Supporting Information for

Accelerated discovery of potential ferroelectric perovskite via active learning

Kyoungmin Min[1, *] and Eunseog Cho[2, *]

[1]School of Mechanical Engineering, Soongsil University, 369 Sangdo-ro, Dongjak-gu, Seoul, 06978, Republic of Korea

[2]Autonomous Material Development Lab, Samsung Advanced Institute of Technology, 130 Samsung-ro, Suwon, Gyeonggi-do, 16678, Republic of Korea.

*Corresponding author: kmin.min@ssu.ac.kr (K. Min), eunseog.cho@samsung.com (E. Cho)

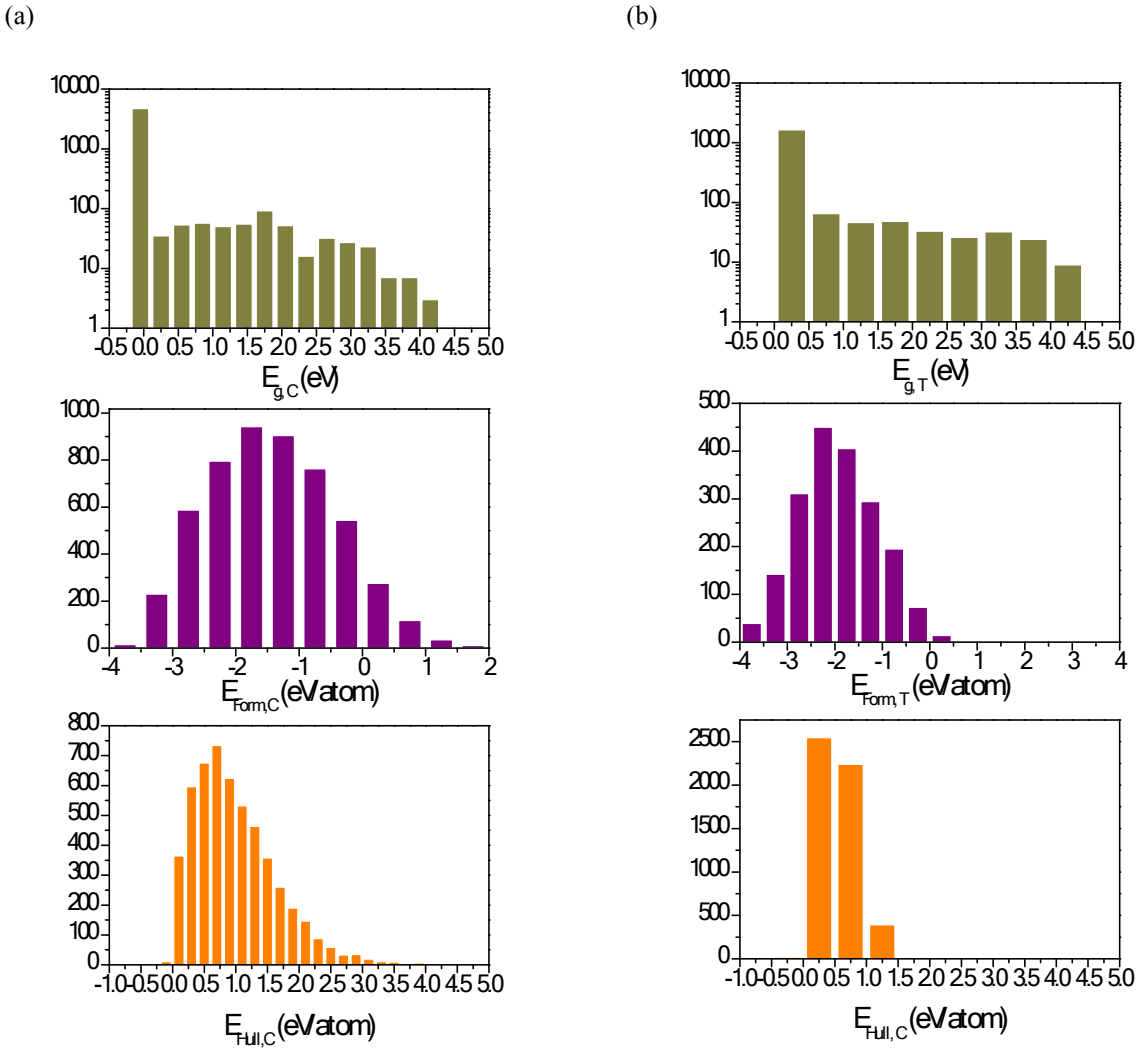
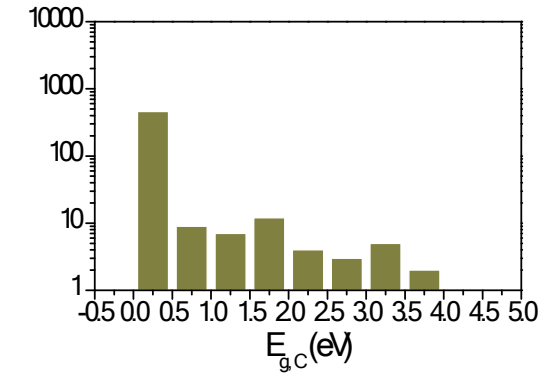
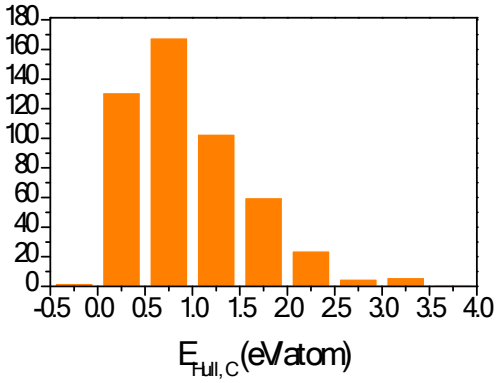


