Deep learning combined with IAST to screen thermodynamically feasible MOFs for adsorption-based separation of multiple binary mixtures

Ryther Anderson, Diego A. Gómez-Gualdrón[a*]

[a]Department of Chemical and Biological Engineering, Colorado School of Mines, Golden CO 80401, USA
*email: dgomezgualdron@mines.edu

ABSTRACT: The structures of metal-organic frameworks (MOFs) can be tuned to reproducibly create adsorption properties that enable the use of these materials in fixed-adsorption beds for non-thermal separations. However, with millions of possible MOF structures, the challenge is to find the MOF with the best adsorption properties to separate a given mixture. Thus, computational, rather than experimental, screening is necessary to identify promising MOF structures that merit further examination, a process traditionally done using molecular simulation. However, even molecular simulation can become intractable when screening an expansive MOF database for their separation properties at more than a few composition, temperature, and pressure combinations. Here, we illustrate progress towards an alternative computational framework that can efficiently identify the highest-performing MOFs for separating various gas mixtures at a variety of conditions and at a fraction of the computational cost of molecular simulation. This framework uses a "multipurpose" multilayer perceptron (MLP) model that can predict single component adsorption of various small adsorbates which, upon coupling with ideal adsorbed solution theory (IAST), can predict binary adsorption for mixtures such as Xe/Kr, CH_4/CH_6, N_2/CH_4 and Ar/Kr at multiple compositions and pressures. For this MLP+IAST framework to work with sufficient accuracy we found it critical for the MLP to make accurate predictions at low pressures (0.01-0.1 bar). After training a model with this capability, we found that MOFs in the 95th and 90th percentiles of separation performance determined from MLP+IAST calculations were 65% and 87%, respectively, the same as MOFs in the simulation-predicted 95th percentile across several mixtures at diverse conditions (on average). After validating our MLP+IAST framework, we used a clustering algorithm to identify "privileged" MOFs that are high performing for multiple separations at multiple conditions. As an example, we focused on MOFs that were high performing for the industrially relevant separations 80/20 Xe/Kr at 1 bar and 20/80 N_2/CH_4 at 5 bar. Finally, we used the MOF free energies (calculated on our entire database) to identify privileged MOFs that were also likely synthetically accessible, at least from a thermodynamically perspective.

I. INTRODUCTION

Approximately half of the energy consumed by the U.S. chemical industry is used to separate mixtures.[1] This is partly due to the dominant role of distillation, a thermal separation process, which accounts for nearly half of the energy used in chemical separations.[1] Therefore, substitution of distillation with non-thermal methods to separate molecules could greatly save energy and reduce CO_2 emissions.[1] Using fixed adsorption beds is one way to separate molecules non-thermally. However, one of the challenges for widespread implementation of these methods is the need for new porous materials that can carry the separation of interest economically. Nanoporous crystals are well suited for adsorption-based separations because their highly ordered structures allow for the rational tuning of their pore sizes, shapes, and chemistries to reproducibly optimize separation properties.

Among nanoporous crystals, metal-organic frameworks (MOFs) are particularly promising (not only for separations) due to their large "design space," whose size arises from the millions of possible MOF variations that there can be.[2] However, this also makes identifying the best possible MOFs for a given application an overwhelming task experimentally. Consequently, over the past decade, high throughput computational screening (HTCS) has emerged as a powerful strategy to identify promising MOFs. HTCS has been especially prominent for adsorption-based gas storage. Some of the best MOFs for hydrogen and methane storage were identified, synthesized, and tested following HTCS—using grand canonical Monte Carlo (GCMC) simulations.[3–7] Additionally, the extensive data from these screenings revealed structure-property relationships (with unprecedented clarity) and apparent "ceilings" for MOF hydrogen and methane adsorption (deliverable) capacities.[2,5,8] In principle, a similar screening strategy can be applied to screen MOFs for adsorption-based separations, but there are additional challenges associated with screening for separation performance. For example, while gas storage depends on single-component adsorption at established T, P values, chemical

separation depends on the adsorption of mixtures, whose composition, and T, P values may vary widely depending on the mixture source. Therefore, the best MOF for a given separation could be different depending on the exact operating conditions, likely making coupled MOF-process optimization necessary.

To date, MOF screenings for separation have tended to focus on one or two operating conditions at a time. For instance, Sikora et al. used GCMC to screen a database of 137,000 MOFs for separation of Kr/Xe mixtures at compositions relevant to separation from air (80% Xe, 20% Kr) for three pressures (1, 5, and 10 bar) at 273 K.[9] On the other hand, a separate Monte Carlo (specifically using the Widom insertion method) screening of 125,000 MOFs was conducted by Banerjee et al. for separation of Kr/Xe mixtures at conditions relevant to used nuclear fuel reprocessing, i.e., 400 ppm Xe and 40 ppm Kr at 1 bar and 298 K.[10] Note that these Monte Carlo-based adsorption calculations are relatively affordable—due to the monoatomic, nonpolar gases involved— but even then the computational cost would rapidly become prohibitive if one were to repeatedly screen hundreds of thousands of MOFs for different combinations of composition, T, and P in a single study.

The problem of computational cost is even more serious for other separations of interest, where the relevant Monte Carlo simulations are inherently more expensive. One example is separations involving long hydrocarbons. For instance, the increased cost of simulating the adsorption these types of molecules led to Chung et al. performing (configurational bias) GCMC simulations only for 318 MOFs (at one T, P and composition) as part of hierarchical screening of ~5000 MOFs for separation of hexane and heptane isomers.[11] Similar issues are encountered when considering separations involving water, where several works[12–15] have devised clever hierarchical screening strategies to minimize the number of MOFs for which adsorption of water-containing mixtures is directly simulated.

The above indicates that the development of alternative methods that can assess the separation ability of MOFs faster than GCMC simulations could revolutionize MOF screening for separations. Accordingly, there is an ever growing interest in developing machine learning (ML) models, trained on GCMC data, to predict MOF adsorption-based separation capabilities.[16–18] As examples, in early efforts, Simon et al. trained random forests (RFs) to predict selectivity for 80/20 molar Kr/Xe mixtures at 298 K and 1 bar.[16] In more recent efforts, Dureckova et al. used tree-based models to predict selectivity and working capacity for separation of 40/60 $CO_2/H_2$ mixtures for one operating conditions (a 40 bar ↔ 1 bar swing at 313 K), training a different model for each metric.[17] Burner et al. also trained separate artificial neural networks (ANN) to predict selectivity and working capacity for 15/85 CO_2/N_2 mixtures at 1 bar and 298 K.[18]

The caveat is that ML models trained as described above have suffered from a lack of "reusability." Briefly, the ML models are trained from GCMC simulation to predict separation metrics for a given MOF database for specific components, compositions, and operating conditions. However, since the separation properties of the given MOF database were already known from the "training" GCMC data, the motivation to reuse the ML model is nullified and the computational time "investment" to train the model is not recovered. In contrast, a given ML model could find widespread application if it were transferable to components, compositions, and operating conditions on which it was not trained. Thus, most recently, studies using "transfer learning" and developing "multipurpose" models are emerging.[19–22]

Anderson et al.[19] showed that an ANN could predict H_2 adsorption in MOFs at T, P values different from those included in the training data. Sun et al. demonstrated a similar capability for an ANN to predict sorption of pentanediol/water on zeolite MFI, and also illustrated the reuse of some ANN layers on a new model that predicts sorption when changing either the adsorbent or one of the adsorbates.[22] Ma et al. reused some layers of an ANN trained to predict H_2 sorption in a new ANN trained to predict CH_4 sorption.[20] Using genetic algorithm regression and adsorption data for multiple molecules in a subset of previously synthesized MOFs, Gharagheizi et al. obtained an equation capable of predicting isotherms for a diversity of $C_xH_yS_uO_vN_w$ molecules.[21] These authors used this equation to screen MOFs for many chemical separations. However, since the equation did not predict mixture adsorption, adsorption selectivities were approximated as ratios of single-component adsorption loadings at the corresponding single component vapor pressure.

