SUPPORTING INFOMRATION

Disentangling Structural Confusion through Machine Learning: Structure Prediction and Polymorphism of Equiatomic Ternary Phases ABC

Anton O. Oliynyk,[†,‡] Lawrence A. Adutwum,[‡] Brent W. Rudyk,[‡] Harshil Pisavadia,[‡] Sogol Lotfi,[†] Viktor Hlukhyy,[§] James J. Harynuk,[‡,*] Arthur Mar,[‡,*] Jakoah Brgoch[†,*]

† Department of Chemistry, University of Houston, Houston, TX 77204 USA
‡ Department of Chemistry, University of Alberta, Edmonton, AB T6G 2G2 Canada
§ Department Chemie, Technische Universität München, Garching, 85747, Germany

Table S1. Data extraction for machine-learning

Prototype	Number of entries	Number of entries with three elements	Coordinates for element *A*	Coordinates for element *B*	Coordinates for element *C*
TiNiSi	1102	1018	0.02; 1/4; 0.68	0.19; 1/4; ~0.07	~0.28; 1/4; ~0.39
PbFCl	674	332	1/4; 1/4; ~0.62	3/4; 1/4; 0	1/4; 1/4; ~0.23
ZrNiAl	1008	922	~0.58; 0; 0	~0.24; 0; 1/2	1/3; 2/3; 1/2 0; 0; 0
YPtAs	64	64	0; 0; 1/4 0; 0; 0	1/3; 2/3; ~0.13	1/3; 2/3; ~0.62
LiGaGe	94	92	0; 0; 0	1/3; 2/3; ~0.25	1/3; 2/3; 0.70
LaSiPt	74	71	0; 0; ~0.58	0; 0; ~0.16	0; 0; 0
CeScSi (UGeTe)	145	63	0; 0; ~0.32	0; 0; ~0.12	0; 1/2; 0

Table S2. List of descriptors. 33 properties of elements combined with 30 formulae each gives 990 individual variables.

33 properties of elements, used to calculate descriptors	
1. Atomic number	17. Number of outer shell electrons
2. Atomic weight	18. Period number
3. Atomic radius	19. Group number
4. Covalent radius	20. Family number
5. Metallic radius	21. L quantum number
6. Single bond radius	22. Melting point
7. Zunger radii sum	23. Boiling point
8. Ionic radius	24. Density
9. Crystal radius	25. First ionization energy
10. Pauling electronegativity	26. Electrical conductivity
11. Martynov-Batsanov electronegativity	27. Specific heat
12. Gordy electronegativity	28. Heat of fusion
13. Mulliken electronegativity	29. Heat of vapourization
14. Allred-Rochow electronegativity	30. Thermal conductivity
15. Metallic valence	31. Heat atomization
16. Number of valence electrons	32. Polarizability
	33. Mendeleev number

30 formulae, used to calculate descriptors	
1. Average number (of 3)	16. Number average *B* and *C*
2. Number *A*	17. Sum of two largest numbers
3. Number *B*	18. Sum of two smallest numbers
4. Number *C*	19. Difference of two largest numbers
5. Number sum of *A* and *B*	20. Difference of two smallest numbers
6. Number sum of *A* and *C*	21. Ratio of two largest numbers
7. Number sum of *B* and *C*	22. Ratio of two smallest numbers
8. Number difference of *A* and *B*	23. Average of two largest numbers
9. Number difference of *A* and *C*	24. Average of two smallest numbers
10. Number difference of *B* and *C*	25. Sum of two extremes
11. Number ratio *A*/*B*	26. Difference of two extremes
12. Number ratio *A*/*C*	27. Ratio of two extremes
13. Number ratio *B*/*C*	28. Average of two extremes
14. Number average *A* and *B*	29. Smallest number
15. Number average *A* and *C*	30. Largest number

Tables S3. Machine-learning model metrics for classification of seven ternary equiatomic structure types.

	Predicted structure type							
Before feature selection	TiNiSi	ZrNiAl	PbFCl	LiGaGe	YPtAs	UGeTe	LaPtSi	**average**
Correct rate	0.967	0.973	0.992	0.998	0.999	0.997	0.980	**0.987**
Sensitivity	0.964	0.980	0.993	0.998	0.999	0.997	0.979	**0.987**
Specificity	0.971	0.958	0.990	1.000	1.000	1.000	1.000	**0.988**
Validated model								
Correct rate	0.906	0.921	0.971	0.983	0.992	0.998	0.969	**0.963**
Sensitivity	0.912	0.935	0.974	0.992	0.992	0.998	0.968	**0.967**
Specificity	0.897	0.892	0.941	0.800	1.000	1.000	1.000	**0.933**
After feature selection								
Correct rate	0.998	0.996	0.999	0.998	1.000	1.000	1.000	**0.999**
Sensitivity	0.998	0.996	0.999	0.998	1.000	1.000	1.000	**0.999**
Specificity	0.998	0.997	1.000	1.000	1.000	1.000	1.000	**0.999**
Validated model								
Correct rate	0.931	0.950	0.967	0.973	0.990	0.992	0.983	**0.969**
Sensitivity	0.939	0.957	0.968	0.978	0.990	0.996	0.982	**0.973**
Specificity	0.920	0.934	0.961	0.880	1.000	0.875	1.000	**0.939**

Tables S4. Ambiguous compounds (<0.7 probability for their native structure type, or >0.2 probability for competitor structure type) from the machine-learning model that could potentially be polymorphs.