Figure S1. Distribution of band gap (E_g), formation energy (E_{Form}), and convex hull (E_{Hull}) database for (a) cubic and (b) tetragonal phase of ABO_3.

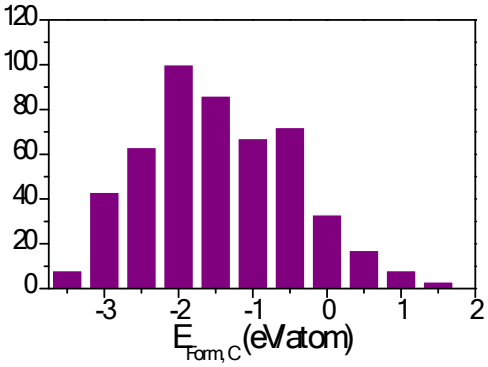
(a)



(c)



(b)

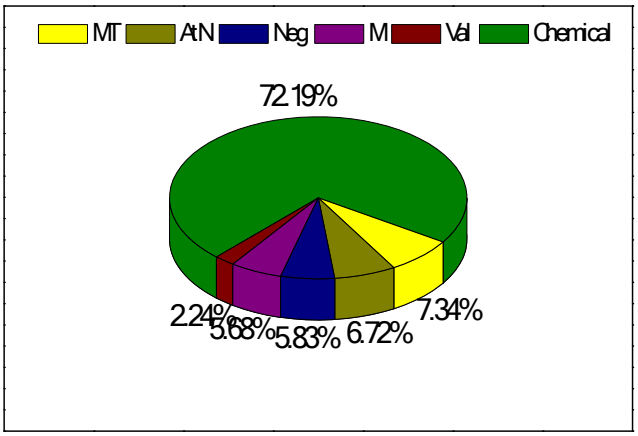


(d)

	Initial	Total
Category	**Number**	
Total	500	5218
$E_{Hull,\ C} < 0.5$ eV/atom	14	1285
$E_{Form,\ C} < 0.0$ eV/atom	456	4780
$E_{g,\ C} > 0.2$ eV	47	511
$E_{g,\ C} > 1.0$ eV	33	366

Figure S2. Initial database (500) before applying the active learning process for (a) band gap, (b) formation energy, (c) convex hull energy, and (d) database number satisfying the given condition.