In our own efforts to develop "reusable" models, we trained a deep learning, multipurpose model (a multilayer perceptron or MLP) capable of predicting single-component adsorption for a diversity of small, nonpolar adsorbates across different MOFs and pressures.[23] The key to the development of this model is the inclusion of adsorbate properties and fugacity (along with MOF properties) as model inputs, allowing the model to learn the underlying physics of adsorption, agnostic to the chemistry of the adsorbates, and the use of adsorption data for alchemical species to increase the diversity of the training data. In fact, no data for real molecules was used in training. The caveat is that despite its versatility, much like the

multipurpose model by Gharagheizi et al.,[21] our model could not directly predict mixture adsorption data.
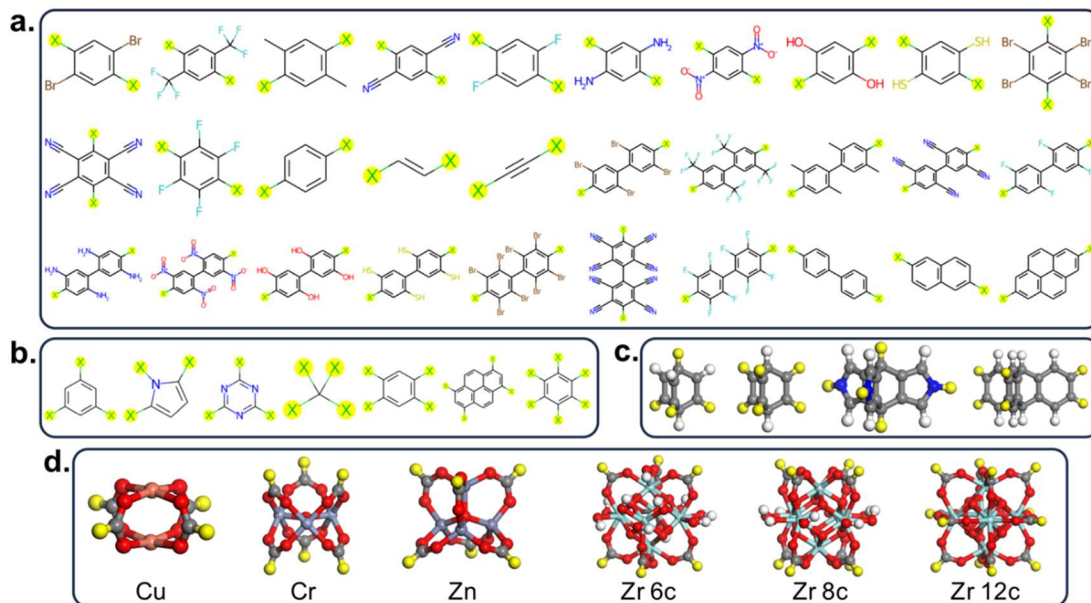
Figure 1. The computational MOF building blocks used to construct our 51,520 MOF database. a) Connecting (so-called "edge") building blocks. b). Planar organic nodular building blocks. c) Non-planar organic nodular building blocks. d) Nodular inorganic building blocks. Connection sites "x", where bonds are formed with other building blocks, are highlighted in yellow.

Intriguingly, however, if one were to use single-component isotherm predictions from ML models as inputs for established thermodynamic theories such as Ideal Adsorbed Solution Theory (IAST), one would be able to generate mixture adsorption data for any composition at unprecedented speed. The potential issue is that while the difficulty in measuring mixture adsorption has made IAST popular with experimentalists, the accuracy of IAST is a subject of debate.[24–27] Indeed, there are assumptions in IAST (such as interactions of all mixture components with the adsorbent being identical) that one would expect to break under the selective adsorption scenario that one pursues for adsorption-based separations. However, considering the transferability of IAST, and the numerous demonstrations of its sufficient accuracy for calculating multicomponent loadings of small molecules in MOFs and other porous crystals (albeit with some error),[24–26,28,29] combining IAST with a machine learning model that can predict full isotherms for multiple adsorbates has would allow us to (almost) instantly screen MOFs for hundreds of separation conditions. This motivated to test and demonstrate the efficacy of this approach.

To achieve this goal, we leveraged a herein updated version of our previously trained multipurpose MLP model,[23] which predicts single-component isotherms for components of a variety of mixtures relevant to chemical separations, and our ability to directly verify the accuracy of IAST via GCMC simulations of mixture adsorption. We chose to focus on the accuracy of the MLP+IAST framework to screen a 51,520 MOF database for their ability to separate the binary mixtures Ar/Kr, Kr/Xe, CH_4/C_2H_6, and N_2/CH_4 at different molar compositions and pressures. The components in these mixtures present different degrees of similarity, providing different scenarios under which to test IAST validity. Additionally, these mixtures are industrially/experimentally relevant. As noted earlier, Kr/Xe separation is relevant to Xe extraction from air[30] and used nuclear fuel reprocessing.[31,32] N_2/CH_4 separation is relevant to upgrading natural gas, as many deposits have >10% N_2 content,[33,34] which needs to be reduced below 5% before use. CH_4/C_2H_6 separation is also relevant to natural gas processing.[35] Natural gas often contains large amounts of ethane, which is a precursor for the synthesis of ethene, an important industrial chemical.[36] Finally, Ar/Kr separation is relevant to separating noble gases from air[25,30] and has niche applications in radioisotope dating, as isotopes of Kr and Ar are excellent environmental tracers, mainly due to their inertness.[37,38]

To assess MOF separation performance, here we calculated the hybrid selectivity-capacity metric

$\Delta Q_{B/A}$ defined by Krishna,[39] which aims to anticipate relative MOF performance in breakthrough experiments normally done during the lab testing phase of MOFs for separation. Thus, based on this metric, we specifically investigated the ability of MLP+IAST to *i)* correctly capture top-performing MOFs within our database for direct material discovery, and *ii)* faithfully reproduce structure-performance relationships that can guide rational material design. Subsequently, using a clustering algorithm, we examined the similarity among top-performing MOFs for different separations and different operating conditions, thus revisiting the concept of "privileged" adsorbents introduced earlier by Tang et al.[40] Finally, we complemented our MOF performance predictions with large-scale free energy calculations, using our recently developed computational framework,[41] to identify which MOFs among top performers are more synthetically likely, at least from a thermodynamic point of view.

II. COMPUTATIONAL METHODS

A. Database Construction. We used our graph-theory based ToBaCCo-3.0 code[42,43] to build a 51,520-MOF database featuring MOFs in 424 different topologies. The building blocks for MOF construction are shown in **Fig. 1**. The geometry of each MOF built by ToBaCCo was optimized using LAMMPS (version 31 March 2017)[44] with interactions described by the UFF4MOF forcefield.[45,46] We used an iterative approach for the optimizations, where in each iteration the atom coordinates were first optimized keeping the MOF lattice parameters fixed, and then the atom positions and lattice parameters were optimized together. Iterations continued until the energy per atom changed by less than 1.0×10^{-4} kcal/mol. The individual optimizations were considered converged when no atom experienced a force large than 1.0×10^{-6} kcal/(mol Å^{-1}) and the energy between one geometry and the next changed by less than 1.0×10^{-4} %. The fast inertial relaxation engine algorithm of Bitzek at al.[47] was used when optimizing only the atom positions and a steepest descent algorithm was used when optimizing the atom positions and lattice parameters together.

B. Adsorption Simulations. Adsorption loadings were simulated using GCMC as implemented in RASPA-2.0.[48] In these simulations, the chemical potential (μ), volume (V) and temperature (T) are kept constant, while the number of molecules (N) varies. To obtain single-component adsorption loadings used to train our neural network (see below), we directly used adsorbate fugacity as input, which RASPA then

used to calculate chemical potential. To obtain mixture adsorption data to evaluate our MLP+IAST model, target pressures were first converted to fugacities using the Peng-Robinson equation of state, which were then used as input for the simulations. Single-component GCMC simulations used 2,000 equilibration cycles, followed by 2,000 production cycles to calculate ensemble averages. Binary mixture GCMC simulations used 3,000 equilibration cycles followed by 3,000 production cycles. Adsorbate-adsorbate interactions were modeled using Coulomb and Lennard-Jones (LJ) interactions, while adsorbate-framework interactions were modeled using only LJ interactions. Lorentz-Berthelot mixing rules were used. Long-range electrostatic interactions between adsorbates were modeled using Ewald summation with a precision of 1.0×10^{-5}, while long-range LJ interactions were neglected beyond a cutoff of 12.8 Å. The LJ parameters for all real adsorbates are given in **Tables S1** and **S2**. The LJ parameters for MOF atoms were assigned from the Dreiding forcefield[49] if available, otherwise UFF[50] parameters were used.

C. MLP+IAST Model Construction. We retrained our previously published single-adsorption-predicting multi-layer perceptron (MLP) model,[23] which was originally trained with adsorption loadings at fugacities between 1 and 100 bar, to now include lower fugacities (as here IAST required adsorption data starting at sub-atmospheric pressures). We used broadly the same training framework detailed in our previous work[23] with a few key differences (noted here). Specifically, while we used adsorption loadings of the same 200 alchemical adsorbates on the same 2,400 MOFs, we now included fugacities as low as 0.01 bar (see **Table S3** for the training, validation, and testing fugacities used). We also employed the same 200-MOF validation set, the same adsorbate features as MLP inputs, and similar MOF features. However, while we used the same six MOF textural properties, namely void fraction (VF), gravimetric surface area (GSA), largest pore diameter (LPD), pore-limiting diameter (PLD), and pore size standard deviation (PSSD), as in ref. [23], we updated the chemical features to use the UFF4MOF atom types found in our 51,520-MOF databases. We use this typing scheme as it gives more implicit information about the chemical environment of the MOF atoms than the Dreiding/UFF atom types. The MOF atom types considered (with a brief description) are listed in **Table S4**. LPDs and PLDs for every MOF were calculated using zeo++ (version 0.2.2).[51] The GSAs, VFs, and pore size distributions were calculated using RASPA-2.0 with a N$_2$ sized probe for surface area calculations and a helium probe for VF calculations. Each chemical motif number density was calculated by dividing the

number of each motif type present in a given MOF by the MOF unit cell volume in Å^3.

