

# Quantitative Structure–Property Relationship Analysis for the Prediction of Propylene Adsorption Capacity in Pure Silicon Zeolites at Various Pressure Levels

Li Zhao, Qi Zhang, Chang He, Qinglin Chen, and Bing J. Zhang\*

Cite This: *ACS Omega* 2022, 7, 33895–33907

Read Online

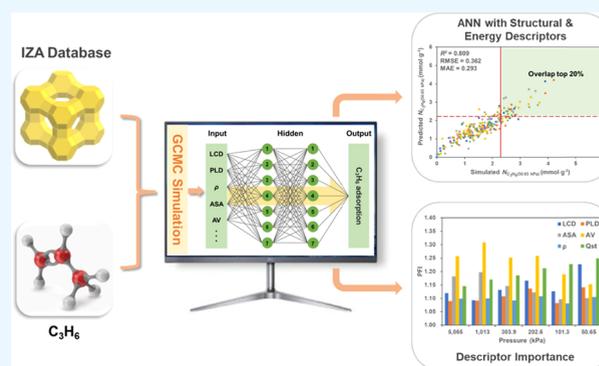
ACCESS |

Metrics &amp; More

Article Recommendations

Supporting Information

**ABSTRACT:** This work is devoted to the development of quantitative structure–property relationship (QSPR) models using various regression analyses to predict propylene ( $C_3H_6$ ) adsorption capacity at various pressures in zeolites from a topologically diverse International Zeolite Association database. Based on univariate and multilinear regression analysis, the accessible volume and largest cavity diameter are the most crucial factors determining  $C_3H_6$  uptake at high and low pressures, respectively. An artificial neural network (ANN) model with five structural descriptors is sufficient to predict  $C_3H_6$  uptake at high pressures. For combined pressures, the prediction of an ANN model with pore size distribution is pleasing. The isosteric heat of adsorption ( $Q_{st}$ ) has a significant impact on the improvement of the prediction of low-pressure gas adsorption, which finely classifies zeolites into high or low  $C_3H_6$  adsorbers. The conjunction of high-throughput screening and QSPR models contributes to being able to prescreen the database rapidly and accurately for top performers and perform further detailed and time-consuming computational-intensive molecular simulations on these candidates for other gas adsorption applications.



## 1. INTRODUCTION

Nanoporous materials are defined as materials with interpenetrating channels and pore sizes less than 100 nm which is often comparable to the size of a molecule.<sup>1</sup> Nanoporous materials composed of countless molecular building blocks in their synthesis, such as zeolites, porous carbons, metal–organic frameworks (MOFs), zeolitic imidazolate frameworks (ZIFs), and covalent organic frameworks (COFs), exhibit superior chemical and geometrical tunability, a diverse range of surface areas, pore surfaces, and void fractions, which perceive them as the next-generation technology.<sup>2</sup> A wide range of properties have successfully promoted diverse applications of nanoporous materials including, but not limited to gas storage, separation, catalysis, drug delivery, and sensing.<sup>3</sup> In the realm of nanoporous materials, it is always expected that an optimal material is tailored to a specific application. In recent years, an exponential increase in published lab-synthesized and computer-generated hypothetical nanoporous materials has provided us with a library of tens of thousands of potentially interesting new materials.<sup>4</sup> These new materials provide an ideal platform for understanding thoroughly how to tailor-make an optimal material for a given application.

High-throughput computational screening techniques using brute-force experimental and simulated methods could generate the thermodynamic data needed to predict the

performance of these materials for specific applications.<sup>5,6</sup> It is impossible to synthesize and characterize materials sequentially for identifying promising candidates by traditional extensive trial and error experimental methods with the advent and development of large databases due to time-consuming and expensive characteristics. To accelerate high-throughput screening, several outstanding theoretical computing tools have been widely adopted for characterization of novel materials over the last few years, for instance, first principles (ab initio) methods, density functional theory, and molecular simulations (Monte Carlo and molecular dynamics simulations).<sup>7</sup> In particular, grand canonical Monte Carlo (GCMC) simulation is an excellent example for adsorption studies among them, and its adsorption capacities are in good agreement with experimental results for many systems, which have been confirmed in prior work.<sup>8</sup> So far, the GCMC simulation methods have been extensively applied to methane storage,<sup>9</sup> hydrogen storage,<sup>10</sup> carbon dioxide ( $CO_2$ ) capture,<sup>11</sup>

Received: May 5, 2022

Accepted: August 31, 2022

Published: September 14, 2022



ethanol purification,<sup>12</sup> propylene/propane (C<sub>3</sub>H<sub>6</sub>/C<sub>3</sub>H<sub>8</sub>) separation,<sup>13</sup> and other aspects<sup>14–16</sup> by many groups. Additionally, preliminary structure–property relationships have been revealed. Undoubtedly, the time, cost, and human effort required for the characterization of material properties using GCMC simulation are greatly reduced compared to experimental methods. Although high-throughput screening studies based on GCMC simulation have been helpful in guiding experimental synthesis, this brute-force approach is limited to an almost unlimited number of new structures due to the required expensive computational costs. At present, the lack of efficient computational tools has gradually become a bottleneck in the rapid development of novel materials; consequently, the development of alternative screening methodologies is urgent.<sup>2</sup>

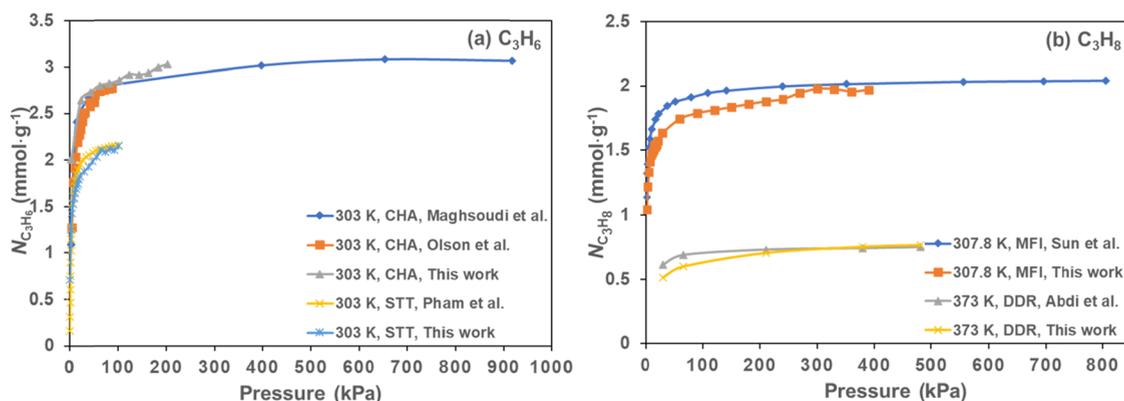
Machine learning methods such as decision tree, support vector machine, and neural networks have become a powerful tool to prescreen high-performing materials and accelerate large-scale simulation in the material field. It saves a lot of time to perform further detailed and time-consuming calculations only on candidates prescreened by machine learning. Explicitly, machine learning methods devise complex models to produce reliable predictions about unknown data through learning from relationships in the dataset provided, which makes the screening of countless nanoporous materials practicable. Quantitative structure–property relationship (QSPR) models<sup>17,18</sup> trained by data-driven machine learning methods can systematically correlate structural features of materials (referred to as descriptors) to their functional properties in quantitative terms, which would be expected to play a crucial role in material screening. For nanoporous materials, one of the most desirable properties to predict is the adsorption capacity of guest molecules at the required temperature and pressure.<sup>19</sup> Naturally, the morphology of the pores described by various descriptors is essential for the adsorption behavior of guest molecules, where adsorbates are located and interact with the material surface.<sup>20</sup> The selection of highly predictive descriptors for determining adsorption capacity is prominent.

Standard structural descriptors for pore morphology, such as mass density, surface area, void fraction, largest cavity diameter, and others, have been used to construct feature vectors for systems under certain thermodynamic conditions and provided satisfactory results. For instance, Durá et al.<sup>21</sup> innovatively made a reasonable fitting for CO<sub>2</sub> adsorption capacity in porous carbons through a simple regression approach derived from microporous and mesoporous volumes. Subsequently, artificial neural network (ANN) methods using microporous and mesoporous volumes and Brunauer–Emmett–Teller surface area were used to predict CO<sub>2</sub> uptake,<sup>22</sup> nitrogen (N<sub>2</sub>) uptake under ambient conditions, and CO<sub>2</sub>/N<sub>2</sub> selectivity<sup>23,24</sup> of porous carbons. Fernandez et al.<sup>25</sup> accurately predicted methane (CH<sub>4</sub>) uptake of MOFs at 100 bar based only on the dominant pore diameter, the maximum pore diameter, the void fraction, the gravimetric surface area, the volumetric surface area, and the framework density. Furthermore, results for N<sub>2</sub>,<sup>26</sup> CO<sub>2</sub> working capacities, and the CO<sub>2</sub>/CH<sub>4</sub> selectivity<sup>27</sup> of MOFs were also studied using machine learning algorithms and standard structural descriptors. Similarly, simple textural descriptors were regarded as the MOF fingerprints to predict hydrogen (H<sub>2</sub>) adsorption uptake and CO<sub>2</sub>/H<sub>2</sub> selectivity.<sup>28</sup> Recently, Lin et al.<sup>29</sup> systematically screened hypothetical pure-silica zeolites and identified 230 pre-eminent zeolites for effective removal of

linear siloxanes and derivatives using a random forest method based on simple structural descriptors. Analogously, high-performing zeolites for anion removal from water were identified.<sup>30</sup> Guest molecules occupy almost the entire void space of nanoporous materials at high pressures, so these descriptors capturing global porosity characteristics are often popular. Although the prediction of high-pressure gas adsorption performance is encouraging, there is no clear principle guiding the selection of appropriate descriptors, especially for low-pressure gas adsorption with poor predictive performance.