(a)



(b)

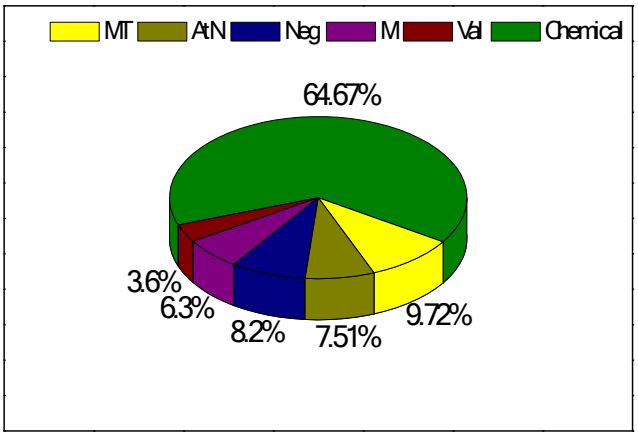


Figure S3. Portion of descriptors in feature importance for prediction model of (a) band gap and (b) formation energy. Each descriptor is abbreviated as follows: melting point (MT), atomic number (AtN), electro-negativity (Neg), Mendeleev number (M), and valence (Val).

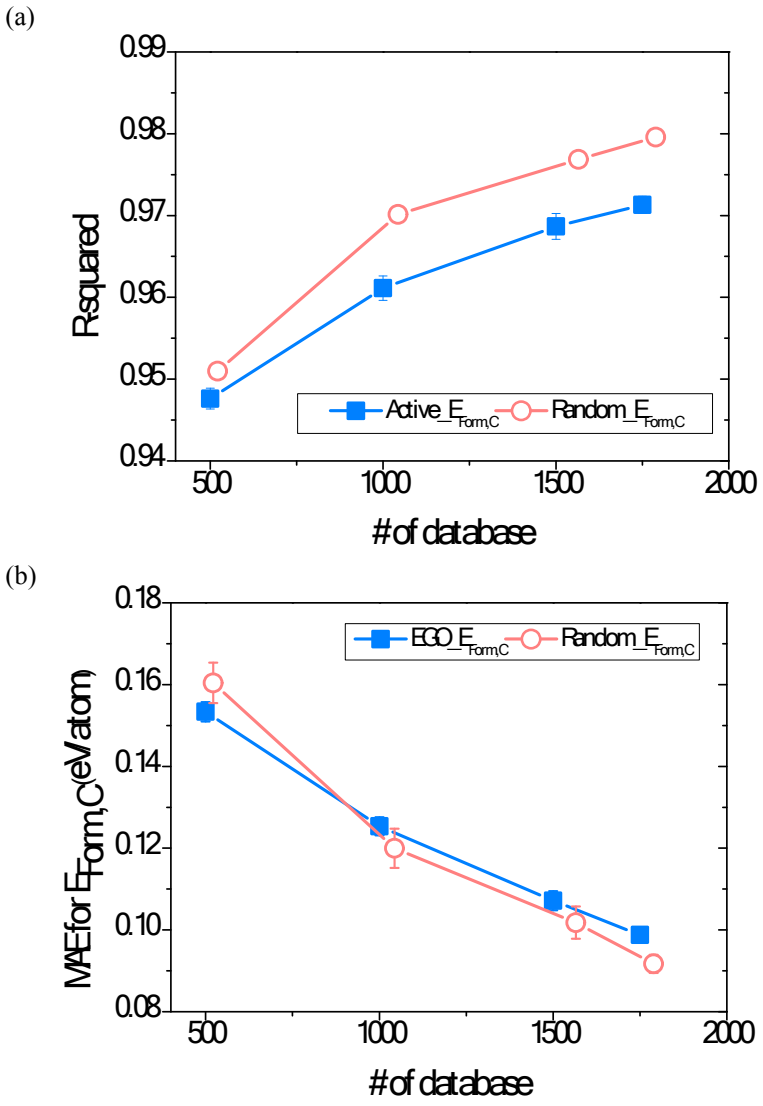


Figure S4. Prediction accuracy change in terms of (a) R^2 and (b) mean absolute error (MAE) with respect to the increase in the database number for the formation energy ($E_{form,\,C}$) of the cubic phase. The database is suggested from the active learning for $E_{g,\,C}$.