Another difference with our previous MLP training was that here we used log of fugacity as input instead of fugacity itself. We used the `TensorFlow Python` library to build and train our MLP. All MLP training details and hyperparameters are given in **Section S2** and **Table S5**, respectively. The learning curve for our MLP is shown in **Fig. S2** and a comparison of our MLP validation and test predictions is shown in **Fig. S3**. Co-adsorption loadings were calculated from predicted single component isotherms using the `pyIAST Python` package written by Simon et al.[52] IAST requires single-component adsorption isotherms of each component of the mixture as input to estimate mixture adsorption. Traditionally, a model (e.g., a Langmuir model) with an analytical integral is used to provide single component loadings at any required pressure, but it is also possible to linearly interpolate single-component isotherms and calculate integrals numerically.[25,52] This latter method has the advantage of not making assumptions about the shapes of the isotherms but requires sufficient "resolution" in adsorption data for linear interpolation and integral calculation to be accurate. We use our retrained MLP-model to do just that, providing single-component isotherms consisting of adsorption loadings at 73 pressures between 0.01 and 100 bar. Given that our MLP works with (logs of) fugacity as input, pressures used in IAST calculations were converted to fugacities using the Peng-Robinson equation of state before being input to the MLP. All code versions used for our MLP training and MLP+IAST calculations are given in **Table S6**.

D. Model-based clustering. For the clustering study discussed in **Section 3.4**, we used a model-based clustering approach as implemented in the `R` package `mclust`[53] (version 5.4.7 with `R` version 3.6.2). In model-based clustering the data are assumed to be sampled from a mixture of probability distributions, typically from the same parametric family. The `mclust` package is specifically for Gaussian Mixture Modeling (GMM),[53] in which each observation (data point) is assumed to be sampled from a multivariate-normal distribution. Modeling the data with k distributions corresponds to categorizing the data in k clusters. One of the main advantages of model-based over k-means or hierarchical clustering is that each observation can be assigned a probability of belonging to each cluster, i.e., the cluster assignment is "soft", which allows for more refined analysis of the data structure. More details on model-based clustering and our use of the `mclust` algorithm are given in **Section S3**.

E. Free-energy calculations. We calculated the free energies used in **Section 3.5** using the Frenkel-Ladd (FL) path method.[54] The FL path is a thermodynamic integration approach used to calculate the absolute free energies of crystalline solids (in our case MOFs). Essentially, the free energy difference between an Einstein crystal (EC) representation of a MOF (in which each atom is tethered to its equilibrium position with a harmonic restraint and does not interact with the other framework atoms) to a flexible forcefield representation of the MOF is calculated using thermodynamic integration. Then, since the free energy of an EC is known analytically, the absolute free energy of the MOF is obtained as the sum of the calculated free energy difference and the free energy of the EC. The specifics of this calculation are given in our previous work in ref [41]. We used the same approach as described therein for all free energy calculations.

III. RESULTS AND DISCUSSION

A. Predicting single-component loadings at low fugacities. There are two major reasons why an MLP+IAST framework could potentially give significantly different adsorption loadings than direct GCMC simulation of mixture adsorption: *i)* shortcomings in the IAST framework and *ii)* inaccurate isotherms used as input for IAST. Our previously trained MLP was capable of accurately predicting single-component adsorption loadings of a variety of small-gas adsorbates in MOFs at fugacities between 1 and 100 bar.[23] But, numerous small-gas separations depend on mixture adsorption in the $1-10$ bar range which, if it is to be predicted with IAST, requires accurate single-component loadings for pressures around the 0.01-0.1 bar range to be accurate This is because IAST calculations heavily depend on the shape of the single component isotherm, especially at (relatively) low pressures. Therefore, we retrained our previous MLP model to make accurate predictions at fugacities as low as 0.01 bar (as described in Section 2.3), which required adding adsorption data at low fugacities to the original 1-100 bar data.

To test the prediction capabilities of the new MLP model, we compared MLP predictions for six adsorbates in 1,000 test MOFs at 15 fugacities with the results from GCMC simulations. We started by comparing individual loading predictions for 15 different fugacities between 0.02 and 100 bar as shown in the parity plots in **Fig. 2**. The plots are presented in log scale to emphasize the comparison for the new predictions at low fugacities, although it should be noted that we retained similar prediction accuracy at high fugacities as in our previous MLP. Four prediction performance metrics were considered for

the data in **Fig. 2:** the median absolute error (MdAE) median absolute percentage error (MdAPE), the Pearson correlation coefficient (R), and Spearman rank correlation coefficient (S), all calculated for each fugacity (see **Fig. S4**).
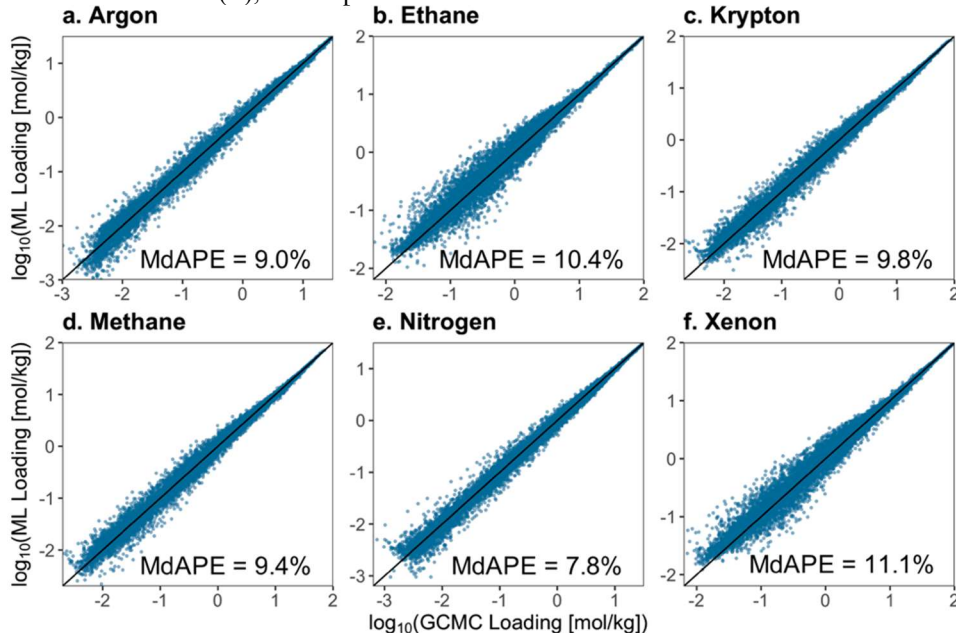
Figure 2. Machine learning (ML) predictions versus GCMC predictions for single-component loadings of the indicated adsorbates in our 1,000 test MOFs. Loading values are compared for each MOF corresponding to 15 different fugacities between 0.02 and 100 bar. The MdAPE calculated across all fugacities is shown for each plot.

We used median errors instead of mean errors because the error distributions herein have long tails, making the medians a better measure of central tendency for the error. Furthermore, for small loadings, percentage errors can be very large even for accurate predictions, e.g., predicting 0.01 mol/kg as 0.02 mol/kg would correspond to a 100% error, despite being an accurate prediction. Thus, MdAPE is a more reliable metric than mean absolute percentage error (MAPE) in cases (such as this one) where predictions can span orders of magnitude. As for the correlation coefficients, R values near unity indicate a strong linear relationship between the MLP- and GCMC-predicted values, whereas S values near unity indicate that the ranking of MOFs using MLP predictions is largely similar to that obtained using GCMC data.

Our MLP model MdAE increased with fugacity for all adsorbates, the MdAPE error decreased, and the correlation coefficients got closer to unity. This makes intuitive sense as at low fugacities (0.02-0.08 bar) loadings tend to be small (< 1 mol/kg) and with low variance (e.g., the spread of loading values across MOFs is smaller at 0.02 bar than at 1 bar). This means that even relatively small absolute errors can result in large percentage errors and significant deviations from linearity/rank correspondence. Nonetheless, for each fugacity (including low fugacities) the MdAEs for each adsorbate are all considerably less (generally by order of magnitude) than the corresponding median GCMC loadings. Accordingly, we were satisfied with the predictions of the retrained MLP model. Notably, this illustrates the advantage of multipurpose models. In our case, the ability of our original MLP to already predict adsorption at different fugacities made it so that to make the model transferable to lower fugacities we could still reuse old training data only needing to *add* data for new fugacities (we expect an analogous scenario if were to add data for, say, new types of adsorbates).

