Supplementary Information

Thermodynamic assessment of machine learning models for solid-state synthesis prediction

Jane Schlesinger[1], Simon Hjaltason[1], Nathan J. Szymanski[1], Christopher J. Bartel[1*]

[1] University of Minnesota, Chemical Engineering and Materials Science, Minneapolis, MN 55455

[*] correspondence to cbartel@umn.edu

Table S1. Target formula, reaction string, temperature (K), E_{hull} (eV/atom), ΔG_{rxn} (eV/atom), C_1 (eV/atom), C_2 (eV/atom), Γ_{obs} (for the listed reaction and temperature, eV/atom), Γ_{opt} (for the listed target and temperature, eV/atom) for each text-mined reaction considered here. The full dataset is provided in the accompanying file table_s1.csv.

Table S2. Target formula, E_{hull} (eV/atom), Γ_{opt} (eV/atom), and synthesizability predictions for each of the ML predictors for each generated material considered here. E_{hull} and Γ_{opt} were computed using CHGNet-calculated energies for the generated material and DFT-calculated materials from the Materials Project for the other materials in the chemical space. The full dataset is provided in the accompanying file table_s2.csv.

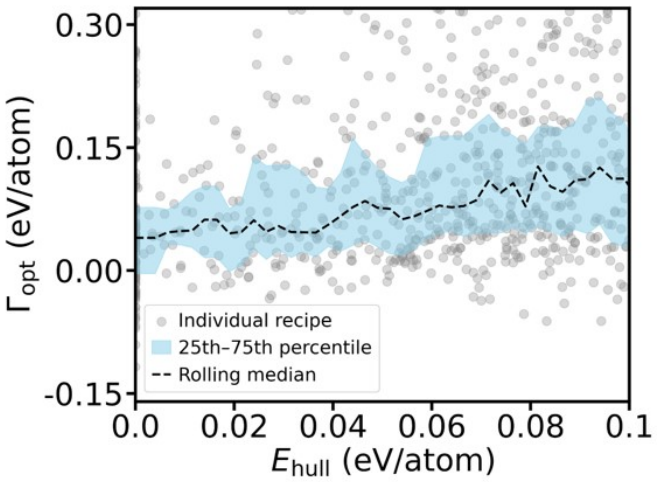

Figure S1. Γ_{opt} vs. target E_{hull} for each Chemeleon-generated materials (grey points) and a rolling median of Γ_{opt} as a function of target E_{hull}, with the median evaluated over a range of 8 meV/atom every 2.5 meV/atom. The distance from the median to the 25th and 75th percentile values of Γ_{opt} is shaded in light blue. Material energetics were computed using CHGNet,[11] discussed further in **Methods**.

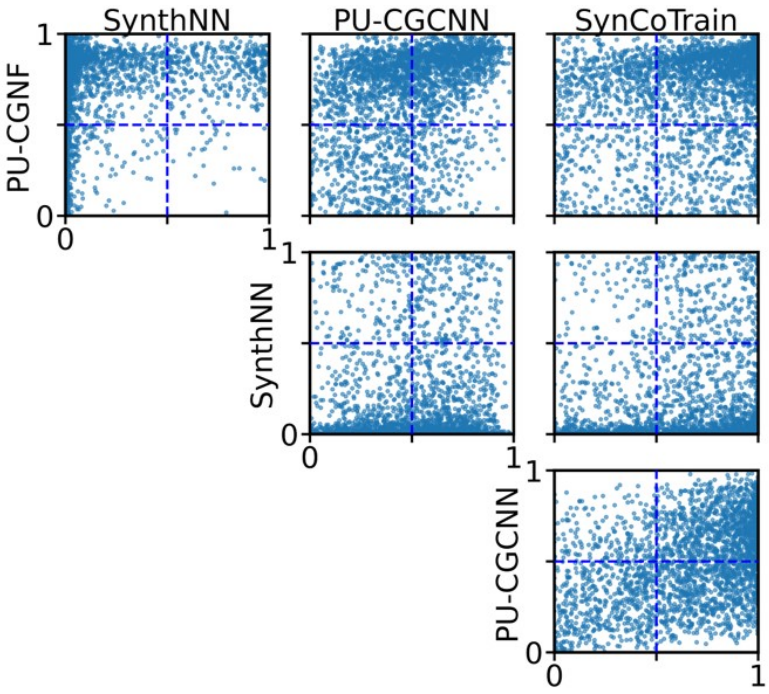

Figure S2. Pairwise score comparisons for each of the four ML predictors compared to one another. Each point represents a generated material, with its coordinates representing the score ascribed to it by each model. The blue dashed lines indicate the score cutoffs for synthesizability classification (0.5), such that points that fall in the lower-left and upper-right quadrants are materials classified as unsynthesizable and synthesizable, respectively, by both models.