At low pressures, guest molecules are usually adsorbed in the strongly binding regions of the material's pore, which cannot be captured well by simple structural descriptors.<sup>31</sup> To address this problem, some specific descriptors have been continuously developed by researchers for obtaining a universal model for adsorption behavior prediction at different pressures. A novel atomic property weighted radial distribution function descriptor accounting for the topological diversity was tailored by Fernandez et al.<sup>26,25,32</sup> and combined with traditional structural descriptors for the prediction of CH<sub>4</sub>, CO<sub>2</sub>, and H<sub>2</sub> uptakes at low pressures with pleasing results. Lee and co-workers<sup>2,33,34</sup> developed a new descriptor for nanoporous materials by using topological data analysis to quantify similarity of pore structures and successfully predicted CH<sub>4</sub> uptake at low pressures. A vectorized persistence diagram for topology analysis was also expected to be applied to the screening of various materials.<sup>19,35</sup> Although detailed pore structure information is captured by a topological descriptor, it is incapable of reflecting the relative proportion of pores with a special pore size, let alone the guest–host interaction. Accordingly, the Voronoi energy descriptor<sup>36</sup> takes into account both geometrical structural information and the energetics information, to be highly predictive of xenon/krypton (Xe/Kr) separation performance. Later, a histogram of the guest–host energy was regarded as the feature for machine learning, and the predicted gas adsorption capacities in MOFs were in good agreement with GCMC simulations.<sup>8,10</sup> Fanourgakis et al.<sup>37</sup> proposed to treat the probabilities of a set of different probe atoms adsorbed by materials as new descriptors for fast screening of large databases. Recently, heat of adsorption and Henry's coefficient of special adsorbates were combined with traditional structural descriptors for gas adsorption and separation study.<sup>38,39</sup> For nanoporous materials with a wide variety of elements, especially MOFs, chemical descriptors considering types and contents of chemical elements were used to predict gas adsorption performance at low pressures.<sup>40–42</sup> In spite of the continuous generation of new descriptors, the selection of appropriate descriptors for a particular application remains an open scientific issue.

As an essential component of various household plastic products, C<sub>3</sub>H<sub>6</sub> is obtained by an energy-intensive cryogenic distillation process.<sup>43</sup> The development of alternative separation technologies with low energy consumption is of great value. A physical adsorption process, especially pressure-swing adsorption (PSA) technology with high gas purity, is a promising choice.<sup>44</sup> In the PSA process, guest molecules are adsorbed at high pressures and desorbed at low pressures. The selection of high-performance adsorbents is the key to achieving an efficient separation process. Various porous materials have been reported for the PSA process so far, such as zeolites,<sup>45,46</sup> MOFs,<sup>47</sup> and ZIFs.<sup>48</sup> Considering the uniform system of pores, high porosity, and excellent thermal



**Figure 1.** Comparison between simulated and open published adsorption isotherms of (a) pure  $C_3H_6$  and (b) pure  $C_3H_8$  in zeolites.

and chemical stability, zeolites have been proven to be promising for gas adsorption.<sup>49</sup>

In this work, we present a comprehensive QSPR analysis of the database of 232 zeolites. Correlations between various descriptors and  $C_3H_6$  adsorption capacity ( $N_{C_3H_6}$ ) at pressures of 5,065, 1,013, 303.9, 202.6, 101.3, and 50.65 kPa, at 300 K, were determined using multilinear regression analysis, quadratic regression analysis, and ANN models. In addition, pore size distribution (PSD) and isosteric heat of adsorption ( $Q_{st}$ ) were introduced to predict  $N_{C_3H_6}$  at low pressures, and the prediction performance was further evaluated by receiver-operator-curve (ROC) analysis. These QSPR models allowed for the accurate identification of high-performing zeolites, and rapid material prescreening significantly reduced the number of computational-intensive GCMC simulations. Finally, we described the relative importance of various descriptors determining  $N_{C_3H_6}$  at different pressures, which provided insights into the understanding of structural performance relationships at the atomic level.

## 2. MODELS AND METHODS

**2.1. Molecular Models.** First, 232 ordered pure silicon zeolites (Si/O = 1:2) considered in this study were obtained from the International Zeolite Association (IZA) database. The zeolite framework types were generated using a library of 49 composite building units. The framework atoms were described by Lennard-Jones (LJ) and electrostatic potentials.<sup>50,51</sup>

$$U_{LJ+elec}(r) = \sum 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (1)$$

where  $\epsilon_{ij}$  and  $\sigma_{ij}$  are the well depth and collision diameter, respectively,  $r_{ij}$  is the distance between atoms  $i$  and  $j$ ,  $q_i$  is the charge of atom  $i$ , and  $\epsilon_0 = 8.8542 \times 10^{-12} \text{ C}^2 \text{ N}^{-1} \text{ m}^{-2}$  is the permittivity of vacuum. The adsorbate  $C_3H_6$  was represented by a united-atom model with  $CH_x$  as a single interaction site. The LJ potential parameters and atomic charges of zeolites and  $C_3H_6$  were adopted from the COMPASS force field, which fairly well predicted gas adsorption in a wide variety of zeolites.<sup>52,53</sup> The Lorentz-Berthelot combining rules were employed to calculate the cross LJ parameters.

**2.2. Simulation Methods.** Before adsorption simulation,  $C_3H_6$  molecules and zeolite frameworks were geometrically optimized to obtain the configurations with stable structures,

and the optimized structures with minimum energy were used in the subsequent adsorption simulation process. GCMC<sup>54</sup> simulation in the sorption module of Materials Studio 2018<sup>55</sup> was conducted to evaluate the adsorption performance of 232 zeolites toward  $C_3H_6$ . GCMC is a statistical mechanical approach, in which the adsorption process is explored depending on random sampling and probabilistic interpretation in the adsorbent framework. The adsorption was assumed to be conducted at 300 K and 6 different pressure levels (5,065, 1,013, 303.9, 202.6, 101.3, and 50.65 kPa). During simulation, the  $C_3H_6$  molecule was considered as an ideal gas with negligible interactions, whose fugacity was equal to pressure.<sup>56</sup> Zeolite atoms were assumed to be rigid, and their positions remained constant. A spherical cut-off of 15.5 Å was used to calculate the LJ interactions, whereas the electrostatic interactions were calculated using the Ewald summation method. The cell lengths of each zeolite were expanded to at least 31 Å (twice the cut-off distance) along all three dimensions, and the periodic boundary conditions were exerted. In each zeolite, the GCMC simulation was run for  $1.1 \times 10^6$  cycles with  $1 \times 10^5$  for equilibration and the remaining for ensemble average. Each cycle consisted of  $n$  trial moves ( $n$ : the number of adsorbate molecules), including translation, rotation, regrowth, and swap. To verify the suitability of the COMPASS force field and the above assumptions used in this study, Figure 1 shows the adsorption isotherms of pure  $C_3H_6$  in CHA and STT and  $C_3H_8$  in MFI and DDR, respectively. Good agreement is observed between the simulation and open published data,<sup>46,57–60</sup> which suggests the reliability of the force field selected. Besides GCMC simulation for the adsorption of pure  $C_3H_6$ , the  $Q_{st}$  of  $C_3H_6$  at infinite dilution was estimated. For this case, one adsorbate molecule ( $C_3H_6$ ) was added into a zeolite and simulation was conducted in a canonical ensemble.

**2.3. Descriptor Selection.** In this work, in addition to  $C_3H_6$  adsorption data simulated, five general 1D structural descriptors including the largest cavity diameter (LCD), pore-limiting diameter (PLD), accessible surface area (ASA), accessible volume (AV), density ( $\rho$ ) and a 2D structural descriptor of PSD, and an energy descriptor of  $Q_{st}$  were selected in our QSPR analysis. As described in Figure S1, based on principles of moderate correlation between each descriptor and  $N_{C_3H_6}$  and no strong correlation among descriptors, we selected the above five general 1D structure descriptors from the initial seven descriptors to quantitatively describe the structure of zeolites.

The LCD corresponds to the maximum of the PSD, and the PLD refers to the largest characteristic guest molecule size for which there is a nonzero AV. LCD and PLD determine whether a specific guest molecule can enter the zeolite window; furthermore, ASA and AV reflect the void space that guest molecules can reach. All void space of a zeolite is reflected by  $\rho$  indirectly. The PSD provides information about the fraction of void space that is occupied by pores of a certain size, and the  $Q_{st}$  value reflects the energy information related to the adsorption process. These diverse descriptors with strong structure–performance relationships between guest molecules and zeolites reveal the features of zeolites from various aspects, which could be applied in accuracy prediction of machine learning.<sup>17,27,38,61</sup> These descriptors are relatively easy to measure, and they can be used directly to guide the synthesis and application of zeolites. In this work,  $\rho$  was obtained directly through zeolite crystalline structure, and LCD, PLD, ASA, AV, and PSD were determined by the Zeo++ program, in which the radius of a probe (1.2 Å)<sup>29</sup> was used for ASA and AV, and a bin size of 0.1 Å<sup>62</sup> was used to obtain PSD histograms. The  $Q_{st}$  value was calculated by the NVT-Monte Carlo simulation.

**2.4. Multilinear and Quadratic Regression Models.** Multilinear regression analysis is performed when the relationship between multiple descriptors and  $N_{C_3H_6}$  of 232 zeolites is assumed to be linear. The general form of the multilinear regression model is as follows:

$$y_k = \sum_i \beta_i x_{ik} + \beta_0 + \gamma_k \quad (2)$$

The quadratic regression is performed through adding binary interaction terms to the multilinear regression model. The general form of the quadratic regression model is as follows:

$$y_k = \sum_{i,j} \beta_{ij} x_{ik} x_{jk} + \sum_i \beta_i x_{ik} + \beta_0 + \gamma_k \quad (3)$$

where  $y_k$  and  $x_{ik}$  refer to the target value and input value of sample  $k$ ;  $\beta_{ij}$  and  $\beta_i$  are binary and linear coefficients, respectively;  $\beta_0$  and  $\gamma_k$  are the constant and error term, respectively.

**2.5. Neural Network Models.** To clarify the role of structural descriptors and an energy descriptor on adsorption capacity, the above descriptors were chosen as neurons, imported to the input layer, and passed in an orderly manner into the hidden layers and output layer. The information obtained from the ANN model was finally stored and transferred via a feed-forward process to predict  $N_{C_3H_6}$  of 232 zeolites. As a typical machine learning algorithm, the ANN model was trained countless times by comparing the simulated and calculated output values and then adjusting the weights and thresholds to decrease the error, where the mean squared error was used as the cost function. The optimization of the cost function was carefully monitored to determine the optimal number of epochs so that overfitting to the training data was avoided. Figure 2 exhibits the architecture of the ANN model used in this paper and the ANN model was performed on a MATLAB R2020a platform. ANN models with different descriptors as nodes, two hidden layers, and seven nodes for each layer were built to predict the GCMC-simulated  $N_{C_3H_6}$  of zeolites.