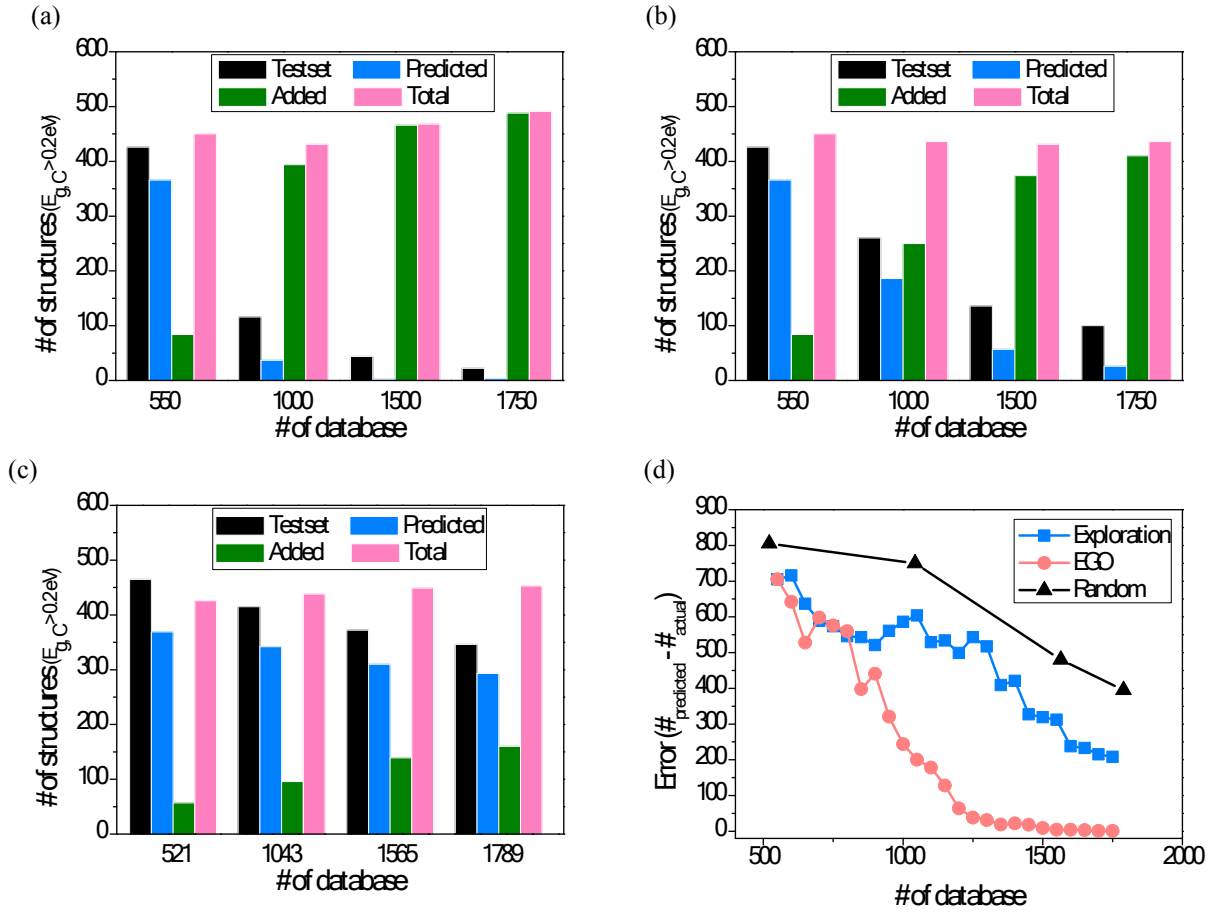


Figure S5. Number of structures satisfying the given condition for the band gap when a new database is added based on (a) EGO, (b) exploration, (c) random active learning, and (d) random selection.

