

Supporting Information: Multi-fidelity Bayesian Optimization of Covalent Organic Frameworks for Xenon/Krypton Separations

Nickolas Gantzler¹, Aryan Deshwal², Janardhan Rao Doppa², and Cory M. Simon^{3*}

¹Department of Physics. Oregon State University. Corvallis, OR.

²School of Electrical Engineering and Computer Science. Washington State University.
Pullman, WA

³School of Chemical, Biological, and Environmental Engineering. Oregon State University.
Corvallis, OR. USA.

*cory.simon@oregonstate.edu, jana.doppa@wsu.edu

October 13, 2023

S1 Analysis of surrogate model and observed low- vs. high-fidelity Xe/Kr selectivity correlation at the iteration before the optimal COF is acquired

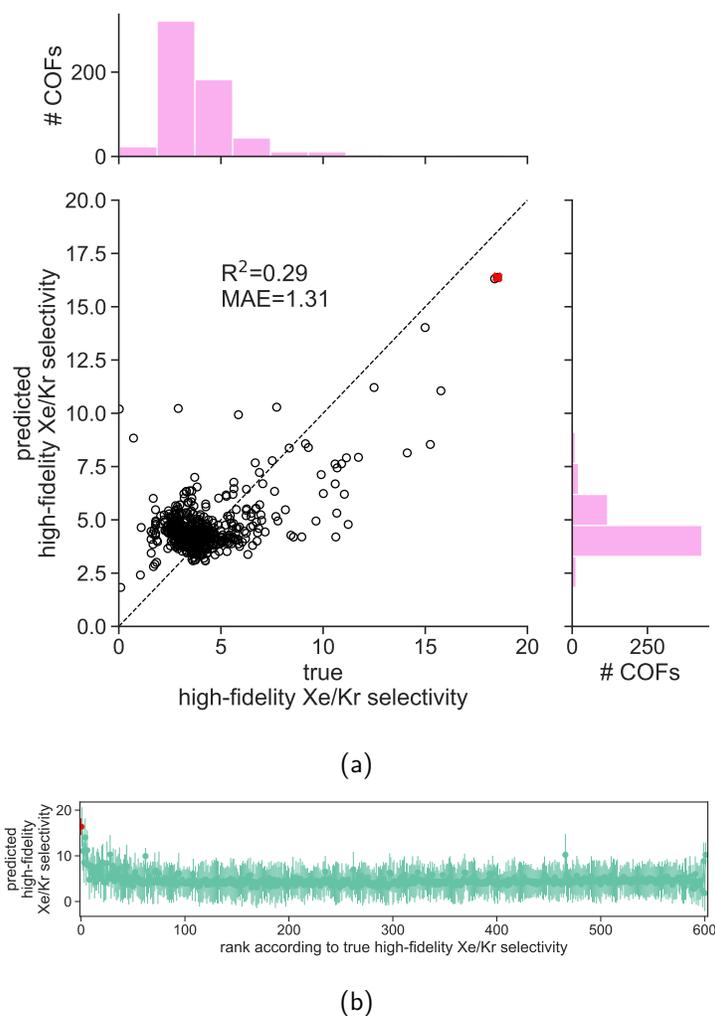


Figure S1: **Predictivity of high-fidelity Xe/Kr selectivity, on un-acquired COFs, by the surrogate model the iteration before the optimal COF was acquired.** (a) Parity plot between true simulated and predicted high-fidelity Xe/Kr selectivity. Each COF is a point. The red x marks the optimal COF acquired in the next iteration. The histograms show the marginal distributions. (b) Posterior distribution of predicted high-fidelity Xe/Kr selectivity of COFs not yet acquired. The COFs are sorted in descending order, by their true simulated Xe/Kr selectivity. Red highlights the optimal COF acquired in the next iteration.