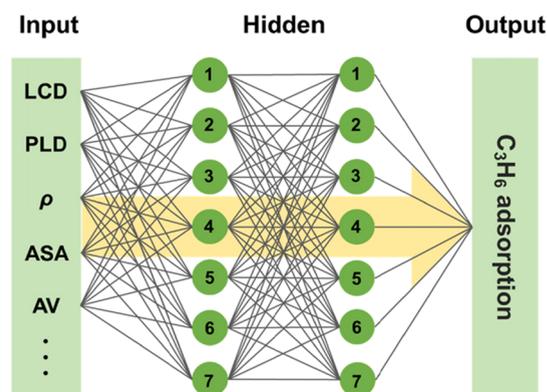


Figure 2. Architecture of an ANN model.

**2.6. Performance Criteria.** During the training process, the data sets were primary randomly divided into two parts, 80% of which was used for training, and the remaining 20% was used to test the generalization ability of the model. Moreover, to reduce input data set partition uncertainties and minimize overfitting issues, a fivefold cross validation (CV) approach was also employed for the ANN model, and the average of the results of the five calculations was taken as the model performance. All input and output data sets were preprocessed by a normalization method to speed up the training process. The quality of the training and test results was evaluated by the determinate coefficient ( $R^2$ ), the root mean square error (RMSE), and the mean absolute error (MAE) as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{i=n} (y_i - \bar{y})^2} \quad (4)$$

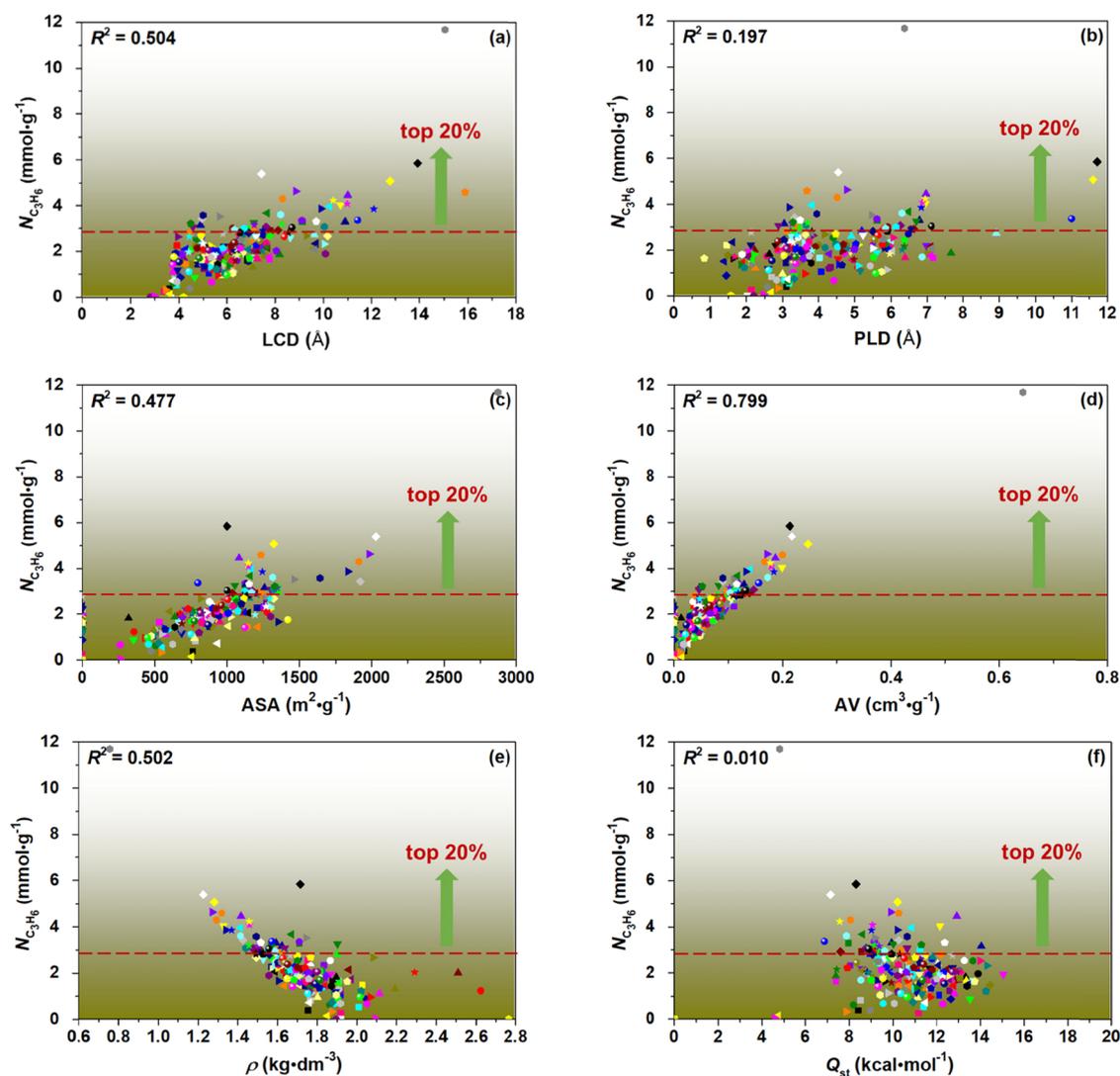
$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (5)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (6)$$

where  $n$ ,  $y_i$ ,  $\hat{y}_i$ , and  $\bar{y}$  refer to the number of samples, target value, predicted value, and average target value, respectively.

### 3. RESULTS AND DISCUSSION

**3.1. Univariate Analysis.** A comprehensive understanding of the relationship between the structural and energy descriptors and performances of the zeolites used for  $C_3H_6$  adsorption would be conducive to the identification of potential materials. We initially performed a simple univariate analysis where we looked for correlations between a single descriptor and the simulated  $N_{C_3H_6}$  at 300 K, 5,065 and 50.65 kPa, in which the top 20% of the data was classified as zeolites with high adsorption capacity, and the remaining 80% as zeolites with low adsorption capacity. Figures 3a–d, and S2a–d show that four structural descriptors (LCD, PLD, ASA, and AV) are positively corrected with  $N_{C_3H_6}$ . When LCD is less than 3.75 Å, adsorption of guest molecules is impeded due to unfavorable potential overlap with the framework, and  $N_{C_3H_6}$  could be considered to be almost zero. With the gradual increase of LCD, the large void space makes the adsorption capacity increase overall at 5,065 and 50.65 kPa. ASA and AV



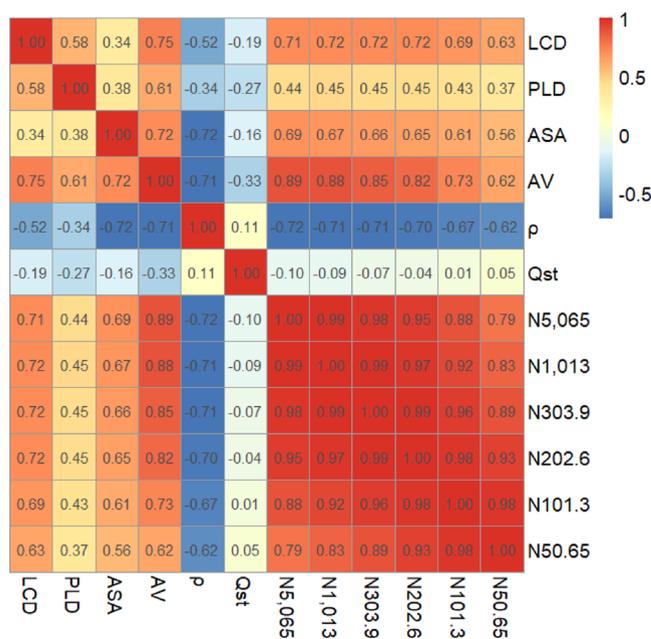
**Figure 3.** Relationships between (a)  $N_{C_3H_6} \sim LCD$ , (b)  $N_{C_3H_6} \sim PLD$ , (c)  $N_{C_3H_6} \sim ASA$ , (d)  $N_{C_3H_6} \sim AV$ , (e)  $N_{C_3H_6} \sim \rho$ , and (f)  $N_{C_3H_6} \sim Q_{st}$  at 300 K and 5,065 kPa.

are also strongly correlated with  $N_{C_3H_6}$  at 5,065 and 50.65 kPa. The relationship between  $N_{C_3H_6}$  and  $\rho$  also shows a similar linear trend in Figures 3e and S2e, but  $N_{C_3H_6}$  gradually decreases as  $\rho$  increases. There is no evident trend in the relationship between  $N_{C_3H_6}$  and  $Q_{st}$ , which shows that  $Q_{st}$  cannot be a good linear explanation of  $N_{C_3H_6}$ . The interpretation of  $N_{C_3H_6}$  by a single descriptor at 50.65 kPa is all reduced compared to that of 5,065 kPa even though similar linear trends are observed. From the perspective of univariate analysis, the correlation of simple structure descriptors is  $AV > LCD > \rho > ASA > PLD$  at 5,065 kPa (high pressure), which is different from that at 50.65 kPa (low pressure) with  $LCD > AV > \rho > ASA > PLD$ . It can also be observed that  $N_{C_3H_6}$  is not uniquely determined by a single descriptor. Overall, a single structural or energy descriptor can only determine the individual performance relationship and cannot explain synergies between various descriptors that might contribute significantly to the performance of zeolites.

A heat map of the correlation coefficient matrix for various descriptors and  $N_{C_3H_6}$  at various pressures is shown in Figure 4.

Upon observing the correlations among the set of descriptors, a strong positive correlation is observed between LCD and AV, with the Pearson correlation coefficient value of  $r$  equal to 0.75; moreover, a strong negative correlation exists between  $\rho$  and ASA or AV, with an absolute value of  $r$  greater than 0.71. However, no significant correlation is observed between  $Q_{st}$  and structural descriptors. For  $N_{C_3H_6}$ , three structural descriptors (AV, LCD, and  $\rho$ ) are strongly related to it with the absolute value of  $r$  greater than 0.62, and two structural descriptors (ASA and PLD) have moderate correlation with  $r$  greater than 0.37. Nevertheless, there is a little correlation between the energy descriptor ( $Q_{st}$ ) and  $N_{C_3H_6}$  with an absolute value of  $r$  lower than 0.10. It can also be clearly observed from Figure 4 that with the decrease of adsorption pressure, the contribution of a single descriptor to  $N_{C_3H_6}$  gradually decreases.