Originally reported TiNiSi-type structure			*Originally reported ZrNiAl-type structure*		
Formula	*Probability for TiNiSi*	*Probability for ZrNiAl*	*Formula*	*Probability for TiNiSi*	*Probability for ZrNiAl*
TbSnPd	0.009683	0.981555	TiCoGe	0.996496	0.001609
ZrOsSi	0.038687	0.953313	MoNiP	0.985543	0.00622
ZrMnGe	0.081407	0.908584	WNiP	0.969773	0.002422
ZrMnSi	0.137845	0.843764	HfFeGe	0.952227	0.045835
YbCdAu	0.161259	0.826731	MnRhGe	0.796265	0.178295
ZrRhP	0.175141	0.818258	TiMnP	0.662975	0.332066
CrRuAs	0.304456	0.678101	TiCrP	0.653831	0.336883
ZrSnCu	0.26303	0.663791	CrRhAs	0.621201	0.340561
ZrPdGe	0.384383	0.614109	TaFeB	0.590548	0.390514
EuCdPd	0.394562	0.599028	USbRu	0.563186	0.424357
CeZnAu	0.328769	0.596663	HfAlPt	0.524163	0.44996
HfMnSi	0.382308	0.589644	CrCoAs	0.455718	0.531695
NbFeAs	0.454005	0.532515	SmSnNi	0.39786	0.583435
CaAgP	0.342277	0.497151	GdBiRh	0.364817	0.624215
SmAlPd	0.54782	0.44311	LaCuPd	0.339884	0.560965
FeCoP	0.563451	0.423128	ThGaPt	0.333809	0.638717
PrAlRh	0.591373	0.389065	HfMoP	0.289979	0.690126
NaCdSb	0.147401	0.347832	CuAlY	0.287975	0.522724
CaCdPd	0.667898	0.322757	LuPdAs	0.283142	0.696985
YbMgAg	0.660556	0.289252	FeCoAs	0.267756	0.705953
YbMgAg	0.660556	0.289252	MnFeAs	0.267646	0.71934
YbZnPd	0.707739	0.287249	YbPdAs	0.249745	0.726831
NaCdAs	0.136046	0.26739	MnRhAs	0.246065	0.721822
NbPtGe	0.73492	0.259294	YbAgGe	0.238043	0.756783
LuInRh	0.724665	0.246755	TbBiRh	0.235532	0.753582
ScFeSi	0.728737	0.245	NbCrGe	0.227652	0.670293
ZrCrSi	0.669289	0.244117	ThSnCo	0.224516	0.751693
VNiAs	0.701671	0.212203	ThGaIr	0.222801	0.752545
EuSnPd	0.785584	0.207798	YSiAg	0.195055	0.570171
MnCoAs	0.777291	0.204026			
TbZnRh	0.791666	0.202749			
ZrMoSi	0.669843	0.201896			

Table S5. Experiment summary for 35 polymorphs compared to machine learning predictions; * – this study.

#	Formula	Experiment		Ref.	Machine learning prediction		
					Predicted probability		Most probable structure
		TiNiSi-type	*ZrNiAl-type*		*TiNiSi-type*	*ZrNiAl-type*	*(>0.7 probability)*
1	YbMnGe	low-temperature	high-temperature	[26]	0.986	0.010	TiNiSi-type
2	YZnPd	annealed	not annealed	[37]	0.975	0.015	TiNiSi-type
3	YAlPd	low-temperature	high-temperature	[27]	0.971	0.018	TiNiSi-type
4	TmMnGe	low-temperature	high-temperature	[36]	0.967	0.025	TiNiSi-type
5	LuPbNi	low-temperature	high-temperature	[28]	0.950	0.010	TiNiSi-type
6	CrNiP	low-temperature	high-temperature	[38]	0.905	0.073	TiNiSi-type
7	GdAlPd	low-temperature	very-low-temp.	[27]	0.823	0.173	TiNiSi-type
8	HfRhGe	low-temperature	high-temperature	[30]	0.806	0.193	TiNiSi-type
9	TbAlPd	low-temperature	high-temperature	[27]	0.781	0.215	TiNiSi-type
10	ScRhGe	identical synthesis conditions		[43]	0.680	0.317	TiNiSi/ZrNiAl
11	MnNiP	identical synthesis conditions		[44]	0.672	0.302	TiNiSi/ZrNiAl
12	TiFeP	identical synthesis conditions		*	0.661	0.334	TiNiSi/ZrNiAl
13	YSnPt	identical synthesis conditions		[31]	0.592	0.399	TiNiSi/ZrNiAl
14	LaSnPd	ambient-pressure	high-pressure	[40]	0.397	0.579	TiNiSi/ZrNiAl
15	LuZnPd	annealed	not annealed	[37]	0.387	0.605	TiNiSi/ZrNiAl
16	TiPdGe	identical synthesis conditions		[45]	0.344	0.652	TiNiSi/ZrNiAl
17	ScRuSi	identical synthesis conditions		[43]	0.332	0.650	TiNiSi/ZrNiAl
18	LaSnPt	ambient-pressure	high-pressure	[41]	0.331	0.635	TiNiSi/ZrNiAl
19	ScCuSi	identical synthesis conditions		[46]	0.307	0.653	TiNiSi/ZrNiAl
20	NdSnNi	ambient-pressure	high-pressure	[40]	0.253	0.720	ZrNiAl-type
21	CeSnNi	ambient-pressure	high-pressure	[40]	0.174	0.790	ZrNiAl-type
22	SmSnPt	ambient-pressure	high-pressure	[41]	0.135	0.846	ZrNiAl-type
23	NdSnPt	ambient-pressure	high-pressure	[41]	0.132	0.843	ZrNiAl-type
24	MnNiAs	identical synthesis conditions		[44]	0.120	0.853	ZrNiAl-type
25	CeSnPt	ambient-pressure	high-pressure	[41]	0.111	0.864	ZrNiAl-type
26	TmZnPd	annealed	not annealed	[37]	0.082	0.911	ZrNiAl-type
27	CeSnPd	ambient-pressure	high-pressure	[42]	0.056	0.919	ZrNiAl-type
28	GdSnPt	not annealed	annealed	[31]	0.053	0.931	ZrNiAl-type
29	PrSnPt	ambient-pressure	high-pressure	[41]	0.044	0.928	ZrNiAl-type
30	YbSnPd	high-temperature	low-temperature	[32]	0.037	0.958	ZrNiAl-type
31	ZrRuP	sintered	flux	[33]	0.024	0.968	ZrNiAl-type
32	CeZnPd	high-temperature	low-temperature	[34]	0.017	0.952	ZrNiAl-type
33	PrZnPd	annealed	not annealed	[37]	0.013	0.956	ZrNiAl-type
34	TbSnPt	high-temperature	low-temperature	[31]	0.012	0.974	ZrNiAl-type
35	ZrOsP	sintered	flux	[35]	0.010	0.978	ZrNiAl-type