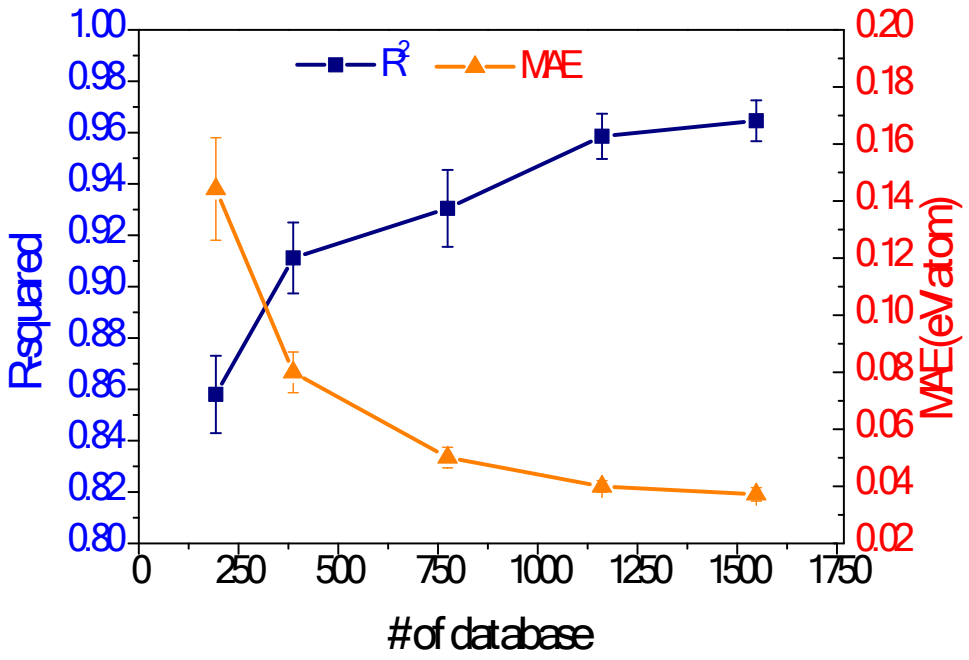


Figure S6. Prediction accuracy for formation energy for tetragonal phase of perovskite with respect to different database number.

Descriptor	Bandgap ($E_{g, C}$)		Formation energy ($E_{Form, C}$)	
	R_2	MAE (eV)	R_2	MAE (eV/atom)
[1)] D_0	0.605	0.259	0.765	0.312
D_0 + [2)] D_P	0.822	0.179	0.977	0.104
D_0 + [3)] M	0.812	0.184	0.971	0.110
D_0 + M + [4)] D_A	0.840	0.110	0.991	0.061
All (D_0 + D_P + M + D_A)	0.831	0.130	0.987	0.070

[1)] D_0: Chemical descriptors [2)] D_P: ionic radius (rA, rB, rArO, rBrO), tolerance factor (t)

[3)] M: Mendeleev number for A and B

[4)] D_A: atomic number, valence, electronegativity, group number, melting point, periodic number

Table S1. Prediction accuracy of machine learning model for band gap and formation energy depending on different combinations of descriptors.