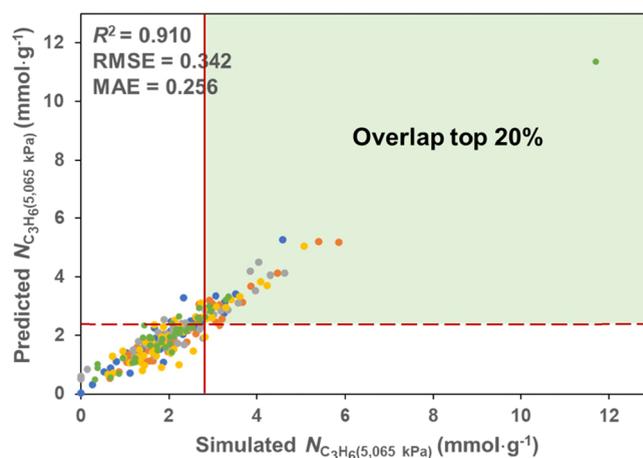
Due to the limitations of univariate analysis in identifying synergies, five structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) with strong linear correlation with  $N_{C_3H_6}$  in univariate analysis were used for multilinear regression and quadratic regression analysis. The predicted results of the two models for



**Figure 4.** Pearson correlation matrix for all the set of descriptors and  $N_{C_3H_6}$  at various pressures (the color bars represent the size of the Pearson correlation coefficients).

$N_{C_3H_6}$  at different pressures are included in the [Supporting Information](#). Although the prediction accuracy of the two prediction models is not excellent, these two models confirm that AV plays the most critical role in  $N_{C_3H_6}$  at 5,065 kPa. Furthermore, multilinear regression analysis further confirms that LCD is the most important descriptor for determining  $N_{C_3H_6}$  at 50.65 kPa.

**3.2. ANN Models.** **3.2.1. Five Structural Descriptors.** Machine learning has been widely used to predict the adsorption performance of materials. In this section, ANN models with five input nodes (LCD, PLD, ASA, AV, and  $\rho$ ), two hidden layers, and seven nodes for each layer were built to predict the GCMC-simulated  $N_{C_3H_6}$  of zeolites. During the training process, both random validation and fivefold CV were implemented, and the results of random validation are shown in the [Supporting Information](#). At 300 K and 5,065 kPa, the predicted outcomes of the ANN model are in good agreement with the GCMC-simulated  $N_{C_3H_6}$  with  $R^2 = 0.910$  for fivefold CV, as shown in [Figure 5](#), indicating that these structural descriptors correlate well with  $N_{C_3H_6}$  at high adsorption pressures. The explanation for this satisfactory situation is that adsorption is mainly a molecular stacking mechanism at high pressures, and these five descriptors describing the global porosity characteristics of the zeolites could fully determine  $N_{C_3H_6}$  of zeolites. To examine this more closely, we select the actual top 20% (46) zeolites in the database as determined by simulations, which are the subset to the right of the solid red line. Likewise, the top 20% zeolites as predicted by the ANN model are also picked up, which are the zeolites above the dashed red line in [Figure 5](#). The intersection part produces top performing zeolites that are recovered by the model. For getting the real top 20% zeolites with excellent  $N_{C_3H_6}$ , one only needs to identify the top 31.0% zeolites predicted by the ANN model for further research, which greatly shortens the time to

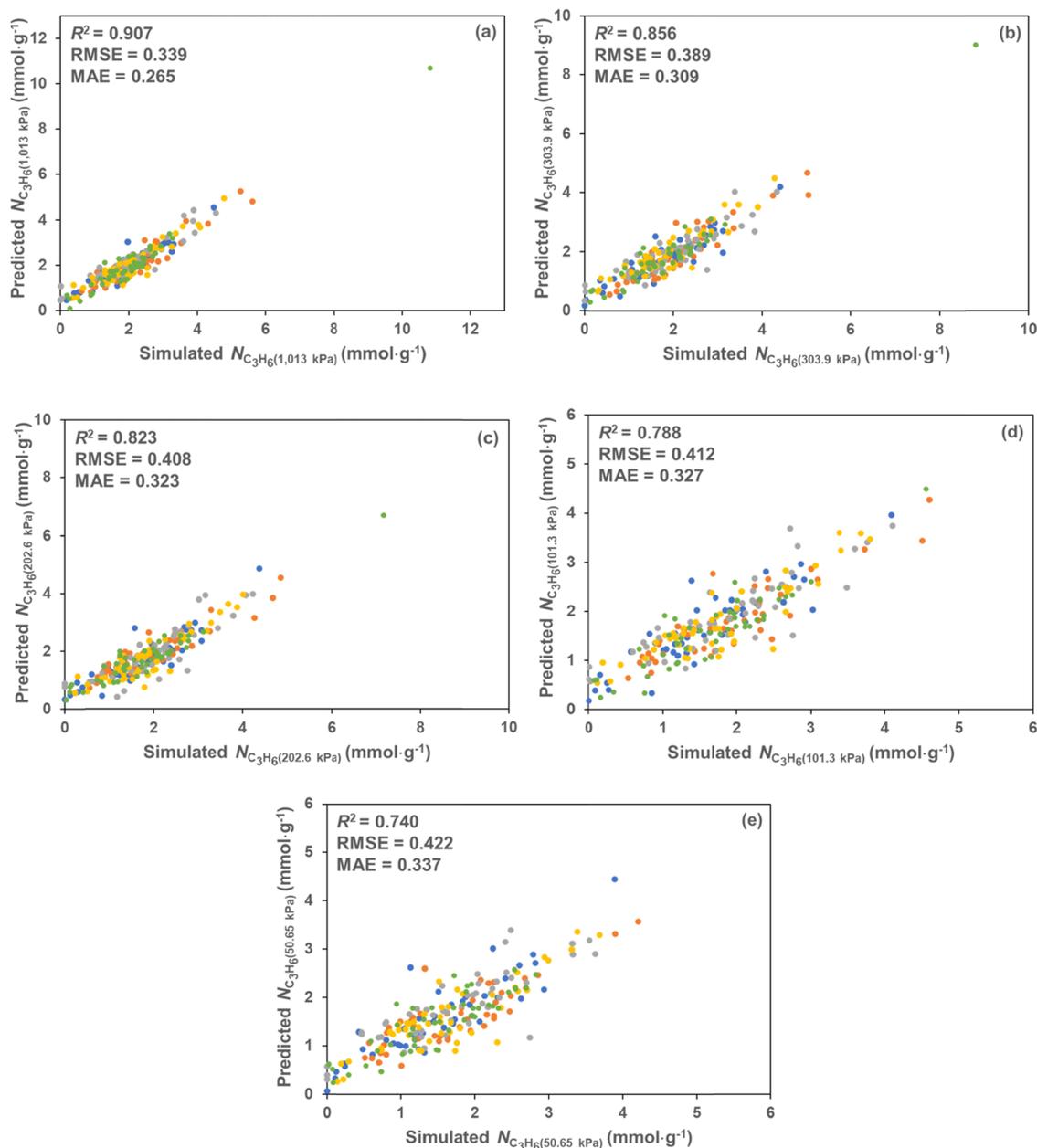


**Figure 5.** Comparison of ANN model predictions using five structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) with  $N_{C_3H_6}$  obtained by GCMC simulation (fivefold CV approach) at 300 K and 5,065 kPa.

find optimal zeolites for  $C_3H_6$  adsorption. A significant acceleration of the screening process for identifying high-performing candidates from millions of materials is also made possible, whereas it is extremely time-consuming by means of molecular simulations alone.

As demonstrated in [Figure 6a–e](#), the significantly enhanced prediction performance does not change the fact that it deteriorates as pressure decreases in comparison with multilinear regression and quadratic regression models. With the gradual decrease of adsorption pressure, the external force is no longer large enough for the guest molecules to fill the entire void space of zeolites. At low pressures, guest molecules are usually adsorbed in a portion of the void regions with strong binding sites that are not sufficiently captured by these general global descriptors, causing undesirable performance at 50.65 kPa in [Figure 6e](#) ( $R^2 = 0.740$ ). Accordingly, the search for a specific descriptor to enhance the prediction accuracy of the model for low-pressure gas adsorption seems to be a must. Considering the combined pressures (see [Figure 7a](#)), the  $R^2$  (0.887) of the ANN model increases greatly compared with the previous models ( $R^2 = 0.722$  for a multilinear regression model, and  $R^2 = 0.803$  for a quadratic regression model), which indicates that strong nonlinear capacity of the ANN model makes the  $P$  descriptor affecting  $N_{C_3H_6}$  work at various pressures. Additionally, the fine predictive performance of the ANN model is further demonstrated by the residual histogram with a standard normal distribution shown in [Figure 7b](#).

**3.2.2. Six Descriptors with PSD.** Simple structural descriptors provide global porosity characteristics of the zeolites, which are not enough to predict low-pressure gas adsorption. Consequently, one tries to explore descriptors with more implicit information, and PSD is a good choice, which refers to the rate of change of pore volume with pore size. PSD describes the pore morphology with upper and lower bounds of pore diameters and their relative proportions, and it is extremely sensitive to small changes in the pore diameter, while it could not embody subtle changes in pore surface texture and other features. Notably, we just need to calculate the PSD histogram of the pore diameter range of 3~6 Å as the powerful attraction potential of  $C_3H_6$  molecules and O atoms in frameworks emerges in this range, as demonstrated in [Figure 8](#), and corresponding LJ parameters are shown in [Table S4](#).



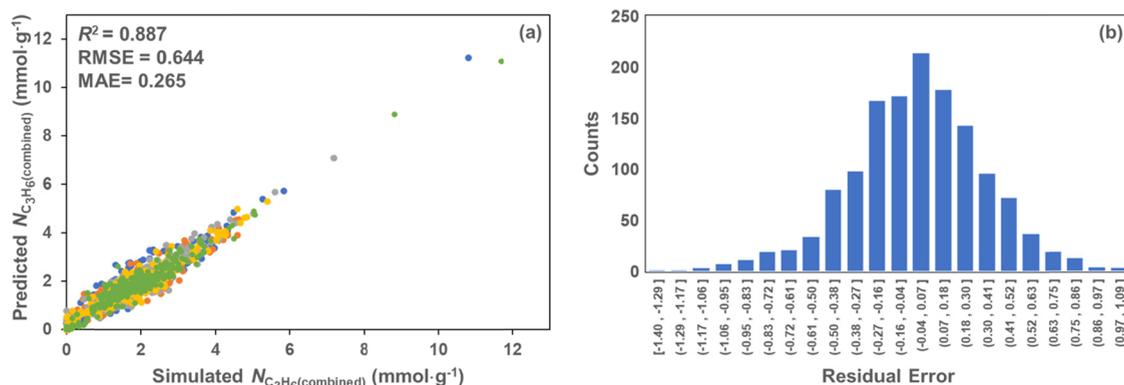
**Figure 6.** Comparison of ANN model predictions using five structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) with  $N_{C_3H_6}$  obtained by GCMC simulation (fivefold CV approach) at 300 K and (a) 1,013 kPa, (b) 303.9 kPa, (c) 202.6 kPa, (d) 101.3 kPa, and (e) 50.65 kPa.