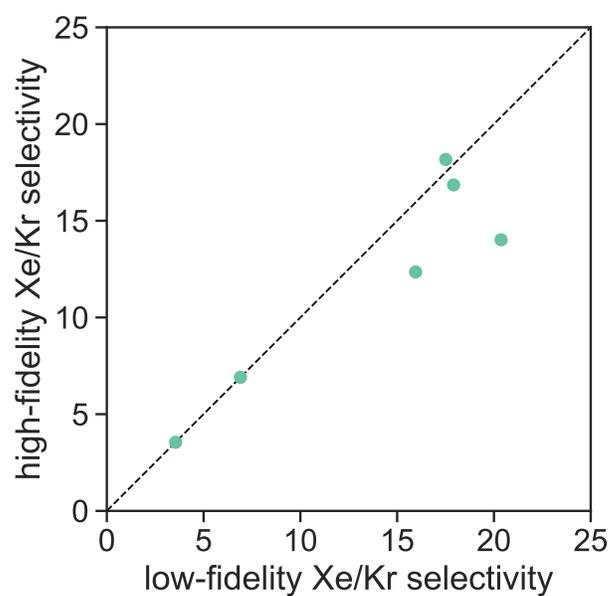


Figure S2: **Comparison of Xe/Kr selectivity predicted by low- and high-fidelity simulations, on the COFs acquired the iteration before the optimal COF was acquired.**

S2 Search efficiency over multiple runs

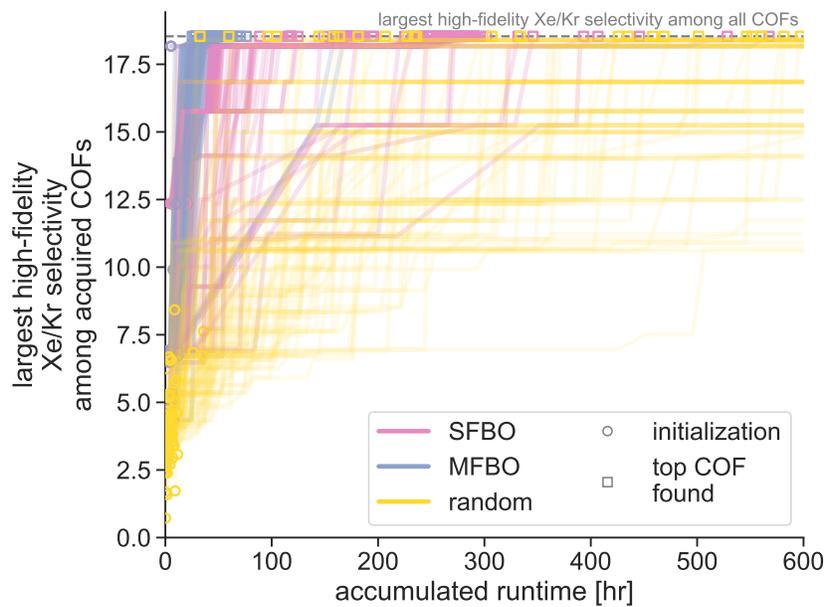


Figure S3: Search efficiency traces for SFBO, MFBO, and random search over different sets of initializing COFs.

S3 MFBO with permuted features

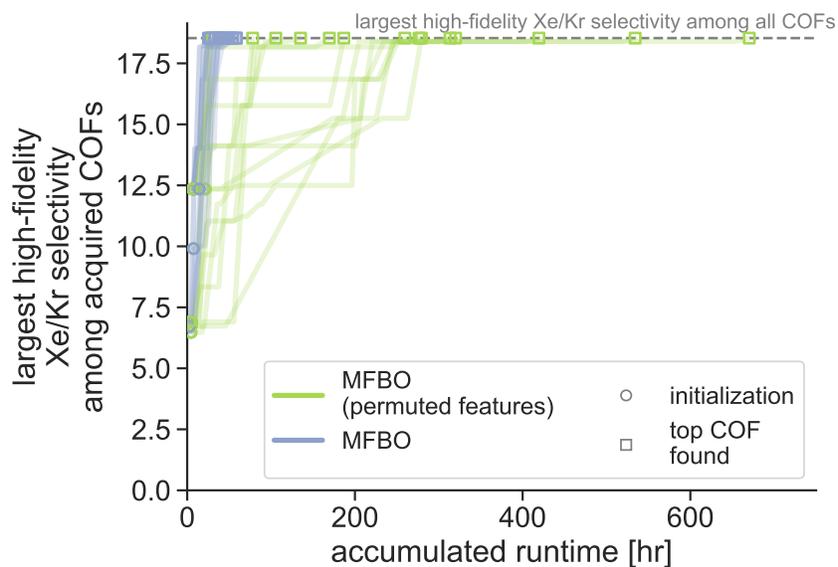


Figure S4: **MFBO with permuted features.** We compare the search efficiency of 15 runs of MFBO with preserved and (each-run-)permuted features.

