GTT-Technologies
Kaiserstraße 103
52134 Herzogenrath, Germany
Phone: +49-(0)2407-59533
Fax: +49-(0)2407-59661
E-mail: info@gtt-technologies.de

The *ab initio* Materials Project (aiMP) and OQMD (aiOQ) databases Version 7.0 Documentation

GTT-TECHNOLOGIES

Table of contents

Introduction 1

Database Files 2

Validity and Applicability 2

Discarded Phases 3

Stable and metastable databases 3

Entropy and Heat Capacity 4

Formation Enthalpies and Corrections 6

Solid Solutions 8

What is new? 8

References 9

Contact 10

Introduction

As of June 2025, the Materials Project (https://next-gen.materialsproject.org) [1], [2], [3], [4] contains data from 200.487 solid phases that were calculated by various groups using *ab initio* methods. Derived from this data, a total of 136.854 compounds are introduced in aiMP v7.0, with a range of compounds having multiple calculated crystallographic structures that are introduced as separate phases into the database. The Materials Project repository contains results from *ab initio* calculations at 0 K and 0 atm. The models used to estimate thermodynamic properties at temperatures above 298 K are described in the later chapters after remarks on validity and possible application areas.

Similar to aiMP, the aiOQ database contains results from ground state *ab initio* calculations calculated by Chris Wolverton's group at Northwestern University (www.oqmd.org) [5], [6]. As of May 2025, the latest Open Quantum Materials Database (OQMD) version was v1.7. The aiOQ database contains data for 475.887 compounds.

aiMP and aiOQ are developed by GTT-Technologies, using data from Materials Project and OQMD as well as own data as input. They are complemented by the aiMP solutions database containing data for metallic FCC, BCC, HCP solid solutions. Phase stability of these solid solutions has been calculated from *ab initio* calculations by GTT-Technologies. The models used are described below.

Database Files

The aiMP and aiOQ databases are split into smaller database files to make setup and evaluation of Equilib calculations easier. The following table shows the different database files, their contents and the respective use cases.

Database file and **FactSage nickname**	Contains	Number of phases	Use cases
AIMPsoln.SDC and AIMPsoln.FDB **AIMP**	Solid solutions: FCC, BCC, HCP	3	Application calculations, materials informatics, database development
AIMPbase.CDB **AIMP**	Stable phases of aiMP	102.171	Application calculations, materials informatics, database development
AIMMbase.CDB **AIMM**	Metastable phases of aiMP	34.683	Database development
AIOQbase.CDB **AIOQ**	Stable phases of aiOQ	341.526	Application calculations, materials informatics, database development
AIOMbase.CDB **AIOM**	Metastable phases of aiOQ	134.361	Database development

Validity and Applicability

Unlike all other databases available in FactSage, aiMP and aiOQ contain mostly *non-curated* data. Therefore, *ab initio* databases cannot be expected to lead to as accurate results as in is the case when using other FactSage databases.

Using data analytics, all formation enthalpies, entropies as well as heat capacities have been checked to be generally reasonable and acceptable given the inaccuracies of the first principles methods that were used. As mentioned below, most formation enthalpies have been corrected based on data in existing FactSage databases.

There are three major applications for these databases:

- Using as a starting point for a CALPHAD assessment.
- Combining standard FactSage databases with aiMP and aiOQ to estimate thermochemical properties in parts of chemical compound space where otherwise no data is available to describe the behavior of minor elements.
- Materials informatics screening of chemical space, especially in connection with ChemApp for Python.

Discarded Phases

There are 76 elements included in both Materials Project and OQMD compounds. These elements are Ag, Al, As, Au, B, Ba, Be, Bi, Br, C, Ca, Cd, Ce, Cl, Co, Cr, Cs, Cu, Dy, Er, Eu, F, Fe, Ga, Gd, Ge, H, Hf, Hg, Ho, I, In, K, La, Li, Lu, Mg, Mn, Mo, N, Na, Nb, Nd, Ni, O, P, Pb, Pd, Pr, Pt, Pu, Rb, Re, Rh, Ru, S, Sb, Sc, Se, Si, Sm, Sn, Sr, Ta, Tb, Te, Th, Ti, Tm, U, V, W, Y, Yb, Zn, Zr. The exclusion of additional elements arises primarily from insufficient CALPHAD data to benchmark or the inability to train the machine learning models.

