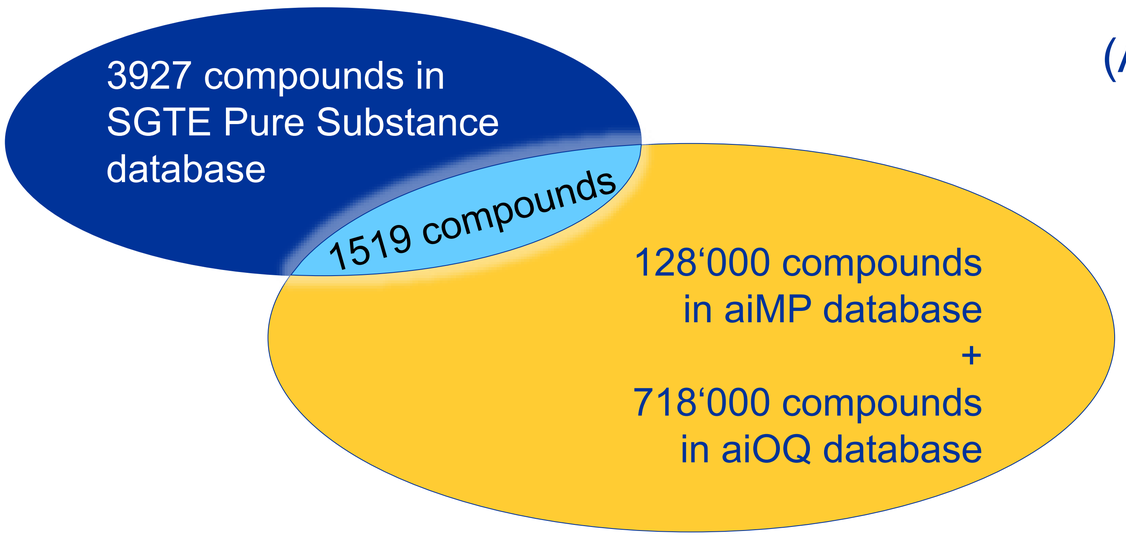
Data-centric detection of compounds containing wrong data in large Calphad databases

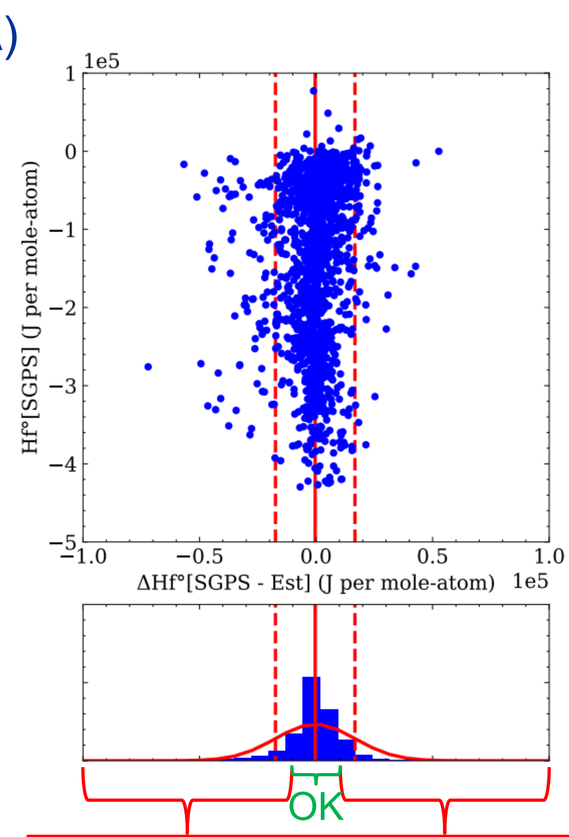
M. to Baben, C. Aras
GTT-Technologies, Herzogenrath

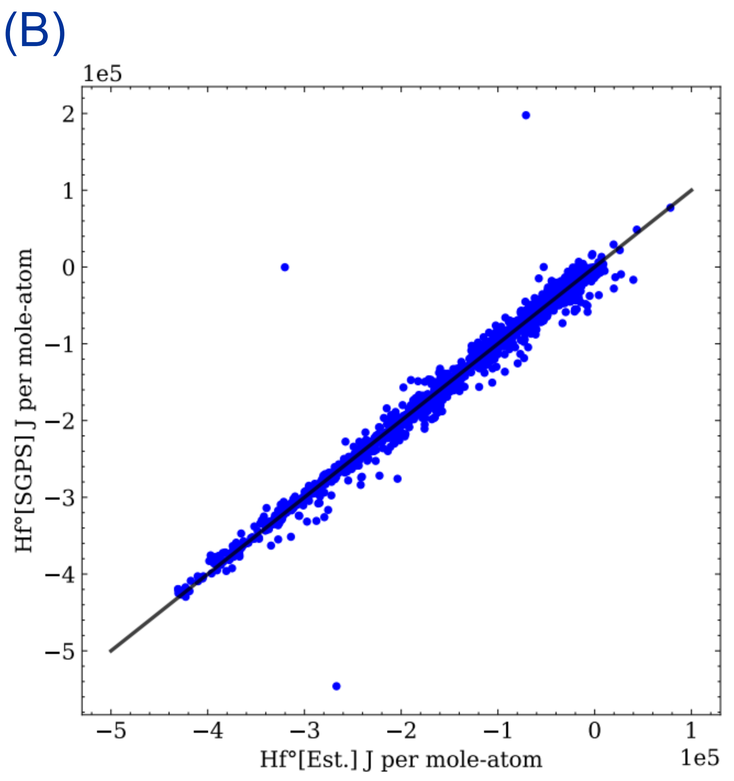
Motivation: During development of the aiMP (ab initio Materials Project) database, compounds in SGTE Pure Substance database with suspicious data had been identified.