S4 Post-MFBO analysis of simulated adsorption data

S4.1 Correlation between low- and high-fidelity simulated Xe/Kr selectivity

In Fig. S5 we show the parity plot between predicted xenon/krypton selectivity for the binary grand-Canonical Monte Carlo (BGCMC) simulations and the Henry coefficient calculations—the high- and low-fidelity simulations, respectively. We observe that there is relatively high agreement between the high- and low-fidelity simulations for materials with lower predicted selectivity; however, for the top performing materials, we see that the selectivity determined by Henry coefficients becomes less reliable and less correlated the BGCMC predictions.

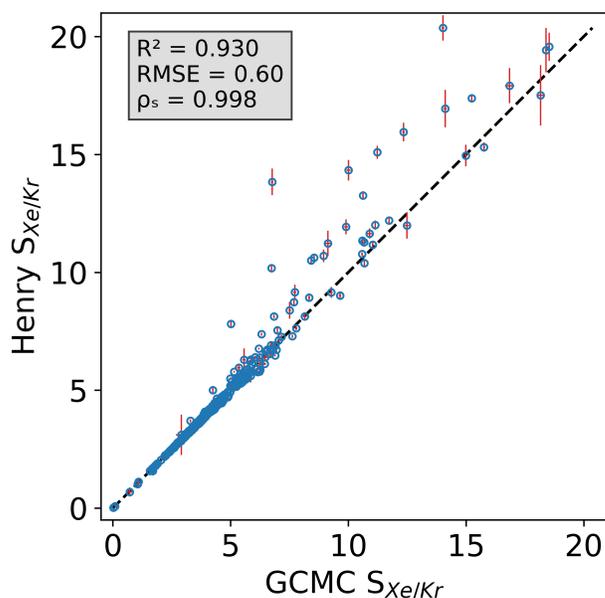


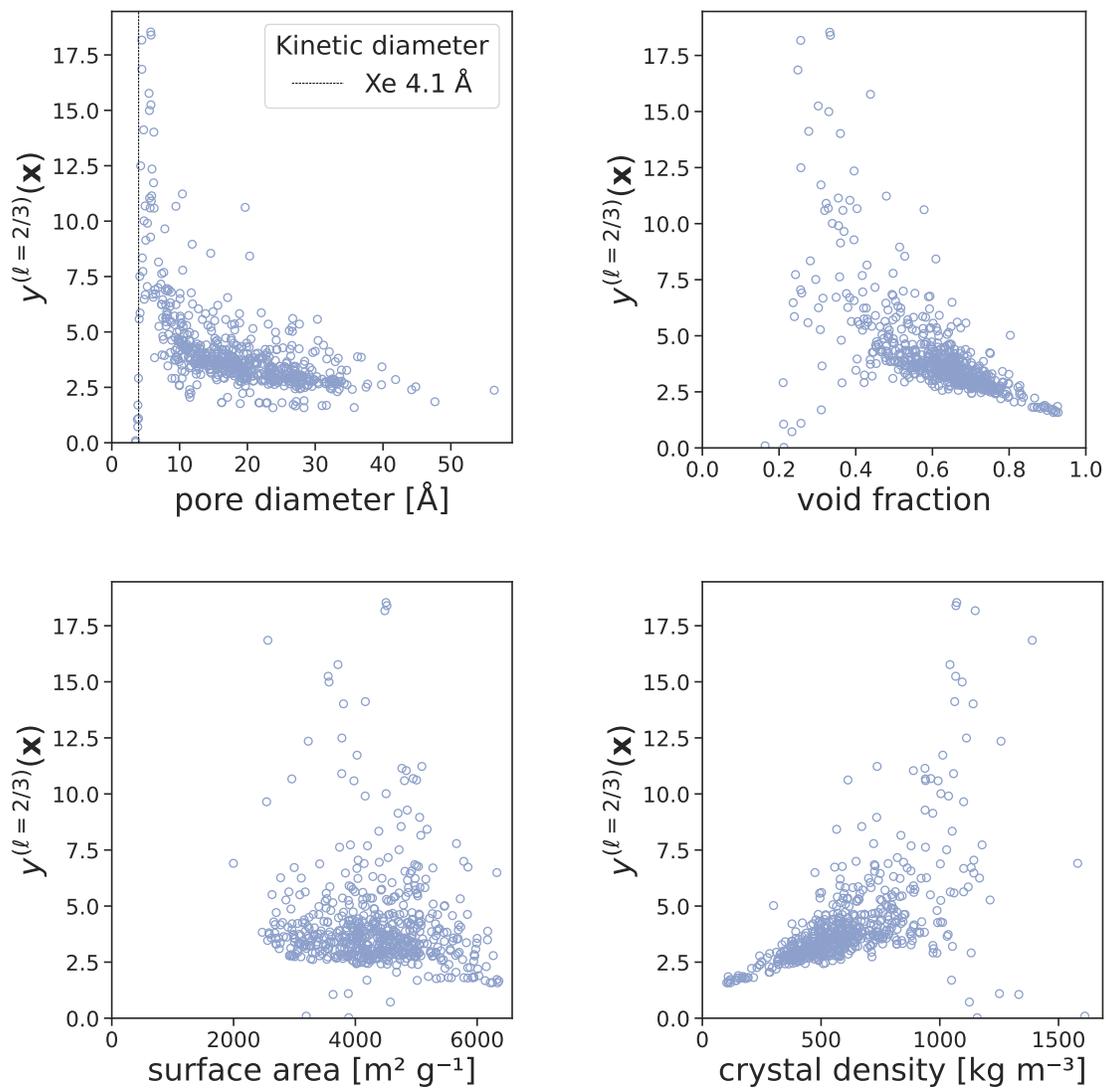
Figure S5: **Comparison of Xe/Kr selectivity predicted by low- and high-fidelity simulations.** Correlation between the Xe/Kr selectivity using high-fidelity (BGCMC) and low-fidelity (Henry) calculations including error bars along both axes. Notice that the top performing COFs, located in the upper right-hand quadrant, have a poor correlation between fidelities with the low-fidelity over estimating material performance.