Additionally,

- Phases which have larger than 96 number of sites in their input cell are discarded.
- Phases which have amorphous tag are discarded.
- Phases with O and P from the Materials Project, which had been recalculated with R2SCAN functional, have been replaced with the previous GGA calculations due to worse agreement with experimental data for the compounds when using the R2SCAN functional
- Phases which have volume larger than 120 $Å^3$ and smaller than 4 $Å^3$ per atom are discarded.
- If more than 50 phases exist for a compound, the first 50 with lowest formation enthalpy are included.
- If more than 7 elements exist in a compound, it is discarded.

Stable and metastable databases

As stated above, aiMP and aiOQ databases were split into two different databases: AIMP/AIOQ contain the stable phases and should be used for application calculations; AIMM/AIOM contain the metastable phases and should be used for thermodynamic database development or if a known metastable polymorph is of interest. The criteria for a phase to be in the stable database are:

- The phase with the most negative enthalpy of formation at 298 K
- All phases with lowest Gibbs energy between 300K and 5000 K for each unique composition
- All phases that have a “exp” tag in materialsproject.org or oqmd.org, i.e. that are considered experimentally confirmed

Entropy and Heat Capacity

Both the entropy and heat capacity machine learning models were trained with pure compounds which exist in any FactSage databases and have an entry in the Materials Project database. Metastable phases are avoided, i.e., only stable phases under standard conditions are used for the training. For the entropy model, the respective entropies at 12 different finite temperatures were trained between 298 K and 5000 K for all phases. 3190 compounds were used in training and for each temperature, a separate Gaussian process was trained and the resulting data for different temperatures were fitted to an entropy vs temperature curve. To prevent unrealistic results, the models for the entropies at 4000 K and 5000 K were specifically trained using the liquid phase. After training, the models were tested with 480 compounds not used in the training data set and the calculated root mean square difference is 3.09 J/K per mole atom. Note that the uncertainty in experimentally determined entropy at 298 K is approximately 1.1 J/K per mole atom as reported in Kubaschewski [7].

The heat capacity is estimated at 10 different finite temperatures between 298 K and 5000 K for all phases. Similarly to the entropy model, at each temperature a different Gaussian process is trained. The same training dataset used for the entropy model was applied here, and the liquid entropies at 4000 K and 5000 K were considered when fitting the heat capacity function. The entropy at 298 K and the heat capacity function:

$$C_p(t) = a + bT + cT^2 + dT^{-2}$$

are fitted to reproduce the predictions of the different entropy and heat capacity models using the Levenberg-Maquardt algorithm. With the modified formation enthalpy, entropy at 298 K and the heat capacity function, the Gibbs energy can be calculated at arbitrary temperatures. Three examples are presented here to compare the predicted entropy and heat capacity from the machine learning models in aiMP against traditionally modelled entropy and heat capacity that exist in other FactSage databases. Note that the selected examples were not part of the training data set.

Figure 1 summarizes how the entropy and heat capacity of MAX phase Ti_2AlC is obtained. (A) compares the aiMP entropy to traditional modelling from SpMCBN database showing a reasonable agreement between the two datasets [8]. While (B) shows the aiMP heat capactiy (black dashed line) compared to the traditionally modelled heat capacity value from the SpMCBN database. The diagrams also show reasonable agreement between the aiMP model and the traditionally modelled heat capacity values.

