Machine Learning on Experimental Crystal Structures to parametrize Models of the Gibbs Energy in Computational Crystallography

Liudmila N Kuleshova

FlexCryst, Uttenreuth, Germany

dwmhofmann@aol.com

A severe problem in crystal structure prediction is the calculation of the absolute Gibbs lattice energy. DFT fails for the computational afford, especially if the crystal contains several heavy atoms, relativistic effects are important, or a multi-reference method is required. So called "universal force fields" can be calibrated to the Gibbs energy, however parameters are only for common organic substances available. We present a force field, (the data mining force field) which overcomes both problems. It has been applied to all crystal structures (1,000,000!) contained in the Cambridge Structure Database and the calculation requires a few hours. The parametrization has been achieved by a combination of two methods of machine learning. A combination of multidimensional regression and data mining [1] allows do integrate in parametrization scheme experimental structures and experimental sublimation energies as well. The obtained force field was validated regarding the three equations: firstly, the experimental lattice energies of the reference structures have been compared with the calculated energies. The correlation coefficient between experimental and calculated lattice energies is 99.5%. Secondly, the Gibb's lattice energies were calculated for all structures available in CSD. Their energies were found below zero in 99.4 %. Finally, 500 random structures and the change in density and energy was checked. The mean errors for density was found below 5%, for the energy below 2%. The very high speed, around 5 s per minimization, makes the model attractive for more complex tasks: for crystal structure structure prediction. Crystal structure structure prediction requires several hundred minimization and a proper similarity index between crystal structures [2]. Even more complex is an in silico co-crystal screening [3]. It requires hundreds of crystal structure predictions in reasonable time. On the high-performance computing cluster of CRS4 this can be done with a few days.
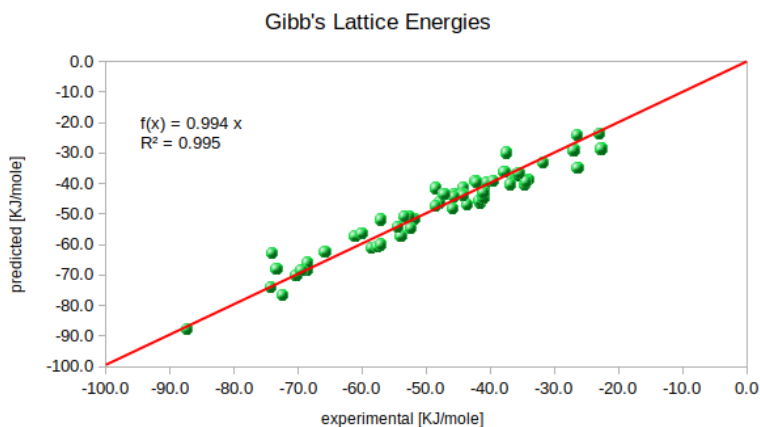

Figure 1. The correlation between experimental and calculated Gibbs lattice energies.

[1] Hofmann, Detlef WM. "Data mining in organic crystallography." Data Mining in Crystallography. Springer, Berlin, Heidelberg, 2009. 89-134.

[2] Hofmann, Detlef Walter Maria, and Ludmila Kuleshova. "New similarity index for crystal structure determination from X-ray powder diagrams." Journal of applied crystallography 38.6 (2005): 861-866.

[3] Stepanovs, Dmitrijs, et al. "Cocrystals of pentoxifylline: In silico and experimental screening." Crystal Growth & Design 15.8 (2015): 3652-3660.Keywords: Crystal Structure Prediction, Co-crystal screening, Gibbs Lattice Energy, Data Mining, Machine Learning