Table S6. Selected interatomic distances (Å) in (a) TiFeP (TiNiSi-type) and (b) TiFeP (ZrNiAl-type)

(a) TiFeP (TiNiSi-type)

Fe—P (×2)	2.255(1)	Ti—Fe	2.735(2)
Fe—P	2.285(2)	Ti—Fe (×2)	2.747(1)
Fe—P	2.298(2)	Ti—Fe (×2)	2.777(1)
Fe—Fe (×2)	2.654(2)	Ti—Fe	2.790(2)
Fe—Ti	2.735(2)	Ti—Ti (×2)	3.037(2)
Fe—Ti (×2)	2.747(1)	P—Fe (×2)	2.256(1)
Fe—Ti (×2)	2.777(1)	P—Fe (×2)	2.298(2)
Fe—Ti	2.790(2)	P—Ti (×2)	2.549(2)
Ti—P	2.533(2)	P—Ti (×2)	2.569(2)
Ti—P (×2)	2.549(2)		
Ti—P (×2)	2.569 (2)		

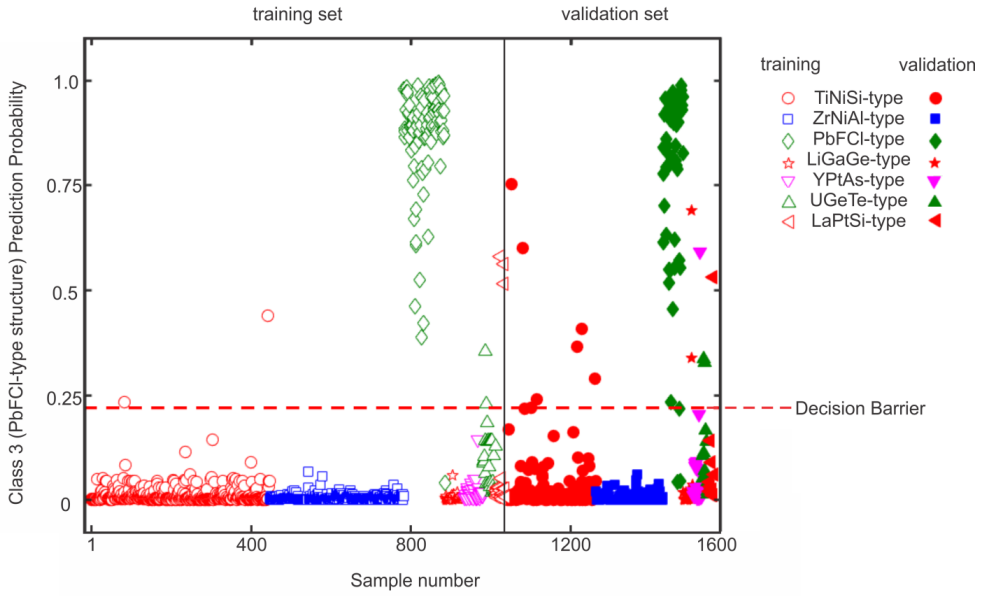
(b) TiFeP (ZrNiAl-type)

Fe—P1 (×2)	2.3140(2)	P2—Fe (×3)	2.3379(2)
Fe—P2 (×2)	2.3379(2)	P2—Ti (×6)	2.5398(1)
Fe—Fe (×2)	2.6008(6)		
Fe—Ti (×2)	2.7060(5)		
Fe—Ti (×4)	2.8212(3)		
Ti—P1	2.5307(5)		
Ti—P2 (×4)	2.5398(1)		
Ti—Fe (×2)	2.7061(5)		
Ti—Fe (×4)	2.8212(3)		
Ti—Ti (×4)	3.1706(3)		
P1—Fe (×6)	2.3140(2)		
P1—Ti (×3)	2.5307(5)		

Table S7. Binding energies for eight ternary phosphides from XPS experiment.

Sample / Composition	*P 2p3/2 BE (eV)*	*Fe 2p3/2 BE (eV)*	*M 2p3/2 BE (eV)*
1.TiFeP (both)	128.9	706.6	453.9
2.TiFeP (ZrNiAl-type)	128.8	706.5	453.8
3.TiFeP (TiNiSi-type)	128.8	706.4	453.7
4.VFeP	129.0	706.6	512.3
5.CrFeP	128.9	706.5	573.6
6.MnFeP	129.0	706.5	638.3
7.Fe_2P	129.8	707.2	
8.CoFeP	130.1	707.4	778.8
9.NiFeP	129.3	706.6	870.1