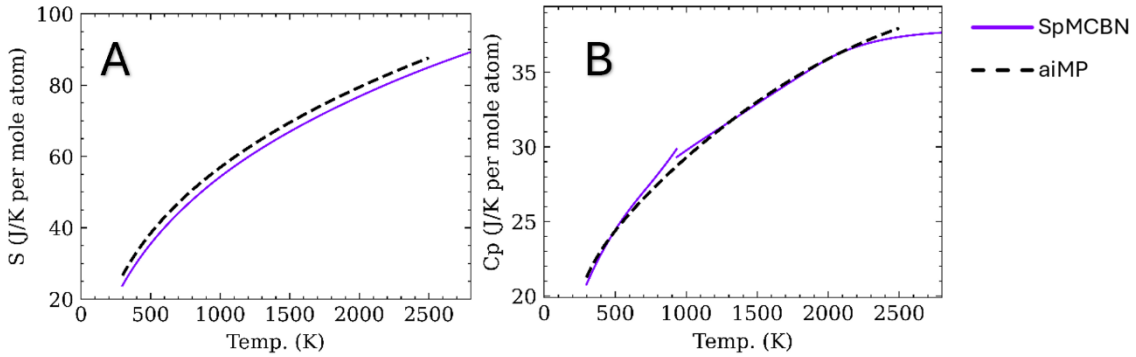

Figure 1: Heat Capacity and Entropy of Ti_2AlC. (A) shows aiMP entropy curve (dashed black line) compared to the entropy curves from SpMCBN (purple line) [8]. (B) shows the heat capacity curve in aiMP (black dashed line) compared to the traditionally modelled heat capacity curves from SpMCBN (puple line) [8].

$NaAlSi_3O_8$, shown in Figure 2 (A), illustrates a comparison between the aiMP entropy (dashed black line) to the SGPS and GTOx databases [8] (blue and red lines respectively). The aiMP curve agrees near-perfectly with GTOx while showing a larger deviation from SGPS. Figure 2 (B) shows the comparison of the heat capacity from aiMP (black dashed line) with the traditionally obtained heat capacity values from GTOx and SGPS databases [8]. The aiMP data shows an overall deviation between 500-800 K but still reasonable agreement overall.

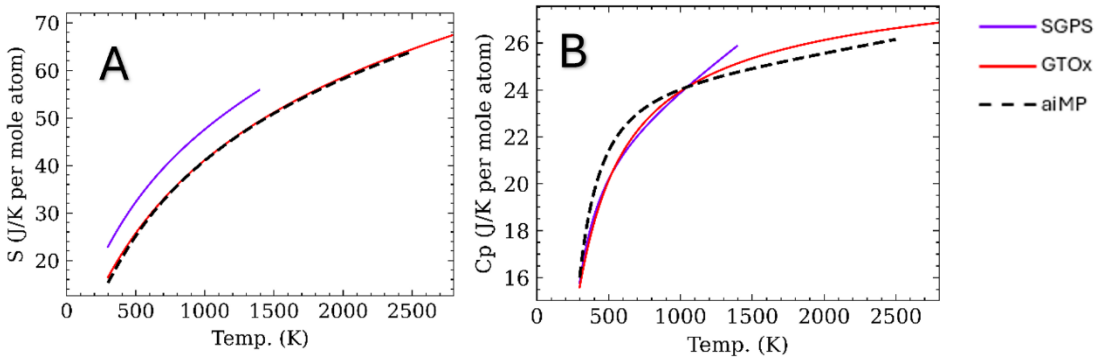

Figure 2: Heat Capacity and Entropy of $NaAlSi_3O_8$. (A) shows aiMP entropy curve (dashed black line) compared to the entropy curves from SGPS and GTOx [8] in blue and red respectively. (B) shows the heat capacity curve in aiMP (black dashed line) compared to the traditionally modelled heat capacity curves from SGPS and GTOx [8] in blue and red respectively.