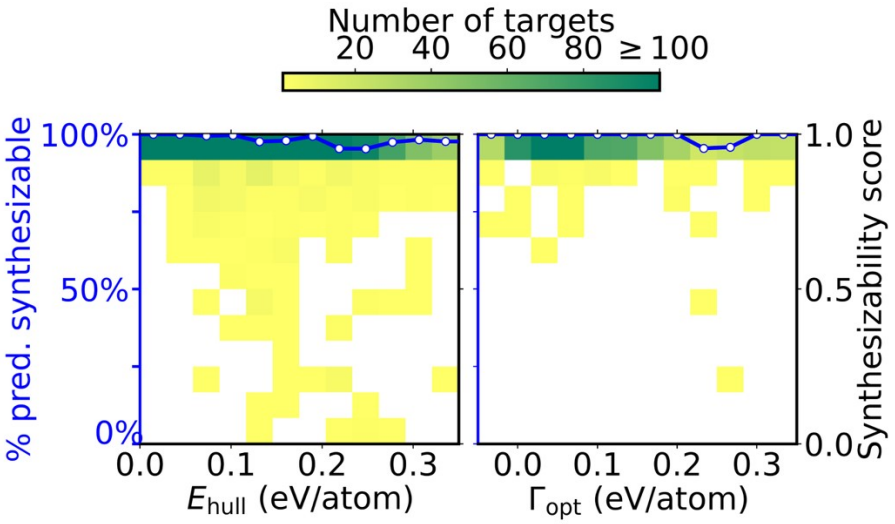

Figure S3. TSDNN predictions and CHGNet-calculated E_{hull} (left) and Γ_{opt} (right) for Chemelon-generated materials. Γ_{opt} is shown only for the 1057 materials with E_{hull} < 100 meV/atom. The heatmap color indicates the number of these target materials having a certain synthesizability score (right-hand *y*-axis) and E_{hull} or Γ_{opt} (*x*-axis). Moving averages of the fraction of materials predicted synthesizable (score > 0.5) as a function of E_{hull} and Γ_{opt} are shown by the blue lines and corresponding left-hand *y*-axes. The moving averages were computed over 30 (left) and 33 (right) meV/atom intervals. White squares indicate no targets falling in a given region.

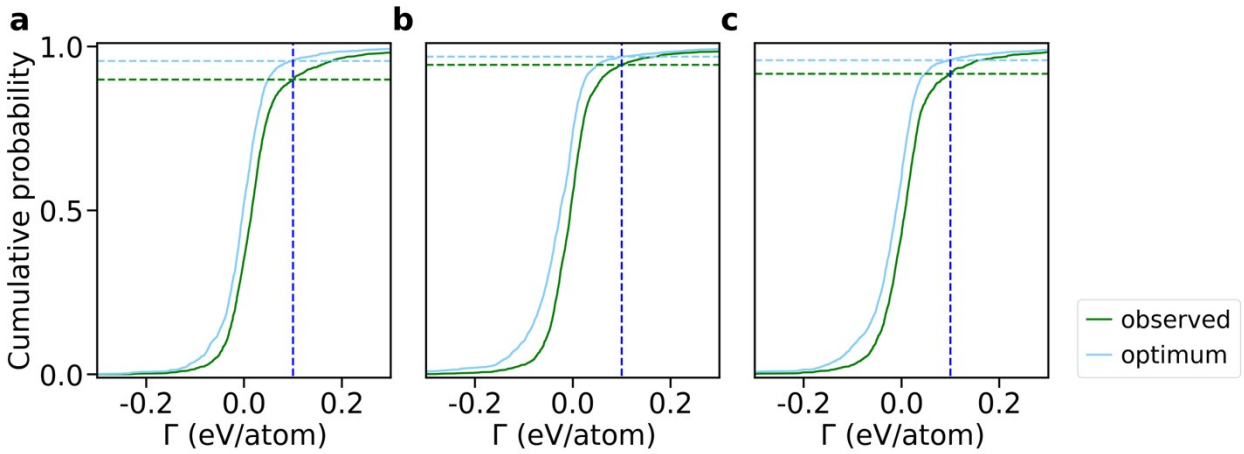

Figure S4: Cumulative distributions of Γ computed with **(a)** McDermott *et al.*'s proposed weights (w_0 = 0.1, w_1 = 0.45, w_2 = 0.45). 90% of observed reactions have Γ < 0.1 eV/atom and 96% of optimized reactions have Γ < 0.1 eV/atom. **(b)** Equal weights on each quantity ($w_0 = w_1 = w_2$ = 0.33). 94% of observed reactions have Γ < 0.1 eV/atom and 97% of optimized reactions have Γ < 0.1 eV/atom. **(c)** Our "optimized" weights (w_0 = 0.16, w_1 = 0.56, w_2 = 0.28). 92% of observed reactions have Γ < 0.1 eV/atom and 96% of optimized reactions have Γ < 0.1 eV/atom. In each panel, the vertical blue dashed line indicates Γ = 100 meV/atom, the dashed green horizontal line indicates the cumulative probability of Γ_{obs} < 100 meV/atom, and the dashed light blue horizontal line indicates the cumulative probability of Γ_{opt} < 100 meV/atom.