Additionally, it is also found that expansion of the pore diameter range does not enhance the prediction of the ANN model. The implementation of this strategy significantly reduces the complexity of the input data while minimizing the impact on the accuracy.

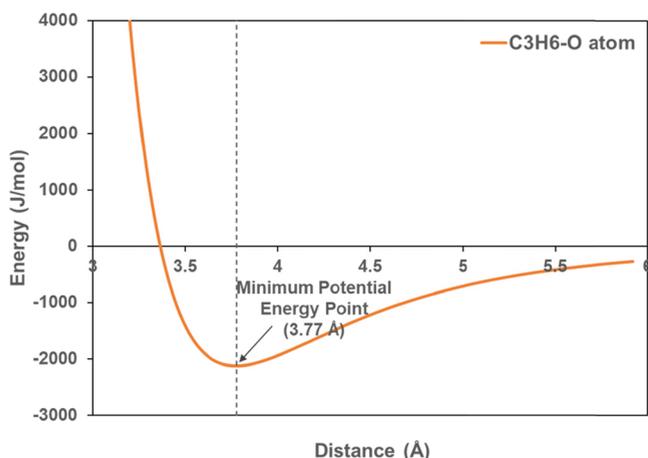
The five structural descriptors mentioned above and the present PSD serve as inputs of ANN models to make predictions to  $N_{C_3H_6}$  at diverse pressures. As shown in Figure S9, the addition of PSD does not improve the prediction accuracy of the model compared with the ANN model using five simple structural descriptors alone, which shows that the supplement of PSD has little effect on the generalization of the model at diverse pressures. Surprisingly, for combined pressures, the  $R^2$  value of the ANN model with six structural descriptors and  $P$  descriptor is 0.967 (see Figure 9a) with an enhancement of 9.0% compared with no PSD, which is a

pleasant outcome. Moreover, the residual histogram with normal distribution shown in Figure 9b also verifies the excellent prediction performance. It is promising that the PSD with many data points may be well matched with a lot of data sets to obtain good fitting accuracy.

**3.2.3. Six Descriptors with  $Q_{st}$ .** The ideal zeolite for  $C_3H_6$  adsorption has a high uptake at adsorption pressure, as well as a low uptake at desorption pressure. To date, the prediction of high-pressure gas adsorption has been very pleasing; however, the desired expectations have not been achieved at low pressures. Consequently, we continued to search for the descriptor that could explain the adsorption mechanism at low pressures. Given the regional features of gas adsorption at low pressures, the chemistry of pores is likely to be a dominating factor in determining the low-pressure adsorption of  $C_3H_6$  and likely other gases, which is also Burner's opinion.<sup>63</sup> As an



**Figure 7.** (a) Comparison of ANN model predictions using five structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) and  $P$  with combined  $N_{C_3H_6}$  obtained by GCMC simulation (fivefold CV approach) at six different pressures; (b) residual statistical histogram for all sets.

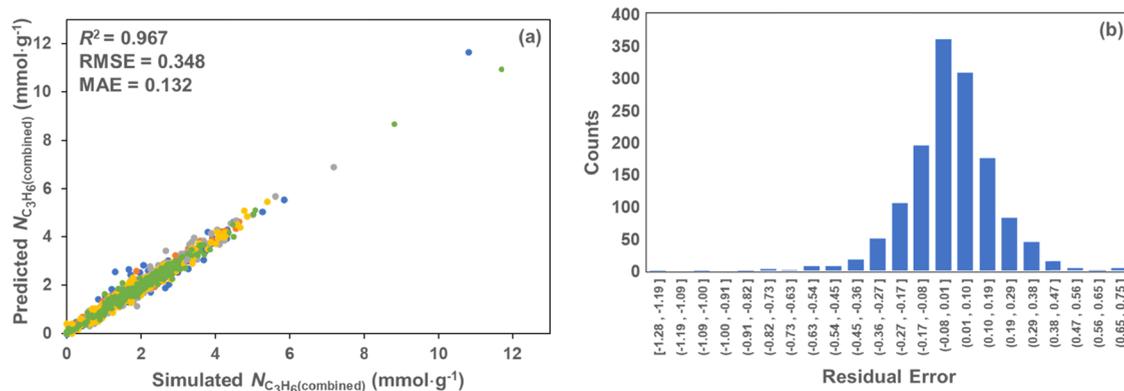


**Figure 8.** LJ potential surfaces for  $C_3H_6 \cdots O$  as a function of distance, where the O atom is assumed to be a concentration of LJ potential energy for zeolites.

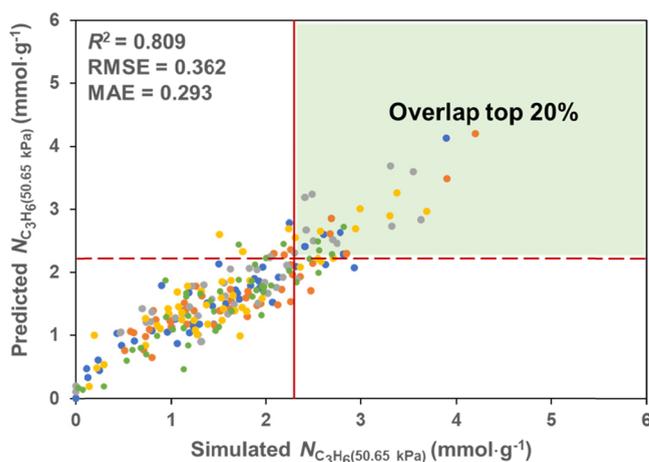
essential parameter to characterize the heterogeneity of the adsorption surface,  $Q_{st}$  could disclose essential information about chemical interactions between the adsorbate and adsorbent, and  $C_3H_6$  molecules with an excessively large  $Q_{st}$  value in a zeolite indicate very strong interaction with the framework, which is a complement to PSD.

Five general structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) together with an energy descriptor ( $Q_{st}$ ) are used as

inputs to train the ANN model. As shown in Figure S10a, the accuracy of the model ( $R^2 = 0.929$ ) is improved by 2.1% with the addition of  $Q_{st}$  at 300 K and 5,065 kPa, which means that the five structural descriptors alone are sufficient to predict  $N_{C_3H_6}$  if concise input data are required at high pressures. The proportion of void space filled by guest molecules decreases with the decline of pressure, and a gradual improvement in  $R^2$  (1.3%  $\sim$  5.1%) is observed, compared to the case where only five structural descriptors are used (see Figure S10b–e). At 50.65 kPa, the model's  $R^2$  (0.809) is enhanced by 9.3% in Figure 10, which is quite remarkable in contrast with no energy descriptor ( $Q_{st}$ ), and this further confirms the indispensability of  $Q_{st}$  for low-pressure gas adsorption prediction. The actual top 20% and predicted top 20% zeolites are all selected to further observe the deviation predicted for the ANN model from Figure 10. Predicted top 41.8% zeolites are necessary to be identified for subsequent research for the purpose of gaining the real top 20% zeolites, which is satisfactory. Accordingly, the  $Q_{st}$  value does accurately capture the local porosity characteristics of gas adsorption at low pressures. For combined pressures, while the prediction performance ( $R^2 = 0.902$ ) of the ANN model containing  $Q_{st}$  in Figure S11a is inferior to the ANN model with PSD ( $R^2 = 0.967$ ), the superiority of it is not negligible compared with the ANN model with five general structural descriptors ( $R^2 = 0.887$ ). Furthermore, the residual histogram shown in Figure S11b is a standard normal distribution. For clarity,  $R^2$ , RMSE, and MAE values corresponded to 232 zeolites for three different ANN models



**Figure 9.** (a) Comparison of ANN model predictions using six structural descriptors (LCD, PLD, ASA, AV,  $\rho$ , and PSD) and  $P$  with combined  $N_{C_3H_6}$  obtained by GCMC simulation (fivefold CV approach) at six different pressures; (b) residual statistical histogram for all sets.



**Figure 10.** Comparison of ANN model predictions using five structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) and an energy descriptor ( $Q_{st}$ ) with  $N_{C_3H_6}$  obtained by GCMC simulation (fivefold CV approach) at 300 K and 50.65 kPa.

using the fivefold CV approach are summarized in Table 1, and the contribution of various descriptors to the predictive performance of ANN models is quite clear. Additionally,  $R^2$  on the training and test sets of three ANN models using the random validation approach at various pressures is summarized in Table S5 in the Supporting Information.

A comparison of the predicted labels from the ANN model and the labels from the GCMC simulation at different pressures is depicted using an ROC plot, which illustrates the ability of the model to correctly label zeolites. The ROC depicts relative trade-offs between true positives (benefits) and false positives (costs). The perfect prediction method would yield a point in the upper left corner of the ROC plot, with a 100% sensitivity (no false negatives) and 100% specificity (no false positives) and an area under the curve (AUC) of 1. In order to examine the fine predicted performance of the ANN model with the five structural descriptors and  $Q_{st}$  on  $N_{C_3H_6}$  at low pressures closely, we initially selected a cut-off criterion of 20% in light of simulated  $N_{C_3H_6}$  and then labeled zeolites as “positive” or “negative” depending on whether the predicted  $N_{C_3H_6}$  was in the top 20% or in the remaining 80% on the 232 zeolites using random validation and fivefold CV approaches at 300 K and 50.65 kPa, as shown in Figure 11. The ANN model displays outstanding performance, with an AUC of 0.955 for random validation and 0.951 for fivefold CV at 300 K and 50.65 kPa. Taking into account the possible impact of the criteria for dividing labels, AUC values for random validation

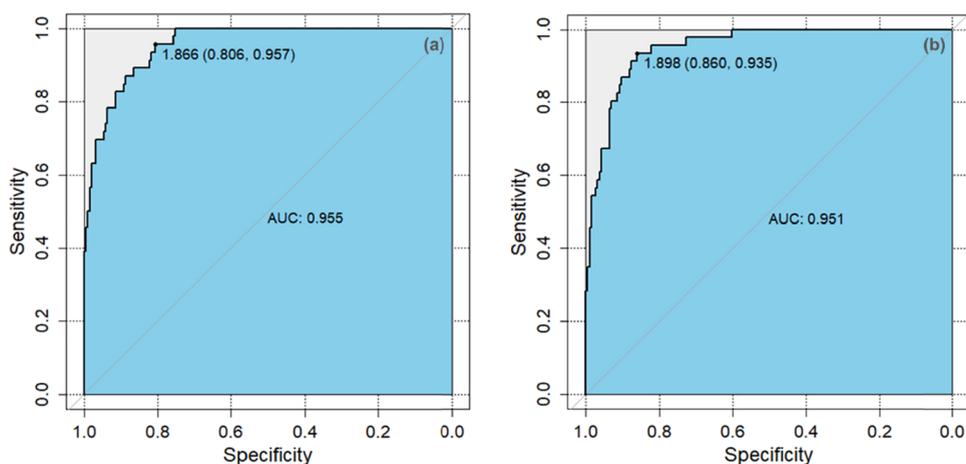
and fivefold CV at 50.65 kPa were calculated according to several classification criteria. As illustrated in Figure 12, the criteria do not affect the fact that the AUC values exceed 0.9 for these two validation methods. This excellent performance certifies the superiority of the ANN model in classifying zeolites into high or low  $C_3H_6$  adsorbers at 50.65 kPa based on five structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) and an energy descriptor ( $Q_{st}$ ).