Before CR-FS



After CR-FS

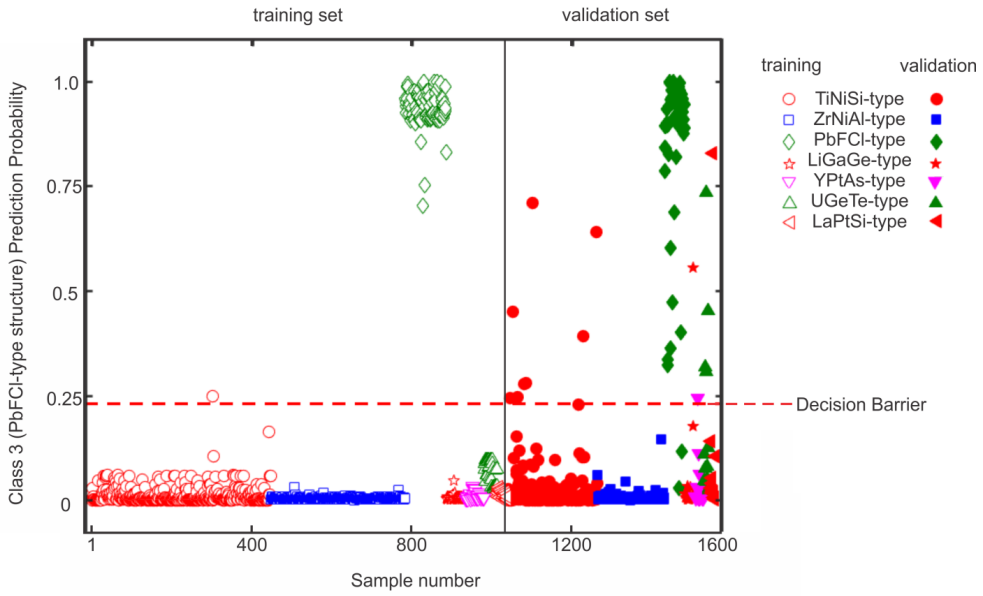
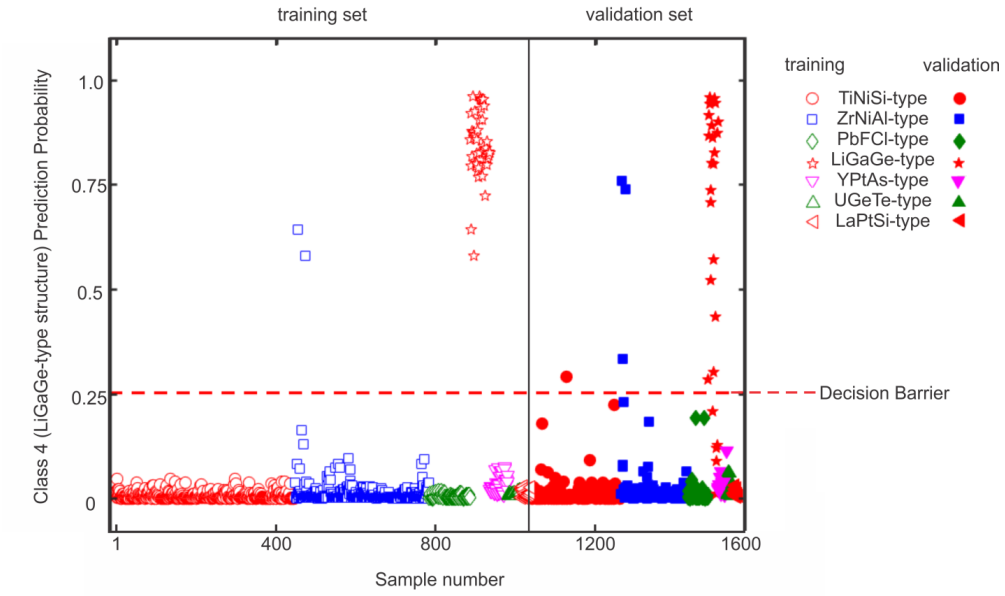


Figure S1. Model for PbFCl-type structure prediction probability.

Before CR-FS



After CR-FS

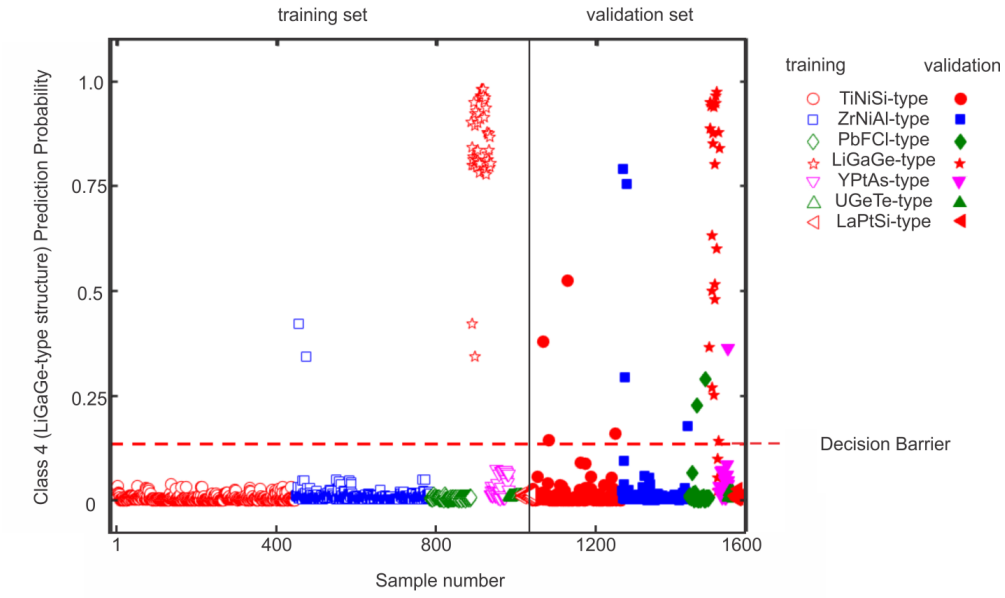
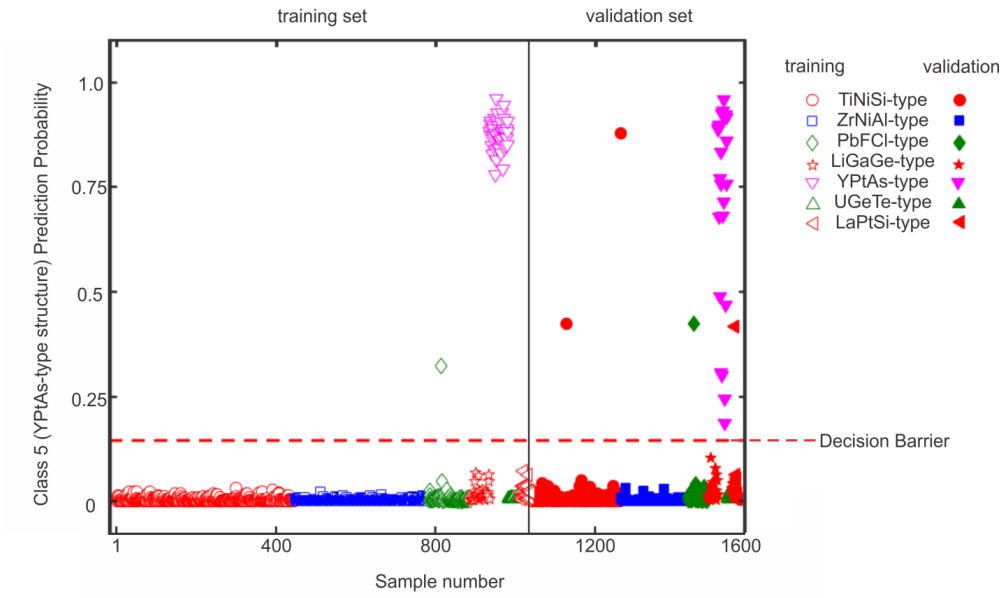


Figure S2. Model for LiGaGe-type structure prediction probability.