Crucially, while an MLP model that can accurately predict loadings at individual pressures for a variety of adsorbates is useful, to be able to couple the model with IAST, full isotherms must also be accurately predicted (especially capturing the isotherm curvature). Thus, to understand how well our model predicts full isotherms we calculated the isotherm mean absolute percentage error (IMAPE) for each of our 1000 test MOFs. IMAPE is defined as the mean of the absolute percentage errors (MAPE) of each point in an isotherm (in this case the 15-point isotherms we used to test our model). Contrary to the performance metrics for individual predictions, we use means instead of medians for the isotherm percentage errors

6

because in this case we are not interested in the central tendency of the errors in each isotherm, but rather we are seeking an aggregate measure of the quality of the "fit" of each isotherm. The latter can be interpreted as the distance between the simulated isotherm and the corresponding predicted isotherm.
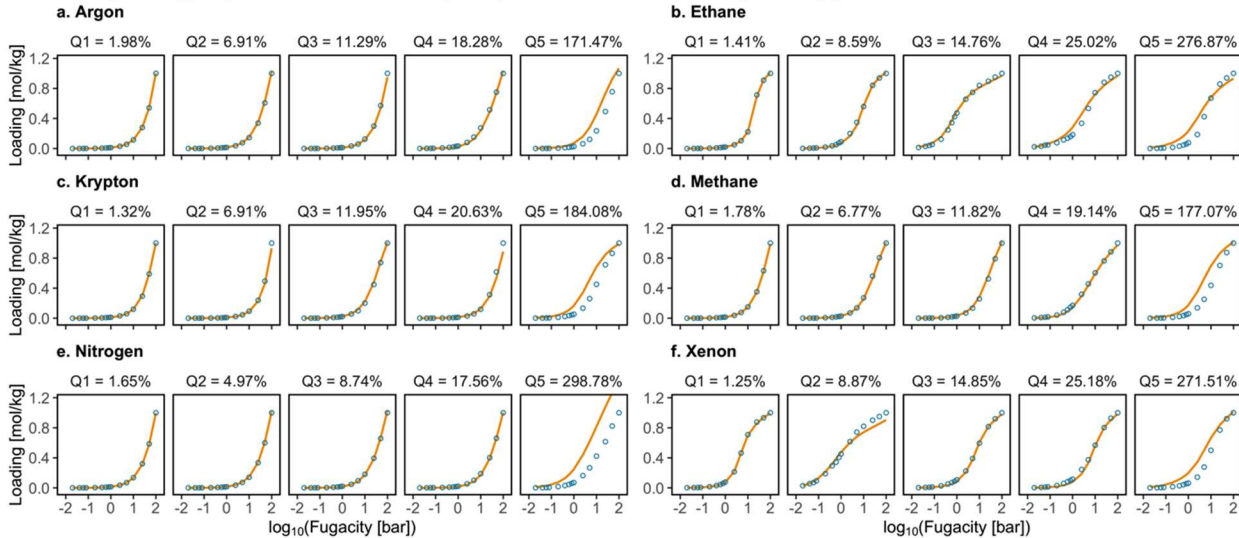
Figure 3. Comparison of our MLP-predicted isotherms for the six indicated adsorbates to GCMC simulated isotherms for the 0.00 (Q_1), 0.25 (Q_2), 0.50 (Q_3), 0.75 (Q_4), and 1.00 (Q_5) quantiles of the IMAPE. The open, blue points are GCMC simulated values and the orange lines are MLP predictions.

Next, for each adsorbate, we calculated the 0.00 (Q_1), 0.25 (Q_2), 0.50 (Q_3), 0.75 (Q_4), and 1.00 (Q_5) quantiles of IMAPE and plotted the isotherms of the MOFs nearest each quantile (**Fig. 3**). The Q_1 isotherm can be thought of as the "best" prediction, the Q_5 as the worst, and the Q_3 as an average (median) prediction (according to the IMAPE values). From **Fig. 3** we can observe our model generally reproduces isotherm shapes well, without noise or unphysical behavior (i.e., loading decreasing with fugacity). Even the "worst" predictions tend to be qualitatively correct, albeit with large systematic errors. **Fig. S5** shows the same data as **Fig. 3** but on a log-log scale, which is better for observing deviations for small loading values.

At this point, it is worth emphasizing that the prediction performance metrics discussed so far are for 1,000 MOFs that were not present in the training or validation set and have many organic building blocks and hundreds of topologies that were not present in the training set. Thus, we expect our MLP model will achieve similar accuracy when applied to other MOFs with a diversity of chemistries and topologies that were not present in the current work (although in our experience ML predictions seem to be relatively more sensitive to the chemistry of the inorganic nodes).

B. Combining neural network predictions with IAST. With a promising MLP model completed, we then proceeded to evaluate the efficacy of combining it with IAST to screen MOFs for separation processes using fixed adsorption beds. For such application we are concerned with both the selectivity and adsorption capacity for the strongly adsorbed component. Screening materials for either metric individually often results in contradicting rankings as it has become clear that materials with the highest selectivities tend to have low capacities (and vice versa).[10,39] Consequently, screening is best done using a combined selectivity-capacity metric. Here we use the separation potential ΔQ, a hybrid metric recently defined by Krishna.[39] For a binary mixture, ΔQ can be calculated from the mixture adsorption loadings for each component as:

$$\Delta Q = Q_A \frac{x_B}{1 - x_B} - Q_B \qquad (1)$$

where Q_A is the loading of the strongly adsorbed component, Q_B is the loading of the weakly absorbed component, and x_B is the mol fraction of component B. ΔQ can also be expressed in terms of selectivity for the strongly adsorbed component

$$\Delta Q_{B/A} = Q_A \frac{x_B}{1 - x_B} \left(1 - \frac{1}{S_{A/B}} \right) \qquad (2)$$

where $S_{A/B}$ is the adsorption selectivity defined as:

$$S_{A/B} = \frac{Q_A/Q_B}{x_A/x_B}. \qquad (3)$$

We can observe from Eq 2 that

$$\Delta Q_{B/A} \approx Q_A \frac{x_B}{1 - x_B} \qquad (4)$$

for $S_{A/B} \gg 1$. Therefore, a material with very large $S_{A/B}$ is likely to be ranked lower according to $\Delta Q_{B/A}$, than a material with a moderate $S_{A/B}$ and high capacity. A detailed derivation of ΔQ is given in ref.

[39]. To assess the accuracy of MLP+IAST, then, we compared MLP+IAST predictions of loadings and ΔQ for four binary gas mixtures (iAr/Kr, Kr/Xe, CH$_4$/C$_2$H$_6$, and N$_2$/CH$_4$) at four compositions (20/80, 40/60, 60/40, and 80/20) and at 10 pressures from 1 to
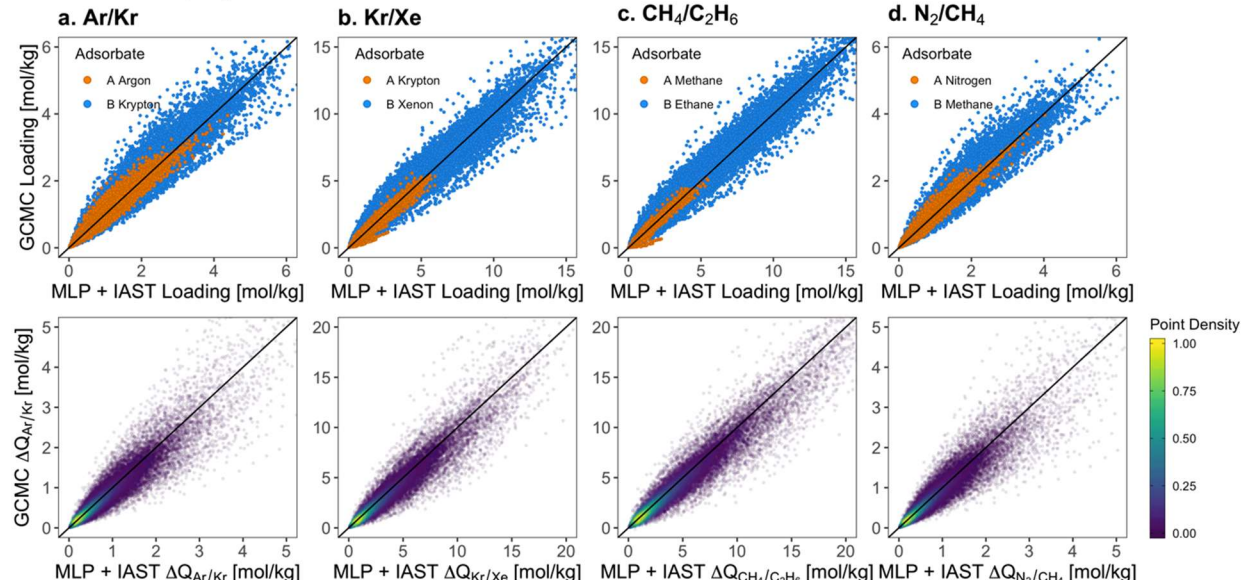
Figure 4. MLP+IAST predictions of the loadings of each component (top) and $\Delta Q_{B/A}$ (bottom) verses GCMC predictions for the indicated binary mixtures. The loading plots are colored according to the component and the $\Delta Q_{B/A}$ plots are colored by point density. These plots show values for all 10 pressures (1 to 10 bar) and mixture compositions (20/80, 40/60, 60/40, 80/20) considered.