The final example shows the comparison of S and Cp for Mg_2Si (Figure 3). (A) compares the aiMP entropy (dashed black line) to results from two databases, SGPS and SGTE. The aiMP data exhibits good agreement with the other two sources. (B) shows the comparison of aiMP (dashed black line) with the traditionally modelled heat capacity values from the SGTE and SGPS databases. The low temperature aiMP heat capacity (up to the melting point of Mg_2Si) curve shows excellent agreement with the SGPS data. In the range above the melting point SGTE assumes a constant Cp value while the aiMP approach uses a further increase of Cp. In the SGTE (solution) database an independent Cp curve has been stored which shows a large negative deviation from both the SGPS and the aiMP data. It is worth noting that the deviation between SGTE and aiMP is the largest as compared to the other two cases discussed above. The disagreement between the SGTE and the SGPS Cp curve occurs, since the SGTE description is based on more recent experimental evidence.

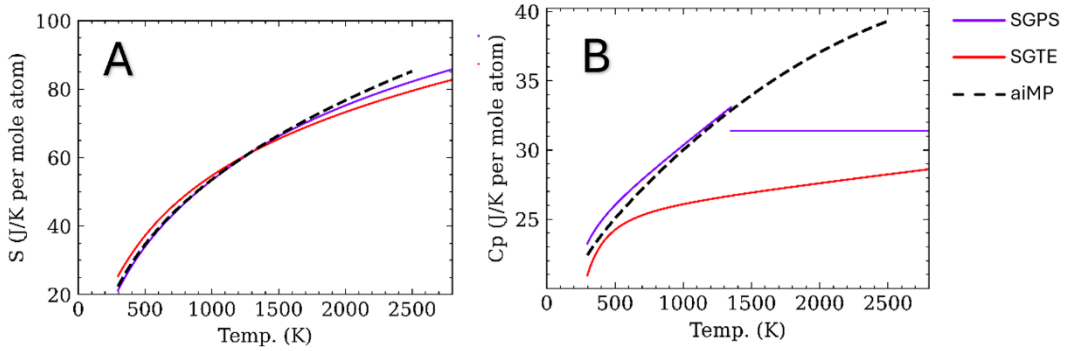

Figure 3: Heat Capacity and Entropy of Mg_2Si. (A) shows aiMP entropy curve (dashed black line) compared to the entropy curves from SGPS and SGTE in blue and red respectively. (B) shows the heat capacity curve in aiMP (black dashed line) compared to the traditionally modelled heat capacity curves from SGPS and SGTE in blue and red respectively [8].

Formation Enthalpies and Corrections

Approximate DFT functionals result in systematic errors, and a correction of DFT formation enthalpies is required. Even though Materials Project and OQMD are already applying such corrections, systematic errors are still observed. Thus, we apply additional corrections to the DFT values. Similar to the enthalpy modification schemes used by Materials Project [9] and OQMD [10], one linear enthalpy modification term is applied per element. **Fehler! Verweisquelle konnte nicht gefunden werden.** shows a comparison between the Materials Project or OQMD calculated enthalpies and the corrected enthalpies of aiOQ and aiMP. The formation enthalpies for all compounds at room temperature are assumed to be the same as in the Materials Project or OQMD

at 0 K if all constituting elements' ground states are the same at 0 K and 298 K and when the crystals do not contain functional groups.

If a constituting element's ground state changes between 0 K and 298 K or when the crystals contain functional groups, we follow a similar approach to Materials Project correction scheme [9]. But instead of a few hundred compounds, our training set contains 3547 compounds. In these cases, the formation enthalpies are corrected by element-specific corrections. We verify that the mean absolute error between calculated and experimental formation enthalpies is around 15 kJ per mole atom for both MP and OQMD datasets. We reduce this number to 10 kJ per mole atom by increasing the corrected number of ions in the model. We train two different models for Materials Project and OQMD datasets.

It should be noted that 10 kJ per mole atom is still significantly higher than the desired accuracies. Shifting the formation enthalpy of a compound by only 1 kcal per mole atom while keeping the liquid free energy curve the same might even lead a 500 K shift of congruent melting point. Thus, even though the thermodynamical data provided by the databases are reasonable, it is not expected to acquire accurate phase diagrams as is the case in other FactSage databases.