S4.2 Correlation between high-fidelity simulated Xe/Kr selectivity and COF features

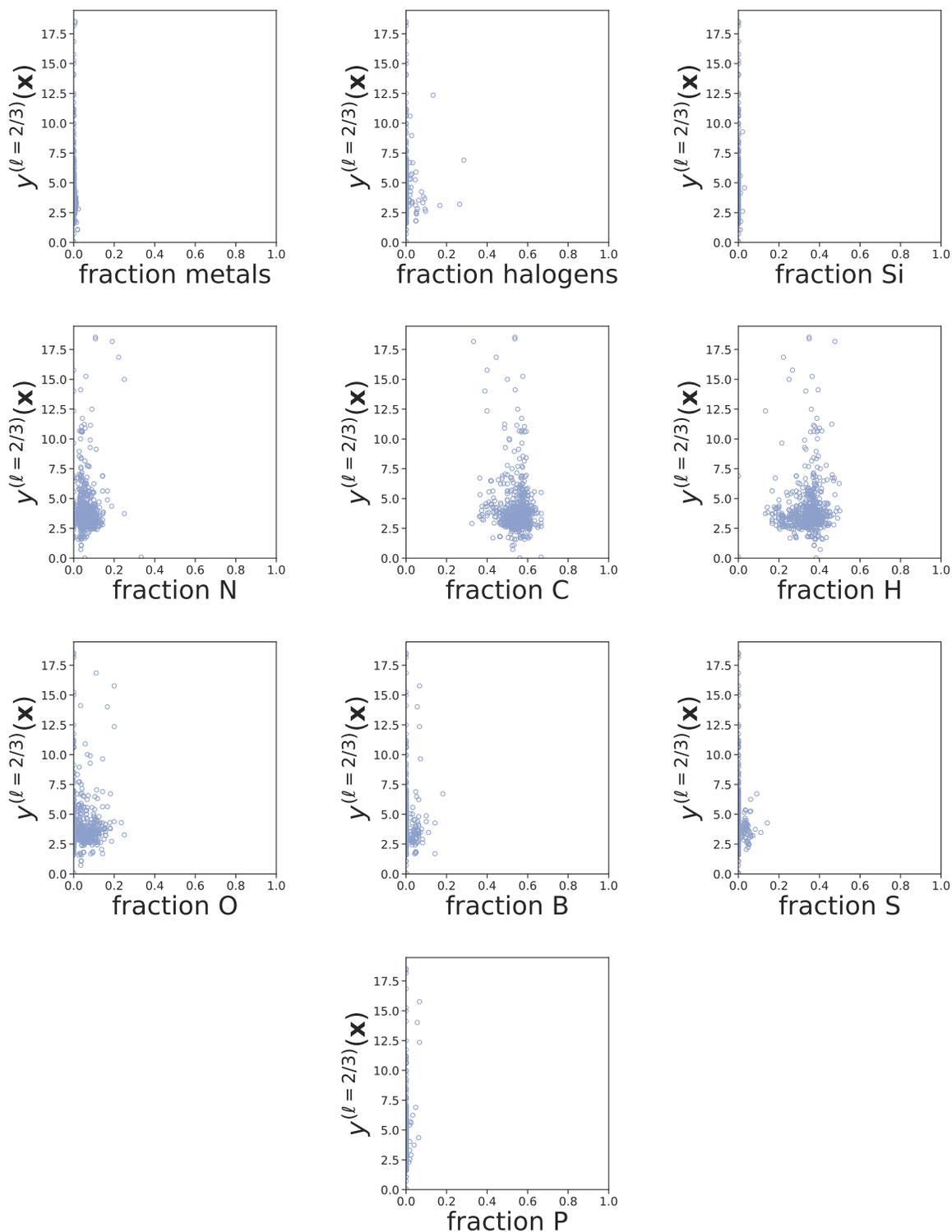
Fig. S6 displays the relationship between the high-fidelity Xe/Kr selectivity of the COFs and their features. Regarding the pore size, we see that COFs that exhibit the highest selectivity have a pore diameter close to the kinetic diameter of Kr and Xe. This is a known trend for metal-organic frameworks (MOFs) [1].