Before CR-FS



After CR-FS

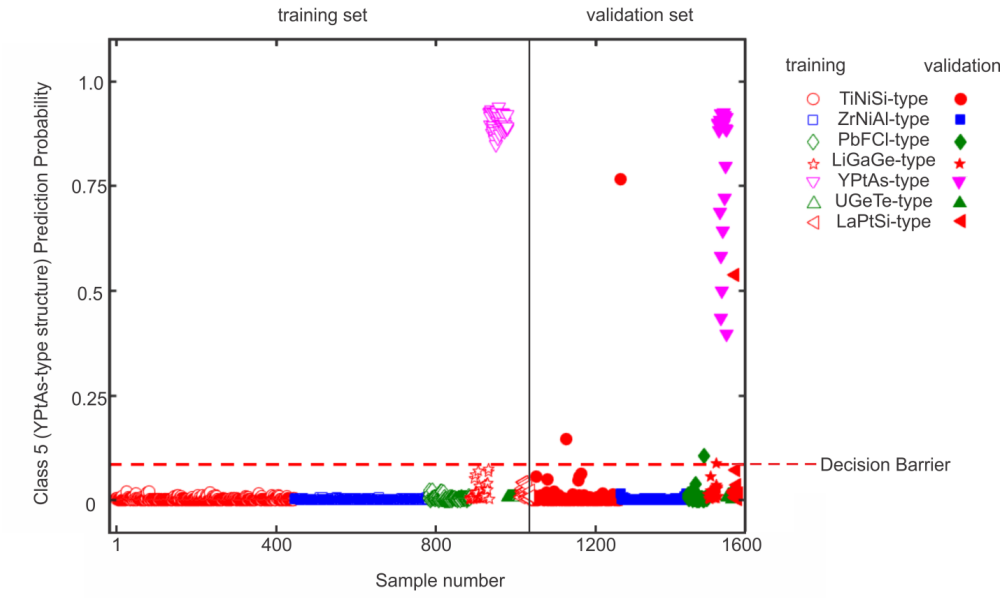
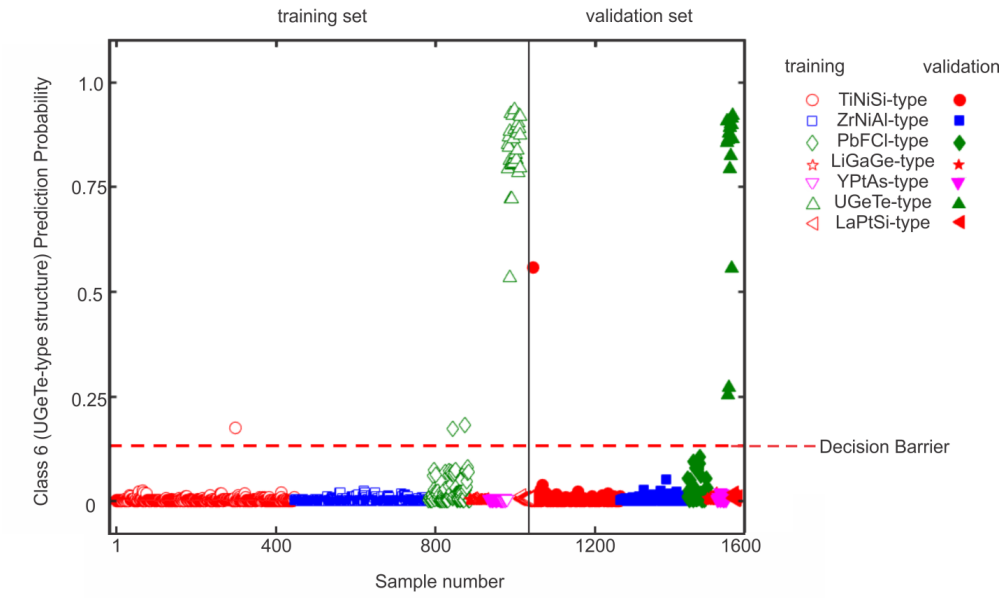


Figure S3. Model for YPtAs-type structure prediction probability.