→SGTE project to detect suspicous compounds using data analysis

→Detailed analysis confirmed that there are 29 solid phases with wrong data (1%)







(A) Comparison between the SGPS and curated DFT formation enthalpies. The dashed red line shows the standard deviation (17 kJ per mole-atom). The curves in the lower plots correspond to normal distribution computed from the mean and standard deviation.
(B) SGPS vs curated DFT results. Black line represents the x=y line.

300 suspicious compounds

Comparison to FactPS formation enthalpies confirms suspicion

Comparison to FactPS and aiMP S^{298} / C_P^{298} confirms suspicion

Composition	SGPS ΔHf° (J per mole atom)	FactPS ΔHf°	DFT ΔHf°	Possible Reason
Ba2 Si3 O8	-322	-321908	-320262	Mistake in units.
Ca1 Al2	-73220	-28000	-33281	DFT result favors FactPS.
Ca1 Al4	-43095	-19000	-20412	DFT result favors FactPS.
Mo1 Se2	-51167	-65549	-68017	DFT result favors FactPS.
U1 S2	-146440	-175728	-171628	DFT result favors FactPS. In SGPS, real stoichiometry is reported as U1 S1.9
Li1 Nb1 O3	-546096		-266900	Unrealistic value.
La1 B6	-14883		-57748	[14] reports -57.1 kJ.
Pt1 Zr1	-150624		-105921	[16]reports -104.1 kJ.
Se1 O3	-41	-42572	-52700	Unrealistic value.
In1 N1	-48272	-8577	-9393	DFT result favors FactPS. SGPS reference paper also states that the value is not very reliable.
Al1 B2	-50332	-22314	-7485	DFT result agrees neither SGPS nor FactPS. [19] reports Hf as -7.7 kJ.
B5 Mo2	-36714	-50928	4253	Mo over-stoichiometry in phase diagram [22].
K2 O2	-110750	-123951	-124341	DFT result favors FactPS. Additionally, [25] reports -124 kJ.
Fe2B1	-34169		-21060	DFT result does not agree with SGPS. Additionally, [27] reports Hf as -22.6 kJ.
Co1 B1	-47070	-47070	-30886	DFT result agrees neither SGPS nor FactPS. [27] reports Hf as -34.8 kJ.
Co2 B1	-41840	-41840	-19668	DFT result agrees neither SGPS nor FactPS. [27] reports Hf as -19.4 kJ.
Ti1 Te1	197694		-71008	DFT result does not agree with SGPS

Composition	SGPS S° (J/K per mole atom)	FactPS S°	Est. S°	Possible Reason
Tb1	32.1	73.3	69.6 ± 3.1	[30] additionally supports FactPS.
Ni9 S8	5.44	27.57	29.7 ± 3.0	Unrealistic value.
Ca1 Si2	15.1	25.8	24.5 ± 4.1	Model supports FactPS.
La1 Mg1	51.6	42.2	43.4 ± 3.5	Model supports FactPS.
Si3 N4	9.2	15.3	18.1 ± 3.6	Model supports FactPS.
Li3 N1	9.4	15.6	17.1 ± 4.1	Model supports FactPS.
H1 I1	103.4		36.7 ± 4.2	Unrealistic value.
Ti1 Te1	133.3		37.4 ± 4.0	Unrealistic value, probably gaseous species.

Composition	SGPS Cp° (J/K per mole atom)	FactPS Cp°	Est. Cp°	Possible Reason
In1 N1	233.7	20.84	21.1± 0.6	Unrealistic value.
Pt1 I4	195.2	25.1	27.1± 0.6	Unrealistic value.
Ni9 S8	63.5	24.1	23.9± 0.3	Model supports FactPS.
Sr2 Mn1 O4	-9.5		20.8± 0.3	Unrealistic value.
Gd1	13.9	26.78	26.2± 1.1	Model supports FactPS.
Sr1 Mn1 O3	-0.3		20.1± 0.2	Unrealistic value.
U1 F6	42.5	23.9	21.9± 0.7	Model supports FactPS.

- aiMP (ab initio materialsproject.org): 128‘000 compounds
- aiOQ (ab initio OQMD.org): 718‘000 compounds
  - ΔH_f^{0K} from materialsproject.org / oqmd.org [1,2]
  - consistency modifications for ΔH_f^{298K}
  - machine learning for C_p
  - machine learning for S^{298K}
  - DFT calculated enthalpy of mixing in FCC, BCC, HCP for ~1400 binary systems
- First commercial CalPhaD database relying on ab initio + ML [3]

[1]: Jain et al., APL Materials, 1 (2013) 011002.
[2]: Saal et al., JOM 65 (2013) 1501.
[3]: https://gtt-technologies.de/data/#aimp-ab-initio-materials-project

Work was financially supported by SGTE in 2022.

GTT-Technologies
Kaiserstraße 103
52134 Herzogenrath, Germany
Phone: +49-(0)2407-59533
E-mail: info@gtt-technologies.de