10 bar with the corresponding metric calculated from binary adsorption GCMC simulations in the same 1,000 test MOFs we focused on in **Section 3.1**. From this point on we refer to a pressure, composition and mixture combination as a "condition," in which case in this section we considered a total of 160 conditions for our tests. Parity plots comparing the binary loadings and ΔQs predicted by our MLP+IAST model with the corresponding GCMC results are shown in **Fig. 4** for all the conditions considered. Further, we show four error metrics for each condition in **Figs. S6-S9**. These metrics include the previously described MdAPE, R, and S, as well as an additional ranking metric described below. The MdAPEs for each condition were all (generally, well) below 25% and R and S values were all above 0.80 for both loading and ΔQ predictions.

Ultimately, for HTCS studies, provided that ΔQ predictions are "reasonable," one is arguably more interested in correctly capturing the highest performing MOFs so they can be suggested for synthesis and experimental testing. While to some extent S can gauge whether this is true for our models, we find that it is more informative to directly compare the top N MOFs predicted by our MLP+IAST framework with the top M MOFs predicted by GCMC for increasing values of N (with $N \geq M$, and N and M being integers). Specifically, we considered how many

GCMC-predicted top 50 ΔQ MOFs (i.e., $M = 50$ corresponding to the 95th percentile) were identified when considering the MLP+IAST-predicted top N MOFs for each condition. This equates to 8,000 top performing MOFs (50 for each of the 160 studied condition). These results are summarized in **Fig. 5**.

When $N = 50$ we found that between 60 and 70% of GCMC-predicted top 50 MOFs were identified (depending on the condition). By increasing N so $N = 100$ (i.e. considering the 90th percentile), we found that between 80 and 95% of the GCMC-predicted top 50 MOFs were identified, with an average (mean) of 87.5% correctness across conditions. Thus, in screening MOFs for chemical separations, one can expect (on average) to capture 87.5% of the MOFs in the actual 95th percentile (top 50 for 1000 MOFs) if one considers MOFs in the 90th percentile of ΔQ values predicted by MLP+IAST. In other words, 90% of MOFs in a database could be discarded by our MLP+IAST framework, if this were to be used in a first step of a hierarchical screening study.

C. Predicting structure-performance relationships.
While direct discovery of top-performing MOFs is the central goal of HTCS, it is not the only goal. Structure-performance relationships are an invaluable secondary result of HCTS, which can provide "clear-cut" design rules to guide rational material design experiments.

The clarity of the derived rules hinges on a large amount of data from which structure-performance relationships are visualized, even to the point of revealing potential physical limits for a given performance metric.[2,4,6,8,9] Thus, a predictive model that replaces GCMC simulations should—in addition to being quantitively accurate enough for use in hierarchical screening—correctly reproduce these relationships. Here, we tested that ability for our MLP+IAST framework to do this. For various conditions, we plotted MLP+IAST and GCMC ΔQ values versus MOF properties on the same graph. Additionally, we compared the mean and standard deviation of ΔQ for MOFs in the 95th percentile for each condition using MLP+IAST and GCMC.
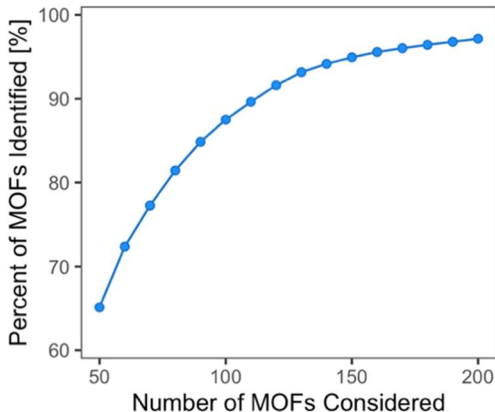
Figure 5. The percent of GCMC-predicted top 8000 ΔQ MOFs (y-axis) identified (50 top MOFs for each of the 160 conditions considered) when considering the top N MLP+IAST-predicted MOFs for each condition (x-axis).

As an example, **Fig. 6** shows MLP+IAST and GCMC $\Delta Q_{Kr/Xe}$ versus MOF void fraction (VF) for various pressures and mixture compositions. Clearly, the MLP+IAST predictions closely reproduce the GCMC-predicted relationships for these conditions. **Figs. S10-S12** show analogous examples for the rest of gas mixtures we considered in this work. We also found that the mean and standard deviations of VF, LPD, PLD, GSA, and density for MOFs in the 95th percentile of MLP+IAST ΔQs matched the analogous values for GCMC ΔQs very closely for all our considered conditions (see **Figs. S13-S17**), indicating that our MLP+IAST model reproduces the relationships of ΔQ with other MOF properties with high fidelity. Accordingly, one would draw the same conclusions about which MOF properties generally yield high performance regardless of whether one uses our MLP+IAST model or GCMC simulations to

screen a MOF database. Notice that both the spread and central tendency of the properties of top MOFs are informative because relationships with sharp peaks (as in the $\Delta Q_{Kr/Xe}$ verses VF relationship) potentially indicate that the MOF property under consideration is more important for high performance than another property with a broad peak in the structure-performance relationship.

D. Identifying for what separations there may be privileged MOFs. After validating the predictive ability of our MLP+IAST model in the previous sections, in this final section we use our model in an example screening study. We highlight some of the unique results that can be garnered by harnessing the transferability of our model to a wide variety of conditions. We screened a database of 51,520 MOFs for separation performance (assessed with ΔQ) for four gas mixtures (Ar/Kr, Kr/Xe, CH_4/C_2H_6, and N_2/CH_4) at two compositions (80/20 and 20/80) and five pressures (1.0, 2.5, 5.0, 7.5 and 10.0 bar). This equates to 2,060,800 data points, which would take ~2,000,000 CPU hours to calculate using GCMC (assuming 1 CPU hour per simulation). In contrast, we were able the generate these data in 12.6 CPU hours using our MLP+IAST framework. We decided to use this data to identify MOFs that are expected to be "privileged" and be high performing for multiple conditions (recall we define a condition as a pressure, composition, and mixture combination). We accomplished this by calculating the mean properties of MOFs in the 95th percentile of ΔQ for each condition and then clustered the conditions based on these mean properties.

Conditions that are in the same (or similar) clusters are expected to have top-performing MOFs with similar properties, and the property distributions of each cluster yield information useful for designing materials that are useful for multiple separations. Specifically, we considered the MOF largest pore diameter (LPD), pore limiting diameter (PLD), gravimetric surface area (GSA), void fraction (VF), density, and pore size standard deviation (PSSD) as the clustering variables and used a model-based clustering algorithm as implemented in the mclust R package. We used these variables as we expect the ΔQ values for the conditions we considered to be primarily influenced by MOF textural properties and because these variables have underlying distributions that can be modeled.

The results of the clustering are summarized in **Fig. 7** which show the clustered conditions projected onto the first three directions resulting from the
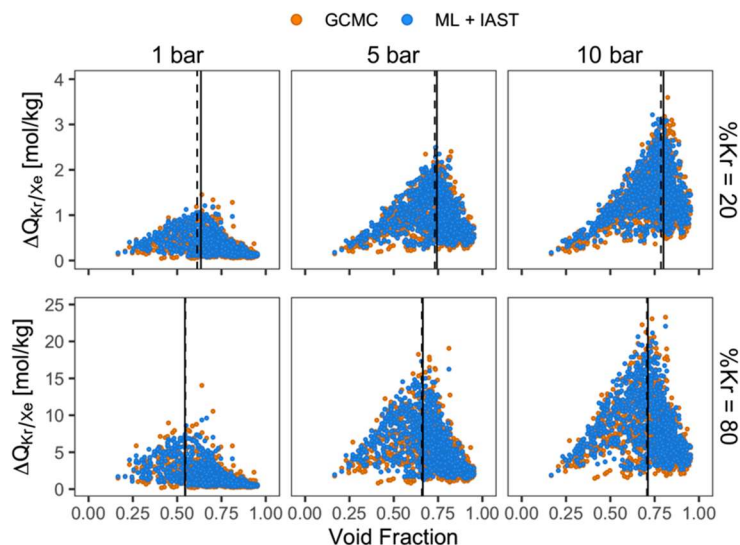
Figure 6. Plots comparing the relationship of $\Delta Q_{Kr/Xe}$ versus MOF void fraction as predicted by GCMC (orange) and our MLP+IAST model (blue). The mean VF of MOFs in the 95th percentile of $\Delta Q_{Kr/Xe}$ is shown by the solid line for GCMC $\Delta Q_{Kr/Xe}$, and by the dashed line for MLP+IAST $\Delta Q_{Kr/Xe}$.

dimension reduction algorithm of Scrucca (these directions account for the majority of the structure in the clustered data).[55] This algorithm finds a set of linear combinations of the features considered during clustering that results in maximal cluster separation. **Fig. 7** also shows the boundaries between the classification regions defined for each cluster, colored according to the boundary uncertainty (darker areas indicate higher uncertainty). For example, we can observe from **Fig. 7** that clusters *2* and *3* have some close points, and that two conditions lie on the boundary between clusters *2* and *3*. The condition 20/80 N_2/CH_4 at 5 bar is in cluster two (it is also one of the points at the boundary between clusters *2* and *3*) and the condition 80/20 Kr/Xe at 1 bar is in cluster *3*. This means that MOFs with high ΔQs for these separations are likely to be similar, a significant result as both these conditions are relevant to industrial separation conditions. In other words, this type of analysis can be used to identify privileged MOFs, i.e. MOFs that are high performing for multiple separations and/or multiple T, P combinations. An additional plot justifying this conclusion using an alternative method is shown in **Figs. S18**.