Before CR-FS



After CR-FS

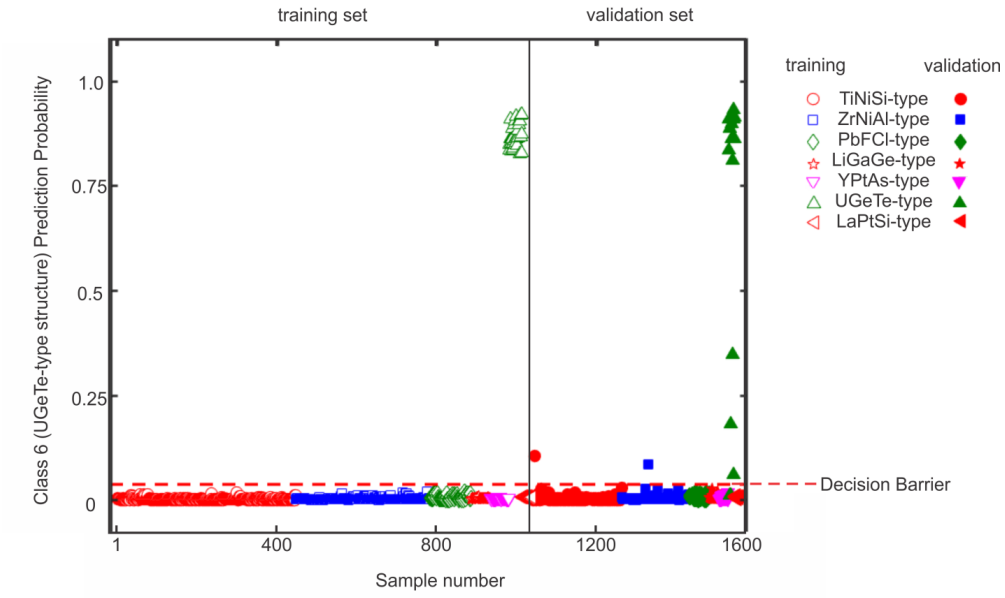
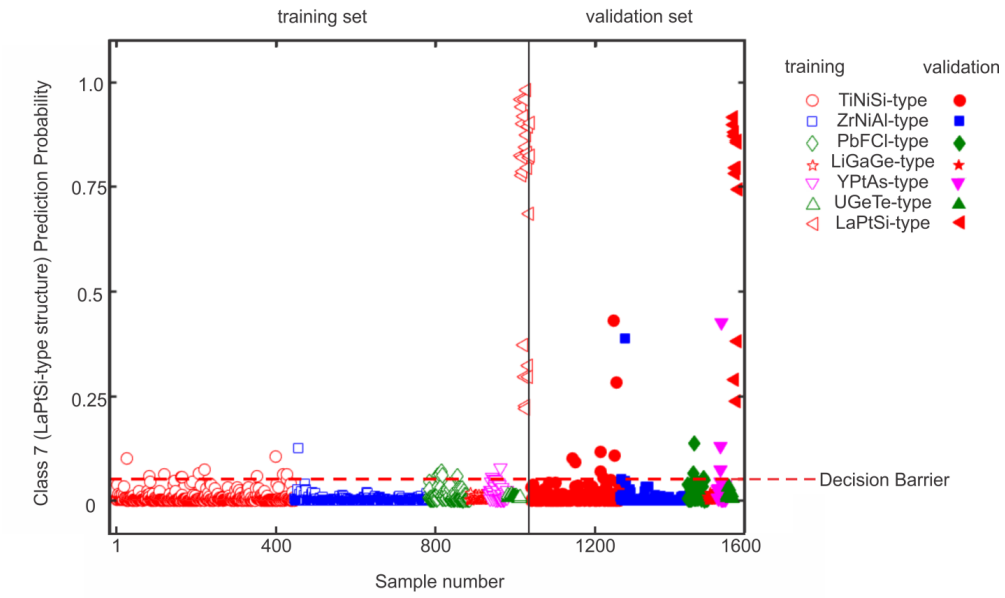


Figure S4. Model for UTeGe-type structure prediction probability.

Before CR-FS



After CR-FS

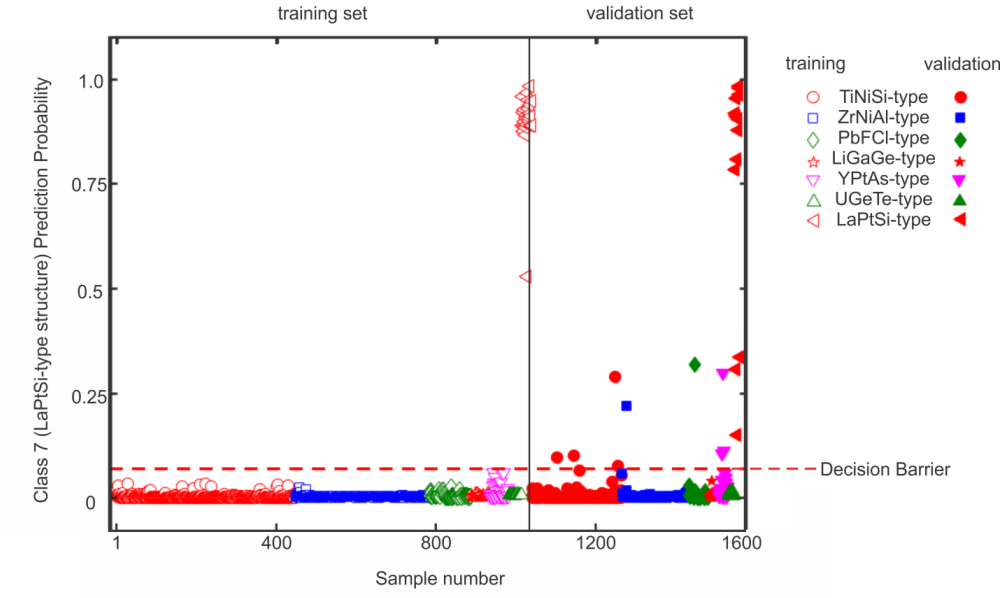


Figure S5. Model for LaPtSi-type structure prediction probability.

