_3mc$	✓ (✓)	✓ (✓)	✓	✓
BN	4	$P6_3/mmc$	× (×)	✓ (✓)	×	×
$LiCoO_2$	16	$R\bar{3}m$	✓ (✓)	✓ (✓)	×	×
Bi_2Te_3	5	$R\bar{3}m$	✓ (✓)	✓ (✓)	×	×

Element substitution (top 10) ※ : 82.5%
Wyckoff generator (top 10) ※ : 47.5%
USPEX with known space groups: 32.5%
USPEX with unknown space groups: 12.5%
※ 15,000 candidate crystals

Success of the space group prediction (top 10): 84.4%
→ Fail to predict 19 structures
→ Unable to generate the true Wyckoff labels of 16 structures

Composition	Number of atoms	Space group	Wyckoff position generation	Element substitution	USPEX R1	USPEX R2
$Ba(FeAs)_2$	5	$I4/mmm$	✓ (✓)	✓ (✓)	×	✓
SiO_2	6	$I\bar{4}2d$	× (−)	× (✓)	×	×
VO_2	6	$P4_2/mnm$	✓ (×)	✓ (✓)	×	✓
La_2CuO_4	7	$I4/mmm$	× (×)	✓ (✓)	×	×
$LiPF_6$	8	$R\bar{3}$	✓ (✓)	✓ (✓)	×	✓
Al_2O_3	10	$R\bar{3}c$	✓ (✓)	✓ (✓)	✓	✓
$SrTiO_3$	10	$I4/mcm$	× (×)	✓ (✓)	×	×
$CaCO_3$	10	$R\bar{3}c$	✓ (✓)	✓ (✓)	×	✓
TiO_2	12	$C2/m$	× (×)	× (✓)	×	✓
ZrO_2	12	$P2_1/c$	× (×)	✓ (✓)	×	×
$ZrTe_5$	12	$Cmcm$	× (×)	✓ (✓)	×	×
V_2O_5	14	$Pmmn$	× (×)	× (×)	×	×
Si_3N_4	14	$P6_3/m$	× (×)	✓ (✓)	✓	✓
Fe_3O_4	14	$Fd\bar{3}m$	× (✓)	✓ (✓)	×	×
$Mn(FeO_2)_2$	14	$Fd\bar{3}m$	✓ (✓)	✓ (✓)	×	✓
ZnSb	16	$Pbca$	✓ (−)	✓ (✓)	×	✓
$CoSb_3$	16	$Im\bar{3}$	✓ (✓)	✓ (✓)	×	×
$LiBF_4$	18	$P3_121$	× (✓)	✓ (✓)	×	✓
Y_2Co_{17}	19	$R\bar{3}m$	✓ (✓)	✓ (✓)	×	×
GeH_4	20	$P2_12_12_1$	× (−)	✓ (✓)	−	×
$CsPbI_3$	20	$Pnma$	× (×)	✓ (✓)	×	×
$NaCaAlPHO_5F_2$	24	$P2_1/m$	× (×)	−	×	×
$LiFePO_4$	28	$Pnma$	✓ (✓)	✓ (✓)	×	×
$Cu_{12}Sb_4S_{13}$	29	$I\bar{4}3m$	✓ (✓)	✓ (✓)	×	×
MgB_7	32	$Imma$	× (×)	× (×)	−	×
Li_3PS_4	32	$Pnma$	× (×)	× (×)	×	×
Cd_3As_2	80	$I4_1/acd$	✓ (✓)	✓ (✓)	−	×
$Li_4Ti_5O_{12}$	42	$C2/c$	× (×)	✓ (✓)	−	×
$Ba_2CaSi_4(BO_7)_2$	46	$I\bar{4}2m$	× (×)	× (×)	−	×
Ag_8GeS_6	60	$Pna2_1$	× (✓)	✓ (✓)	−	×
$Nd_2Fe_{14}B$	68	$P4_2/mnm$	× (×)	✓ (✓)	−	−
$Y_3Al_5O_{12}$	80	$Ia\bar{3}d$	✓ (✓)	✓ (✓)	−	×
$Ca_{14}MnSb_{11}$	104	$I4_1/acd$	× (×)	✓ (✓)	−	−
Overall			19/40 = 47.5%	33/40 = 82.5%	5/40 = 12.5%	13/40 = 32.5%

Dataset II. 50 benchmark crystals randomly selected from the Materials Project database