E. Using MOF thermodynamic feasibility calculations in HTCS. After identifying potentially privileged MOFs using the methods outlined in the previous section, we can proceed to use a free energy threshold to identify MOFs more likely to be synthesizable on account of having similar free energy per atom as other previously synthesized MOFs.[41] While the exact criterion we outlined in ref. [41] may not necessarily be transferable to all MOFs in general, it does apply to all the structures in our 51,520 MOF database as they are based on the same type of inorganic nodes. This single, global criterion, however, is not the complete picture, as if there is isomorphism (same nodes and linkers, but a different topology) observed for a particular MOF composition, the isomorph with the lowest, calculated free energy has around 80% likelihood (based on 32 isomorphic families in ref. [41]) of being the one observed experimentally.

Having calculated the ΔQs across our database using our MLP+IAST model for 80/20 Kr/Xe mixtures at 1 bar and 20/80 N_2/CH_4 mixtures at 5 bar, we can proceeded to identifying MOFs that are high-performing for both separations (recalling the results of the clustering analysis in the previous section), while simultaneously considering the thermodynamic feasibility of these high-performing MOFs. To accomplish this, we calculated the free energies of each structure in our database using the Frenkel-Ladd path method. Following findings from our previous work we considered MOFs with free energy below a 4 kJ/mol per atom threshold thermodynamically feasible.
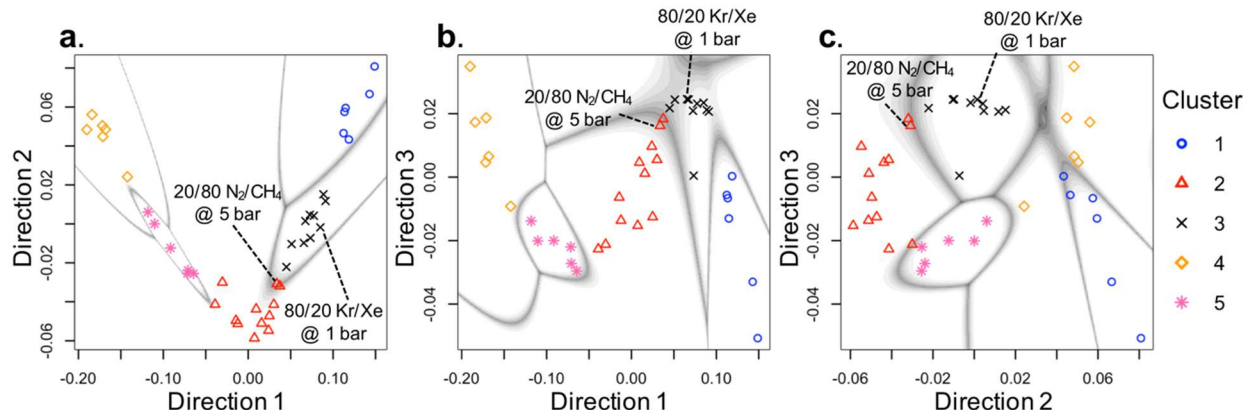
Figure 7. Clusters of conditions as clustered by their mean high-performing MOF properties. Each cluster is shown as a different shape and color. The clusters are projected onto directions found from a cluster-based dimensionality reduction algorithm that seeks to find linear combinations of the mean MOF properties that results in maximal cluster separation. Here the first three of four directions are shown in *a*, *b*, and *c*. These three directions account for the majority of the structure in the clustered data. Uncertainty boundaries between the clusters are shown in grayscale, the darker the area the higher uncertainty. The specific conditions 20/80 N_2/CH_4 at 5 bar and 80/20 Kr/Xe at 1 bar are directly labeled
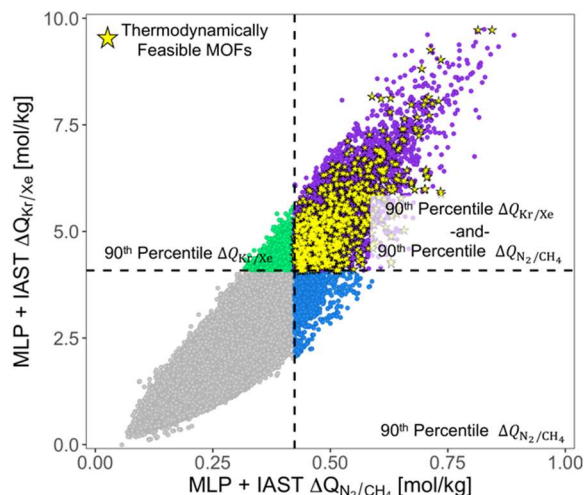

Figure 8. $\Delta Q_{Kr/Xe}$ for an 80/20 composition at 1 bar verses $\Delta Q_{N_2/CH_4}$ for an 20/80 composition at 5 bar, both values were predicted using our MLP+IAST model. Cutoffs for to be in the 90th percentile for both ΔQs are shown and thermodynamically accessible MOFs in both regions are highlighted as yellow stars.

Using the free energy criteria described above, in **Fig. 8** we show $\Delta Q_{Kr/Xe}$ versus $\Delta Q_{N_2/CH_4}$, with MOFs that are predicted to be simultaneously thermodynamically feasible and in the 90th percentile of both ΔQs, these MOFs correspond to the "star" points. We considered the 90th percentiles of our MLP+IAST predictions as our results from **Section 3.2** suggest that this would allow us to capture the vast

majority of the 95th percentile of MOFs that would be identified using GCMC. Using this data, we were able to identify several promising structures that were ranked highly for both separations and were thermodynamically accessible. We used the average of the Kr/Xe and N_2/CH_4 rankings to obtain a single ranking for both separations. Six of the thermodynamically feasible MOFs with high rankings for both separations are shown in **Fig. 9**. Reassuringly for the validity of our criterion, we were to connect these structures with previously synthesized MOFs.

The MOF in **Fig. 9a** (average rank = 165) of **edq** topology is similar to the existing MOF Co-ABTC synthesized by Duan et al (note that **edq** topology MOFs are often reported as having **soc** topology).[56] Various isoreticular versions of this **edq** MOF, with similarly sized linkers, also have high performance for both separations. The MOFs in **Figs. 9b** (average rank = 12) and **9c** (average rank = 58) are isomorphs that were determined to both be thermodynamically accessible (their free energies were with 1% of each other) and are isoreticular with the existing MOF-150 (**pyr**)[57] and MOF-177 (**qom**),[58] but with a contracted linker leading to smaller pores. Conveniently, both thermodynamically feasible isomorphs are ranked highly in this case. The MOFs in **Figs. 9d** (average rank = 12.5) and **9e** (average rank = 8.5) are also isomorphic and both considered thermodynamically feasible. The **jjt** structure is isoreticular with the existing MOF JUC-101,[59] but with a smaller linker. Finally, the MOF in **Fig. 9f** (average rank = 74) is an existing MOF previously synthesized by Chen et al.[60] We also found that the analogs of the existing MOF
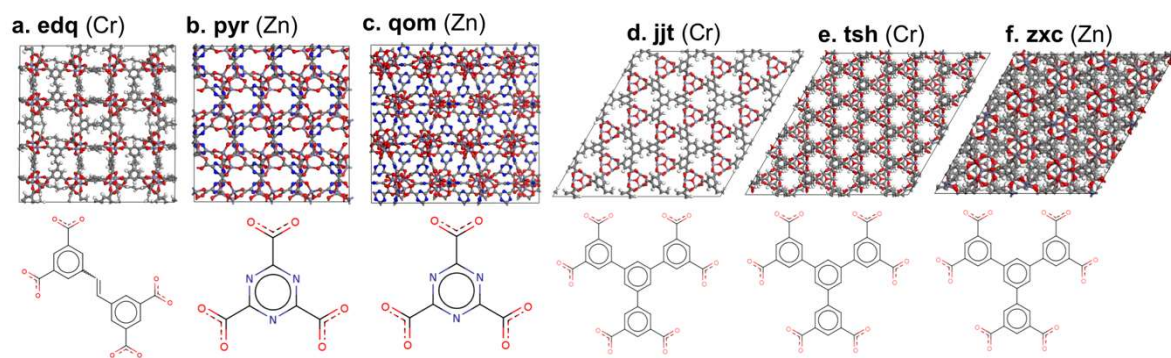
Figure 9. Six high-performing and thermodynamically feasible MOFs identified from our screening study. Topologies are shown in bold with the metal in parenthesis. The corresponding linker molecules (carboxylate form) are shown below each MOF.