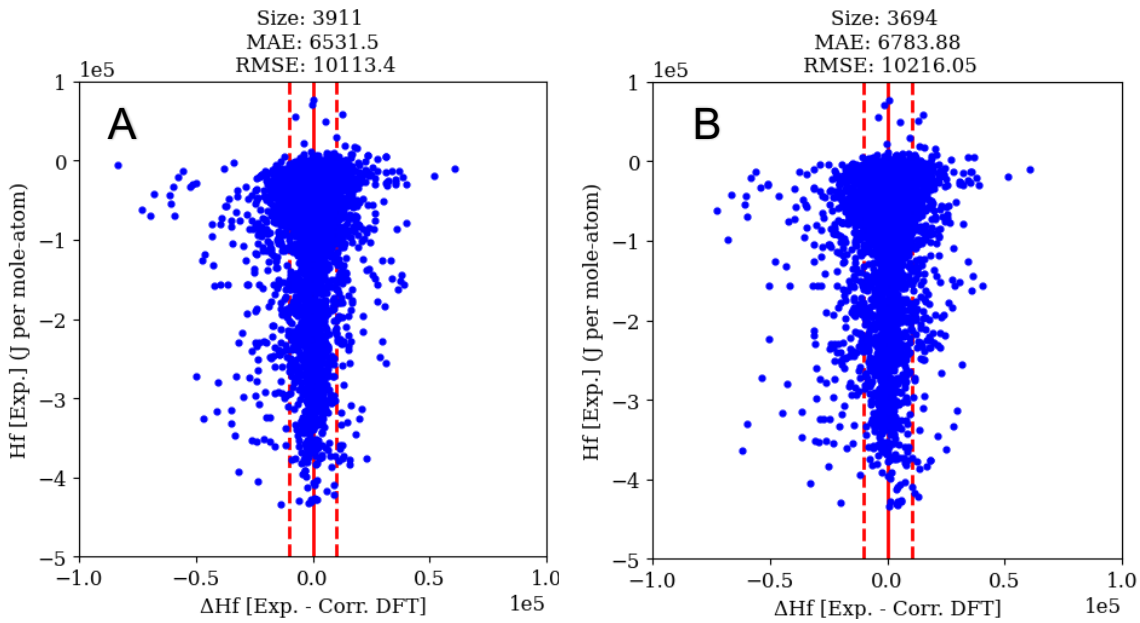

Figure 4: Comparison between Materials Project calculated enthalpy of formation and the corrected enthalpy of formation used in aiMP (A). The solid red line represents the average error and the dashed red lines show the standard deviation. There exists a systemic error even though enthalpies provided by Materials Project already have a correction scheme that our correction scheme tries to fix. (B) shows the corrected enthalpies of formation for aiOQ vs the enthalpies of formation calculated with OQMD.

Solid Solutions

Enthalpies of mixing at 0 K have been systematically calculated by GTT-Technologies for the FCC_A1, BCC_A2 and HCP_A3 solutions. Based on these, 1900 binary element interaction parameters have been derived.

In the phase FCC_A1, 1146 interaction parameters have been derived for binary systems combining any metal with atomic number between 3 (Li) and 83 (Bi) with one of the following elements: Al, Ca, Ni, Cu, Sr, Rh, Pd, Ag, Ir, Pt, Au, Pb

In the phase BCC_A2, 608 interaction parameters have been derived for binary systems combining any metal with atomic number between 3 (Li) and 83 (Bi) with one of the following elements: Li, Na, K, V, Fe, Nb, Mo, Ta, W

In the phase HCP_A3, 146 interaction parameters have been derived for binary systems combining any metal with atomic number between 3 (Li) and 83 (Bi) with one of the following elements: Mg, Ti, Zr.

What is new?