(a)

Rank	Importance	Feature
1	100.00	MTA
2	83.28	AtNB
3	71.93	NegB
4	65.27	mean_CovalentRadius
5	53.52	mean_MeltingT
6	51.80	NegA
7	50.26	MB
8	47.15	maxdiff_GSvolume_pa
9	46.75	MTB
10	46.11	mean_GSvolume_pa
11	45.41	mean_NUnfilled
12	44.80	MA
13	44.34	mean_Electronegativity
14	42.06	ValB
15	38.41	maxdiff_MeltingT
16	38.13	frac_dValence
17	35.50	maxdiff_Electronegativity
18	34.73	maxdiff_CovalentRadius
19	33.68	mean_MendeleevNumber
20	30.01	dev_NUnfilled
21	29.98	AtNA
22	29.69	mean_SpaceGroupNumber
23	26.53	dev_GSvolume_pa
24	24.66	frac_pValence
25	24.27	dev_NValance
26	21.62	maxdiff_Number
27	20.27	MeanIonicChar
28	19.19	maxdiff_MendeleevNumber
29	19.10	dev_MeltingT
30	18.99	mean_Number
31	18.67	mean_Column
32	17.64	frac_sValence
33	16.75	CanFormIonic
34	16.10	frac_fValence
35	15.30	mean_NpUnfilled
36	14.57	max_NUnfilled
37	14.45	dev_MendeleevNumber
38	14.11	mean_NfUnfilled
39	13.36	maxdiff_SpaceGroupNumber
40	13.07	mean_NdValence
41	13.06	max_NValance
42	12.77	min_MendeleevNumber
43	12.57	mean_NdUnfilled
44	12.47	mean_AtomicWeight
45	12.30	ValA
46	12.16	maxdiff_NUnfilled
47	10.99	dev_NdValence
48	10.52	mean_NValance
49	10.40	maxdiff_NValance
50	10.30	mean_NsValence

(b)

Rank	Importance	Feature
1	100.00	mean_MeltingT
2	96.62	MTB
3	96.02	AtNB
4	87.94	mean_GSvolume_pa
5	86.65	mean_Electronegativity
6	85.55	mean_CovalentRadius
7	81.95	mean_SpaceGroupNumber
8	76.81	MTA
9	75.65	mean_MendeleevNumber
10	72.56	frac_dValence
11	69.38	NegA
12	68.32	NegB
13	67.53	MB
14	66.65	MA
15	62.77	mean_NUnfilled
16	62.77	AtNA
17	55.81	mean_Number
18	52.05	dev_NValance
19	49.76	MeanIonicChar
20	49.22	frac_sValence
21	46.98	maxdiff_MeltingT
22	46.95	dev_GSvolume_pa
23	44.60	dev_MeltingT
24	43.31	dev_NUnfilled
25	40.53	maxdiff_Number
26	40.08	frac_pValence
27	37.75	mean_AtomicWeight
28	36.63	maxdiff_GSvolume_pa
29	36.47	dev_MendeleevNumber
30	33.37	CanFormIonic
31	33.18	maxdiff_MendeleevNumber
32	31.34	maxdiff_Electronegativity
33	29.82	maxdiff_CovalentRadius
34	29.13	maxdiff_NValance
35	28.66	dev_SpaceGroupNumber
36	28.20	frac_fValence
37	27.67	maxdiff_NUnfilled
38	27.65	ValA
39	25.67	maxdiff_SpaceGroupNumber
40	25.35	ValB
41	24.88	dev_Number
42	24.10	mean_NValance
43	23.87	mean_NdUnfilled
44	23.39	dev_AtomicWeight
45	22.67	mean_NpUnfilled
46	21.86	mean_Column
47	18.61	dev_Column
48	18.48	mean_NdValence
49	15.04	mean_NfUnfilled
50	12.19	mean_GSbandgap

Table S2. List of 50 (descending order) most important features for prediction model of (a) band gap and (b) formation energy. Features other than standard chemical descriptors (145) are colored as red. Each descriptor is abbreviated as follows: melting point (MT), atomic number (AtN), electro-negativity (Neg), Mendeleev number (M), and valence (Val).