NOTT-100[61] and UTSA-20[62] in our database have reasonably high performance with average ranks of 697.5 and 517.5, respectively (note that these are ranks out of over 50,000 MOFs).

IV. CONCLUSIONS

We have demonstrated the combination of Ideal Adsorbed Solution Theory (IAST) with multilayer perceptron (MLP) predictions of single component isotherms to predict binary adsorption data in MOFs. Given that our MLP model is transferable across multiple adsorbates, and IAST is inherently general, the MLP+IAST framework is transferable (without modification) to many binary gas mixtures and operating conditions. This allows for the rapid screening of MOFs for performance in a variety of fixed bed adsorption applications. We showed that our MLP+IAST framework can *i)* effectively capture the majority of high-performing MOFs (e.g., the 95th percentile) with calculations that are many orders of magnitude faster than GCMC simulations, *ii)* faithfully reproduce the same structure-performance relationships that would result from GCMC simulations, and *iii)* screen MOFs for many different chemical separations, enabling us to identify MOFs that are simultaneously high performing for multiple separations/conditions, i.e. privileged MOFs. We demonstrated that these privileged MOFs can be efficiently identified using clustering algorithms. Finally, by combining our MLP+IAST results with large-scale free energy calculations we identify structures that are both high-performing and likely to be synthesizable from a thermodynamic standpoint. Overall, the theoretical framework presented in this work illustrates a computational discovery "pipeline" where MOFs are built *in-silico*, their separations properties are rapidly predicted for conditions of interest (screening a diversity of conditions in this step would allow one to identify privileged MOFs), and promising synthetic targets are proposed based on their expected performance *and* on their synthetic likelihood.

V. SUPPLEMENTAL MATERIAL

See Supplemental Material to find more detailed information about computational methods and additional data validating our MLP+IAST model.

VII. DATA AVAILABILITY

The data that supports the findings of this contribution and all the MOF structures considered are available at the Open Science Framework (OSF), available at the time of writing here: https://doi.org/10.17605/OSF.IO/7DGVY
Any additional materials are available from the corresponding author upon reasonable request.

REFERENCES

(1) Sholl, D. S.; Lively, R. P. Seven Chemical Separations to Change the World. *Nature* **2016**, *532*, 435–437.

(2) Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; et al. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.*

2017, *29*, 2844–2854.

(3)   Mondloch, J. E.; Yildirim, T.; Hupp, J. T.; Gutov, O. V.; Farha, O. K.; Krungleviciute, V.; Snurr, R. Q.; Gomez-Gualdron, D. A.; Borah, B. Computational Design of Metal–Organic Frameworks Based on Stable Zirconium Building Units for Storage and Delivery of Methane. *Chem. Mater.* **2014**, *26*, 5632–5639.

(4)   Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y.-S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q. Evaluating Topologically Diverse Metal–Organic Frameworks for Cryo-Adsorbed Hydrogen Storage. *Energy Environ. Sci.* **2016**, *9*, 3279–3289.

(5)   Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional Hydrogen Storage Achieved by Screening Nearly Half a Million Metal-Organic Frameworks. *Nat. Commun.* **2019**, *10*, 1568.

(6)   Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal–Organic Frameworks. *Nat. Chem.* **2011**, *4*, 83–89.

(7)   Chen, Z.; Li, P.; Anderson, R.; Wang, X.; Zhang, X.; Robison, L.; Redfern, L. R.; Moribe, S.; Islamoglu, T.; Gómez-Gualdrón, D. A.; et al. Balancing Volumetric and Gravimetric Uptake in Highly Porous Materials for Clean Energy. *Science.* **2020**, *368*, 297 LP – 303.

(8)   Gómez-Gualdrón, D. A.; Wilmer, C. E.; Farha, O. K.; Hupp, J. T.; Snurr, R. Q. Exploring the Limits of Methane Storage and Delivery in Nanoporous Materials. *J. Phys. Chem. C* **2014**, *118*, 6941–6951.

(9)   Sikora, B. J.; Wilmer, C. E.; Greenfield, M. L.; Snurr, R. Q. Thermodynamic Analysis of Xe/Kr Selectivity in over 137 000 Hypothetical Metal–Organic Frameworks. *Chem. Sci.* **2012**, *3*, 2217–2223.

(10)   Banerjee, D.; Simon, C. M.; Plonka, A. M.; Motkuri, R. K.; Liu, J.; Chen, X.; Smit, B.; Parise, J. B.; Haranczyk, M.; Thallapally, P. K. Metal–Organic Framework with Optimally Selective Xenon Adsorption and Separation. *Nat. Commun.* **2016**, *7*, ncomms11831.

(11)   Chung, Y. G.; Bai, P.; Haranczyk, M.; Leperi, K. T.; Li, P.; Zhang, H.; Wang, T. C.; Duerinck, T.; You, F.; Hupp, J. T.; et al. Computational Screening of Nanoporous Materials for Hexane and Heptane Isomer Separation. *Chem. Mater.* **2017**, *29*, 6315–6328.

(12)   Li, S.; Chung, Y. G.; Snurr, R. Q. High-Throughput Screening of Metal–Organic Frameworks for CO2 Capture in the Presence of Water. *Langmuir* **2016**, *32*, 10368–10376.

(13)   Daglar, H.; Keskin, S. Computational Screening of Metal–Organic Frameworks for Membrane-Based CO2/N2/H2O Separations: Best Materials for Flue Gas Separation. *J. Phys. Chem. C* **2018**, *122*, 17347–17357.

(14)   Boyd, P. G.; Chidambaram, A.; García-Díez, E.; Ireland, C. P.; Daff, T. D.; Bounds, R.; Gładysiak, A.; Schouwink, P.; Moosavi, S. M.; Maroto-Valer, M. M.; et al. Data-Driven Design of Metal–Organic Frameworks for Wet Flue Gas CO2 Capture. *Nature* **2019**, *576*, 253–256.

(15)   Rogacka, J.; Seremak, A.; Luna-Triguero, A.; Formalik, F.; Matito-Martos, I.; Firlej, L.; Calero, S.; Kuchta, B. High-Throughput Screening of Metal – Organic Frameworks for CO2 and CH4 Separation in the Presence of Water. *Chem. Eng. J.* **2021**, *403*, 126392.

(16)   Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. What Are the Best Materials To Separate a Xenon/Krypton Mixture? *Chem. Mater.* **2015**, *27*, 4459–4475.

(17)   Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust Machine Learning Models for Predicting High CO2 Working Capacity and CO2/H2 Selectivity of Gas Adsorption in Metal Organic Frameworks for Precombustion Carbon Capture. *J. Phys. Chem. C* **2019**, *123*, 4133–4139.

(18)   Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P. G.; Woo, T. K. High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO2 Adsorption Properties of Metal–Organic Frameworks. *J. Phys. Chem. C* **2020**, *124*, 27996–28005.

(19)   Anderson, G.; Schweitzer, B.; Anderson, R.; Gómez-Gualdrón, D. A. Attainable Volumetric Targets for Adsorption-Based Hydrogen Storage in Porous Crystals: Molecular Simulation and Machine Learning. *J. Phys. Chem. C* **2019**, *123*, 120–130.

(20)   Ma, R.; Colón, Y. J.; Luo, T. Transfer Learning Study of Gas Adsorption in Metal–Organic Frameworks. *ACS Appl. Mater. Interfaces* **2020**, *12*, 34041–34048.

(21)   Gharagheizi, F.; Tang, D.; Sholl, D. S. Selecting Adsorbents to Separate Diverse Near-Azeotropic Chemicals. *J. Phys. Chem.*

C **2020**, *124*, 3664–3670.

(22) Sun, Y.; DeJaco, R. F.; Siepmann, J. I. Deep Neural Network Learning of Complex Binary Sorption Equilibria from Molecular Simulation Data. *Chem. Sci.* **2019**, *10*, 4377–4388.

(23) Anderson, R.; Biong, A.; Gómez-Gualdrón, D. A. Adsorption Isotherm Predictions for Multiple Molecules in MOFs Using the Same Deep Learning Model. *J. Chem. Theory Comput.* **2020**, *16*, 1271–1283.

(24) Cessford, N. F.; Seaton, N. A.; Düren, T. Evaluation of Ideal Adsorbed Solution Theory as a Tool for the Design of Metal–Organic Framework Materials. *Ind. Eng. Chem. Res.* **2012**, *51*, 4911–4921.

