1. Predicted and Fig data
   1. All\_Predictions\_Figs.rar: A compressed file containing figures showing all the model predicted activation energies relative to host for all host-impurity pairs in the periodic table for which descriptor data was readily available. Predictions in this file came from the model optimized based on descriptors determined with LOG CV (see Table 1) and using the GKRR and GPR model.
   2. Figs.rar: A compressed file containing all the data used to make Figures in the manuscript.
2. Data. All\_Model\_Data.xlsx: An Excel file with the following tabs
   1. Full\_Data: Contains all the original activation energy data (Y1), the activation energies relative to host (Y2), and all the feature data X.
   2. Red\_Feature\_GKRR\_5-fold: Contains Y1, Y2 (see Full\_Data tab description) and the feature data for the features obtained from the GKRR 5-fold selection process (see Table 1).
   3. Red\_Feature\_GKRR\_LOG: Contains Y1, Y2 (see Full\_Data tab description) and the feature data for the features obtained from the GKRR LOG selection process (see Table 1).
   4. Red\_Feature\_GPR\_5-fold: Contains Y1, Y2 (see Full\_Data tab description) and the feature data for the features obtained from the GPR 5-fold selection process (see Table 1).
   5. Red\_Feature\_GPR\_LOG: Contains Y1, Y2 (see Full\_Data tab description) and the feature data for the features obtained from the GPR LOG selection process (see Table 1).
   6. Red\_Feature\_GPR\_LOG\_with\_barrier: Contains Y1, Y2 (see Full\_Data tab description) and the feature data for the features obtained from the GPR LOG selection process (see Table 1).
   7. Red\_Feature\_GKRR\_LOG\_with\_barrier: Contains Y1, Y2 (see Full\_Data tab description) and the feature data for the features obtained from the GPR LOG selection process (see Table 1).
3. Model files and model
   1. GKKR\_LOG\_Model.pkl: This is the exported model from scikit-learn used to create All\_Predictions\_Figs\_GKRR.rar.
   2. GPR\_LOG\_Model.pkl: This is the exported model from scikit-learn used to create All\_Predictions\_Figs\_GPR.rar.
   3. Hyperparameters: Tables of all hyperparams used in every model in the paper to enable model reuse.
4. Original data set: We also remind the reader that the full data set of impurity activation energies, host activation energies, and all DFT calculated energies used to get the activation energy values is available on Figshare with DOI [https://doi.org/10.6084/m9.figshare.1546772.v1](https://doi.org/10.6084/m9.figshare.1546772.v9)0.