Fig. S7 visualizes the raw feature vector of the 15 COFs with the largest high-fidelity Xe/Kr

selectivity (green) and the 15 lowest (red). The poorest-performing COFs tend to have large void fraction and pore diameter.



(a)



(b)

Figure S6: **The relationship between high-fidelity simulated Xe/Kr selectivity and (a) geometry and (b) compositional features of the COFs.** Each COF is represented by a point. The dashed/dotted line in the top-left panel represents the kinetic diameters of Xe and Kr being 4.1 Å and 3.6 Å respectively.

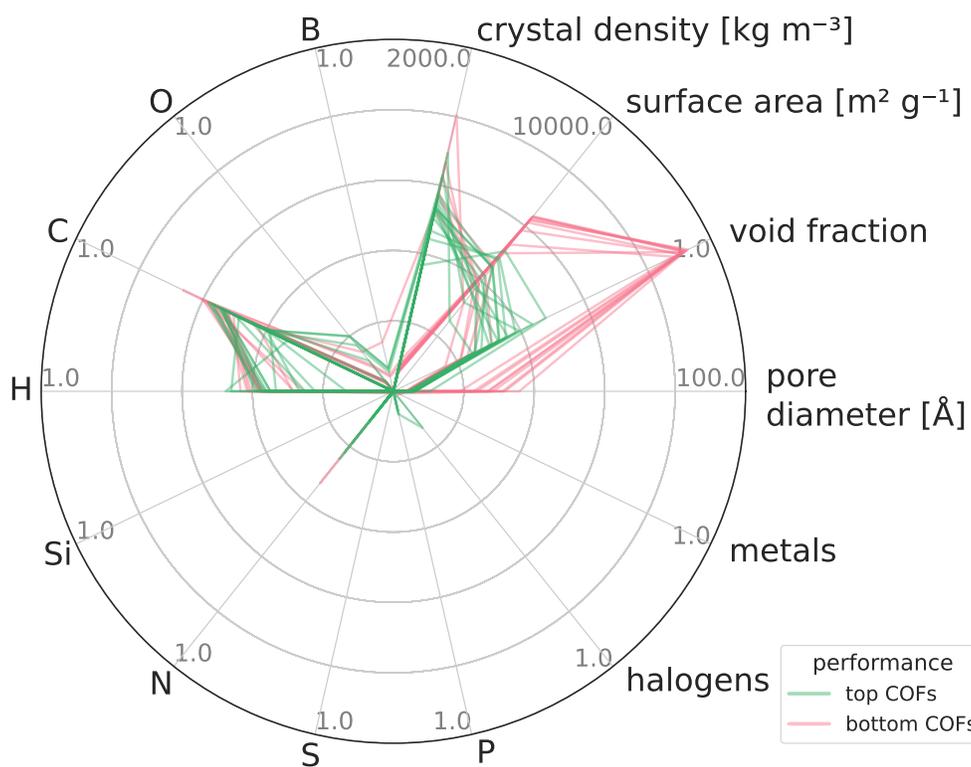


Figure S7: Radar plot displaying COF feature vectors for top 15 and bottom 15 ranking COFs in terms of high-fidelity Xe/Kr selectivity.

S4.3 GP predictivity of high-fidelity Xe/Kr selectivity

Here, we examine the effectiveness of the hand-engineered features of the COFs in \mathbf{x} for predicting the high-fidelity Xe/Kr selectivity via a Gaussian process (GP)—including the case where we treat the low-fidelity Xe/Kr selectivity as a feature (so the input is $[\mathbf{x} \ y^{(1/3)}]$). We randomly partition the COFs into an 80%/20% train/test set. We fit a GP on the train set, then apply the GP to make predictions on the test set of COFs i.e. predict their