Materials Project and OQMD databases are constantly being updated. Naturally, these changes are also applied to the aiMP/aiOQ v7.0. The databases have been split into stable respectively metastable databases to make the selection of compounds easier and the calculations more efficient. Additionally, 212 compounds in aiMP have been updated based on assessed phase stabilities, these compounds have the ending “opt” and can be used in conjunction with the SLIQ-phase of the GTOx database to calculate reasonably accurate phase diagrams. A list of the systems, which include phases with optimized phase stabilities is presented in the following table.

Systems which include compounds with optimized phase stabilities				
Li-Al	Al-V	Ni-Ti	Zn-Ti	Pb-Li
Li-Si	Na-Cr	Ni-Fe	Zn-V	Pb-Al
Na-Si	Cr-P	Ni-Cr	Zn-Ni	Pb-Si
Mg-Si	Al-Mn	Ni-V	Sr-Li	Pb-Ca
Mg-P	Na-Fe-P-O	Ca-Cu	Sr-Mg	Pb-Cr
P-Si	K-Fe-P-O	Cu-Ti	Sr-Al	Pb-Ni
K-Na	Ca-Co-Si-O	Li-Zn	Sr-Si	Pb-Zn
Ca-Li	Al-Ni	Zn-Na	Sr-Co-O	Pb-Sr
Ca-Al	Ni-Si	Zn-P	Sr-Cu	
Ca-Si	Ni-P	Zn-K	Sr-Zn	
Ca-P	Ni-Ca	Zn-Ca	Sr-Zn-P-O	

References

[1] A. Wang *et al.*, “A Framework for Quantifying Uncertainty in DFT Energy Corrections,” May 2021, doi: 10.26434/CHEMRXIV.14593476.V1.

[2] M. Aykol, S. S. Dwaraknath, W. Sun, and K. A. Persson, “Thermodynamic limit for synthesis of metastable inorganic materials,” *Sci Adv*, vol. 4, no. 4, Apr. 2018, doi: 10.1126/SCIADV.AAQ0148.

[3] A. Jain *et al.*, “Formation enthalpies by mixing GGA and GGA + U calculations,” *Phys Rev B Condens Matter Mater Phys*, vol. 84, no. 4, Jul. 2011, doi: 10.1103/PHYSREVB.84.045115.

[4] A. Jain *et al.*, “Commentary: The materials project: A materials genome approach to accelerating materials innovation,” *APL Mater*, vol. 1, no. 1, 2013, doi: 10.1063/1.4812323.

[5] S. Kirklin *et al.*, “The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies,” *npj Computational Materials 2015 1:1*, vol. 1, no. 1, pp. 1–15, Dec. 2015, doi: 10.1038/npjcompumats.2015.10.

[6] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, “Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD),” *JOM*, vol. 65, no. 11, pp. 1501–1509, Nov. 2013, doi: 10.1007/S11837-013-0755-4.

[7] O. Kubaschewski, C. B. Alcock, and P. J. Spencer, *Materials Thermochemistry*, 6th ed. New York: Pergamon Press, 1993.

[8] C. W. Bale *et al.*, “FactSage thermochemical software and databases, 2010–2016,” *Calphad*, vol. 54, pp. 35–53, Sep. 2016, doi: 10.1016/J.CALPHAD.2016.05.002.

[9] A. Wang *et al.*, “A framework for quantifying uncertainty in DFT energy corrections,” *Scientific Reports 2021 11:1*, vol. 11, no. 1, pp. 1–10, Jul. 2021, doi: 10.1038/s41598-021-94550-5.

[10] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, “Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD),” *JOM*, vol. 65, no. 11, pp. 1501–1509, Nov. 2013, doi: 10.1007/S11837-013-0755-4/METRICS.

Contact

In case of questions or to provide feedback, please open a new ticket in our support center: https://support.gtt-technologies.de/ or contact us via

GTT-Technologies
Kaiserstraße 103
52134 Herzogenrath, Germany
Phone: +49-(0)2407-59533
Fax: +49-(0)2407-59661
E-mail: info@gtt-technologies.de