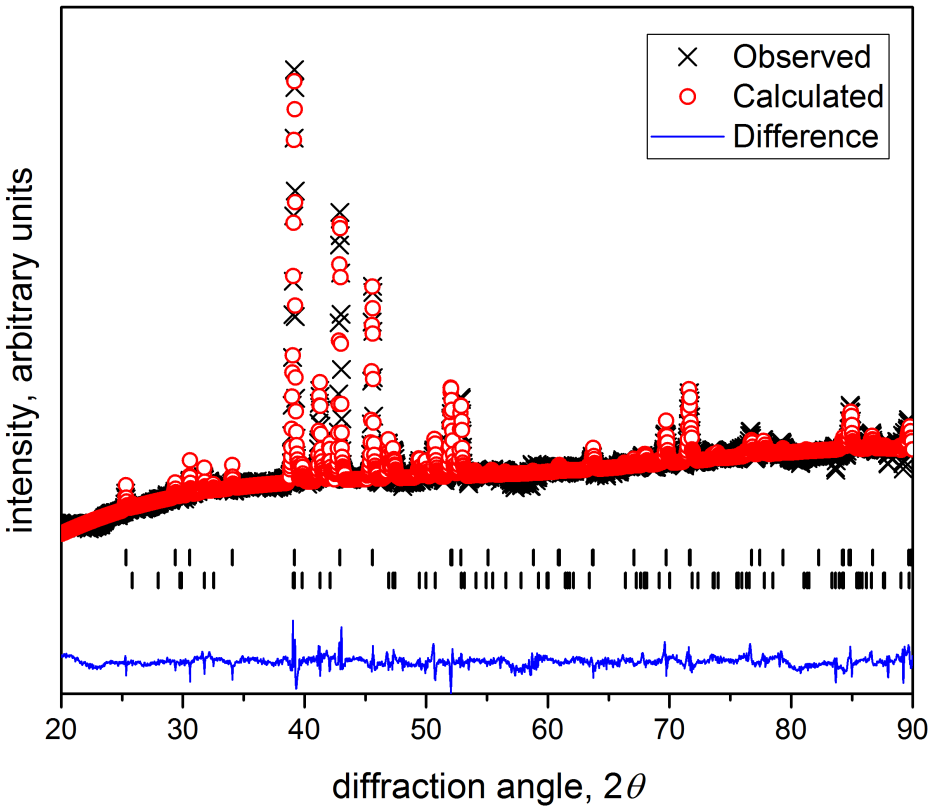


Figure S6. Powder diffraction of the sample containing TiFeP phases in TiNiSi- and ZrNiAl-type structures, present in one sample. Sample content is 47(5)% of TiFeP (TiNiSi-type) and 53(3)% of TiFeP (ZrNiAl-type). Refined cell parameters for TiFeP (TiNiSi-type): a = 5.9999(4) Å, b = 3.5947(2) Å, c = 6.8867(4) Å; for TiFeP (ZrNiAl-type): a = b = 6.0736(2) Å, c = 3.5124(1) Å

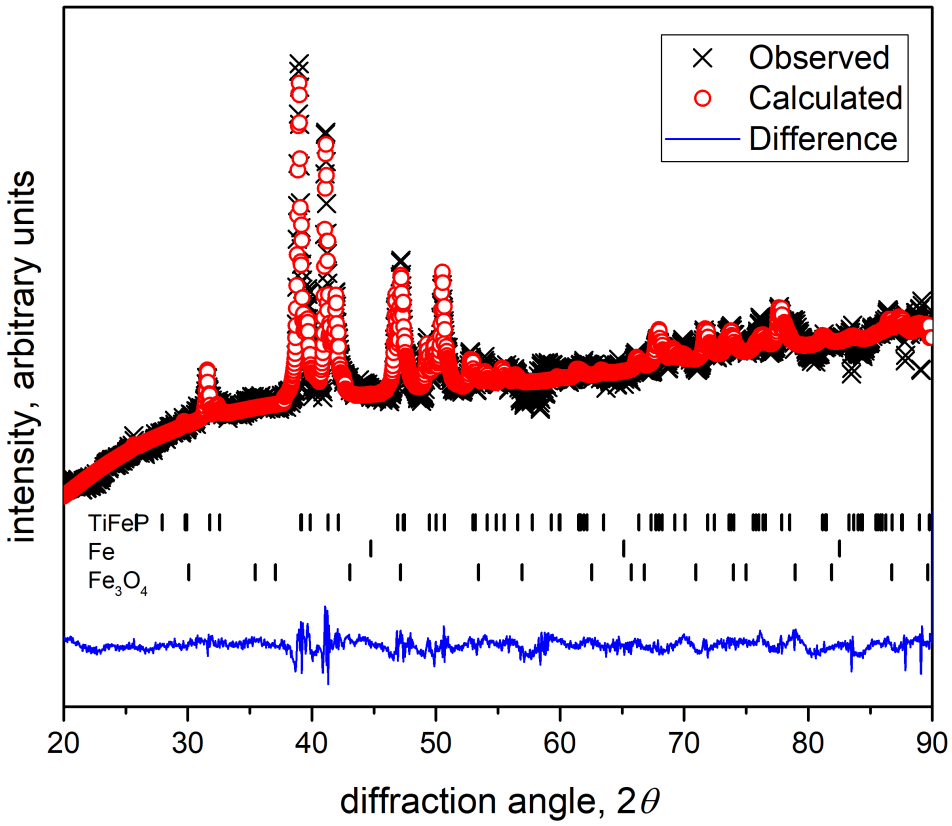


Figure S7. Powder diffraction of the sample containing TiFeP (TiNiSi-type). Rietveld refinement indicated that the sample contains 99.99(5.3)% of TiFeP (TiNiSi-type), 0.01(0.12)% Fe, and 0.00(0.00)% Fe_3O_4.