(25) Van Heest, T.; Teich-McGoldrick, S. L.; Greathouse, J. A.; Allendorf, M. D.; Sholl, D. S. Identification of Metal–Organic Framework Materials for Adsorption Separation of Rare Gases: Applicability of Ideal Adsorbed Solution Theory (IAST) and Effects of Inaccessible Framework Regions. *J. Phys. Chem. C* **2012**, *116*, 13183–13195.

(26) Gurdal, Y.; Keskin, S. Predicting Noble Gas Separation Performance of Metal Organic Frameworks Using Theoretical Correlations. *J. Phys. Chem. C* **2013**, *117*, 5229–5241.

(27) Furmaniak, S.; Koter, S.; Terzyk, A. P.; Gauden, P. A.; Kowalczyk, P.; Rychlicki, G. New Insights into the Ideal Adsorbed Solution Theory. *Phys. Chem. Chem. Phys.* **2015**, *17*, 7232–7247.

(28) Dickey, A. N.; Yazaydın, A. Ö.; Willis, R. R.; Snurr, R. Q. Screening CO2/N2 Selectivity in Metal-Organic Frameworks Using Monte Carlo Simulations and Ideal Adsorbed Solution Theory. *Can. J. Chem. Eng.* **2012**, *90*, 825–832.

(29) Krishna, R.; van Baten, J. M. Using Molecular Simulations for Screening of Zeolites for Separation of CO2/CH4 Mixtures. *Chem. Eng. J.* **2007**, *133*, 121–131.

(30) Kerry, F. G. *Industrial Gas Handbook: Gas Separation and Purification*; CRC Press, 2007.

(31) Liu, J.; Thallapally, P. K.; Strachan, D. Metal-Organic Frameworks for Removal of Xe and Kr from Nuclear Fuel Reprocessing Plants. *Langmuir* **2012**, *28*, 11584–11589.

(32) Anderson, R.; Schweitzer, B.; Wu, T.; Carreon, M. A.; Gómez-Gualdrón, D. A. Molecular Simulation Insights on Xe/Kr Separation in a Set of Nanoporous Crystalline Membranes. *ACS Appl. Mater. Interfaces* **2018**, *10*, 582–592.

(33) Hugman, R. H.; Vidas, E. H.; Springer, P. S. Chemical Composition of Discovered and Undiscovered Natural Gas in the Lower-48 United States. Project Summary. Final Report, 1 November 1988-31 March 1990. November 1, 1990.

(34) Hu, J.; Sun, T.; Liu, X.; Guo, Y.; Wang, S. Separation of CH4/N2 Mixtures in Metal–Organic Frameworks with 1D Micro-Channels. *RSC Adv.* **2016**, *6*, 64039–64046.

(35) Jia, J.; Wang, L.; Sun, F.; Jing, X.; Bian, Z.; Gao, L.; Krishna, R.; Zhu, G. The Adsorption and Simulated Separation of Light Hydrocarbons in Isoreticular Metal–Organic Frameworks Based on Dendritic Ligands with Different Aliphatic Side Chains. *Chem. – A Eur. J.* **2014**, *20*, 9073–9080.

(36) Ren, T.; Patel, M.; Blok, K. Olefins from Conventional and Heavy Feedstocks: Energy Use in Steam Cracking and Alternative Processes. *Energy* **2006**, *31*, 425–451.

(37) Dong, X.-Z.; Ritterbusch, F.; Chu, Y.-Q.; Gu, J.-Q.; Hu, S.-M.; Jiang, W.; Lu, Z.-T.; Yang, G.-M.; Zhao, L. Dual Separation of Krypton and Argon from Environmental Samples for Radioisotope Dating. *Anal. Chem.* **2019**, *91*, 13576–13581.

(38) Péron, S.; Mukhopadhyay, S.; Huh, M. A New Dual Stainless Steel Cryogenic Trap for Efficient Separation of Krypton from Argon and Xenon. *J. Anal. At. Spectrom.* **2020**, *35*, 2663–2671.

(39) Krishna, R. Screening Metal–Organic Frameworks for Mixture Separations in Fixed-Bed Adsorbers Using a Combined Selectivity/Capacity Metric. *RSC Adv.* **2017**, *7*, 35724–35737.

(40) Tang, D.; Wu, Y.; Verploegh, R. J.; Sholl, D. S. Efficiently Exploring Adsorption Space to Identify Privileged Adsorbents for Chemical Separations of a Diverse Set of Molecules. *ChemSusChem* **2018**, *11*, 1567–1575.

(41) Anderson, R.; Gómez-gualdrón, D. A. Large-Scale Free Energy Calculations on a Computational MOF Database : Toward Synthetic Likelihood Predictions. 1–16.

(42) Anderson, R.; Gómez-Gualdrón, D. A. Increasing Topological Diversity during Computational "Synthesis" of Porous Crystals: How and Why. *CrystEngComm* **2019**, *21*, 1653–1665.

(43) Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. *Cryst. Growth Des.* **2017**, *17*,

14

5801–5810.

(44) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

(45) Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. Extension of the Universal Force Field to Metal–Organic Frameworks. *J. Chem. Theory Comput.* **2014**, *10*, 880–891.

(46) Coupry, D. E.; Addicoat, M. A.; Heine, T. Extension of the Universal Force Field for Metal–Organic Frameworks. *J. Chem. Theory Comput.* **2016**, *12*, 5215–5225.

(47) Bitzek, E.; Koskinen, P.; Gähler, F.; Moseler, M.; Gumbsch, P. Structural Relaxation Made Simple. *Phys. Rev. Lett.* **2006**, *97*, 170201.

(48) Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials. *Mol. Simul.* **2016**, *42*, 81–101.

(49) Mayo, S. L.; Olafson, B. D.; Goddard, W. A. DREIDING: A Generic Force Field for Molecular Simulations. *J. Phys. Chem.* **1990**, *94*, 8897–8909.

(50) Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

(51) Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134–141.

(52) Simon, C. M.; Smit, B.; Haranczyk, M. PyIAST: Ideal Adsorbed Solution Theory (IAST) Python Package. *Comput. Phys. Commun.* **2016**, *200*, 364–380.

(53) Scrucca, L.; Fop, M.; Murphy, T. B.; Raftery, A. E. Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *R J.* **2016**, *8*, 289–317.

(54) Frenkel, Daan; Smith, B. *Understanding Molecular Simulations From Algorithms to Applications*; Academic Press, 2002.

(55) Scrucca, L. Dimension Reduction for Model-Based Clustering. *Stat. Comput.* **2010**, *20*, 471–484.

(56) Duan, X.; Zheng, W.; Yu, B.; Ji, Z. A Microporous Metal–Organic Framework with Soc Topology for Adsorption and Separation Selectivity of C2H2/CO2. *Chem. Pap.* **2019**, *73*, 2371–2375.

(57) Chae, H. K.; Kim, J.; Friedrichs, O. D.; O'Keeffe, M.; Yaghi, O. M. Design of Frameworks with Mixed Triangular and Octahedral Building Blocks Exemplified by the Structure of [Zn4O(TCA)2] Having the Pyrite Topology. *Angew. Chemie Int. Ed.* **2003**, *42*, 3907–3909.

(58) Furukawa, H.; Ko, N.; Go, Y. B.; Aratani, N.; Choi, S. B.; Choi, E.; Yazaydin, A. Ö.; Snurr, R. Q.; O'Keeffe, M.; Kim, J.; et al. Ultrahigh Porosity in Metal-Organic Frameworks. *Science.* **2010**, *329*, 424 LP – 428.

(59) Jia, J.; Sun, F.; Borjigin, T.; Ren, H.; Zhang, T.; Bian, Z.; Gao, L.; Zhu, G. Highly Porous and Robust Ionic MOFs with Nia Topology Constructed by Connecting an Octahedral Ligand and a Trigonal Prismatic Metal Cluster. *Chem. Commun.* **2012**, *48*, 6010–6012.

(60) Chen, Z.; Xiang, S.; Liao, T.; Yang, Y.; Chen, Y.-S.; Zhou, Y.; Zhao, D.; Chen, B. A New Multidentate Hexacarboxylic Acid for the Construction of Porous Metal−Organic Frameworks of Diverse Structures and Porosities. *Cryst. Growth Des.* **2010**, *10*, 2775–2779.

(61) Lin, X.; Telepeni, I.; Blake, A. J.; Dailly, A.; Brown, C. M.; Simmons, J. M.; Zoppi, M.; Walker, G. S.; Thomas, K. M.; Mays, T. J.; et al. High Capacity Hydrogen Adsorption in Cu(II) Tetracarboxylate Framework Materials: The Role of Pore Size, Ligand Functionalization, and Exposed Metal Sites. *J. Am. Chem. Soc.* **2009**, *131*, 2159–2171.

(62) Guo, Z.; Wu, H.; Srinivas, G.; Zhou, Y.; Xiang, S.; Chen, Z.; Yang, Y.; Zhou, W.; O'Keeffe, M.; Chen, B. A Metal–Organic Framework with Optimized Open Metal Sites and Pore Spaces for High Methane Storage at Room Temperature. *Angew. Chemie Int. Ed.* **2011**, *50*, 3178–3181.