high-fidelity Xe/Kr selectivity based on their features. Fig. S8 shows a parity plot comparing the predictions of the GP on the test set with the true, held-out high-fidelity Xe/Kr selectivity. Note, here we use *all* simulated data for all 608 COFs. The mean square error (MSE) of the GP is 0.99 when using the standard features \mathbf{x} and 80% of all data for training, and it decreases to 0.29 when the input is augmented with the low-fidelity Xe/Kr selectivity. This dramatic improvement is explained by the strong correlation between the low- and high-fidelity Xe/Kr selectivities (see Fig. S5).

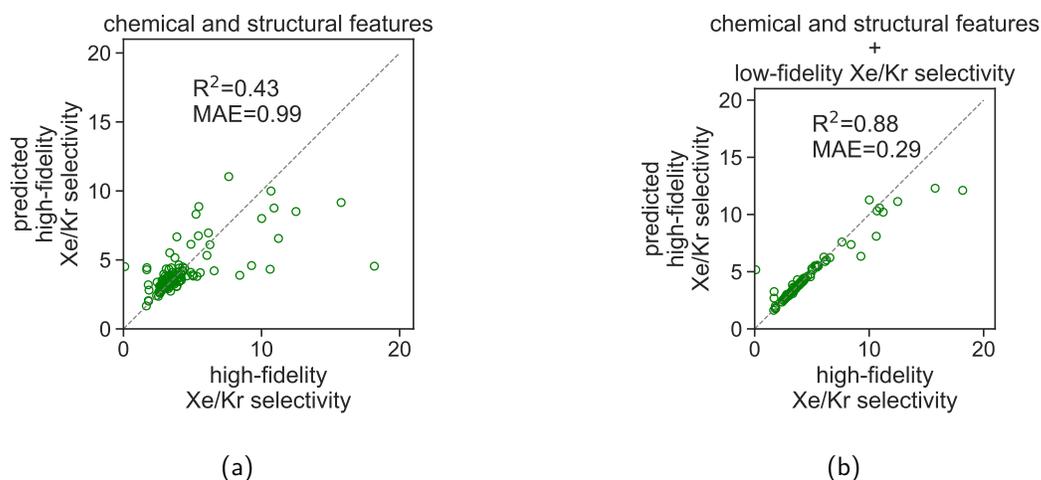


Figure S8: **Parity plot showing GP predictivity of high-fidelity Xe/Kr selectivity** on the test set of COFs (20% of them), with 80% of all COFs for training, using as input (a) the 14 chemical and structural features in \mathbf{x} (same as for MFBO) (b) the same features in (a) but augmented with the low-fidelity Xe/Kr selectivity, so the input is $[\mathbf{x} \ y^{(1/3)}]$.

References

- [1] Benjamin J Sikora, Christopher E Wilmer, Michael L Greenfield, and Randall Q Snurr. Thermodynamic analysis of Xe/Kr selectivity in over 137 000 hypothetical metal–organic frameworks. *Chemical Science*, 3(7):2217, 2012.