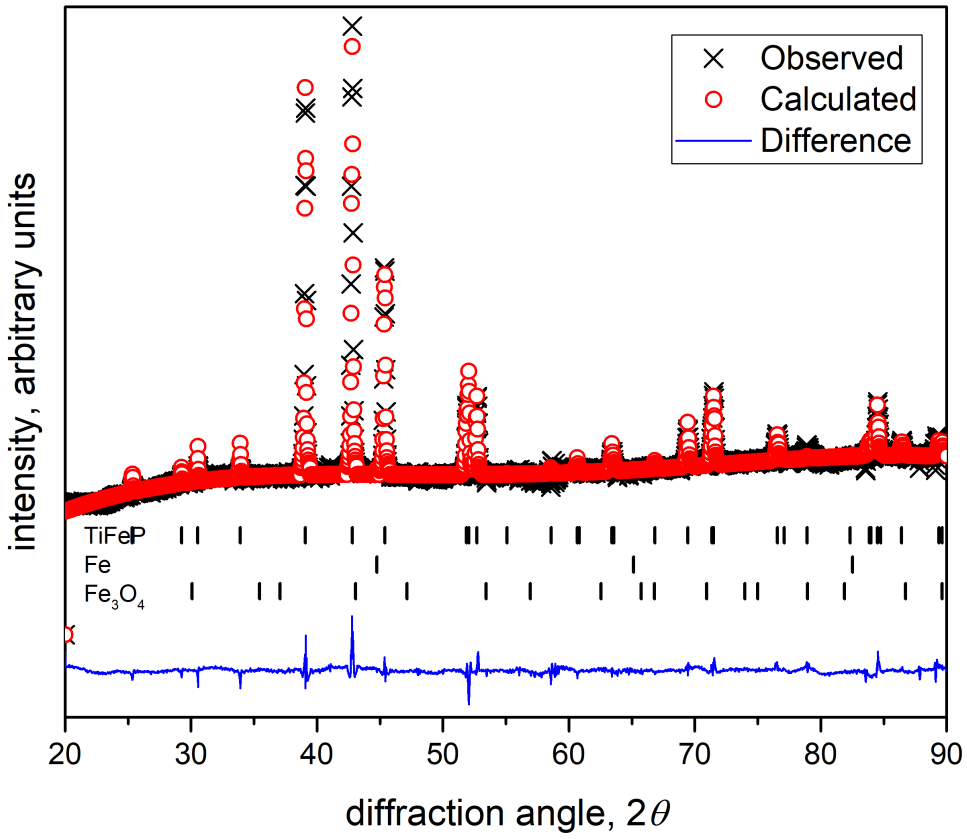


Figure S8. Powder diffraction of the sample containing TiFeP (ZrNiAl-type). Rietveld refinement indicated that the sample contains 99.78(3.6)% of TiFeP (TiNiSi-type), 0.22(0.17)% Fe, and 0.00(0.00)% Fe_3O_4.

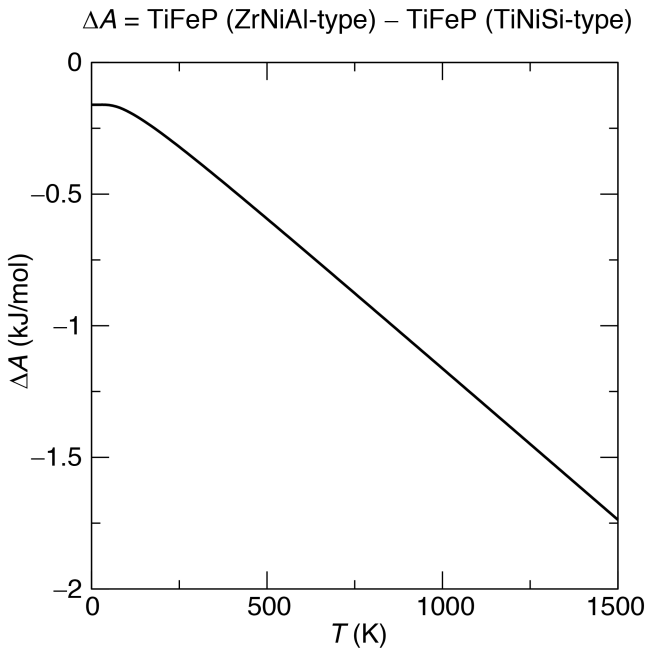


Figure S9. Helmholtz free energy comparison for TiFeP polymorphs (TiNiSi- and ZrNiAl-type structures)

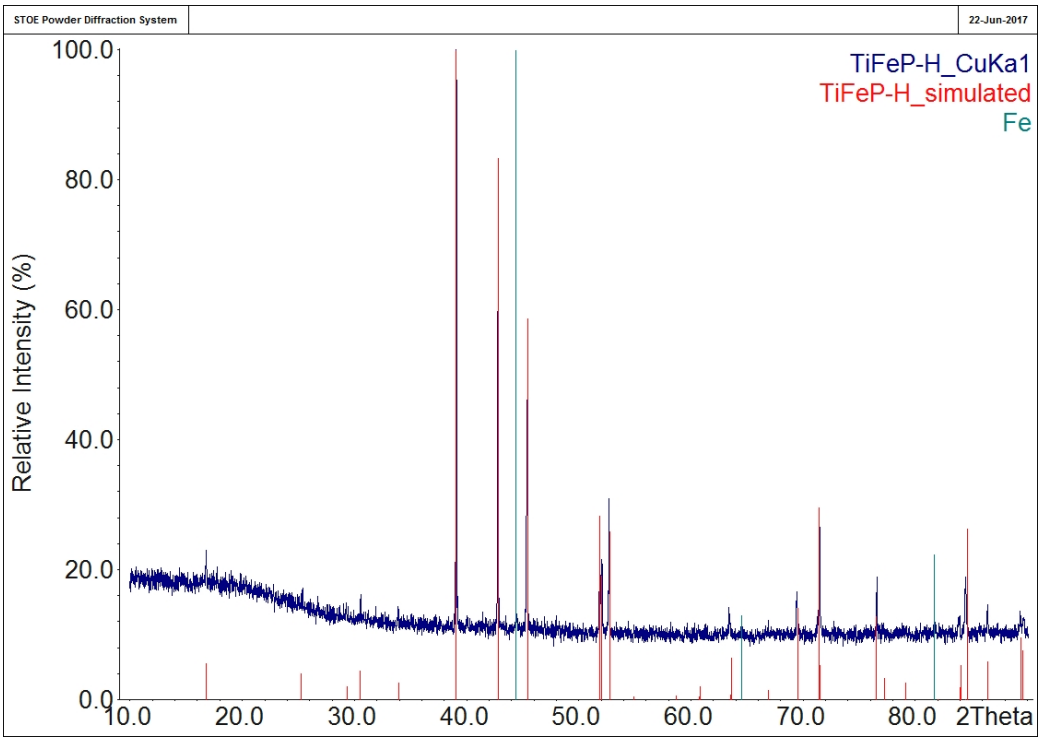


Figure S10. Powder diffraction of the TiFeP (ZrNiAl-type structure) sample powder after magnetic property measurements. The sample contains negligible amount of Fe, which causes ferromagnetic behavior at high temperature.