Subsequently, the influence of six descriptors (LCD, PLD, ASA, AV,  $\rho$ , and  $Q_{st}$ ) on  $N_{C_3H_6}$  in the ANN model at various pressures was further explored. A common feature importance measure, namely, permutation feature importance (PFI),<sup>64</sup> was calculated. To get the PFI of a descriptor such as  $Q_{st}$ , we randomly permute  $Q_{st}$  and use it to predict  $N_{C_3H_6}$  while keeping all other input features nonpermuted, resulting in reduced prediction accuracy. The value of PFI for  $Q_{st}$  is given by the ratio in error measures between the original accuracy and the accuracy resulting from having that  $Q_{st}$  randomly permuted.<sup>65</sup> The above steps are repeated 5 times, and the average value is taken as the final PFI value of  $Q_{st}$  to avoid the influence of the uncertainty of random permutation. All input features have been considered for PFI with the error measure being MAE. The PFI values of each descriptor at different pressures are depicted in Figure 13. The PFI values of the ANN model are of an order of  $AV > ASA > Q_{st} > LCD > \rho > PLD$  at 300 K and 5,065 kPa. Undoubtedly, AV is the most crucial porosity characteristic determining  $N_{C_3H_6}$  at high pressures, and this is also revealed by the previous multilinear regression and quadratic regression models, which confirms the mechanism of molecular filling again for high-pressure gas adsorption. With the decline of pressure, the importance of AV decreases, while the significance of  $Q_{st}$  and LCD becomes increasingly prominent. The order of PFI values of the ANN model is  $Q_{st} > LCD > AV > PLD > \rho > ASA$  at 300 K and 50.65 kPa. When the pressure is not high enough to squeeze the guest molecules into channels, LCD becomes the vital element preventing the guest molecules from entering the window. In this situation, guest molecules are no longer adsorbed into the void space in a stacked state but selectively adsorbed in the regions with strong interaction, and  $Q_{st}$  just reflects this local porosity characteristic. In short, AV is the most critical feature at high pressures, while  $Q_{st}$  and LCD become the two most meaningful features at low pressures, which coincides with the comments in the literature.<sup>19,38</sup>

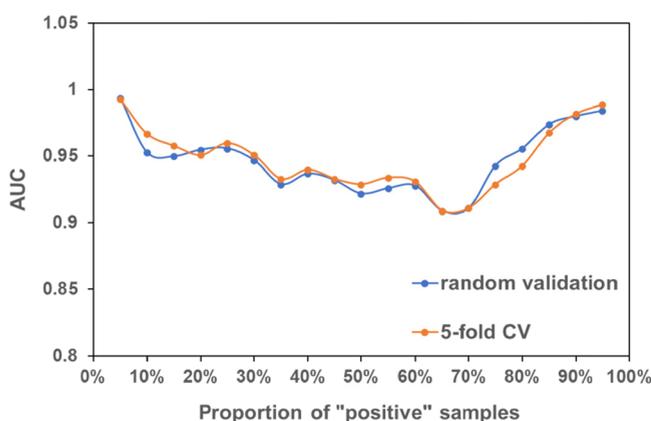
The prediction performance of the ANN model after the addition of  $Q_{st}$  is indeed enhanced for low-pressure gas adsorption, yet it is worth exploring carefully which factor is responsible for this improvement, the change in freedom of

**Table 1.**  $R^2$ , RMSE, and MAE Values Corresponded to 232 Zeolites of Three ANN Models (Fivefold CV Approach) at Various Pressures with S Representing the General Structural Descriptor

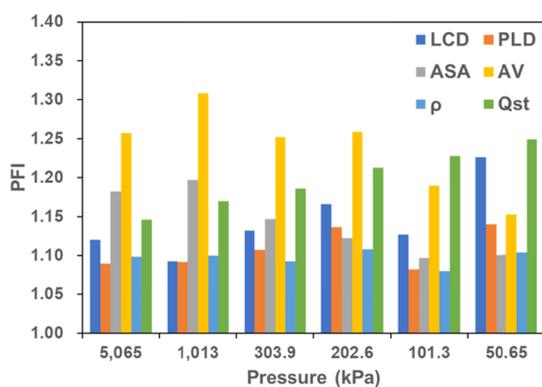
| pressure (kPa) | ANN (SS) |       |       | ANN (SS + PSD) |       |       | ANN (SS + $Q_{st}$ ) |       |       |
|----------------|----------|-------|-------|----------------|-------|-------|----------------------|-------|-------|
|                | $R^2$    | RMSE  | MAE   | $R^2$          | RMSE  | MAE   | $R^2$                | RMSE  | MAE   |
| 5,065          | 0.910    | 0.342 | 0.256 | 0.902          | 0.356 | 0.276 | 0.929                | 0.304 | 0.236 |
| 1,013          | 0.907    | 0.339 | 0.265 | 0.898          | 0.354 | 0.278 | 0.919                | 0.316 | 0.241 |
| 303.9          | 0.856    | 0.389 | 0.309 | 0.861          | 0.382 | 0.313 | 0.878                | 0.358 | 0.283 |
| 202.6          | 0.823    | 0.408 | 0.323 | 0.815          | 0.418 | 0.328 | 0.851                | 0.375 | 0.293 |
| 101.3          | 0.788    | 0.412 | 0.327 | 0.771          | 0.429 | 0.351 | 0.828                | 0.372 | 0.291 |
| 50.65          | 0.740    | 0.422 | 0.337 | 0.734          | 0.428 | 0.358 | 0.809                | 0.362 | 0.293 |
| combined       | 0.887    | 0.644 | 0.265 | 0.967          | 0.348 | 0.132 | 0.902                | 0.601 | 0.246 |



**Figure 11.** Receiver-operator-curve (ROC) plots that show how well the ANN model labels zeolites as “positive” or “negative” depending on whether the  $N_{C_3H_6}$  at 300 K and 50.65 kPa is in the top 20% or in the remaining 80% for (a) random validation and (b) fivefold CV.

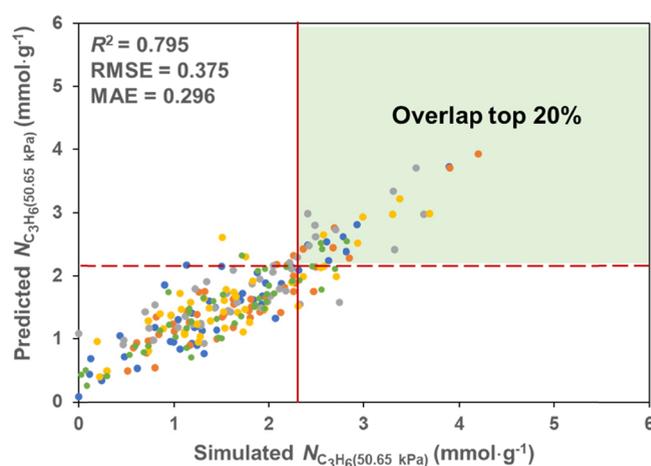


**Figure 12.** Relationships between AUC values and proportion of “positive” samples for random validation and fivefold CV at 300 K and 50.65 kPa.



**Figure 13.** PFI maps of different descriptors with  $N_{C_3H_6}$  at different pressures based on ANN prediction, where  $PFI = MAE_{perm}/MAE_{orig}$ .

input data or the  $Q_{st}$  itself. For clarifying this confusion, ASA with a minimum PFI value was deleted while introducing  $Q_{st}$  to keep the freedom of inputs unchanged for ANN model prediction at 300 K and 50.65 kPa. As illustrated in Figure 14, in the absence of ASA, the prediction performance of the ANN model is still pleasing ( $R^2 = 0.795$ ), and the prediction performance decreases slightly by 1.7% compared with the ANN model with six descriptors, while the performance is



**Figure 14.** Comparison of ANN model predictions using four structural descriptors (LCD, PLD, AV, and  $\rho$ ) and an energy descriptor ( $Q_{st}$ ) with  $N_{C_3H_6}$  obtained by GCMC simulation (fivefold CV approach) at 300 K and 50.65 kPa.

enhanced by 7.4% in contrast with no energy descriptor ( $Q_{st}$ ). Furthermore, we only need to identify the predicted top 49.6% of zeolites for further research to get the real top 20% zeolites, demonstrating this ANN model’s slightly strong impact on zeolites with low  $N_{C_3H_6}$ . The above analysis proves that the enhancement of prediction performance for low-pressure gas adsorption is due to  $Q_{st}$  rather than the increase of freedom of input data, and ASA is not completely ineffective. In light of the simplicity and non-negligible contribution of ASA for low-pressure gas adsorption, the ANN model with five structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) and an energy descriptor ( $Q_{st}$ ) is still considered suitable for predicting  $N_{C_3H_6}$  at low pressures.

#### 4. CONCLUSIONS

A comprehensive QSPR analysis for  $N_{C_3H_6}$  in 232 ordered pure silicon zeolites (Si/O = 1:2) at different pressures was investigated with the aid of a variety of regression models. Univariate analysis and multilinear regression analysis consistently illustrated that AV and LCD were the most significant factors determining  $N_{C_3H_6}$  at 5,065 and 50.65 kPa, respectively.

Due to the involvement of complex binary interaction terms, the prediction performance of the quadratic regression model was significantly superior to multilinear regression, with the enhancements of 10.1% at 5,065 kPa and 31.2% at 50.65 kPa.