Structure	$E_{Hull, T}$ [eV/atom]	$E_{Form, T}$ [eV/atom]	$E_{g, T}$ [eV]	$E_{Hull, C}$ [eV/atom]	$E_{Form, C}$ [eV/atom]	$E_{g, C}$ [eV]	$\lvert E_{Form,T} - E_{Form,C} \rvert$ [eV/atom]	T_c [18] [K]	P_s [18] [$\mu C \cdot cm^{-2}$]
$KNbO_3$	-0.221	-2.697	1.71	0.011	-2.686	1.74	0.011	708	30.0
$BaTiO_3$	-0.030	-3.288	1.68	0.007	-3.281	1.82	0.007	408	26.0
$PbTiO_3$	-0.021	-2.517	1.69	0.024	-2.492	1.48	0.025	765	>50
$LiNbO_3$	0.071	-2.666	2.05	0.167	-2.570	1.66	0.096	1480	71
$LiTaO_3$	0.072	-2.895	2.50	0.140	-2.826	2.70	0.069	938	50

Table S3. List of widely known ferroelectric materials. The name of structures, Curie temperature (Tc), and spontaneous polarization (Ps) are obtained from previous reference [18].

Category	Conditions	Number
1	• $E_{Form, T} < E_{Form, C}$ • $\lvert E_{Form, T} - E_{Form, C} \rvert$ < 100 meV/atom • $E_{g, C} > 1.0$ eV	63
2	• $E_{Form, T} < E_{Form, C}$ • $\lvert E_{Form, T} - E_{Form, C} \rvert$ < 100 meV/atom • 0.2 eV < $E_{g, C}$ < 1.0 eV	16
3	•$E_{Form, T} > E_{Form, C}$ •$\lvert E_{Form, T} - E_{Form, C} \rvert$ < 100 meV/atom •$E_{g, C} > 0.2$ eV	65
4	• $E_{Form, T} > E_{Form, C}$ • $\lvert E_{Form, T} - E_{Form, C} \rvert$ > 100 meV/atom • $E_{g, C} > 0.2$ eV	70
5	•Not available for tetragonal phase ($RbPaO_3$, $BaHfO_3$, $TlPaO_3$, $LiPaO_3$ $CeGaO_3$, $GaPaO_3$, $NpScO_3$, $NpGaO_3$ $NdScO_3$, $SmScO_3$, $InPaO_3$, $NpAlO_3$ $NpInO_3$, $NdInO_3$, $BaPaO_3$, $NpCrO_3$ $CuTaO_3$, $CuPaO_3$)	18
	Total	**232**

Table S4. Number of structures satisfying each condition. Categories 1 and 2 are the target conditions in this study.

Table S5. List (232) of cubic phase satisfying the target conditions ($E_{g,\,C} > 0$ eV, $E_{Hull,\,C} < 0.5$ eV/atom). The name of tab in excel file is 'Cubic_232'

Table S6. List (63) of structures satisfying the target conditions ($E_{g,,C} > 1$ eV, $E_{Hull,\,C} < 0.5$ eV/atom, $E_{Form,\,T} < E_{Form,\,C}$ whose energy difference is < 100 meV/atom). The name of tab in excel file is 'T-C-Satisfied'

Table S7. List of those partially satisfying target conditions. The name of tab in excel file is 'T-C-Unsatisfied'. Condition in red means that the structure does not qualify.

Chemical Formula	**Stable Phase**
$RbPaO_3$	Orthorhombic
$BaHfO_3$	Cubic
$TlPaO_3$	Cubic
$LiPaO_3$	Rhombohedral
$CeGaO_3$	Orthorhombic
$GaPaO_3$	Orthorhombic
$NpScO_3$	Orthorhombic
$NpGaO_3$	Orthorhombic
$NdScO_3$	Orthorhombic
$SmScO_3$	Orthorhombic
$InPaO_3$	Cubic
$NpAlO_3$	Rhombohedral
$NpInO_3$	Orthorhombic
$NdInO_3$	Orthorhombic
$BaPaO_3$	Orthorhombic
$NpCrO_3$	Orthorhombic
$CuTaO_3$	Orthorhombic
$CuPaO_3$	Rhombohedral

Table S8. Materials whose properties are not available at tetragonal phase but available at their most stable phase from OQMD database.