Composition	Number of atoms	Space group	Wyckoff position generation	Element substitution	USPEX
CsCl	2	$Fm\bar{3}m$	✓ (✓)	✓ (✓)	−
MnAl	2	$P4/mmm$	✓ (✓)	✓ (✓)	−
HoHSe	3	$P\bar{6}m2$	✓ (✓)	✓ (✓)	✓
$ErCdRh_2$	4	$Fm\bar{3}m$	✓ (✓)	✓ (✓)	×
Eu_2MgTl	4	$Fm\bar{3}m$	✓ (✓)	✓ (✓)	×
Pm_2NiIr	4	$Fm\bar{3}m$	✓ (✓)	✓ (✓)	×
VPt_3	4	$I4/mmm$	✓ (✓)	✓ (✓)	✓
$Gd(SiOs)_2$	5	$I4/mmm$	✓ (✓)	✓ (✓)	✓
$LaAl_3Au$	5	$I4mm$	✓ (✓)	✓ (✓)	×
U_2SbN_2	5	$I4/mmm$	× (✓)	✓ (✓)	×
$MnGa(CuSe_2)_2$	8	$I\bar{4}$	× (−)	✓ (✓)	×
SmZnPd	9	$P\bar{6}2m$	✓ (✓)	✓ (✓)	×
$Sn(TePd_3)_2$	9	$I4mm$	× (−)	✓ (✓)	×
V_2S_4	9	$I4/m$	× (✓)	✓ (✓)	×
Cs_3InF_6	10	$Fm\bar{3}m$	✓ (✓)	✓ (✓)	×
$Eu(CuSb)_2$	10	$P4/nmm$	× (✓)	✓ (✓)	×
$Rb_2TlAgCl_6$	10	$Fm\bar{3}m$	✓ (✓)	✓ (✓)	×
$Ca_3Ni_2B_2$	12	$R\bar{3}m$	× (✓)	✓ (✓)	×
$DyPO_4$	12	$I4_1/amd$	✓ (✓)	✓ (✓)	✓
LaSiIr	12	$P2_13$	× (✓)	✓ (✓)	×
$SmVO_4$	12	$I4_1/amd$	✓ (✓)	✓ (✓)	×
VCl_5	12	$P\bar{1}$	× (✓)	✓ (✓)	×
YbP_5	12	$P2_1/m$	− (−)	✓ (✓)	×
$Eu(Al_2Cu)_4$	13	$I4/mmm$	− (✓)	✓ (✓)	−
Zr_4O	15	$R\bar{3}$	× (−)	× (×)	×
$K_2Ni_3S_4$	18	$Fddd$	× (✓)	✓ (✓)	×
$Sr(ClO_3)_2$	18	$Fdd2$	× (✓)	✓ (✓)	×
$LiSm_3IrO_6$	20	$P2_1/c$	× (✓)	✓ (✓)	×
Pr_2ZnPtO_6	20	$P2_1/c$	× (✓)	✓ (✓)	×
$Sc_2Mn_{12}P_7$	21	$P\bar{6}$	− (×)	✓ (✓)	×
$LaSi_2Ni_9$	24	$I4_1$/amd	✓ (✓)	✓ (✓)	×
$CeCu_4Sn$	28	$Pnma$	✓ (✓)	✓ (✓)	−

Composition	Number of atoms	Space group	Wyckoff position generation	Element substitution	USPEX
$LiP(HO_2)_2$	32	$Pna2_1$	× (−)	✓ (×)	×
$Mg_3Si_2H_4O_9$	36	$P6_3cm$	× (−)	× (×)	×
$Y_4Si_4Ir_9$	36	$P6_3/mmc$	× (✓)	−	×
$Na(WO_3)_9$	37	$R\bar{3}$	− (−)	−	−
$Sm_6Ni_{20}As_{13}$	39	$P\bar{6}$	− (✓)	✓ (✓)	−
$BaCaGaF_7$	40	$P2/c$	× (✓)	✓ (✓)	×
$Tm_{11}Sn_{10}$	42	$I4/mmm$	− (×)	✓ (✓)	−
$AlH_{12}(ClO_2)_3$	44	$R\bar{3}c$	× (−)	✓ (✓)	×
$K_2ZrSi_2O_7$	48	$P2_1/c$	× (✓)	✓ (×)	×
$Ba_3Ta_2NiO_9$	60	$P\bar{3}m1$	× (×)	× (×)	×
$LiZr_2(PO_4)_3$	72	$P2_1/c$	× (×)	✓ (✓)	−
$K_5Ag_2(AsSe_3)_3$	76	$Pnma$	× (×)	−	×
$Be_{17}Ru_3$	80	$Im\bar{3}$	✓ (✓)	✓ (✓)	✓
$Cu_3P_8(S_2Cl)_3$	80	$Pnma$	× (×)	✓ (✓)	×
Al_2CoO_4	84	$P\bar{3}m1$	× (−)	✓ (✓)	×
$Li_6V_3P_8O_{29}$	92	$P1$	− (×)	✓ (×)	−
$ReBi_3O_8$	96	$P2_13$	× (×)	✓ (✓)	×
$Na_5FeP_2(O_4F)_2$	288	$Pbca$	− (−)	−	−
Overall			16/50 = 32.0%	43/50 = 86.0%	5/50 = 10.0%

Element substitution (top 10) ※ : 86.0%
Wyckoff generator (top 10) ※ : 32.0%
USPEX with unknown space groups: 10.0%

高精度・高性能を同時に実現！