For high-pressure gas adsorption with a molecular stacking mechanism, the predicted  $N_{C_3H_6}$  of the ANN model with only five structural descriptors (LCD, PLD, ASA, AV, and  $\rho$ ) was in good agreement with the GCMC-simulated  $N_{C_3H_6}$ . A pleasant prediction ( $R^2 = 0.967$ ) was obtained with the introduction of PSD for combined pressures. At low pressures, the prediction performance of the ANN model increased by 9.3% with the addition of  $Q_{st}$  reflecting chemical interactions between the adsorbate and adsorbent. ROC analysis certifies the superiority of the ANN model with  $Q_{st}$  in classifying zeolites into high or low  $C_3H_6$  adsorbents at 50.65 kPa. Based on the relative weight analysis, AV is the most critical feature at high pressures, while LCD and  $Q_{st}$  become the two most meaningful features at low pressures. Moreover, the enhancement of model performance for low-pressure gas adsorption is due to the addition of  $Q_{st}$  rather than the increase of freedom of input data. Our comprehensive insights into QSPR analysis in this work will provide new ideas for understanding of structure–performance relationships at the atomic level and the design of high-performance zeolites for  $C_3H_6$  adsorption.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acsomega.2c02779>.

Descriptor selection, relationships between  $N_{C_3H_6}$  and zeolite descriptors, predicted  $N_{C_3H_6}$  by multilinear regression models versus simulated results, predicted  $N_{C_3H_6}$  by quadratic regression models versus simulated results, predicted  $N_{C_3H_6}$  by ANN models with six descriptors (LCD, PLD, ASA, AV,  $\rho$ , and PSD) versus simulated results, and predicted  $N_{C_3H_6}$  by ANN models with six descriptors (LCD, PLD, ASA, AV,  $\rho$ , and  $Q_{st}$ ) versus simulated results (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Bing J. Zhang** – School of Materials Science and Engineering, Guangdong Engineering Center for Petrochemical Energy Conservation, The Key Laboratory of Low-carbon Chemistry & Energy Conservation of Guangdong Province, Sun Yat-sen University, Guangzhou 510006, P. R. China; [orcid.org/0000-0001-9830-9193](https://orcid.org/0000-0001-9830-9193); Email: [zhbjing@mail.sysu.edu.cn](mailto:zhbjing@mail.sysu.edu.cn)

### Authors

**Li Zhao** – School of Materials Science and Engineering, Guangdong Engineering Center for Petrochemical Energy Conservation, The Key Laboratory of Low-carbon Chemistry & Energy Conservation of Guangdong Province, Sun Yat-sen University, Guangzhou 510006, P. R. China

**Qi Zhang** – School of Materials Science and Engineering, Guangdong Engineering Center for Petrochemical Energy Conservation, The Key Laboratory of Low-carbon Chemistry & Energy Conservation of Guangdong Province, Sun Yat-sen University, Guangzhou 510006, P. R. China

**Chang He** – School of Materials Science and Engineering, Guangdong Engineering Center for Petrochemical Energy Conservation, The Key Laboratory of Low-carbon Chemistry & Energy Conservation of Guangdong Province, Sun Yat-sen University, Guangzhou 510006, P. R. China; [orcid.org/0000-0002-9024-9269](https://orcid.org/0000-0002-9024-9269)

**Qinglin Chen** – School of Materials Science and Engineering, Guangdong Engineering Center for Petrochemical Energy Conservation, The Key Laboratory of Low-carbon Chemistry & Energy Conservation of Guangdong Province, Sun Yat-sen University, Guangzhou 510006, P. R. China

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acsomega.2c02779>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This research was supported by the National Natural Science Foundation of China (No. 21978330).

## ■ REFERENCES

- (1) Li, H.; Wang, K.; Sun, Y.; Lollar, C. T.; Li, J.; Zhou, H.-C. Recent advances in gas storage and separation using metal–organic frameworks. *Mater. Today* **2018**, *21*, 108–121.
- (2) Zhang, X.; Cui, J.; Zhang, K.; Wu, J.; Lee, Y. Machine learning prediction on properties of nanoporous materials utilizing pore geometry barcodes. *J. Chem. Inf. Model.* **2019**, *59*, 4636–4644.
- (3) Thommes, M.; Schlumberger, C. Characterization of nanoporous materials. *Annu. Rev. Chem. Biomol. Eng.* **2021**, *12*, 137–162.
- (4) Song, Y.; Bao, H.; Shen, X.; Li, X.; Liang, X.; Wang, S. Emerging nanoporous materials for biomolecule separation. *Adv. Funct. Mater.* **2022**, *32*, No. 2113153.
- (5) Ga, S.; Lee, S.; Kim, J.; Lee, J. H. Isotherm parameter library and evaluation software for CO<sub>2</sub> capture adsorbents. *Comput. Chem. Eng.* **2020**, *143*, No. 107105.
- (6) Mohamed, A. M. O.; Bicer, Y. A comprehensive methodology to screen metal–organic frameworks towards sustainable photofixation of nitrogen. *Comput. Chem. Eng.* **2021**, *144*, No. 107130.
- (7) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Froudakis, G. Fast screening of large databases for top performing nanomaterials using a self-consistent, machine learning based approach. *J. Phys. Chem. C* **2020**, *124*, 19639–19648.
- (8) Li, Z.; Bucior, B. J.; Chen, H.; Haranczyk, M.; Siepmann, J. I.; Snurr, R. Q. Machine learning using host/guest energy histograms to predict adsorption in metal–organic frameworks: application to short alkanes and Xe/Kr mixtures. *J. Chem. Phys.* **2021**, *155*, No. 014701.
- (9) Lee, S.; Kim, B.; Cho, H.; Lee, H.; Lee, S. Y.; Cho, E. S.; Kim, J. Computational screening of trillions of metal–organic frameworks for high-performance methane storage. *ACS Appl. Mater. Interfaces* **2021**, *13*, 23647–23654.
- (10) Bucior, B. J.; Bobbitt, N. S.; Islamoglu, T.; Goswami, S.; Gopalan, A.; Yildirim, T.; Farha, O. K.; Bagheri, N.; Snurr, R. Q. Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks. *Mol. Syst. Des. Eng.* **2019**, *4*, 162–174.
- (11) Farmahini, A. H.; Krishnamurthy, S.; Friedrich, D.; Brandani, S.; Sarkisov, L. Performance-based screening of porous materials for carbon capture. *Chem. Rev.* **2021**, *121*, 10666–10741.
- (12) Bai, P.; Jeon, M. Y.; Ren, L.; Knight, C.; Deem, M. W.; Tsapatsis, M.; Siepmann, J. I. Discovery of optimal zeolites for challenging separations and chemical transformations using predictive materials modeling. *Nat. Commun.* **2015**, *6*, 5912.
- (13) Li, L.; Duan, Y.; Liao, S.; Ke, Q.; Qiao, Z.; Wei, Y. Adsorption and separation of propane/propylene on various ZIF-8 polymorphs: insights from GCMC simulations and the ideal adsorbed solution theory (IAST). *Chem. Eng. J.* **2020**, *386*, No. 123945.