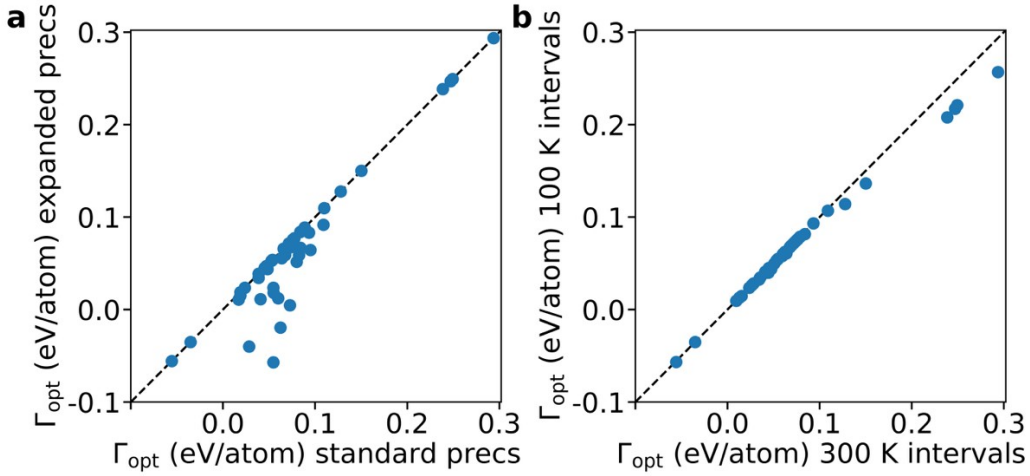

Figure S5: Parity plots of the minimum (optimum) Γ for each of the 44 generated targets in the Ba-Cu-O chemical space, considering **(a)** reactions with only precursors for the text-mined dataset as allowed precursors (14 precursors) vs. those with allowed precursors including experimentally realized materials listed in MP within 50 meV/atom of the hull (18 precursors) and **(b)** reactions over 100 K temperature intervals between 300 and 1800 K vs. reactions over 300 K intervals between 600 and 1800 K.

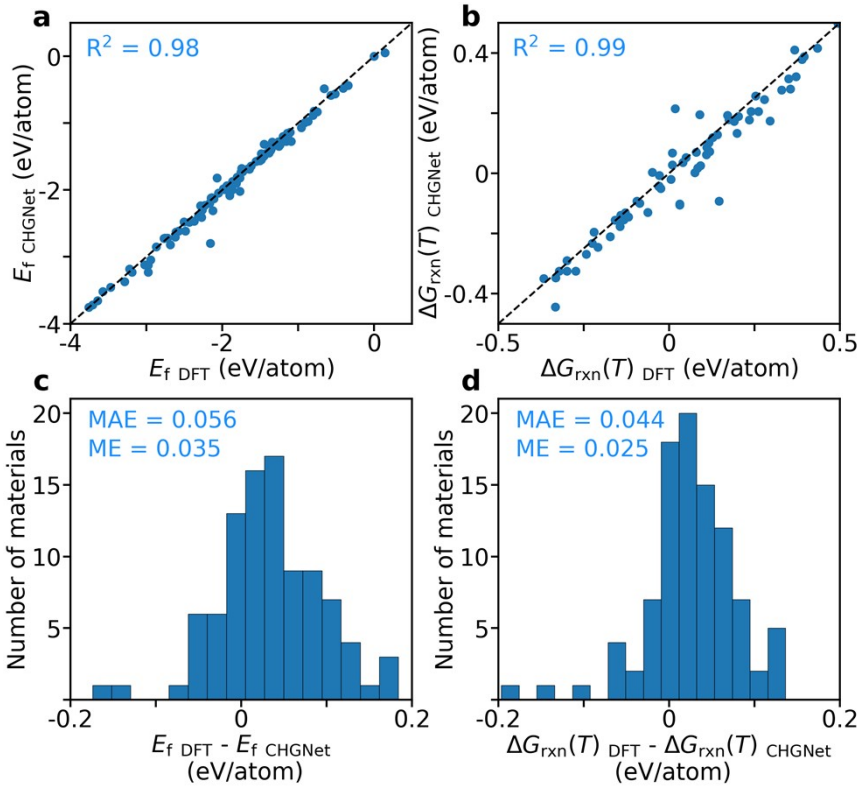

Figure S6: A comparison of heuristics determined using CHGNet vs. using DFT for 100 randomly selected materials from the dataset of generated materials. **(a)** A parity plot of E_f computed using a CHGNet total energy for the target material vs. using DFT with GGA/GGA+U functionals for each material with the R^2 associated with the plot listed. **(b)** A parity plot of $\Delta G_{rxn}(T)$ for the minimum-Γ reaction computed using a CHGNet total energy for the target material vs. using DFT with GGA/GGA+U functionals for all relevant species with the R^2 associated with the plot listed. **(c)** The distribution of error $E_{f\,DFT}$ - $E_{f\,CHGNet}$ for each material and associated mean absolute error (MAE) and mean error (ME) for the whole dataset listed. **(d)** The distribution of error $\Delta G_{rxn}(T)_{\,DFT}$ - $\Delta G_{rxn}(T)_{\,CHGNet}$ for each material and associated MAE and ME for the whole dataset listed.