- (14) Lu, X.; Wu, Y.; Wu, X.; Cao, Z.; Wei, X.; Cai, W. High-throughput computational screening of porous polymer networks for natural gas sweetening based on a neural network. *AIChE J.* **2022**, *68*, No. e17433.
- (15) Hiraide, S.; Sakanaka, Y.; Kajiro, H.; Kawaguchi, S.; Miyahara, M. T.; Tanaka, H. High-throughput gas separation by flexible metal–organic frameworks with fast gating and thermal management capabilities. *Nat. Commun.* **2020**, *11*, 3867.
- (16) Tang, H.; Jiang, J. In silico screening and design strategies of ethane-selective metal–organic frameworks for ethane/ethylene separation. *AIChE J.* **2021**, *67*, No. e17025.
- (17) Alshehri, A. S.; Gani, R.; You, F. Deep learning and knowledge-based methods for computer-aided molecular design-towards a unified approach: state-of-the-art and future directions. *Comput. Chem. Eng.* **2020**, *141*, No. 107005.
- (18) Zhang, L.; Mao, H.; Liu, L.; Du, J.; Gani, R. A machine learning based computer-aided molecular design/screening methodology for fragrance molecules. *Comput. Chem. Eng.* **2018**, *115*, 295–308.
- (19) Krishnapriyan, A. S.; Haranczyk, M.; Morozov, D. Topological descriptors help predict guest adsorption in nanoporous materials. *J. Phys. Chem. C* **2020**, *124*, 9360–9368.
- (20) Simon, C. M.; Kim, J.; Lin, L.-C.; Martin, R. L.; Haranczyk, M.; Smit, B. Optimizing nanoporous materials for gas storage. *Phys. Chem. Chem. Phys.* **2014**, *16*, 5499–5513.
- (21) Durá, G.; Budarin, V. L.; Castro-Osma, J. A.; Shuttleworth, P. S.; Quek, S. C. Z.; Clark, J. H.; North, M. Importance of micropore-mesopore interfaces in carbon dioxide capture by carbon-based materials. *Angew. Chem., Int. Ed.* **2016**, *128*, 9319–9323.
- (22) Zhang, Z.; Schott, J. A.; Liu, M.; Chen, H.; Lu, X.; Sumpster, B. G.; Fu, J.; Dai, S. Prediction of carbon dioxide adsorption via deep learning. *Angew. Chem., Int. Ed.* **2019**, *131*, 265–269.
- (23) Wang, S.; Zhang, Z.; Dai, S.; Jiang, D.-E. Insights into CO<sub>2</sub>/N<sub>2</sub> selectivity in porous carbons from deep learning. *ACS Mater. Lett.* **2019**, *1*, 558–563.
- (24) Wang, S.; Li, Y.; Dai, S.; Jiang, D.-E. Prediction by convolutional neural networks of CO<sub>2</sub>/N<sub>2</sub> selectivity in porous carbons from N<sub>2</sub> adsorption isotherm at 77 K. *Angew. Chem.* **2020**, *132*, 19813–19816.
- (25) Fernandez, M.; Trefiak, N. R.; Woo, T. K. Atomic property weighted radial distribution functions descriptors of metal–organic frameworks for the prediction of gas uptake capacity. *J. Phys. Chem. C* **2013**, *117*, 14095–14105.
- (26) Fernandez, M.; Barnard, A. S. Geometrical properties can predict CO<sub>2</sub> and N<sub>2</sub> adsorption performance of metal–organic frameworks (MOFs) at low pressure. *ACS Comb. Sci.* **2016**, *18*, 243–252.
- (27) Aghaji, M. Z.; Fernandez, M.; Boyd, P. G.; Daff, T. D.; Woo, T. K. Quantitative structure–property relationship models for recognizing metal organic frameworks (MOFs) with high CO<sub>2</sub> working capacity and CO<sub>2</sub>/CH<sub>4</sub> selectivity for methane purification. *Eur. J. Inorg. Chem.* **2016**, *2016*, 4505–4511.
- (28) Dureckova, H.; Krykunov, M.; Aghaji, M. Z.; Woo, T. K. Robust machine learning models for predicting high CO<sub>2</sub> working capacity and CO<sub>2</sub>/H<sub>2</sub> selectivity of gas adsorption in metal organic frameworks for precombustion carbon capture. *J. Phys. Chem. C* **2019**, *123*, 4133–4139.
- (29) Lin, S.; Wang, Y.; Zhao, Y.; Pericchi, L. R.; Hernández-Maldonado, A. J.; Chen, Z. Machine-learning-assisted screening of pure-silica zeolites for effective removal of linear siloxanes and derivatives. *J. Mater. Chem. A* **2020**, *8*, 3228–3237.
- (30) Wanyonyi, F. S.; Fidelis, T. T.; Louis, H.; Mutua, G. K.; Orata, F.; Rhyman, L.; Ramasami, P.; Pembere, A. M. S. Simulation guided prediction of zeolites for the sorption of selected anions from water: machine learning predictors for enhanced loading. *J. Mol. Liq.* **2022**, *355*, No. 118913.
- (31) Lin, L.-C.; Berger, A. H.; Martin, R. L.; Kim, J.; Swisher, J. A.; Jariwala, K.; Rycroft, C. H.; Bhowan, A. S.; Deem, M. W.; Haranczyk, M.; Smit, B. In silico screening of carbon-capture materials. *Nat. Mater.* **2012**, *11*, 633–641.
- (32) Fernandez, M.; Boyd, P. G.; Daff, T. D.; Aghaji, M. Z.; Woo, T. K. Rapid and accurate machine learning recognition of high performing metal organic frameworks for CO<sub>2</sub> capture. *J. Phys. Chem. Lett.* **2014**, *5*, 3056–3060.
- (33) Lee, Y.; Barthel, S. D.; Dlotko, P.; Moosavi, S. M.; Hess, K.; Smit, B. Quantifying similarity of pore-geometry in nanoporous materials. *Nat. Commun.* **2017**, *8*, 15396.
- (34) Lee, Y.; Barthel, S. D.; Dlotko, P.; Moosavi, S. M.; Hess, K.; Smit, B. High-throughput screening approach for nanoporous materials genome using topological data analysis: application to zeolites. *J. Chem. Theory Comput.* **2018**, *14*, 4427–4437.
- (35) Krishnapriyan, A. S.; Montoya, J.; Haranczyk, M.; Hummelshøj, J.; Morozov, D. Machine learning with persistent homology and chemical word embeddings improves prediction accuracy and interpretability in metal–organic frameworks. *Sci. Rep.* **2021**, *11*, 8888.
- (36) 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.
- (37) Fanourgakis, G. S.; Gkagkas, K.; Tylanakis, E.; Froudakis, G. A generic machine learning algorithm for the prediction of gas adsorption in nanoporous materials. *J. Phys. Chem. C* **2020**, *124*, 7117–7126.
- (38) Shi, Z.; Liang, H.; Yang, W.; Liu, J.; Liu, Z.; Qiao, Z. Machine learning and in silico discovery of metal–organic frameworks: methanol as a working fluid in adsorption-driven heat pumps and chillers. *Chem. Eng. Sci.* **2020**, *214*, No. 115430.
- (39) Yan, Y.; Shi, Z.; Li, H.; Li, L.; Yang, X.; Li, S.; Liang, H.; Qiao, Z. Machine learning and in-silico screening of metal–organic frameworks for O<sub>2</sub>/N<sub>2</sub> dynamic adsorption and separation. *Chem. Eng. J.* **2022**, *427*, No. 131604.
- (40) Pardakhti, M.; Nanda, P.; Srivastava, R. Impact of chemical features on methane adsorption by porous materials at varying pressures. *J. Phys. Chem. C* **2020**, *124*, 4534–4544.
- (41) Halder, P.; Singh, J. K. High-throughput screening of metal–organic frameworks for ethane-ethylene separation using the machine learning technique. *Energy Fuels* **2020**, *34*, 14591–14597.
- (42) Wang, Z.; Zhou, Y.; Zhou, T.; Sundmacher, K. Identification of optimal metal–organic frameworks by machine learning: structure decomposition, feature integration, and predictive modeling. *Comput. Chem. Eng.* **2022**, *160*, No. 107739.
- (43) Yeo, B. C.; Kim, D.; Kim, H.; Han, S. S. High-throughput screening to investigate the relationship between the selectivity and working capacity of porous materials for propylene/propane adsorptive separation. *J. Phys. Chem. C* **2016**, *120*, 24224–24230.
- (44) Martins, V. F. D.; Seabra, R.; Silva, P.; Ribeiro, A. M.; Cho, K. H.; Lee, U.-H.; Chang, J.-S.; Loureiro, J. M.; Rodrigues, A. E.; Ferreira, A. C<sub>2</sub>/C<sub>3</sub> hydrocarbon separation by pressure swing adsorption on MIL-100 (Fe). *Ind. Eng. Chem. Res.* **2020**, *59*, 10568–10582.
- (45) Kim, J.-J.; Hong, S.-H.; Park, D.; Chung, K.; Lee, C.-H. Separation of propane and propylene by desorbent swing adsorption using zeolite 13X and carbon dioxide. *Chem. Eng. J.* **2021**, *410*, No. 128276.
- (46) Maghsoudi, H.; Abdi, H.; Aidani, A. Temperature- and pressure-dependent adsorption equilibria and diffusivities of propylene and propane in pure-silica Si-CHA zeolite. *Ind. Eng. Chem. Res.* **2020**, *59*, 1682–1692.
- (47) Zhang, X.; Zhou, T.; Sundmacher, K. Integrated metal–organic framework and pressure/vacuum swing adsorption process design: descriptor optimization. *AIChE J.* **2022**, *68*, No. e17524.
- (48) Andres-Garcia, E.; Oar-Arteta, L.; Gascon, J.; Kapteijn, F. ZIF-67 as silver-bullet in adsorptive propane/propylene separation. *Chem. Eng. J.* **2019**, *360*, 10–14.
- (49) Bacsik, Z.; Cheung, O.; Vasiliev, P.; Hedin, N. Selective separation of CO<sub>2</sub> and CH<sub>4</sub> for biogas upgrading on zeolite NaKA and SAPO-56. *Appl. Energy* **2016**, *162*, 613–621.

- (50) Qiao, Z.; Xu, Q.; Cheetham, A. K.; Jiang, J. High-throughput computational screening of metal–organic frameworks for thiol capture. *J. Phys. Chem. C* **2017**, *121*, 22208–22215.
- (51) Qiao, Z.; Xu, Q.; Jiang, J. Computational screening of hydrophobic metal–organic frameworks for the separation of H<sub>2</sub>S and CO<sub>2</sub> from natural gas. *J. Mater. Chem. A* **2018**, *6*, 18898–18905.
- (52) Khalkhali, M.; Ghorbani, A.; Bayati, B. Study of adsorption and diffusion of methyl mercaptan and methane on FAU zeolite using molecular simulation. *Polyhedron* **2019**, *171*, 403–410.
- (53) Qin, S. Molecular dynamics simulation of diffusion and separation of CO<sub>2</sub>/CH<sub>4</sub>/N<sub>2</sub> on MER zeolites. *J. Fuel Chem. Technol.* **2021**, *49*, 1531–1539.
- (54) Abdelrasoul, A.; Zhang, H.; Cheng, C.-H.; Doan, H. Applications of molecular simulations for separation and adsorption in zeolites. *Microporous Mesoporous Mater.* **2017**, *242*, 294–348.
- (55) Ahmad, Z. U.; Chao, B.; Konggadinata, M. I.; Lian, Q.; Zappi, M. E.; Gang, D. D. Molecular simulation and experimental validation of resorcinol adsorption on ordered mesoporous carbon (OMC). *J. Hazard. Mater.* **2018**, *354*, 258–265.
- (56) Plaza, M. G.; Ribeiro, A. M.; Ferreira, A.; Santos, J. C.; Lee, U.-H.; Chang, J.-S.; Loureiro, J. M.; Rodrigues, A. E. Propylene/propane separation by vacuum swing adsorption using Cu-BTC spheres. *Sep. Purif. Technol.* **2012**, *90*, 109–119.
- (57) Olson, D. H.; Cambor, M. A.; Villaescusa, L. A.; Kuehl, G. H. Light hydrocarbon sorption properties of pure silica Si-CHA and ITQ-3 and high silica ZSM-58. *Microporous Mesoporous Mater.* **2004**, *67*, 27–33.
- (58) Pham, T. D.; Lobo, R. F. Adsorption equilibria of CO<sub>2</sub> and small hydrocarbons in AEI-, CHA-, STT-, and RRO-type siliceous zeolites. *Microporous Mesoporous Mater.* **2016**, *236*, 100–108.
- (59) Sun, M. S.; Shah, D. B.; Xu, H. H.; Talu, O. Adsorption equilibria of C<sub>1</sub> to C<sub>4</sub> alkanes, CO<sub>2</sub>, and SF<sub>6</sub> on silicalite. *J. Phys. Chem. B* **1998**, *102*, 1466–1473.
- (60) Abdi, H.; Maghsoudi, H. All-silica DD3R zeolite for adsorptive separation of propylene from propane: equilibrium and kinetic data. *Microporous Mesoporous Mater.* **2020**, *307*, No. 110513.
- (61) Long, R.; Xia, X.; Zhao, Y.; Li, S.; Liu, Z.; Liu, W. Screening metal–organic frameworks for adsorption-driven osmotic heat engines via grand canonical Monte Carlo simulations and machine learning. *iScience* **2021**, *24*, No. 101914.
- (62) Pinheiro, M.; Martin, R. L.; Rycroft, C. H.; Jones, A.; Iglesia, E.; Haranczyk, M. Characterization and comparison of pore landscapes in crystalline porous materials. *J. Mol. Graphics Modell.* **2013**, *44*, 208–219.
- (63) Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P. G.; Woo, T. K. High-performing deep learning regression models for predicting low-pressure CO<sub>2</sub> adsorption properties of metal–organic frameworks. *J. Phys. Chem. C* **2020**, *124*, 27996–28005.
- (64) Molnar, C. *Interpretable Machine Learning*, 2nd ed.; Available online: <https://christophm.github.io/interpretable-ml-book/> (accessed on 12 July 2022).
- (65) Janitza, S.; Strobl, C.; Boulesteix, A.-L. An AUC-based permutation variable importance measure for random forests. *BMC Bioinformatics* **2013**, *14*, 119.